How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies

Joanna Bergström
University of Copenhagen

Tor-Salve Dalsgaard
University of Copenhagen

Jason Alexander
University of Bath

Kasper Hornbæk
University of Copenhagen

ABSTRACT

The VR community has introduced many object selection and manipulation techniques during the past two decades. Typically, they are empirically studied to establish their benefits over the state-of-the-art. However, the literature contains few guidelines on how to conduct such studies; standards developed for evaluating 2D interaction often do not apply. This lack of guidelines makes it hard to compare techniques across studies, to report evaluations consistently, and therefore to accumulate or replicate findings. To build such guidelines, we review 20 years of studies on VR object selection and manipulation. Based on the review, we propose recommendations for designing studies and a checklist for reporting them. We also identify research directions for improving evaluation methods and offer ideas for how to make studies more ecologically valid and rigorous.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; HCI design and evaluation methods.

KEYWORDS

virtual reality, object selection and manipulation, experiments

1 INTRODUCTION

Object selection and manipulation are fundamental interactions in virtual reality (VR). In their textbook, LaViola et al. [38] identify selection and manipulation as one of the three types of tasks for interaction techniques with interfaces such as VR (in addition to travel and system control). The literature contains numerous and diverse set of proposed interaction techniques that aim to improve

object selection [e.g., 32, 41, 68, 84] and manipulation [e.g., 94] in VR (for summaries, see [1, 9, 20, 38, 52]).

These techniques are often empirically studied to assess their performance characteristics and understand their advantages over the state-of-the-art. Before doing so, researchers and practitioners are faced with the question of how to fairly, systematically, and comprehensively evaluate them. For non-VR techniques, significant guidance on how to conduct studies of selection and manipulation exist, both in 2D [e.g., 22, 71] and 3D [e.g., 75]. The literature includes standardized layouts of objects, recommendations on data analysis, and validated questionnaires on fatigue.

For immersive VR, we lack such guidance. While many papers present evaluations, we are unaware of any paper that explicitly provides guidance on how to conduct selection and manipulation evaluations in VR. The recommendations for 2D and 3D interaction techniques often do not apply directly to immersive VR. For instance, immersive VR employs stereoscopic displays and often a one-to-one mapping between the physical body motions and the avatar's or the controller's motions. These influence depth perception differently from desktop environments where there is rarely a one-to-one mapping of movements in any direction. Occlusion is also more of an issue from the first-person perspective. In VR, objects of interest may also be distributed across space, making the ISO circle tapping task that is often used for 2D [22] less useful.

We conjecture that the meagre guidance on selection and manipulation evaluation in VR has several harmful consequences. First, it makes designing valid and accurate evaluations difficult for researchers and practitioners. For instance, we have ourselves struggled with simple questions such as 'which arrangement of targets should be used to evaluate a ray-casting technique in VR?'. Second, the absence of standardized approaches to evaluation makes it hard to compare the multitude of techniques introduced every year because each is evaluated differently (e.g., using different tasks, instructions, or settings). As a consequence, accumulating knowledge of good designs and the generalizability of results suffers. Third, it is unnecessarily laborious to replicate and conduct meta-analyses on studies because of the lack of standard established practices in reporting them. VR studies need to cover many more factors than studies of non-immersive surroundings, which exacerbates these difficulties.

To remedy this situation, we review twenty years (2000-2019) of studies of object selection and manipulation in VR across central conferences and journals (IEEE VR, VR journal, VRST, CHI, and UIST) in the field. These studies form the basis for three contributions. First, we *report and analyze the studies* to form an empirical basis of best practices in VR studies. Second, we present *guidelines*

for empirical studies of object selection and manipulation in the form of recommendations for designing studies and a checklist for reporting studies. These aim to help researchers design and replicate studies and compare and generalize from the knowledge gained from existing studies. Third, we discuss *research directions in the study of selection and manipulation techniques*. The purpose of these is to outline open issues in evaluation methodology, to question which techniques we develop, and to reflect on the ecological validity and rigorousness of our empirical work.

2 RELATED WORK

We first summarize earlier work on selection and manipulation techniques in virtual reality. Then we discuss recommendations for evaluating such techniques outside of VR and identify the few study recommendations that do exist specifically for VR.

2.1 Selection and Manipulation Techniques

There are many surveys of the design of selection and manipulation techniques in VR [e.g., 1, 9, 11, 20, 30, 38, 52]. This work covers the design space of such techniques, taxonomies of the tasks they are intended to support, and guidelines for their design.

Bowman and colleagues presented classical reviews of 3D interaction techniques in 2001 and 2004 [9, 11]; they were updated in LaViola et al.'s [38] 2017 book. They separated interaction techniques in virtual environments according to three tasks they can support: selection and manipulation, travel, and system control. Further, they highlighted the importance of low-level, generalizable tasks: if an interaction technique works at a low-level, it may be used across a range of higher-level application areas. LaViola et al. [38] called this set of basic manipulation tasks canonical manipulation tasks. Such tasks should be used in evaluation because a technique's performance in those influences and is part of more complex tasks. LaViola et al. [38] identified four canonical tasks for manipulation (pp. 258-259): selection, positioning, rotation, and scaling. The parameters that shape these tasks could be used in evaluations as independent variables. For example, a selection-task includes distance and direction to target, target size, the density of objects around the target, number of targets to be selected, and target occlusion. A rotation or a scaling task includes equivalent parameters as a positioning task: distance and direction to the initial position, length and direction to the target position, translation distance, and required precision of positioning. Their book [38], however, did not discuss how representative or comprehensive these listings of parameters are, nor did it provide suggestions on how to use those in designing experiments.

Hand [30] presented another early survey of 3D interaction techniques. They covered interaction techniques for the same categories of task noted above (object manipulation, navigation, and application control). Manipulation was discussed mostly from the perspective of what different input devices enable for the techniques. While Hand's work did not cover evaluation, they called for increasing our understanding of how to evaluate alongside developing the 3D techniques.

Three further reviews by Dang [20], Argelaguet and Andujar [1], and Mendes et al. [52] also inform the present paper. Both Dang [20] and Argelaguet and Andujar [1] reviewed interaction techniques

in particular with respect to user representations (such as avatar hands or types of cursors) and the characteristics of selection and manipulation techniques (such as their mapping functions or design parameters). The main contribution in the paper by Dang [20] is a definition and classification of cursor types and the related 3D pointing techniques. They did not cover the evaluation thereof but presented a criticism that many of the techniques they review and classify have been implemented without an evaluation, or with an assessment based only on some specific input device. Argelaguet and Andujar [1] further covered models of human pointing and listed a range of factors (e.g., target geometry, object distance and area of reach, object density, and input and output device's features such as DoFs and latencies) that affect the performance with selection techniques. However, they discussed only how to design for performance factors and did not explain how to evaluate a technique's performance using the factors.

In a recent review, Mendes et al. [52] presented a detailed survey of 3D manipulation techniques in virtual environments. They briefly discussed how to compare such techniques but only in terms of the differences and similarities across techniques, not in terms of how those comparisons were conducted.

Other related reviews exist on 3D selection and manipulation techniques (such as for non-immersive 3D environments [35], for mid-air interaction [37], and for hand-held AR [27]), but those are similarly sparse on the details of evaluations. In sum, whereas the types and designs of selection and manipulation techniques seem well covered, how to compare such techniques is underdiscussed in related reviews.

2.2 Standards for Evaluating Selection and Manipulation Techniques

Numerous recommendations on how to conduct studies using standardised tests (such as the ISO 9241-9) for target selection exist for 2D techniques [e.g., 22, 71]. However, many studies raise concerns of these not fitting well into 3D environments. For example, Teather and Stuerzlinger [77] found that the conventional 2D formulation of Fitts' law models the throughput in planar pointing tasks well and seems externally valid with varying feedback (tactile feedback on or off) and view conditions (on or above a stereo display). However, they found that full 3D motions were less well modeled.

In a more recent synthesis, Stuerzlinger and Teather [75] also explained that "the notion of throughput in ISO9241-9 relies on a (at least approximately) spherical hit distribution for the effective measures", and that "strong deviations from that distribution may invalidate the underlying assumption(s) that enable the combination of speed and accuracy into a single measure." Therefore, the standard does not cover how to deal with non-spherical hit distribution which in particular appears in 3D applications of the task. Nor does it cover how to design for and calculate target IDs in 3D.

Some more recent studies have created new ways of applying the ISO tapping task, so as to remedy parts of the issues that it poses for VR, or for 3D use in general. For example, Qian and Teather [67] applied the same ISO 9241-9 task for which the above recommendations were developed, but adapted it into a spiral-shaped layout for 3D interaction in VR. The 3D layout helps include variation in distance of the targets in depth. However, this layout, and other

similar applications of the 2D tapping task for 3D interaction, have not been assessed beyond using them in empirical studies of techniques. Nor are they synthesised or compared to propose a task suitable for VR.

To summarise, the standard does not provide guidance on the ways to vary and calculate target IDs, and we are not aware of further studies proposing how to do that. The standard also does not address many of the issues in VR, such as the use of distractor and occluder targets, and it does not discuss the representativeness of a floating, centered circle of targets in real-world VR applications. Therefore, we are no closer to using a standardised task when evaluating techniques for VR than we were when ISO 9241-9 was released.

2.3 Testbeds for Evaluating Selection and Manipulation Techniques

The problem of a lack of standard tests is addressed in two works by Poupyrev et al. [66] and Bowman et al. [10]. They developed testbeds focused on evaluating interaction in immersive VR. These testbeds relate closely to our work, because they provide useful lists of task types and their parameters, and they have been empirically evaluated.

Poupyrev et al. [66] presented an early framework for manipulation techniques in virtual reality, on which they base their testbed. They implemented three tasks in the testbed: Select, Position, and Orient. They listed independent variables for the three tasks, and proposed evaluation metrics for each independent variable. For example, for the 'Position' task the independent variables are: initial distance to target, initial horizontal and vertical directions, final distance, final horizontal and vertical directions, vertical precision, and horizontal precision. Poupyrev et al. [66] implemented these in a testbed called VRMAT in virtual cubits (a metric dependent on the participant's body), in degrees of arc, and in percent of overlap. The dependent variables can then also be tested and reported in these units.

Bowman et al. [10] held a similar ambition for creating a testbed for standardized evaluations of manipulation in VR. Their aim was to capture representative sets of tasks and environments that can be found in real VR applications. Bowman et al. [10] identified four categories of factors beyond the interaction technique that may influence object selection and manipulation performance: characteristics of the task (e.g., the required accuracy), environment (e.g., the number of objects), user (e.g., spatial ability), and system (e.g., stereo vs. binocular viewing). They include some of these as parameters in their testbed.

Bowman et al. [10] also recognised a lack of definition of performance in VR interaction. Time and accuracy are common measures of performance across most techniques in HCI. Bowman et al. [10], however, believed the setting for VR techniques is more complex, and that performance can also include experienced presence and a number of usability factors, such as ease of use, ease of learning, and user comfort, as well as task-related performance factors, such as spatial orientation or expressiveness of manipulation. In their book with LaViola et al. [38], they also expand their list of important evaluation metrics for VR techniques, adding, for instance,

system performance, task performance (speed and accuracy), and subjective responses on presence, comfort, and sickness.

LaViola et al. [38] discussed the pros and cons of evaluation types, including the testbed. With respect to quantifying the performance of 3D interaction techniques, they explained that the testbed approach is to include as many of the potential factors influencing performance as possible. Further, as many of these factors as possible should be held constant, but finding a balance between these two extremes is difficult. While this discussion is useful for researchers to think about the threats and issues that different approaches pose to evaluations, the book gives no guidance on finding that balance, nor exemplifies the possible approaches between a testbed and a focused experiment.

The discussions in the works of Poupyrev et al. [66] and Bowman et al. [10] are valuable sources and initial steps toward common guidelines for evaluations of object selection and manipulation in VR. However, it is not clear that these testbeds capture current issues in VR evaluation, nor that they are widely used. For example, the framework of Poupyrev et al. [66] introduced the canonical tasks and task parameters of the evaluation space. Bowman et al. [10] extended the testbed also to cover travel and discussed other interaction qualities beyond performance, such as sickness and presence. Many recent experiments address these qualities, for instance, from the perspective of developed techniques related to avatar appearance, larger FoV, and graphics about the surroundings and the task space therein. These techniques are made possible by current technology, but evaluating their effects on the qualities is not covered in the early testbeds, which instead focused on canonical tasks and performance measures. With these, they do guide further research, including ours, sharing the same ambition: to create common practices in evaluation.

3 METHOD

This review aims to analyze research practices in evaluating object selection and manipulation in virtual reality. Through analyzing those practices we aim to describe how evaluations are currently conducted. Based on that description we will discuss how evaluations might be improved and which research questions are rarely explored. We do so by a structured review of the available literature, following the PRISMA [56] guidelines on reporting systematic reviews and their four-phase procedure, presented with our data on Figure 1.

To meet this purpose, our review includes papers that meet the following three criteria:

(1) **VR technology**. The paper needs to involve immersive VR technology, such as head-mounted displays, CAVEs, or other stereoscopic displays.
(2) **User Study**. The paper needs to report on a user study, such as an evaluation or an experiment of interaction in VR. The study needs to be conducted with human participants.
(3) **Object Selection and Manipulation**. The study needs to measure performance in object selection or manipulation tasks, such as selection speed or rotation accuracy.

With Criterion 1, we exclude studies on augmented and mixed reality technologies when they use only see-through setups because those setups depend on the real world (e.g., concerning targets,
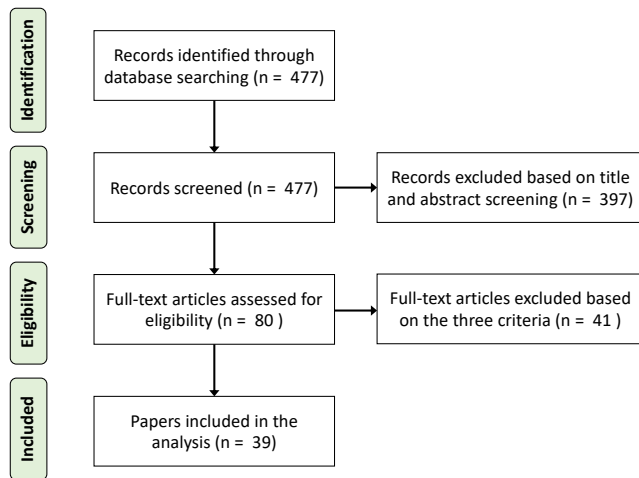
Figure 1: Our literature search, and inclusion phases and rates using the PRISMA procedure.

distractors, and occlusions), unlike in immersive VR. We also exclude 3D interaction, such as mid-air input, when it is performed in physical environments (such as smart homes) or 2D monoscopic projections [e.g., 40] and displays (such as large screens), because they do not provide the similar depth perception cues for object manipulation as in immersive virtual reality.

With Criterion 2, we exclude demonstrations of interaction techniques when they are not evaluated with users [e.g., 90], because our purpose is to identify human study practices. The studies can use any interaction technique. The techniques can be, for instance, based on raycasting or on correcting offsets in inaccuracies of pointing. They can also be interaction techniques intended to support object selection and manipulation instead of directly facilitating it, such as haptic feedback.

With Criterion 3, we exclude studies that do not concern object selection or manipulation performance. For instance, these can be studies whose independent or dependent variables are not focused on object manipulation, but instead cover a higher-level task or experience where object selection or manipulation is merely an incidental part of the task (e.g., temporal navigation [42]). Another example are studies which concern object manipulation, but not task performance (e.g., an observation study [78]).

3.1 Phase 1: Identification

We aimed to identify high-impact papers on object selection and manipulation published in venues on VR and HCI. Using Google Scholar Metrics we identify five such venues: The IEEE Conference on Virtual Reality and 3D User Interfaces (IEEEVR), IEEE; Virtual Reality journal (VR), Springer; The ACM Symposium on Virtual Reality Software and Technology (VRST), ACM; The ACM Conference on Human Factors in Computing Systems (CHI), ACM; and The ACM Symposium on User Interface Software and Technology (UIST), ACM.

To focus our search on *object selection and manipulation* in *virtual reality*, we included these in our search terms. We aimed to

search these terms in all relevant forms, and allowed them to appear anywhere in the title or the abstract. An example query of this on the ACM Digital Library is:

```
Title:((select* OR manipulat*) AND (virtual OR VR)) OR

Abstract:((select* OR manipulat*) AND (virtual OR VR)),
```

Where * denotes any number of unknown characters (wild cards). This way, we include words such as 'select' and 'selection', and 'manipulate' and 'manipulation'. The word 'virtual' was included as an exact search term, but we left the word 'reality' out to include different forms of expressing such settings and related technologies, interactions, user interfaces, and techniques, such as 'virtual reality', 'virtual environment', and 'virtual object manipulation'.

The databases for our five venues differ in their search query options. The Virtual Reality journal (Springer) is only searchable on the full content of a paper (including the body) and we thus filtered the results after the query search to ensure that the search terms appeared in the title or the abstract. For IEEE, we listed the internal conference identifiers specific for the IEEEVR conference and added them to the query to include only that venue. And for CHI, UIST, and VRST (ACM), we included the entire conference proceedings and excluded non-full papers (such as abstracts, posters, and other adjunct publications) in the latter phases.

We included full papers from the past twenty years, 2000–2019, published in English. We chose to include only full papers because posters or adjunct publications often cannot provide the level of details (due to limited paper length) about the experimental methodologies and results. We selected a span of two decades to cover publications with most of the modern technologies for headsets and motion controllers, yet result in an extensive set of papers to learn from. This resulted in 477 results: 229 from IEEEVR, 84 from VRST, 73 from CHI, 59 from VR, and 32 from UIST. We compiled the titles and abstracts of these 477 publications for screening in Phase 2.

3.2 Phase 2: Screening

We screened the titles and abstracts of the 477 papers collected in Phase 1 by using the inclusion criteria presented above. The four authors individually rated the same set of 30 randomly chosen papers for inclusion. The overall percentage agreement on these 30 papers was 91.11%, and the Cohen's Kappa 0.82, 95% CI [0.68, 0.97]. Two of the authors rated the rest of the papers for inclusion. Out of the 477, we included 80 papers for Phase 3 (of which 9 were in the set which was interrated), thereby excluding 397 papers (of which 21 were in the set which was interrated).

3.3 Phase 3: Eligibility

We screened the full-text articles for eligibility with the three criteria. The reasons for exclusion in this phase were either (a) that a paper did not meet one or more of the three inclusion criteria despite the abstract screening, or (b) that a paper was not a full paper. In this phase, we excluded a further 41 publications. Eighteen publications were excluded because they were posters. A further 13 papers were excluded because they did not use VR technology
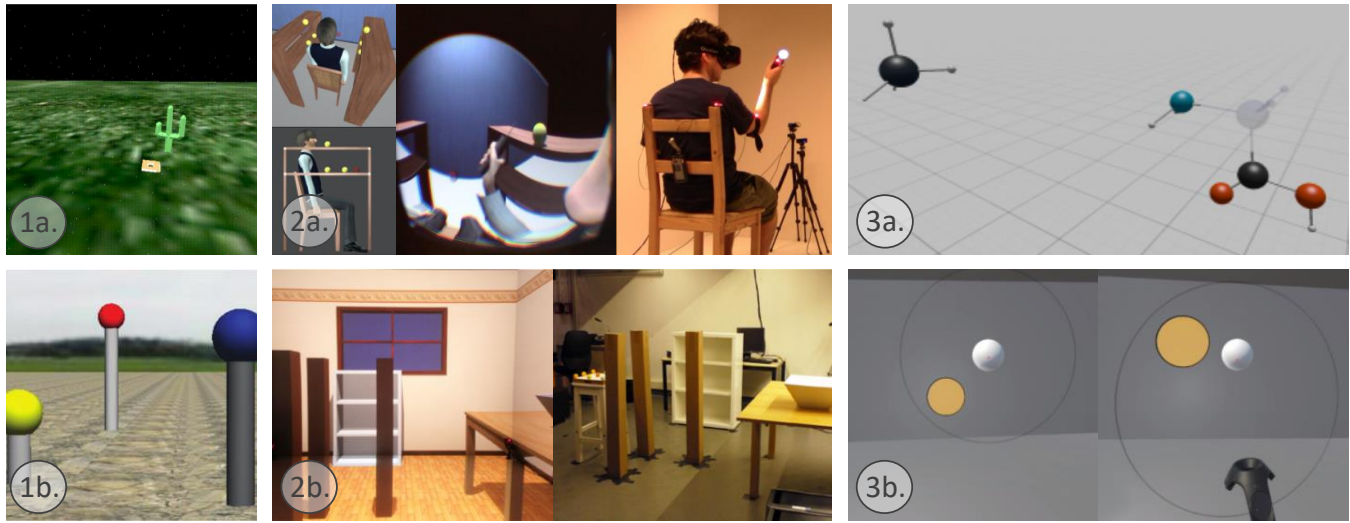
Figure 2: Examples of three types of virtual settings used in the studies: Outdoor scenery ((1a) from [86], and (1b) from [62]); replicas of the physical room with furniture ((2a) from [21], and (2b) from [17]); and plain default VR scenes ((3a) from [54], and (3b) from [47]).

(Criterion 1), two because they did not contain a user study (Criterion 2), and seven because they did not investigate object selection or manipulation performance (Criterion 3).

3.4 Phase 4: Data Set and Coding Process

The remaining 39 publications were included in the review[1]. These 39 publications consist altogether of 48 studies, each of which was coded separately.

Interrater reliability for coding the studies was difficult to establish, because many characteristics of studies were initially coded using open-ended text. To ensure that this was done in a consistent way, three authors all coded a randomly selected sample of five papers from our inclusion set. Subsequently, another author determined whether each pair of authors agreed on the 34 dimensions on which we coded the papers. That agreement was either Fully Agree (84% of cases), Partially Agree (6% of cases), or Do Not Agree (10% of cases). These percentages were taken to indicate good agreement, but we nevertheless further clarified our coding manual based on the observations of imprecise field descriptions. Our final coding manual consists of 36 fields[2].

We initially coded most fields as text fields, by writing or collecting quotations from the papers. Each co-author then took a set of fields for further data processing. Out of these, we coded the fields we could for quantitative analysis by using a fixed set of options or numerical inputs. The open text entries with no quantitative data were analysed qualitatively. These results are reported next as percentages of, and instances in, the 48 studies.

---

[1]These 39 papers are marked on the references of this paper, and a list of those is also available at https://vrevaluation.github.io
[2]The coding manual is available at https://vrevaluation.github.io

4 RESULTS

In this section, we report on the 48 studies (across the 39 papers in our sample) on object selection and manipulation in VR. Thirty of these studies focused on object selection (such as pointing or typing), and 18 focused on object manipulation (such as dragging, docking, rotating, or throwing).

4.1 Study Goals

We identified two primary goals for the studies in our sample. The majority of the studies (77.1%) aimed to compare input techniques, devices, or user representations. For example, the techniques evaluated include retargeting strategies [57] and pointing versus crossing selection [79]. The devices could be alternative controllers (such as the TORC device in comparison to an HTC VIVE controller [39]) or new tracking solutions (such as for finger tracking in comparison to a wand with an immersive cube [78]). User representation comparisons studied, for instance, the presentation of no avatar, an avatar hand, or a full-body avatar [8].

The remaining 11 studies (22.9%) focused on understanding object selection and manipulation performance for particular characteristics of immersive VR. The goals of these studies included describing basic perceptual phenomena (e.g., understanding how visual depth affects 3D target selection in VR [44]) or building models of performance (e.g., understanding pointing offsets and how to correct for them [50]).

4.2 Participants and their Expertise

The studies had an average of 18.8 participants (SD = 9.1), where 27.7% are women. All studies were conducted in laboratories, and none were crowd-sourced, making this number of participants closely aligned with the averages of human-computer interaction in general (20–30 participants [5, 15]).

Participants have mixed experiences with VR. Thirteen studies did not report anything about participants' previous use of VR, and several papers give vague or unclear descriptions (e.g., "some experience"). Twenty-seven studies provide data on expertise in the form of rating scales (e.g., "participants rated their experience in VR [...] on a scale from 1 (novice) to 5 (expert)" [72]) or binary (e.g., "Eleven participants had occasional VR experience, and one used VR daily" [69]). This data suggests that, on average, 53.5% of the participants have some experience with VR before participating in the studies. Other expertise in the content or domain was rarely mentioned and then it mostly involved 3D gaming or 3D software, such as CAD programs.

4.3 Tasks

Thirty studies (62.5%) evaluated object selection, and 18 (37.5%) object manipulation. Among selection studies, 12 used a task based on ISO 9241-400/411 (part 9) with multiple targets on a circle [e.g., 59, 65, 93] or with two targets [e.g., 44]. Typically, studies would state that they used the ISO standard in 3D, or used a 3D version of it (although no study referred to any established 3D version). Just five studies (10.4%), beyond those using the ISO-style pointing tasks, based their tasks at least partially on previous work. Three of those studies used a typing task on a keyboard with established phrase sets: one [72] used the Enron corpus [81], and two (in Yu et al. [91]) used the MacKenzie phrase set [46]. One further study used a pick-and-place task [12], which is based on a real-world peg-transfer task used in training and assessment tools for laparoscopic surgeons [76]. Another study used a 6-DoF object manipulation task [82], originally used by Zhai and Milgram [92], where a tetrahedron is positioned and aligned in space with another, similar tetrahedron. The remaining 31 (64.6%) studies used either new tasks designed for the present study or a task from the authors' own recent work.

Of the 18 manipulation studies, eight (16.7% of all studies) used a task that includes both translation and rotation. Such tasks involve moving and aligning an object with a target object (e.g., a cube in a similar but transparent target cube [89]) or docking an object with another (e.g., a key in a keyhole [39]). Five studies (10.4%) used just translation, for instance, docking an object without an orientation requirement (e.g., a cactus in a hole on the ground [86]), or positioning an object with uniform orientation (e.g., a generic sphere [26]). One study used a task of aligning the orientation of a house [80] (only rotation and not translation). We did not find studies that involved scaling. The remaining four manipulation studies used other tasks including throwing a ball [8], balancing a tray with balls on it from one table to another, navigating between poles on the way [17], and manipulating settings of a virtual car (such as adjusting a sun shield, a mirror, and opening a door [55]). In summary, for both types of study, tasks are rarely drawn from earlier work.

4.4 Physical and Virtual Settings

The physical and virtual settings play an important role in contextualising (or abstracting) selection and manipulation tasks. Using text and figures, we could determine the physical setting used in 89.6% of studies. All experiments are carried out in a laboratory, sometimes

Table 1: Display hardware used in surveyed studies

Display Hardware	% of studies
HTC Vive	34.7%
Occulus Rift	16.3%
CAVE	16.3%
Other HMD	16.3%
TV with Stereo glasses	10.2%
Samsung Gear VR	6.1%

resembling a workplace [e.g., 48, 60] or a living room [e.g., 17, 21]. Figure 2 shows examples of virtual and physical environments. Some studies attempt to mirror the physical and virtual worlds. For example, Debarba et al. [21] placed shelves in the physical world to match the virtual world so that the participant could interact with them during the study. In contrast, the laboratory environment is justified as being calm and controlled [e.g., 47, 48, 54]. If furniture was placed in the physical setting (48.9% of studies), a chair was always present, sometimes with a table and/or shelf (18.2% of studies with furniture). Participants typically had to stand (45.8% of studies) or sit (41.7%), while 8.3% of studies require walking to explore a scene. To complete a task, participants had to utilize their arm (66.7% of studies), their head (12.5%), only their hand (8.3%), or their full body (8.3%).

All studies describe or depict the virtual setting. One in eight studies replicated the physical setting in the virtual world, sometimes for better immersion and sometimes because physical elements, such as furniture, are relevant for the study (e.g., as Debarba et al. [21] described above). In 62.5% of studies, a room is built for the experiment, 20.8% use a scene with floor and sky stretching to the horizon (similar to the default Unity scene), 12.5% use an outdoor scene, and the last 4.2% place complex 3D models of cars or skeletons in the space. The standard room is empty, apart from virtual elements that are required for the task. Outdoor scenes range from grassy plains [86] to hilly landscapes [29, 61].

4.5 VR Displays, Interaction Techniques, and User Representations

The hardware used to render VR environments, and for registering pointing and manipulation, impacts the quality of immersion, precision of input, performance in tasks, and overall user experience. In this section we explore how studies have reported their VR displays, interaction devices and techniques, and user representations.

4.5.1 VR Displays. Table 1 shows the VR hardware used. While all studies report the manufacturer of the hardware used, only 51.2% report any of the capabilities of this display hardware. A breakdown of the capabilities that are described are shown in Table 2. Display resolution and field-of-view are reported in 29.2% and 27.1% of studies respectively, while refresh rate in only 14.6% of studies.

Interaction in VR environments often requires tracking of (a) the head or the body to determine the user's position and orientation in the environment, and (b) either the controller or the hands to determine the user's input. HMD tracking typically uses 'built in' capabilities (50.0% of studies) that are shipped with the device. A

Table 2: Attributes of display hardware reported in studies

Display Attribute	Reported (% of studies)
Resolution	29.2%
Field of View	27.1%
Refresh rate	14.6%
Size (only for CAVEs)	62.5%
Frames-per-second	10.4%

Table 3: Five most frequent combinations of feedback modalities in VR selection and manipulation studies.

Feedback modality combination	% of studies
Visual only	29.2%
Visual & haptic	18.8%
No feedback	16.7%
Not stated	12.5%
Visual & audio	8.3%

further 31.3% use custom external capabilities, while the remainder (18.8%) either do not report or do not require tracking. However, across all studies, only 8.3% provide any insight into the accuracy of this tracking. For example, Pham and Stuerzlinger [65] use the same external system for HMD and interaction device tracking, reporting "there were eight OptiTrack S250e, 250 Hz IR cameras, which were hung above the experimental area. The OptiTrack system was calibrated to sub-millimeter accuracy, which corresponds for the pen also to well below a degree of rotation error". We did not find studies that provided fully evidenced data for the accuracy, refresh rate, or lag of HMD tracking—in most cases this is "assumed" based on manufacturers specifications together with the correct execution of their setup instructions.

4.5.2 Interaction Techniques. We discuss two components related to interaction techniques below: how input was tracked and how feedback was presented.

(1) **Input devices and tracked body parts:** Across the 48 studies, 28 unique input devices were used. Bare-hand interaction was used in 16.7% of studies, while the remainder required the user to hold or wear a device. Of these devices, 83.7% were 'off the shelf' while 16.3% were either augmented or custom fabricated. The most commonly used device was the HTC Vive Controller (22.4% of studies).

Three types of input was made with these devices: (1) Mid-air gestures (67.3%); (2) Constrained 3D movement (e.g. using a haptic device [61], 10.2%) or; (3) Other devices including joysticks [86], touch surfaces [29], and gaze input [59]. Mid-air gestures were clearly the most studied as they provide dimensional movement that matches that possible in the virtual environment. To trigger selection, the majority (42.0%) of techniques used physical buttons on the device, 27.5% used direct 'touch' of objects (either from hands of controllers), and 11.6% used buttons external to the primary pointing controller.

For devices, 48.1% used built-in tracking (with 32.0% reporting some measure of accuracy), with the remainder (except one that did not report) using external tracking to capture device or hand/finger movements (30.8% report accuracy). In 10.4% of cases, tracking of the input mechanism was *different* in different conditions of the study; however 60.0% of these studies did recognise the potential impact this has on the comparative results.

(2) **Feedback on user actions:** A breakdown of the most frequently occurring modalities for user feedback in the surveyed studies is provided in Table 3. In the majority (64.7%)

of cases, at least visual feedback was provided, while haptic and audio modalities were far less applied and investigated.

4.5.3 Visual representation of the user and/or device. Across all studies, 30.8% of techniques provided at least a realistic representation of the user's hands and/or arms, 27.7% a realistic representation of the physical input device, and 30.8% provided a cursor or abstract representation of the user. To assist in pointing tasks, 21.5% of studies showed virtual extensions of the device (i.e., ray-casting). The remaining 15.3% did not represent the user in any way.

4.6 Experimental Design

Next, we analyse design choices made during experimental setup: independent variables, study design, and participant training. Dependent variables are discussed as part of the *Measures and Analysis* section.

4.6.1 Independent Variables. Independent variables describe the elements that experimenters manipulate to understand, for instance, performance. Across studies, we identify three types of independent variable: Interaction Techniques, Targets, and Tasks. Almost all studies (97.9%) varied at least one of these. Within one study, multiple types can occur; some studies manipulated both technique and target (e.g., Pham and Stuerzlinger [65] varied both the controller and target size).

(1) **Interaction Techniques** were varied in 79.2% of studies. This category includes pointing, selection, and manipulation, techniques, for example, when a study compares a novel technique with a baseline [e.g., 47, 60, 69]. Some studies varied the controller as the technique-variable; Lee et al. [39] compared a novel haptic device to the well-known HTC Vive Controller. This category of variables is commonly changed in both manipulation and selection tasks (88.9% and 73.3% respectively).

(2) **Targets** were varied in about half of the studies (56.3%). This includes varying the size, position, and density of targets. More than one of these properties are varied in 55.6% of studies varying target-variables. Most often the target position is varied (70.4%), followed by size (55.6%) and density (33.3%). Tu et al. [79], for instance, varied target position and size during their experiment. Two out of three studies using selection tasks used some target-variable, while only two out of five studies using manipulation tasks employed target-variables.

(3) **Tasks** were varied the least commonly (22.9% of studies) but often the tasks carry much variation in themselves. For

Table 4: An overview of study setups with low complexity (one independent variable), medium complexity (two or three independent variables) and high complexity (more than three independent variables).

Complexity	% of studies	Focus	Examples
Low (1 variable)	8.3%	Techniques	Studies comparing novel techniques with a baseline [69, 72, 80, 91].
Medium (2–3 variables)	72.9%	Techniques	Most often two techniques are compared, while varying target properties or the overall tasks [e.g., 39, 60, 87], only a few [e.g., 55, 59] use more than two techniques in one study.
		Tasks	Studies conceal experiment complexity in the task design, such that varying target properties are covered in the designed tasks [e.g., 44, 53, 54].
		Targets	Studies vary the target, when for instance testing Field of View [21] or display restrictions [44].
High (4+ variables)	18.8%	Targets	Study designs become complex, due to varying target properties (e.g. size, position, orientation, and density) [65, 78, 79, 86, 89].

example, Mendes et al. [53, 54] had six levels of task that require different amounts of translation and rotation.

4.6.2 Study Design. Most studies (92.7%) use a within-subjects study design, while only 8.3% use a between-subjects design. In 62.5% of studies counterbalancing is reported; 29.2% of these make use of a Latin square design. It is possible to reconstruct the number of trials in 56.3% of studies from the number of independent variables and repetitions or by direct reporting. There is a large diversity in the number of trials per participant across studies (median: 80, min: 6 [2], max: 1800 [79]). Less then half (43.8%) of the studies report the time participants spent completing the study. The duration ranges between 10 and 120 minutes (median: 45 minutes).

Table 4 shows the complexity (as the number of independent variables) of the reviewed study designs. The combination of many independent variables can be challenging to analyse thoroughly, while few independent variables might not fully uncover the potential of an interaction technique. Complex designs aim to vary target properties (see 'High' row). For example, Tu et al. [79] varied five task types, three target depths, five target widths, and two target distances. Studies with medium complexity vary a range of factors, not only techniques, tasks and targets, but also sensory feedback [3, 23] or virtual environment [50, 55]. Studies with medium complexity sometimes also vary many levels of a variable, for instance, by comparing six techniques [55] or six tasks [53, 54]. Half of the studies with low complexity were conducted in conjunction with another, more complex study. All of the low complexity studies reviewed use interaction technique as their single independent variable [69, 72, 80, 91].

4.6.3 Training. Many of the reviewed studies (64.6%) report training their participants. Often this involved familiarising the participant with a technique, the environment, and/or the HMD [e.g., 3, 25, 74, 91]. Of the studies that report training, 64.5% trained their participants directly on a similar task or trial than used in the study, 41.9% on the interaction techniques and devices used in the study, and 6.5% asked the participants to, for instance, "try the environment for a few minutes" [80]. The participants were most often allowed to try the task or trial as long as they wanted, typically until they felt comfortable with it (32.3%). In other cases the participants were allowed to train for a fixed maximum amount of time (25.8%, mean: 5.9 minutes) or try a fixed amount of trials (19.4%, mean: 7.3 trials). Some studies (22.58%) reported that participants were allowed to train, but did not detail the amount of time or number of trials.

4.7 Targets

Targets form the core of the study setup for object selection and manipulation. This section is organised by using the parameters that LaViola et al. [38] identified for canonical tasks. We combine these for selection and manipulation tasks into the following parameters: target shape and size or required precision of positioning, distance to target or translation distance, target arrangement or direction to target, the number of targets to be selected, and other objects (distractors) around the target. These parameters of target setups are summarised in Table 5.

4.7.1 Target Shapes. Most studies (83.3%) used abstract target objects. For example, about half of the studies (55.6%) used spherical targets. In 2D and 3D, this allows uniform width of the target. Two studies (5.6%) used a single point as a target, measuring accuracy as the distance selected from the target (e.g., using a cross-hair target [50]). Five studies used cuboid targets. Some of these were in 2D as squares or rectangles, such as keys of a keyboard. Cuboids and other geometric shapes were also employed in object manipulation tasks (e.g., a house-like shape in [80] and a tetrahedron in [82]), where they support object orientation, unlike spheres.

Eight studies (22.2%) used realistic object shapes such as a cactus, but a simple target shape such as a hole in which to dock it [86]. Some used a target shape that appeared more complex, but could be reduced into simpler factors. For example, a keyhole consists of a hole and a required orientation for the docking object [39], and a more complex molecule [54] includes many components, but only one (in other words, a single target position) on which a carbon component can be docked.

4.7.2 Target Sizes. The papers reported target sizes in SI units (such as millimeters, centimeters, or meters), in angles (degrees),

or in both. The target sizes express diameter, or another kind of a width from the approach direction. For example, three out of the 12 studies using the ISO task included target widths of 1.5cm, 2.5cm, and 3.5cm (e.g. [6, 44, 65]). The other nine ISO studies used a set of distinct target widths and sizes. These included, for instance, sphere diameters ranging from 2.9 to 7.5cm [3], discs ranging from 8.5mm to 612mm [79], and three target widths of "1°, 2° and 3° of visual angle" [93]. Other selection studies (wherein the targets were not laid out based on the ISO task) used a variety of similar sizes. In the three typing studies, one reported the size of the entire keyboard to be 6 meters wide (and equaling to a reported 33.4° of the FoV) [91], while the two others did not control the size, letting it be adjustable to each participant [72].

Target sizes are used for controlling the difficulty of the task. The smaller the target, the harder it is to successfully select. Therefore, the target size is used for determining successful trials (e.g., to measure speed in those as in the Fitts paradigm) and errors. However, the target size alone cannot specify errors unless the cursor size (or size of the object that is translated) is also reported. In many instances, it is not. For example, in the ISO tapping tasks, only one study reported the cursor size in units equivalent to targets (a 1cm sphere [44]), as well as that the cursor was required to be completely inside the target for a successful selection. Other selection studies also reported the cursor size sparsely, with a few exceptions such as Mardanbegi et al. [47] who report using a target of 10° of the "visual angle" and a cursor of 15° that is overlaid with the target.

Among manipulation studies, only a few mentioned the equivalent target sizes or ranges of precision both in size and in orientation. For example, Mendes et al. [54] reported that their docking error boundaries are less than 1mm for position and 1° for orientation and Yang et al. [89] report that the thresholds for aligning cubes were 5mm and 0.1 radians. One manipulation study measured accuracy instead of using target sizes for errors [48]. All other manipulation studies left the required target size or precision unstated, or simply conveyed the count of distinct sizes in the types of targets, such as 'easy' and 'difficult' [26], or "seven differently sized, weighted, shaped and colored parts" [23].

4.7.3 Target Distances. The effective target distance depends on the input method. For example, ray-casting techniques allow a much larger distance ranges than direct touch with no motion gain. Therefore, the studies used a large variety of target distances.

In direct (virtual) touch, the smallest distance between targets was 10cm [44]. In contrast, one of the three studies on typing stated their keyboard width to be 6 meters and also stated that the keyboard was 10m away from the user [91]. This 6 meters was the largest distance between the targets among methods based on ray-casting.

From the studies based on the ISO tapping task, seven reported their design's ID range. The ID ranges are dependent on the distance and size of the targets. The reported ID ranges were 2.81 to 3.46 bits in [59], 1.94 to 4.39 bits in [6] and [44], 1.58 to 7.01 bits in [79], 2.5 to 3.5 bits in [3], and 1.31 to 25.74 in [93]. These studies used 6 to 9 levels of IDs per study. The remaining five papers who stated following an ISO or a Fitts design [e.g., 62] did not state their ID range.

Table 5: The common target parameters and the ranges of values used across the sample of studies. The units and values here are reported as they are in the sample of papers.

Target parameter	Common types, ranges, and usage (% of studies)
Shape	Sphere (55.6%); realistic (22.2%); cuboid (10.4%)
Size	Ranging from 8.5mm to 612mm, and from 1° to 3° of the FoV
Distance	Up to 6 meters and 33.4° FoV, IDs ranging from 1.31 to 25.74
Arrangement	Random (20.8%); circular (18.8%); grid (12.5%)
Direction	Single or pair of targets (6.3%); layout in 2D (14.6%); layout in 3D (79.2%)
Number of targets	Range from 1 to 120 individual targets
Distractors	Used in 35.4%

Seven of the manipulation studies reported some parameters of the target distance ranges. For example, the center-to-center distance between pegs in the peg-transfer task was 6cm [12], and the key hole was positioned within ranges of "x-axis from 180 to 360 degrees, y-axis from 0 to 30 degrees, z-axis from 60 to 60 degrees" [39].

4.7.4 Target Arrangements. Targets were arranged in diverse ways across the studies. Figure 3 presents examples of these. Nine studies (18.8%) used a circular layout (1a-b) in Fig. 3, six (12.5%) laid out the targets on a grid, three (6.3%) used a single target or a pair (as in a reciprocal tapping task, (2a-b) in Fig. 3, and three used a keyboard layout (1c) in Fig. 3.

Ten studies (20.8%) reported using a random target layout. These in turn applied a variety of constraints, such as laying out the random targets on a surface (e.g., on a 2D virtual pad [43]), or within a constrained 3D space (e.g., inside a cube [78], 3a in Fig. 3). The reporting practices of the constraints used for random arrangements varied. For example, Tran et al. [78] report that targets were "distributed within a 35cm × 25cm × 30cm virtual box", whereas some merely stated that the positions were random (e.g., "15 predefined random sphere positions distributed uniformly within the workspace" [62]).

The remaining 17 studies (35.42%) had other, unique target layouts, such as positioning tennis balls on six shelves of a virtual room [21], a key in a keyhole [39], a carbon component on a model of a protein compound [54] or a cactus in a hole on the ground [86].

4.7.5 Directions to Targets. The three classes of target arrangements presented in Figure 3 influence movement directions to targets in distinct ways. Eleven (22.92%) selection studies used targets only on the vertical plane in two dimensions, such that the targets were laid out similar to a display surface in front of the participant [e.g., 65, 72], and the depth of targets remained constant. A further four selection studies used targets only on the horizontal plane, such that the height of the targets remained constant, but the depth and the lateral distances varied [e.g., 44]. Most (18.8%) studies using
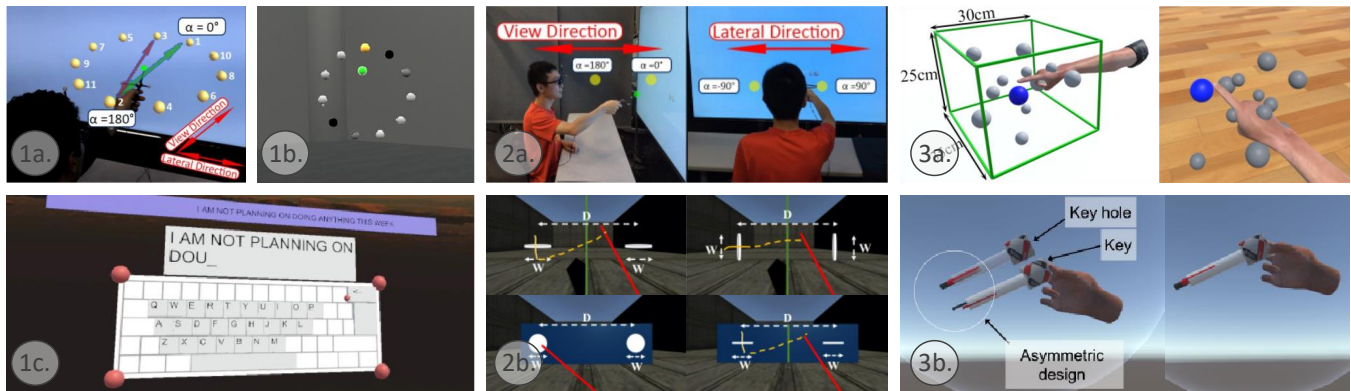
Figure 3: Examples of target arrangements that lead to three types of movement directions to the targets: 2D layouts ((1a) a horizontal ISO task layout from [44], and (1b) a vertical from [65], as well as (1c) a vertical 2D keyboard from [72]); Target pairs ((2a) for tapping on frontal and lateral axis from [44], and (2b) for crossing on lateral axis from [79]); and (3) 3D layouts ((3a) with a random arrangement of multiple targets for selection from [78], and (3b) with a random position of a single target for manipulation from [39]).

the ISO task used these types of 2D layouts of targets on a circle (1 in Fig. 3)

Another way the ISO studies laid out targets was to use two targets on a single axis (e.g., on the lateral [6] or frontal axis [44], 2 in Fig. 3). Regardless of their statements indicating the use of a 3D version of ISO, none of the ISO studies in our sample used targets in all three dimensions (in contrast, for instance, to the spiral layout in [67]).

In the 30 selection studies, 10 (20.1% of all studies) used targets laid out in all three dimensions (3 in Fig. 3). Seven manipulation studies describe directions about their object layouts in three dimensions (e.g., "There were three directions of movement: to the left, towards the user, and away from the user" [86], or "The trials were a mix between horizontal, vertical, and diagonal movements." [43]). However, none report the number of targets along each, nor specify (or report in case of a random distribution) these directions or target locations numerically. From the remaining 11 manipulation studies, it remains unclear which dimensions they include.

4.7.6 Number of Targets. The number of targets was stated in 70.1% of the studies. These studies used an average of 16 target locations per study, ranging from a single location up to 120 unique locations in a study. Among the 12 studies using the ISO task, some used only two targets [e.g., 44], and some used 11 [e.g., 65], 13 [e.g., 59], or 15 [e.g., 93] targets on a circle. The remaining 14 studies did not state how many (distinct) targets there were. Some of these studies described only the complete object layout, not reporting how many and which of these objects were used as targets, and which as distractors.

In studies using keyboards, the complete set of target locations are based on the characters in the phrase set that required transcription. These studies did not report the number of the keys on the keyboard, nor the rates of occurrence of each character in the phrase sets (i.e., it is unknown if every key gets pressed at least ones). Two studies report that these phrases were randomly chosen [72, 91].

4.7.7 Distractors. A distractor is an object that can be selected in addition to the current target. Seventeen of the 30 selection studies (35.4%) included distractors. For instance, in an ISO tapping task, these could be the other spheres on the circle beyond the current target, if all of the spheres are presented at the same time. The remaining 13 selection studies (27.1%) presented only the current target object(s) at any given time and therefore did not include distractors.

Some of the ISO studies implemented distractors by showing all of the targets on the circle and simply highlighting the target to be selected next [e.g., 65], whereas a further 10 used no distractors and showed only the current and next target [e.g., 44] at any point in time. This count also includes the typing studies, however it is unclear how many distinct keys were included in the randomly selected phrase sets which were used as tasks.

Of the remaining seven studies using distractors, two studies used distractors that were external to the task (i.e., those objects were never used as target objects). Both of these studies used complex distractor and occlusion designs. In theses studies the task was to select large numbers of objects in point clouds [74], and a complex 3D model (e.g., a ship, a DNA structure, and a human thorax) in a highly occluded setting [2]. From these studies, it is unclear how the distractor settings could be applied with other object types or arrangements.

4.8 Measures and Analysis

4.8.1 Dependent variables. Table 6 shows an overview of the frequently observed dependent variables, as well as examples of how these are measured. Task completion time and errors are most frequently measured, both in above 70% of the studies. There is a large overlap in these two sets of studies, as 52.1% of studies measure both task completion time and errors.

Task completion time most often describes the time a participant uses to complete a task or trial in seconds. Studies with selection tasks measure the time a participant needs to select the next target in 53.3% of occurances, while studies with manipulation tasks

Table 6: An overview of the dependent variables used in the reviewed studies, organised by following the categories in Hornbæk [33]. * _Words Per Minute_

Category	% of studies	Measure	Units	Usage examples	Example studies
Time	75.0%	Completion time	sec	Time to complete a task	[25, 47, 82, 87]
		Time in mode	sec	Manipulation time; selection time	[2, 21, 29, 80]
Accuracy	70.2%	Error rates	%, #	Percentage false selections; number of completed tasks; number of false selections; number of errors on the way to task completion	[2, 26, 65, 69, 79]
		Spatial accuracy	cm, mm, °	Distance from target; orientational error	[43, 48, 50, 53]
Movement	22.9%	Distance	cm, m, °	Movement/input path; head and gaze movement; travel distance	[57, 60, 62, 69]
Task dependent	16.7%	Obstacle collision	#	Number of virtual collisions; number of physical collisions	[12, 17]
		Input rate	WPM*	Words per minute entered	[72, 91]
		Other	cm	Deviation from optimal path; amount of movement correction	[3]
Throughput	10.4%	Throughput	bits/s	Fitts' law throughput	[3, 6, 44, 65]

measure the time a participant manipulates virtual objects. Veit et al. [80] measures not only manipulation time, but also the time a participant uses "coarse" or "fine" manipulations to complete the given task.

Error measurement is diverse, as it is dependent on the task and the independent variables. For instance, Mendes et al. [54] measured both how far the final position of a virtual object was from the true position and whether the object is within an acceptable margin in terms of position and orientation. Pham and Stuerzlinger [65] employed a selection task and measures the number of missed selections as a percentage of the total number of selections. Both Frees and Kessler [26] and Wang and Lindeman [82] measured the number of completed trials, since participants could fail to complete a trial due to frustration or a fixed time limit.

Apart from these two performance measures, 22.9% of studies measure movement time and movement distance. Movement time is measured as "the time the user spends moving a pointing device" [59], and is often mentioned in connection with the Fitts' law prediction of movement time [6, 44, 62]. The movement distance measure is defined differently across studies: Sidenmark and Gellersen [69] measures the participants head and eye movements, Montano Murillo et al. [57] the participants physical and virtual hands path lengths, Chapoulie et al. [17] the participants movement path across a room, and Park et al. [60] the length of a cursor's trace.

One in six studies report a dependent variable, that is task specific. For example, as Speicher et al. [72] and Yu et al. [91] study the use of virtual keyboards, a dependent variable in their studies is

Words Per Minute (WPM), that describes how fast a user can type. Both Chapoulie et al. [17] and Brickler et al. [12] measure the number of times a participant collides with an obstacle, might that be in the virtual or physical world. Veit et al. [80] presents a new measure that describes "the proportion of time users manipulate one, two and three DOF at the same time". Other task specific dependent variables include "the number of times the target object changed its selection status prior to confirming the selection" [2] and "under/overshooting distance along the vector defined by the positions of the last target and the current target being selected" [3].

Of the reviewed studies 10.4% report "measuring" or "computing" throughput as a dependent variable [3, 6, 44, 65]. Ariza et al. [3] justify the use of throughput as dependent variable, since it "[incorporates] both errors and time into an overall estimate of performance".

In 10.4% of studies it is not clearly stated which dependent variables are measured during the experiment, but they only hint at "performance" as dependent variable.

4.8.2 Questionnaires. Questionnaires are a frequently used approach to gather qualitative data. In the reviewed studies, 70.8% employ this method either during the experiment (e.g., after each condition), after all trials are completed, or both. Table 7 shows seven categories of questionnaires, how often these questionnaires were used, and what they intend to measure. The 34 studies administering questionnaires use 53 questionnaires in total, where close to two in five studies (38.2%) used multiple types of questionnaires (e.g., [3, 43, 72]). Of all questionnaires, 44.1% are based on previous

Table 7: Overview of questionnaire use in the reviewed study, organised by following the categories in Hornbæk [33].

Category	Measures	% of QAs.	% based on prev. work	Examples of prev. work	Example studies
Satisfaction	Preference; ease-of-use; fun factor	35.9%	0.0%	-	[17, 25, 47, 53, 65, 74, 82]
Usability	Usability; perceived usability	17.0%	77.8%	System Usability Scale [13]	[23, 48, 61, 72]
Simulator Sickness	Nausea; dizziness; unpleasantness	15.1%	50.0%	Kennedy et al. [36]	[3, 29, 72, 93]
Presence	Feeling of "being there"; feeling of control	11.3%	100.0%	Slater-Usoh-Steed [70]; Witmer and Singer [88]	[3, 39, 72, 89]
Workload	Mental work	9.4%	100.0%	NASA-TLX [31]	[29, 50, 59, 72]
Fatigue	Physical strain on arm, eye or head	7.6%	0.0%	-	[25, 43, 57]
Immersion	Body ownership; embodiment	3.8%	100.0%	Gonzalez-Franco and Peck [28]	[21, 39]

work or on established questionnaires, most notably (raw) NASA-TLX [31], Slater-Usoh-Steed [70] and System Usability Scale [13]. No study that administers a self-developed questionnaire reports a complete list of asked questions.

4.8.3 Analysis methodology and results. Before analysing the data collected in the studies, 31.9% of studies describe pre-processing by removing or aggregating data. The removal of data is typically driven by the desire to eliminate outliers, which are either identified by a mathematical statement (e.g., data points that are more than two standard deviations from the mean [57, 93]), by wrong selections (e.g., "double-clicks" [44]), or by incorrect sensor data [62]. In the reviewed studies between 2.4% and 2.8% of collected data was discarded due to being classed as 'outlier' data. Zielinski et al. [93] removed all data connected to one participant "as he did not follow the instructions for the selection task", leading to 5.6% of data being removed. When studies aggregate data, most commonly (14.9% of studies) throughput is computed [e.g., 6, 44, 59]. Others remove dwell time from completion time [59], compute the completion time [82] or error [8] of each participant, combine two dependent variables in two trials into a combined score [23], normalize the completion time "to create a normal distribution" [86], or convert a dependent variable to scale independent quantities [74]. Apart from removing data based on standard deviations and computing throughput, no two pre-processing steps were the same.

Most studies (79.2%) conducted repeated measures ANOVAs to determine whether the effect of independent variables on dependent variables was significant. Other significance tests mentioned in the studies are Friedman tests (16.7%), Student's t-test (8.3%), and Kruskal-Wallis test (2.1%). Some studies used Wilcoxon signed-ranks test (14.6% of studies) and the Tukey–Kramer method (2.1%) as post-hoc tests. To counteract the multiple comparisons problem the Bonferroni correction (20.8%) was used, while studies mention

the Greenhouse-Geisser (8.3%) and Huynd-Feldt (2.1%) corrections to be used for correcting for lack of sphericity.

Studies using a Fitts' law style task compute throughput as a function of effective index of difficulty and movement time according to the ISO specification. Many studies use throughput only as a means to compare two techniques and check for significance using ANOVAs (58.3% [e.g., 6, 69, 93]). The remaining 41.7% of studies describe the goodness of fit (R^2), but here the differences were in how the relationship between movement time, target distance, and target width are expressed: the original Fitts' model formulation [24] is used in 16.7% of ISO studies, while the Shannon formulation [45] and the Shannon-Welford variation [83], are used in 41.6% of ISO studies. Machuca and Stuerzlinger [44] proposed a new formulation for 3D object selection, that is, in addition to the previously mentioned factors, accounting for the "change of target depth". This new method was compared to the Shannon-Welford formulation.

Close to third of studies (31.3%) do not analyse one or more dependent variables that are measured during the study. We count variables as analysed if the measure is directly used in a study's analysis or if it is part of an aggregate (e.g., throughput is computed from time and accuracy) that is analysed. Nearly half (46.7%) of these variables are time, 33.3% are accuracy, and 20.0% task related measures.

5 RECOMMENDATIONS AND A CHECKLIST FOR STUDIES

Based on the reviewed papers, we formulated a set of guidelines for researchers planning object selection and manipulation studies in VR. Deriving such recommendations from what researchers currently do is difficult; many philosophers consider this impossible to do validly (e.g., Moore's naturalistic fallacy, Hume's is-ought

problem). Pragmatically, however, learning from best practices in the literature seems a reasonable way to improve VR research.

We present 10 recommendations. These recommendations are developed with the joint goals of supporting better replicability, allowing researchers to build on previous work more easily, and facilitating a more straightforward comparison between studies. Therefore, these recommendations are mostly for studies that seek to compare the usability of interaction techniques or to build fundamental understanding of object selection and manipulation in VR. The 10 recommendations for studies are:

(1) Define the goal of the evaluation: *Choose speed or accuracy as the main dependent variable.*

(2) Estimate the required number of participants with power analysis: *Use a minimum of 20.*

(3) Strive for high control with a simple design: *Use a maximum of three independent variables.*

(4) Use low-level tasks in evaluations: *point and select, or translate and rotate.*

(5) Control the target distance in trials: *Use a fixed starting position.*

(6) Control the target size: *Use spherical objects for selection and polyhedrons for manipulation.*

(7) Aim for a wide and representative range of target sizes and distances: *Use at least three levels of each.*

(8) Include a range of movement directions: *Place targets across all three dimensions.*

(9) Control the physical setting in the study: *Use a fixed user position in space.*

(10) Control the virtual setting in the study: *Use a generic virtual environment.*

The recommendations are organized to cover the main components of empirical studies that researchers should consider during the planning of a study. This set is by no means applicable across all study types; for example, when studying domain-specific techniques or phenomena, alternative methodologies may well be more suitable. The guidelines are also not intended to be fully comprehensive. We identified these as the most inconsistently conducted elements in the current literature and improving them has the largest possibility for positive impact. We believe this set is a good starting point for helping the field to progress.

We also provide a checklist in the appendix, which offers a set of associated questions relevant to both the planning and reporting stages of studies. Both the recommendation set and the checklist are published as open source[3]. We call upon the community to contribute to their development as the field matures and technology and methodologies advance.

Next, we provide the rationale for the inclusion of each of the 10 recommendations, examples of the practices from the reviewed studies, further readings, and descriptions of their trade-offs with other methodological ideals.

(1) **Define the goal of the evaluation.** We recommend deciding early on whether the goal of the study is to evaluate speed or accuracy in object selection or manipulation. The main dependent

---

[3]https://vrevaluation.github.io

variable should clearly follow from this. The purpose of this recommendation is to either eliminate errors from analysing speed (as in the Fitts' paradigms), or to focus on analysing accuracy.

If the goal is to investigate speed (i.e., task completion time) in selection tasks, we recommend using the size difference between the cursor and the target object as the error threshold. It should also be required that the first is completely inside the second for successful task completion. If the cursor is a single-pixel cursor, the target size gives the error threshold (as the width does in the ISO task). If the task is a manipulation task, we recommend using thresholds for object placement (e.g., the thresholds for aligning cubes can be expressed as X mm and Y degrees, as in Yang et al. [89]). That way, the speed can be analysed from (successful) completion times either after a selection or after the threshold of manipulation is met.

If the goal is to measure accuracy, we recommend a free range for possible end-positions, confirmed with some selection trigger. Here, all selections are accepted, and the distance (whether of a single-pixel cursor from a single-pixel crosshair target [50], or the angular difference along the three axes of rotation from the target object's rotation [48]) is measured and analysed as a continuous accuracy variable.

Finally, while speed and accuracy should not be combined in the goal of an evaluation, we recommend doing so in the analysis when it is possible and useful for insights. This can be done by analysing throughput with effective IDs [71]; note, however, the concerns that for instance Stuerlinger and Teather [75] presented for doing this with non-spherical hit distributions in 3D.

(2) **Estimate the required number of participants with power analysis**. We recommend estimating the required sample size with a priori power analysis. If this is not possible, we recommend using at least 20-30 participants. The average number of participants per study in our sample was 18.8, whereas in general in HCI, it is 20–30 [5, 15]. We should strive to reach at least that range. We also recommend reporting effect sizes with statistical tests as this helps others to conduct a priori power analyses to better estimate the required sample size.

Including expert participants can reduce novelty biases and help move towards assessments of VR as a mainstream technology. We recommend the inclusion of expert participants, as VR has strong novelty effects, especially on subjective measures. VR experience can further help with reducing a training time also for canonical tasks like pointing. To measure expertise, we recommend using objective scales, such as options with how often related technology is used (e.g., daily, or N times a week as in Sidenmark and Gellersen [69]) rather than reflective ones without a baseline (e.g., "rate your expertise from 1 to 5" [72]).

(3) **Strive for high control with a simple study design**. We recommend using a maximum of three independent variables or factors. Nearly all studies (92.7%) used a within-subject design, which we also recommend because it helps decrease the influence of interpersonal variability on the performance in low-level tasks. However, counterbalancing the order of study conditions in a within-subject design becomes tricky when there are many conditions (again, a simple design helps). Yet, counterbalancing is usually necessary to decrease learning effects.

Clearly define the independent variables and their levels in the study. When there is more than one independent variable or group, we recommend reporting them and their levels in the N × M × P format. It is essential to include all the factors varied. We noticed that when following this style only partially the study designs are hard to follow and the analysis methods remain unclear (e.g., "for each interaction technique, there were 24 conditions (2 × 2 × 3 × 2)" [86], there were actually two techniques, giving 48 conditions in total).

However, if a study design results in many conditions, two approaches may help simplify the design. First, consider combining target distances and sizes into different IDs: you can include targets at different depths, for instance, by increasing the target size or decreasing the distance so that it matches the IDs (see an example in Tu et al. [79]). Second, consider decreasing *device* × *technique* combinations by holding either constant. For example, control the device type as an independent variable when your study is about a technique and vice versa.

(4) **Use low-level tasks in evaluations.** When the goal of a study is to compare interaction techniques, we recommend using low-level tasks: pointing and selection, or translation and rotation. Such tasks help to generalise the benefits of interaction techniques to higher-level tasks: the performance in those is always dependent on the lower-level components (as LaViola et al. also suggest [38]). Additionally, using lower-level tasks helps isolate and gain insights into the tested techniques' particular strengths and weaknesses.

For selection techniques, the lower-level tasks are pointing and selection. We recommend separating these in the study design. To do this, the study design can fix on of them as a constant. For example, selection can be made via a button press when studying pointing techniques, or by using a single pointing technique (e.g., ray-casting as in [6]) when studying the effects of a selection method on accuracy. If these low-level tasks are combined, we recommend designing dependent measures such that the performance in the two sub-tasks can be separated. Treating both as independent variables is often unnecessary and results in a complex study design (recommendation (3)).

For manipulation tasks, we recommend including at least translation and rotation tasks. We recommend performing the translation tasks with a target object collocated in space. This target object can be a docking point (e.g., Lee et al. [39]), or a transparent version of the object in which it needs to be co-located (e.g.,Yang et al. [89]).

If the task includes rotation, the methods used for translation apply, with suitable thresholds for accurate placement or successful co-location. Alternatively, rotation-only tasks can use a dislocated duplicate model object as a target (e.g., a house [80]). This is particularly useful when accuracy instead of error is measured (not to convey assumptions of successful orientation with visual co-location). In general, we recommend the first approach with co-located targets to prevent other factors, such as size or depth perception from influencing performance in the translation or rotation task.

(5) **Control the target distance in trials.** Object selection and manipulation tasks can be discrete or serial. Discrete tasks have a particular starting position for the task. In contrast, a serial task consists of a sequence of targets with the starting position for the next target being the previous target (see, e.g., the two schemes for implementing the Fitts paradigm in Soukoreff and MacKenzie [71]). We recommend using discrete tasks, as they are simpler to implement. For example, the task can always be started at the same point in space and in the same relation to the target arrangement. Thereby, it is easy to account systematically for variation in the resulting distances of targets, movement directions toward them, and so on.

A serial task is more complex to implement if a circular layout (as in the ISO tapping task) is not used. This is because systematic target distance and size variation, as well as possible distractor targets on the optimal motion path, need to be carefully designed and controlled. Most studies that did not use the ISO task used a discrete task. Those that did use a serial task outside of ISO, randomised the target layout instead of using a systematic control of distance—and then failed to report the target distances resulting from randomisation. The recommendation of a discrete task and starting position holds for selection and translation and for rotation (that is, always starting from the same object orientation).

(6) **Control the target size.** We recommend using spherical targets to enable clear size control using the sphere's diameter (such as for width in Fitts' law studies). This means there is no need to orient targets in the selection task according to the movement direction. We also recommend using either a single-pixel pointer or a fixed-sized cursor across all tasks in a selection study. If a cursor is used, its size should be reported to allow correct calculation of errors (the method for determining errors should also be reported). We recommend defining errors as selections when the pointer is not completely inside the target object.

With manipulation tasks, we recommend using color-coded cuboids or other polyhedrons that express the orientation unambiguously (e.g., tetrahedrons as in Wang and Lindeman [82]). We also recommend using the same object shape both in the manipulated and target object (e.g., both cubes instead of placing a sphere in a cube [43]). This helps to determine the required accuracy for a successful performance (e.g., if a successful translation or rotation is determined by docking an object of size d inside an object sized $1.5d$ [47]).

(7) **Aim for a wide and representative range of target sizes and distances.** Soukoreff and McKenzie recommend using a wide and representative range of ID values for pointing device evaluation [71]. They suggested using ID values ranging from 2 to 8 bits. We recommend taking that as a guideline, but extending the general recommendation of a wide and representative range to target sizes and distances in general, and in both object selection and manipulation. We recommend using at least 6 levels of IDs (or at least 3 distances and 3 sizes). The studies with ISO tasks in our survey that used 6 to 9 levels of IDs, most commonly used three sizes and three distances. However, as mentioned above, the studies using ISO tapping tasks did not arrange targets across all three dimensions.

None of the studies with "true" 3D target arrangement in the sample included 3 × 3 levels in their design. They either did not report distances (and had random targets), or varied those only at a maximum of three levels but did not vary the size. Therefore

we maintain our recommendation at a modest minimum of three levels but with an ambition to include that for both distance and size variables.

We recommend using euclidian distances for selection and translation tasks, and degrees for rotation tasks. We recommend reporting these in SI units (such as meters) and angles in degrees or radians, and in both where possible (e.g., sizes and distances also as angular degrees, such as in Tu et al. [79], but only as absolute ones and not relative to, for instance, FoV which might deviate depending on the used headset).

(8) **Include a range of movement directions**. We recommend placing targets across all three dimensions. The value of a selection or manipulation technique is hard to generalise from fixed, two-dimensional planes, because immersive virtual environments inherently are three-dimensional spaces, and therefore the objects of interest are likely located across the three dimensions. The performance in object selection and manipulation also varies across different directional motions: depth control is more difficult than lateral axis control (such as on a vertical plane). For example, performance in tapping two targets at different depths in front of the user is worse than tapping two targets on the sides with the same width and at the same distance [44]. Therefore, it is important for experimental comprehensiveness to include targets across all the three dimensions. The designs in our sample of studies also agree with the importance of this: over half (58.8%) of the studies that did not restrict themselves to the standard 2D ISO tapping task or involved typing on keyboards laid out targets across three dimensions.

(9) **Control the physical setting in the study.** We recommend using a fixed standing or sitting position in the physical space. Combining walking with object selection or manipulation adds a completely distinct task to the study. For example, LaViola et al. [38] and many papers about interaction techniques for VR (e.g., [9–11, 30, 52]) treat walking as a separate task.

(10) **Control the virtual setting in the study**. We recommend using a generic virtual environment, but with depth cues from shadows and from the surroundings, such as the ground or the walls. Depth cues are important, but realistic settings are difficult to compare with all possible distractors. Consider if your task is too specialised, if you feel a need to design a complex environment.

6 RESEARCH DIRECTIONS

In addition to recommendations for studies, we next discuss higher-level research directions in the study of selection and manipulation techniques. The aim is to use the material we reviewed to highlight open issues in evaluation methodology, the techniques we develop, and our empirical work.

6.1 Techniques for Selection and Manipulation

The studies in the sample contain many original and inspiring techniques for selection and manipulation. Nevertheless, as judged from these papers and not the larger literature on VR, a few neglected areas warrant discussion. Work on *multimodal selection and manipulation techniques* could be much more substantial. Feedback is

predominantly visual only; studies on the impact of state-of-the-art haptics on, say, manipulation performance were absent from the sample.

We discussed user representations in techniques for selection and manipulation; innovative work on *designing useful and pleasant user representations* is another promising direction. Earlier work has suggested that user representations are linked to body ownership [49].

We also expected to see the issues of depth dealt with in more new techniques. It is not, even though depth perception is poor in VR and significantly impacts users' performance [44, 63, 64]. Consequently, a promising research direction is *interaction techniques that minimize the drawbacks of poor depth perception in VR*.

Finally, the sample suggests that an important research direction is *developing techniques for manipulation*—only 37.5% of our surveyed papers addressed manipulation. There are perhaps even fewer methodological approaches or examples of assessing manipulation than pointing, even from the 2D domain. This leaves significant scope for the development of transferable object-manipulation evaluation approaches.

6.2 Study Methodology

These directions concern the empirical studies that we conduct and their methodology. One crucial direction involves *getting studies out of the lab*. All of the studies that we surveyed were conducted in controlled, laboratory environments. It is well known that such environments give researchers full control over the experimental situation but at the cost of realism [51]. However, in 2D pointing, field research has been essential [18], and some of the few crowd-sourced studies on VR have documented surprising variation in the settings where VR is used [e.g., 58].

Another direction concerns a *reference task agenda*. This term was coined by Whittaker et al. [85] to discuss the absence in general within HCI of standardized tasks. In VR research, there have been several attempts to standardize tasks [e.g., 10, 66]. However, 64.6% of the studies use a new task rather than building on an existing study. According to Whittaker et al. [85], standardized tasks helps to focus on what is important to the field, share metrics, and data sets, and develop theory. The studies reviewed shed no light on why previous attempts have not worked. Still, we offer two speculations on future work: (1) the third dimension needs to be an integral and systematically manipulable part of a reference task agenda, and (2) the task agenda needs to be based on what users do in VR, similar to how early studies of web browsing [14, 19] informed much work on hypertext and www-navigation.

In the studies, about half of the participants have some experience with VR. Additionally, those who use and report training time with the studied techniques in VR, spend about 6 minutes training (see more on the results in section 4.6.3). Therefore, an important research direction concerns *longer-term studies*, emphasizing the development of expertise and the wearing-off of novelty effects. Longer-term studies in VR exist (e.g., Steinicke's experiment on 24 hours in VR [73]), but they are rare. Many classic studies on pointing train participants extensively to identify upper bounds of performance, notably those of Card et al. [16]. Such effects of

expertise on performance is another study direction for object manipulation VR, including possible habituation and learning effects of techniques.

Finally, understanding *user experience and satisfaction in VR* with selection and manipulation techniques is an important research direction. As previously discussed, although around 70% of the studies use questionnaires, they are rarely pre-existing or validated. Further, the depth of reporting on user experience and satisfaction is low compared to that used for performance. For 2D pointing, standard questionnaires have been developed: that of Douglas et al. [22] is well known. For VR, it seems that studies have not converged on a similar widely-used questionnaire. Further, although presence is measured in 11% of the studies and immersion is measured in 4%, the centrality of these concepts to our use of avatars would warrant a much more extensive inclusion of these aspects in studies.

6.3 Accumulation and Theory Building

Across the studies, we find a pattern similar to other areas of HCI [34]: few to no replications and an emphasis on novelty. One important research direction is, therefore, to *accumulate empirical findings*. This may happen in different ways, including (a) using standardized test suites so that performance may be compared across studies; (b) perform meta-analyses of findings across a set of studies so as to determine what works; and (c) directly compare to other techniques. This rarely happened in the surveyed studies.

Another research direction that would aid theoretical progress is *more extensive models or theories of VR interaction*. In earlier work, Bederson and Shneiderman [7] separated descriptive, explanatory, predictive, prescriptive, and generative approaches in HCI. In particular, we had expected to see more use of both generative and predictive models. Although the sampled studies do contain models [e.g., 50, 57], we saw no attempts at modelling, for instance, motor learning to be able to predict study results or generate new designs.

7 CONCLUSION

Studies of object selection and manipulation techniques play a central role in VR research. However, few guidelines for conducting such studies exist, making planning them, and comparing their findings unnecessarily hard. We have reviewed 20 years of VR studies to learn about best practices, build recommendations, and to identify open research challenges. In particular, we have discussed how to organize tasks and targets, design the physical and virtual settings, and report results. We also identify topics that are not prioritized in the reviewed studies but that we find essential. They include ideas for new techniques, study methodology, and improving performance modeling in selection and manipulation tasks.

ACKNOWLEDGMENTS

REFERENCES

References marked with • are in the set of reviewed papers.

[1] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136.

[2] • Ferran Argelaguet, Carlos Andujar, and Ramon Trueba. 2008. Overcoming eye-hand visibility mismatch in 3D pointing selection. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* (2008), 43–46. https://doi.org/10.1145/1450579.1450588

[3] • Oscar Ariza, Gerd Bruder, Nicholas Katzakis, and Frank Steinicke. 2018. Analysis of Proximity-Based Multimodal Feedback for 3D Selection in Immersive Virtual Environments. *25th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018 - Proceedings* (2018), 327–334. https://doi.org/10.1109/VR.2018.8446317

[4] • Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique based on Raycasting. (2019), 1–12. https://doi.org/10.1145/3290605.3300331

[5] Louise Barkhuus and Jennifer A Rode. 2007. From mice to men-24 years of evaluation in CHI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vol. 10.

[6] • Anil Ufuk Batmaz and Wolfgang Stuerzlinger. 2019. Effects of 3d rotational jitter and selection methods on 3d pointing tasks. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings* (2019), 1687–1692. https://doi.org/10.1109/VR.2019.8798038

[7] Benjamin B Bederson, Benjamin B Bederson, Ben Shneiderman, et al. 2003. *The craft of information visualization: readings and reflections.* Morgan Kaufmann.

[8] • Bobby Bodenheimer, Sarah Creem-Regehr, Jeanine Stefanucci, Elena Shemetova, and William B. Thompson. 2017. Prism aftereffects for throwing with a self-avatar in an immersive virtual environment. *Proceedings - IEEE Virtual Reality* 0 (2017), 141–147. https://doi.org/10.1109/VR.2017.7892241

[9] Doug Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan P Poupyrev. 2004. *3D user interfaces: Theory and Practice* (1 ed.). Addison-Wesley Professional.

[10] Doug A Bowman, Donald B Johnson, and Larry F Hodges. 2001. Testbed evaluation of virtual environment interaction techniques. *Presence: Teleoperators & Virtual Environments* 10, 1 (2001), 75–95.

[11] Doug A Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan Poupyrev. 2001. An introduction to 3-D user interface design. *Presence: Teleoperators & Virtual Environments* 10, 1 (2001), 96–108.

[12] • David Brickler, Matias Volonte, Jeffrey W. Bertrand, Andrew T. Duchowski, and Sabarish V. Babu. 2019. Effects of stereoscopic viewing and haptic feedback, sensory-motor congruence and calibration on near-field fine motor perception-action coordination in virtual reality. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings* (2019), 28–37. https://doi.org/10.1109/VR.2019.8797744

[13] John Brooke. 1996. SUS: A 'Quick and Dirty' Usability Scale. *Usability evaluation in industry* (1996), 189. https://doi.org/10.1201/9781498710411-35

[14] Michael D Byrne, Bonnie E John, Neil S Wehrle, and David C Crow. 1999. The tangled web we wove: A taskonomy of WWW use. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 544–551.

[15] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[16] Stuart K Card, Thomas P Moran, and Allen Newell. 1983. The psychology of human-computer interaction. 1983.

[17] • Emmanuelle Chapoulie, Maud Marchal, Evanthia Dimara, Maria Roussou, Jean Christophe Lombardo, and George Drettakis. 2014. Evaluation of direct manipulation using finger tracking for complex tasks in an immersive cube. *Virtual Reality* 18, 3 (2014), 203–217. https://doi.org/10.1007/s10055-014-0246-0

[18] Olivier Chapuis, Renaud Blanch, and Michel Beaudouin-Lafon. 2007. Fitts' law in the wild: A field study of aimed movements. (2007).

[19] Andy Cockburn and Bruce J. McKenzie. 2001. What do web users do? An empirical analysis of web use. *Int. J. Hum. Comput. Stud.* 54, 6 (2001), 903–922. https://doi.org/10.1006/ijhc.2001.0459

[20] Nguyen-Thong Dang. 2007. A survey and classification of 3D pointing techniques. In *2007 IEEE international conference on research, innovation and vision for the future*. IEEE, 71–80.

[21] • Henrique G. Debarba, Sami Perrin, Bruno Herbelin, and Ronan Boulic. 2015. Embodied interaction using non-planar projections in immersive virtual reality. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* 13-15-November-2015 (2015), 125–128. https://doi.org/10.1145/2821592.2821603

[22] Sarah A Douglas, Arthur E Kirkpatrick, and I Scott MacKenzie. 1999. Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 215–222.

[23] • Gregory W. Edwards, Woodrow Barfield, and Maury A. Nussbaum. 2004. The use of force feedback and auditory cues for performance of an assembly task in an immersive virtual environment. *Virtual Reality* 7, 2 (2004), 112–119. https://doi.org/10.1007/s10055-004-0120-6

[24] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47, 6 (1954), 381–391. https://doi.org/10.1037/h0055392

[25] • Cédric Fleury, Thierry Duval, Valérie Gouranton, and Anthony Steed. 2012. Evaluation of remote collaborative manipulation for scientific data analysis. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* (2012), 129–136. https://doi.org/10.1145/2407336.2407361

[26] • Scott Frees and G. Drew Kessler. 2005. Precise and rapid interaction through scaled manipulation in immersive virtual environments. *Proceedings - IEEE Virtual Reality 2005* (2005), 99–106. https://doi.org/10.1109/vr.2005.60

[27] Eg Su Goh, Mohd Shahrizal Sunar, and Ajune Wanis Ismail. 2019. 3D object manipulation techniques in handheld mobile augmented reality interface: A review. *IEEE Access* 7 (2019), 40581–40601.

[28] Mar Gonzalez-Franco and Tabitha C. Peck. 2018. Avatar Embodiment. Towards a Standardized Questionnaire. *Frontiers in Robotics and AI* 5 (2018), 74. https://doi.org/10.3389/frobt.2018.00074

[29] • Jan Gugenheimer, David Dobbelstein, Christian Winkler, Gabriel Haas, and Enrico Rukzio. 2016. FaceTouch: Enabling Touch Interaction in Display Fixed UIs for Mobile Virtual Reality. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 49–60. https://doi.org/10.1145/2984511.2984576

[30] Chris Hand. 1997. A survey of 3D interaction techniques. In *Computer graphics forum*, Vol. 16. Wiley Online Library, 269–281.

[31] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. https://doi.org/10.1016/S0166-4115(08)62386-9

[32] Rorik Henrikson, Tovi Grossman, Sean Trowbridge, Daniel Wigdor, and Hrvoje Benko. 2020. 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*. 1–14.

[33] Kasper Hornbæk. 2006. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies* 64 (02 2006), 79–102. https://doi.org/10.1016/j.ijhcs.2005.06.002

[34] Kasper Hornbæk, Søren S Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is once enough? On the extent and content of replications in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3523–3532.

[35] Jacek Jankowski and Martin Hachet. 2015. Advances in interaction with 3D environments. In *Computer Graphics Forum*, Vol. 34. Wiley Online Library, 152–190.

[36] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (1993), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

[37] Panayiotis Koutsabasis and Panagiotis Vogiatzidakis. 2019. Empirical research in mid-air interaction: A systematic review. *International Journal of Human–Computer Interaction* 35, 18 (2019), 1747–1768.

[38] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice* (2 ed.). Addison-Wesley Professional.

[39] • Jaeyeon Lee, Mike Sinclair, Mar Gonzalez-Franco, Eyal Ofek, and Christian Holz. 2019. TORC: A virtual reality controller for in-hand high-dexterity finger interaction. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–13. https://doi.org/10.1145/3290605.3300301

[40] Sangyoon Lee, Jinseok Seo, Gerard Jounghyun Kim, and Chan-Mo Park. 2003. Evaluation of pointing techniques for ray casting selection in virtual environments. In *Third international conference on virtual reality and its application in industry*, Vol. 4756. International Society for Optics and Photonics, 38–44.

[41] Nianlong Li, Teng Han, Feng Tian, Jin Huang, Minghui Sun, Pourang Irani, and Jason Alexander. 2020. Get a Grip: Evaluating Grip Gestures for VR Input using a Lightweight Pen. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[42] Klemen Lilija, Henning Pohl, and Kasper Hornbæk. 2020. Who Put That There? Temporal Navigation of Spatial Recordings by Direct Manipulation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11.

[43] • R. W. Lindeman, J. L. Sibert, and J. N. Templeman. 2001. The effect of 3D widget representation and simulated surface constraints on interaction in virtual environments. *Proceedings - Virtual Reality Annual International Symposium* (2001), 141–148. https://doi.org/10.1109/vr.2001.913780

[44] • Mayra Donaji Barrera Machuca and Wolfgang Stuerzlinger. 2019. The effect of stereo display deficiencies on virtual hand pointing. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery. https://doi.org/10.1145/3290605.3300437

[45] I. Scott MacKenzie. 1992. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human–Computer Interaction* 7, 1 (1992), 91–139. https://doi.org/10.1207/s15327051hci0701_3

[46] I Scott MacKenzie and R William Soukoreff. 2003. Phrase sets for evaluating text entry techniques. In *CHI'03 extended abstracts on Human factors in computing systems*. 754–755.

[47] • Diako Mardanbegi, Benedikt Mayer, Ken Pfeuffer, Shahram Jalaliniya, Hans Gellersen, and Alexander Perzl. 2019. EyeSeeThrough: Unifying tool selection and application in virtual environments. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings* (2019), 474–483. https://doi.org/10.1109/VR.2019.8797988

[48] • Alejandro Martin-Gomez, Ulrich Eck, and Nassir Navab. 2019. Visualization techniques for precise alignment in VR: A comparative study. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings* (2019), 735–741. https://doi.org/10.1109/VR.2019.8798135

[49] Antonella Maselli and Mel Slater. 2013. The building blocks of the full body ownership illusion. *Frontiers in human neuroscience* 7 (2013), 83.

[50] • Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. 2018. The effect of offset correction and cursor on mid-air Pointing in real and virtual environments. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018), 1–13. https://doi.org/10.1145/3173574.3174227

[51] Joseph E McGrath. 1995. Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human–Computer Interaction*. Elsevier, 152–169.

[52] Daniel Mendes, Fabio Marco Caputo, Andrea Giachetti, Alfredo Ferreira, and J Jorge. 2019. A survey on 3D virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, Vol. 38. Wiley Online Library, 21–45.

[53] • Daniel Mendes, Filipe Relvas, Alfredo Ferreira, and Joaquim Jorge. 2016. The Benefits of DOF Separation in Mid-air 3D Object Manipulation. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* 02-04-November-2016 (2016), 261–268. https://doi.org/10.1145/2993369.2993396

[54] • Daniel Mendes, Maurício Sousa, Rodrigo Lorena, Alfredo Ferreira, and Joaquim Jorge. 2017. Using custom transformation axes for mid-air manipulation of 3D virtual objects. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* Part F131944 (2017). https://doi.org/10.1145/3139131.3139157

[55] • Mathias Moehring and Bernd Froehlich. 2011. Effective manipulation of virtual objects within arm's reach. *Proceedings - IEEE Virtual Reality* (2011), 131–138. https://doi.org/10.1109/VR.2011.5759451

[56] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, Prisma Group, et al. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med* 6, 7 (2009), e1000097.

[57] • Roberto A. Montano Murillo, Sriram Subramanian, and Diego Martinez Plasencia. 2017. Erg-O: Ergonomic optimization of immersive virtual environments. *UIST 2017 - Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (2017), 759–771. https://doi.org/10.1145/3126594.3126605

[58] Aske Mottelson and Kasper Hornbæk. 2017. Virtual reality studies outside the laboratory. In *Proceedings of the 23rd acm symposium on virtual reality software and technology*. 1–10.

[59] • Yun Suen Pai, Tilman Dingler, and Kai Kunze. 2019. Assessing hands-free interactions for VR using eye gaze and electromyography. *Virtual Reality* 23, 2 (2019), 119–131. https://doi.org/10.1007/s10055-018-0371-2

[60] • Soonchan Park, Seokyeol Kim, and Jinah Park. 2012. Select Ahead: Efficient object selection technique using tendency of recent cursor movements. *APCHI'12 - Proceedings of the 2012 Asia Pacific Conference on Computer-Human Interaction* (2012), 51–58. https://doi.org/10.1145/2350046.2350060

[61] • Vijay M. Pawar and Anthony Steed. 2009. Evaluating the influence of haptic Force-Feedback on 3d selection tasks using natural egocentric gestures. *Proceedings - IEEE Virtual Reality* (2009), 11–18. https://doi.org/10.1109/VR.2009.4810992

[62] • Vijay M. Pawar and Anthony Steed. 2009. Profiling the behaviour of 3D selection tasks on movement time when using natural haptic pointing gestures. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* (2009), 79–82. https://doi.org/10.1145/1643928.1643947

[63] Alex Peer and Kevin Ponto. 2019. Mitigating Incorrect Perception of Distance in Virtual Reality through Personalized Rendering Manipulation. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 244–250.

[64] Etienne Peillard, Thomas Thebaud, Jean-Marie Norrnand, Ferran Argelaguet, Guillaume Moreau, and Anatole Lécuyer. 2019. Virtual Objects Look Farther on the Sides: The Anisotropy of Distance Perception in Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 227–236.

[65] • Duc Minh Pham and Wolfgang Stuerzlinger. 2019. Is the pen mightier than the controller? A comparison of input devices for selection in virtual and augmented reality. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* (2019). https://doi.org/10.1145/3359996.3364264

[66] Ivan Poupyrev, Suzanne Weghorst, Mark Billinghurst, and Tadao Ichikawa. 1997. A framework and testbed for studying manipulation techniques for immersive VR. In *Proceedings of the ACM symposium on Virtual reality software and technology*. 21–28.

[67] Yuan Yuan Qian and Robert J Teather. 2017. The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*. 91–98.

[68] Ludwig Sidenmark, Christopher Clarke, Xuesong Zhang, Jenny Phu, and Hans Gellersen. 2020. Outline Pursuits: Gaze-assisted Selection of Occluded Objects in

Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[69] • Ludwig Sidenmark and Hans Gellersen. 2019. Eye & Head: Synergetic eye and head movement for gaze pointing and selection. *UIST 2019 - Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (2019), 1161–1174. https://doi.org/10.1145/3332165.3347921

[70] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of Presence in Immersive Virtual Environments. *Teleoperators and Virtual Environments - Presence* 3 (01 1994).

[71] R. William Soukoreff and I. Scott MacKenzie. 2004. 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 (2004), 751–789.

[72] • Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based text entry In Virtual Reality. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2018-April. https://doi.org/10.1145/3173574.3174221

[73] Frank Steinicke and Gerd Bruder. 2014. A self-experimentation report about long-term use of fully-immersive technology. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*. 66–69.

[74] • Rasmus Stenholt. 2012. Efficient selection of multiple objects on a large scale. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* (2012), 105–112. https://doi.org/10.1145/2407336.2407357

[75] Wolfgang Sturzlinger and Robert J Teather. 2014. Considerations for targets in 3D pointing experiments. *Proceedings of HCI Korea* (2014), 162–168.

[76] Youssef S Tanagho, Gerald L Andriole, Alethea G Paradis, Kerry M Madison, Gurdarshan S Sandhu, J Esteban Varela, and Brian M Benway. 2012. 2D versus 3D visualization: impact on laparoscopic proficiency using the fundamentals of laparoscopic surgery skill set. *Journal of Laparoendoscopic & Advanced Surgical Techniques* 22, 9 (2012), 865–870.

[77] Robert J Teather and Wolfgang Sturzlinger. 2011. Pointing at 3D targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 87–94.

[78] • Tanh Quang Tran, Hyun Ju Shin, Wolfgang Sturzlinger, and Jung Hyun Han. 2017. Effects of virtual arm representations on interaction in virtual environments. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* Part F131944 (2017). https://doi.org/10.1145/3139131.3139149

[79] Huawei Tu, Susu Huang, Jiabin Yuan, Xiangshi Ren, and Feng Tian. 2019. Crossing-based selection with virtual reality head-mounted displays. In *Conference on Human Factors in Computing Systems - Proceedings (CHI '19)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3290605.3300848

[80] • Manuel Veit, Antonio Capobianco, and Dominique Bechmann. 2009. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* 1, 212 (2009), 51–58. https://doi.org/10.1145/1643928.1643942

[81] Keith Vertanen and Per Ola Kristensson. 2011. A versatile dataset for text entry evaluations based on genuine mobile emails. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. 295–298.

[82] • Jia Wang and Robert W. Lindeman. 2015. Object impersonation: Towards effective interaction in tablet- and HMD-based hybrid virtual environments. *2015 IEEE Virtual Reality Conference, VR 2015 - Proceedings* (2015), 111–118. https://doi.org/10.1109/VR.2015.7223332

[83] Alan Traviss Welford. 1968. Fundamentals of skill. (1968).

[84] Johann Wentzel, Greg d'Eon, and Daniel Vogel. 2020. Improving Virtual Reality Ergonomics Through Reach-Bounded Non-Linear Input Amplification. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[85] Steve Whittaker, Loren Terveen, and Bonnie A Nardi. 2000. Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI. *Human–Computer Interaction* 15, 2-3 (2000), 75–106.

[86] • Curtis Wilkes and Doug A. Bowman. 2008. Advantages of velocity-based scaling for distant 3D manipulation. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* (2008), 23–29. https://doi.org/10.1145/1450579.1450585

[87] • Chadwick A. Wingrave, Ryan Tintner, Bruce N. Walker, Doug A. Bowman, and Larry F. Hodges. 2005. Exploring individual differences in raybased selection: Strategies and traits. *Proceedings - IEEE Virtual Reality* 2005 (2005), 163–170. https://doi.org/10.1109/vr.2005.35

[88] • Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (1998), 225–240. https://doi.org/10.1162/105474698565686 arXiv:https://doi.org/10.1162/105474698565686

[89] • Jackie Yang, Hiroshi Horii, Alexander Thayer, and Rafael Ballagas. 2018. VR Grabbers: Ungrounded haptic retargeting for precision grabbing tools. In *UIST 2018 - Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, Inc, 889–899. https://doi.org/10.1145/3242587.3242643

[90] Zhong Yang, Yi Li, Yang Zheng, Weidong Chen, and Xiaoxiang Zheng. 2012. An interaction system using mixed hand gestures. In *Proceedings of the 10th asia pacific conference on Computer human interaction*. 125–132.

[91] • Chun Yu, Yizheng Gu, Zhican Yang, Xin Yi, Hengliang Luo, and Yuanchun Shi. 2017. Tap, Dwell or Gesture? Exploring Head-Based Text Entry Techniques for HMDs. (2017), 4479–4488. https://doi.org/10.1145/3025453.3025964

[92] Shumin Zhai and Paul Milgram. 1993. Human performance evaluation of manipulation schemes in virtual environments. In *Proceedings of IEEE Virtual Reality Annual International Symposium*. IEEE, 155–161.

[93] • David J. Zielinski, Hrishikesh M. Rao, Mark A. Sommer, and Regis Kopper. 2015. Exploring the effects of image persistence in low frame rate virtual environments. *2015 IEEE Virtual Reality Conference, VR 2015 - Proceedings* (2015), 19–26. https://doi.org/10.1109/VR.2015.7223319

[94] Tim Zindulka, Myroslav Bachynskyi, and Jörg Müller. 2020. Performance and Experience of Throwing in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

A VR OBJECT SELECTION AND MANIPULATION STUDY CHECKLIST

Method

- ☐ 1. Provide an explicit statement of the goal of the study.
- ☐ 2. Provide information (e.g., a link or an appendix) about preregistration, if applicable.

Participants

- ☐ 3. Report the number of participants and if possible, how this sample size was decided.
- ☐ 4. Describe the aims of recruitment together with demographics (e.g., level of expertise with virtual reality and in the domain, if applicable).

Design

- ☐ 5. State all independent variables or factors and their levels in an explicit sentence. When there are more than one independent variables or factors, report them and their levels in $M \times N \times P$ format. Be sure to include all factors.
- ☐ 6. Report counterbalancing or randomisation of the order of conditions and the method for those. Specify this for each of the independent variables or factors. If none, justify why.
- ☐ 7. Report the number of repetitions if any, and the number of trials in total for the main independent variable (e.g., if M is a technique, report the number of trials with each technique in $N \times P \times$ repetitions format).
- ☐ 8. Define the dependent variables in an explicit sentence (e.g., selection or task completion time, errors, accuracy, and user experience or workload, if applicable). Explain how the values of dependent variables are determined or calculated.

Task

- ☐ 9. Report the task(s) and possible sub-tasks in detail. Classify the tasks as selection or manipulation, and classify manipulation as translation, rotation, or scaling.
- ☐ 10. State if the task is based on previous work, and if so, which parameters were modified, if any.

Procedure

- ☐ 11. Express whether the task is a discrete task or serial. If the task is serial, explain the trials through the sequences (e.g., 10 target selections in a 11-target circle after the initial starting selection).
- ☐ 12. Describe how the participants were trained for the task (if they were). Specify the type of the training (e.g., in the task itself, in getting used to a technique in a setting different from the task). Specify also the duration of the training (e.g., a time or number of trials).

Targets

- ☐ 13. Report the total number of distinct targets.
- ☐ 14. Report whether distractors were used, and whether they were external to the task. Report the number of external distractors in relation to the targets.
- ☐ 15. Report the method of arranging targets. This can be reporting a shape that is followed with an even distribution (e.g., on a matrix), or randomisation. Specify the constraints used in randomisation, such as the dimensions of the area in which the targets are laid out (e.g., a virtual cube, a matrix, or a sphere), and the boundary conditions for randomisation (e.g., minimum distance between targets, the number of targets, or removal of occluded targets). Ideally provide a visual depiction in addition to a text-based description.
- ☐ 16. Report the actual target locations (e.g., in mm and degrees). Do this even when the targets are distributed randomly: log their arrangement and report at least a summary of their locations.
- ☐ 17. Describe the distance to each target as euclidian distance (e.g., meters) and as angles if applicable (e.g., from the participant's point-of-view). Explain also how this distance is defined both from the starting point (e.g., in a discrete task it can be from the reset/starting point) and from the end point (e.g., to the nearest point on the target sphere, or to the center of the target sphere).
- ☐ 18. Report target locations relative to the movement direction (e.g., there were six targets on three axes: frontal, lateral, and vertical, with all together six movement directions: reach and withdraw, left and right, and up and down). Do this numerically.
- ☐ 19. Report the target shape and size. Specify how the size is defined (e.g., the diameter of the target sphere, or the $side \times side \times side$ of a cuboid). Ideally provide a visual depiction in addition to a text-based description.
- ☐ 20. Report the cursor type and size (e.g., a single-pixel ray, or a sphere with a diameter of d), or the size and shape of the object that is manipulated.
- ☐ 21. Connect target distances with size if applicable (e.g., list the target IDs together with sizes and distances).
- ☐ 22. Specify how the cursor or manipulated object and the target object sizes are used to define errors, if any are measured (e.g., the size difference if the cursor needs to fit inside the target for successful performance, or the acceptable threshold of correct positioning). Otherwise, state that accuracy is used as a measure and define how it is calculated.

Materials

☐ 23. Describe the physical setting with respect to the participants and their movement (e.g., what posture were they instructed to maintain, were there any controlled or otherwise limited physical motions such as a use of a chin rest, an instruction to stand on the same spot, or resetting the physical posture in a starting position between trials). Ideally provide a figure in addition to a text-based description.

☐ 24. Specify the physical settings in spatial units, such as distance in meters (e.g., where does the participant sit in relation to a table) or angles in degrees (e.g., the range the participant can move their hand on a haptic device).

☐ 25. Describe the virtual setting. As a minimum, specify which default virtual scene was used or justify the design of a custom virtual scene (e.g., adding walls to make a room) and the light sources in the space. Describe these with spatial measures (e.g., dimensions of a room with meters or a light source position with degrees). Ideally provide a figure in addition to a text-based description.

☐ 26. Specify user representations in the virtual setting. Specify at least where in the virtual setting the user's view-point is located (also in relation to the target space), what kind of representations are presented of users or input devices (e.g., as a full-body avatar, a virtual representation of the controller), and how feedback of the user's actions are given.

☐ 27. Describe the devices used in the study. Detail from virtual reality devices their FoV, refresh rate, and resolution in addition to their brand, type, and version. From the motion tracking or other input devices, specify their spatial and temporal accuracy.

☐ 28. Describe the interaction technique(s) used in the study. Detail how the user's movements are mapped into the virtual setting (e.g., motion gains for cursors and methods for casting rays or pointers) and the mechanism for triggering selection (and if applicable, task completion).

Results and Analysis

☐ 29. Report the effect size of your statistics.

☐ 30. Provide supplementary information (e.g., a link) for accessing data, if open.