This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3320211

Controllers or Bare Hands? A Controlled Evaluation of Input Techniques on Interaction Performance and Exertion in Virtual Reality

Tiffany Luong, Yi Fei Cheng, Max Möbus, Andreas Fender, and Christian Holz

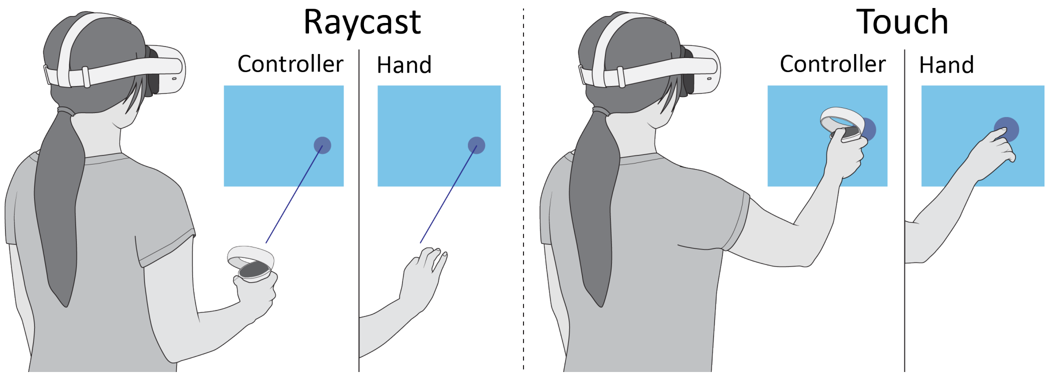

Fig. 1: Our user study ($N = 24$) investigates the impact of VR controllers vs. free-hand interaction on physical exertion, sense of agency, task performance, and motor behavior across different interaction techniques (touch, raycast) and tasks (selection, trajectory tracing). Our apparatus *reliably* tracks participants and their hands outside-in and detects in-air pinch with a wearable sensor to exclude the effect of headset-based input tracking from our results.

Abstract— Virtual Reality (VR) systems have traditionally required users to operate the user interface with controllers in mid-air. More recent VR systems, however, integrate cameras to track the headset's position inside the environment as well as the user's hands when possible. This allows users to *directly* interact with virtual content in mid-air just by reaching out, thus discarding the need for hand-held physical controllers. However, it is unclear which of these two modalities—controller-based or free-hand interaction—is more suitable for *efficient* input, *accurate* interaction, and *long-term* use under *reliable* tracking conditions. While interacting with hand-held controllers introduces weight, it also requires less finger movement to invoke actions (e.g., pressing a button) and allows users to hold on to a *physical* object during virtual interaction.

In this paper, we investigate the effect of VR input modality (controller vs. free-hand interaction) on physical exertion, agency, task performance, and motor behavior across two mid-air interaction techniques (touch, raycast) and tasks (selection, trajectory-tracing). Participants reported less physical exertion, felt more in control, and were faster and more accurate when using VR controllers compared to free-hand interaction in the raycast setting. Regarding personal preference, participants chose VR controllers for raycast but free-hand interaction for mid-air touch. Our correlation analysis revealed that participants' physical exertion increased with selection speed, quantity of arm motion, variation in motion speed, and bad postures, following ergonomics metrics such as consumed endurance and rapid upper limb assessment. We also found a negative correlation between physical exertion and the participant's sense of agency, and between physical exertion and task accuracy.

Index Terms—Virtual Reality, Input Modality, Controllers, Hand Tracking, Raycast, Touch, Physical Exertion, Performance, Behavior

1 Introduction

The recent development of Virtual Reality (VR) headsets enables users to dive into immersive worlds anywhere and at any time, without having to bring much equipment. The ability to operate regardless of the user's surroundings and the present infrastructure makes VR systems appealing to the wider public for usage in their regular environments, such as offices and homes. This, in turn, has spurred much research in the academic community on the topic of comfortable VR interaction modalities [19] that reconcile reliably high interaction performance with the need to use VR systems for more than just brief interactions [52,67], which is a central requirement for most productivity tasks.

In VR, users typically operate two hand-held controllers to interact with virtual interfaces [39, 60]. These controllers typically integrate inertial sensors and physical markers, such that the headset-integrated cameras can track them, and include buttons and/or tactile surfaces that support further user input [53]. More recent head-mounted systems offer hand tracking. Via computer vision [58,72], such headsets translate the user's gestures into commands to interact with virtual objects, thereby limiting all interaction to happen under visual control and inside the headset's field of view [66].

The availability of controller and free-hand input raises a central question for the design of virtual user interfaces: which of these two modalities is more suitable to support elemental interaction, such as selection or tracing, in VR? The trade-off between the two is intriguing; controllers allow users to hold on to a physical object during interaction, stimulate their sense of touch through built-in haptic actuators, and offer reliable input options via physical buttons and other controls. On the other hand, hand tracking removes the need for holding controllers and, thus, additional weight. Free-hand interaction is closer to our everyday interaction with physical objects. This also enhances the user's body ownership toward its virtual representation [3,43].

The choice of input modality determines the performance benefits and level of fatigue incurred during mid-air use. Recent research that has compared both found that controllers outperform hand tracking in task completion time [15,28,29,39,48] as well as in subjective ratings (e.g., usability [15,23,39,48] and satisfaction [42,48] compared to the lack of input affordance in free-hand interaction [12]).

However, previous studies have often relied on the tracking technologies built into the VR headset or added sensors (e.g., Leap Motion),

- *The authors are with the Department of Computer Science at ETH Zürich, Zurich, Switzerland.*
- *E-mail addresses: {firstname.lastname}@inf.ethz.ch.*

which may have affected task performance due to self-occlusion or occasionally inaccurate estimations from the onboard computer-vision algorithms [28,39]. Indeed, current implementations struggle to deliver stable poses with low latency, which makes accurate gesture detection challenging [2]. In past studies, controllers and hands have also primarily been compared for tasks of relatively short duration. It remains relatively unclear how their respective affordances may suit more prolonged tasks. Finally, prior work has mostly focused on the direct manipulation of 3D objects (i.e., grab-and-place, rotation tasks). The suitability of controller-based or free hand for common interactions beyond direct manipulation (i.e., raycasting) has been investigated to a much lesser extent.

In this paper, we present the results of a study in which we compared VR controllers and free-hand input for mid-air interaction tasks —using a *controlled* and *high-accuracy* multi-camera setup for precise tracking and assessment. Our study particularly focused on interactions that exceed quick input. Beyond replicating prior studies with an apparatus that affords higher precision tracking, thereby circumventing technical limitations as potential confounds, we leverage the motion capture data to further derive and analyze ergonomic and behavioral metrics.

1.1 Effect of input modality: hands vs. controllers

Our study of input with 24 participants investigated the effects of modality, i.e., *free-hand interaction* versus *VR controller* on participant performance, subjective preference, and motion behavior. Participants completed two tasks with prolonged duration: *selection* and *trajectory tracing*. We additionally varied the interaction technique they operated, examining the effects of input modality for both direct *touch* and distant *raycast* methods. We assessed participants' subjective preferences and feelings of exertion and agency with questionnaires, performance through quantitative metrics such as selection time, as well as derived behavioral and ergonomic metrics from motion capture data of participants' shoulders, elbows, and hands.

Participants' preference for input modality interestingly differed by interaction modality. For *touch* interactions, participants generally preferred using *free-hand interaction*, but for *raycast* interactions, participants preferred using *VR controllers* for input. For the remaining self-reported metrics (i.e. exertion, agency), as well as task performance measures (i.e. selection time, trajectory offset), significant effects of input modality were only observed in the *raycast* condition. For *raycast*, participants reported lower exertion and a higher sense of agency using *VR controllers*. They were also faster (at *selection*), and more accurate (at *trajectory tracing*). This suggests that for *raycast* based interactions, using a VR controller may be preferable in prolonged usage settings.

Several of our behavioral and ergonomic metrics corroborate this finding. When using *VR controllers* to perform *raycast* interactions, participants kept their hands generally closer to their body, which arguably represents more comfortable, lower exertion positions. They also moved their hands less and applied less pressure during activations (accomplished through pressing a trigger button as opposed to pinch gestures). Across both interaction techniques, our results indicate that when using a VR controller, participants also tended to adopt a position with their elbow lowered and tucked in. Their hands were also placed at a lower height, suggesting a more ergonomic "arms down" posture. Furthermore, they generally moved their upper arm less. These metrics again suggest there are potential ergonomic benefits to having a controller at hand. Participants' preference for using *free-hand interactions* nonetheless in the *touch* condition is thus rather curious, and we speculate on this among other observed behavioral results within our discussion. Ultimately, we conclude from our results that there are distinct differences between free-hand and controller-based input approaches that can not be attributed simply to tracking fidelity. With higher accuracy tracking, controller-based input appears favorable, especially for ray-casting operations. These differences notably manifested in participants' postures and movement behaviors.

Taken together, we contribute an empirical study on the differences between free-hand and controller-based input for interacting with virtual interfaces. Our results provide a set of design guidelines for the effective use of the two modalities in VR applications.

2 Related Work

The controlled experiment in this paper is related to hand input modalities for mid-air interactions in VR.

2.1 Interaction techniques in VR

Several decades of research have brought forward a myriad of interaction techniques relying on input devices and user gestures [5,36,53], with their taxonomies [5, 12, 13, 53], and evaluation methodologies [14,27].

There are two main metaphors for manipulating objects in VR [53]. The *virtual hand* metaphor provides the user with a virtual hand representation that follows and mimics the physical hand movement [62]. While it offers natural input mappings by exploiting well-known real-world actions [5], it also has several shortcomings, such as interacting with objects that are out of reach. On the other hand, the *virtual pointer* metaphor overcomes the physical constraints of the real world and is one of the most popular techniques for selection tasks in VR [5,53]. Using ray-pointing, users can select objects beyond their area of reach by pointing at them with a ray that originates most of the time from their hands [5,49,65]. Many works surveyed this technique and concluded that eye-rooted ray-pointing with hand direction was more efficient than the more simplistic hand-rooted parameters [4,5,49,65]. Other hybrid techniques were also proposed in prior work. For example, Go-Go applies a non-linear amplification to the hand position and allows users to directly manipulate objects at-a-distance [61]. HandyCast uses such non-linear amplification to allow bimanual control over two virtual hand pointers through input on a single hand-held smartphone [40]. Such amplification techniques allow for less fatiguing input across the full range and, in the latter case, operation in space-constrained scenarios from the convenience of one's lap.

The fundamental operations in VR include navigation, selection, and manipulation [13,36]. These operations can be decomposed into lower-level interactions, including colliding, pointing, and selection activation [5]. From a human perspective, most of these interactions require *hand-eye coordination* [37], which is a cognitive proficiency involving the coordinated motor control of eye and hand movements [37]. This skill is essential for performing everyday actions such as grasping a glass to drink. In the context of VR, hand-eye coordination has been extensively studied in selection tasks [10,73] and more complex use-cases like music and surgery training [64]. However, there have been few investigations on hand-eye coordination using controlled continuous tasks, such as pursuit tasks in VR.

In our work, we draw on this literature to inform conditions and tasks included in our experiment. We compare controller and bare hand interaction for both direct *virtual hand* and indirect *pointer* metaphors. In both cases, we examine user's performance in two representative tasks: selection and trajectory tracing.

2.2 Controllers vs. Bare Hand Interactions in VR

In VR, two of the most common interaction modalities are controller-based input and free-hand input [39,60]. For controller-based input, users rely on VR hand controllers, which typically consist of multiple buttons, to operate on and control their virtual environment [53]. For free-hand input, they interact with the VR contents directly with their bare hands. Instead of pushing or pulling buttons, hand gestures (e.g. pinching, grabbing) are typically classified and used as control commands [58]. With recent advancements in computer vision, VR devices are now capable of tracking the position and orientation of both controllers and hands to a high degree of accuracy, enabling flexible and expressive spatial interactions [42].

The prevalence of controller-based and free-hand input in VR naturally begs the question of how the two modalities differ, and what advantages and disadvantages they have respectively. Prior work has therefore extensively investigated the differences between VR controllers and hand tracking techniques in a contexts and tasks ranging from anatomy learning [23, 63] to robot control [18]. The general sentiment of prior work is that while free-hand input is advantageously equipment free [48,59], perceived as more natural and intuitive [39,48], and beneficial for body ownership, realism, enjoyment, presence [3,46], they are associated with diminished performance in comparison to controller-based input [28, 29, 39, 48, 57, 63]. One frequently cited reason for this is the lower reliability and accuracy of hand tracking [29, 39, 39, 48, 51, 57, 63]. Beyond technical limitations, prior work has also suggested that the tactile feedback of controller-based input provides beneficial ergonomics [30].

However, despite the vast body of work that examines the comparative value of controller-based versus free-hand input for VR interactions, we have several reasons to believe that additional studies are warranted. First, the vast majority of prior experiments rely on commercial computer-vision based methods to perform hand-tracking, which suffers from tracking delays, unreliable gesture recognition, and a limited tracking range (e.g., VR HMD-based tracking, Leap Motion) [2, 15, 23, 28, 29, 39, 48, 63]. Poor tracking may confound results in comparisons between controller-based and free-hand input, and as we expect tracking technologies to evolve in the future, it would be valuable to replicate prior experiments with apparatus that support more precise and high-throughput tracking. There are several exceptions to this in prior research, including Lin et al. [46], Adkins et al. [3], and Fahmi et al. [23], who all used motion-tracking gloves. However, these works primarily focused on 3D manipulation tasks [3, 23, 46], which differ from the tasks investigated in our study. More importantly, the current understanding of the comparative advantages and disadvantages of free-hand and controller-based input is still very much informed by the results of experiments conducted with imperfect tracking systems. To address this gap, within our comparison of controller-based and free-hand input modalities, we rely on a sub-millimeter tracking system to both drive the experimental interactions and collect data. This also enabled us to perform detailed analyses of movement and ergonomic metrics (e.g., motor space metrics, Consumed Endurance [32]).

Relating to ergonomics, prior experiments examining controller versus free-hand interactions typically focused on tasks of relatively short time-spans. To our knowledge, there is little to no knowledge on whether controller-based or free-hand input is more suitable for prolonged interactions, where factors like arm fatigue (e.g., gorilla arm [31]) have an arguably more significant impact on usability [22, 32, 35, 55]. The tasks in our study were therefore designed to surface the longer term effects of the selected input modality.

Lastly, the vast majority of prior studies compare controller-based and free-hand input in the context of direct manipulation tasks that rely on the *virtual hand* metaphor (e.g. [48]). As discussed previously, the *virtual pointer* metaphor provides a powerful alternative approach to interact with VR contents, especially for items at a distance. To our knowledge, there is comparatively less work directly comparing the efficacy of controller-based and free-hand input for performing distant, pointer-based interactions. Several example exceptions include Li et al. [45] and Johnson et al.'s [38] works. We nonetheless believe it was valuable to build upon this work and to further understand the comparative advantages of controller-based and free-hand input for distant ray-pointing interactions, and therefore examine this distinction within our work.

All in all, we extend prior work by contributing the results of an experiment that compares controller-based with free-hand interactions using a high-accuracy tracking apparatus, for prolonged tasks, and for both direct and distant interactions.

3 Controlled Study: Controllers vs. Bare Hands

We designed and conducted a controlled user study to investigate the following research questions (**RQ1**) Which of *free-hand interaction* or *VR controller* is more suitable for mid-air interactions in VR? (**RQ2**) What is the relation between perceived physical exertion, perceived agency, task performances, and participants' motion behaviors?

Participants performed two tasks (*selection*, *trajectory tracing*) with each input. We split participants into two groups during the experiment to limit the study duration and its impact on participants' physical exertion. The first group performed the tasks in a direct—*touch* condition i.e., they had to reach out to interact with the virtual content. The second group performed the tasks with the user interfaces (UIs) at a distance using *raycast*.

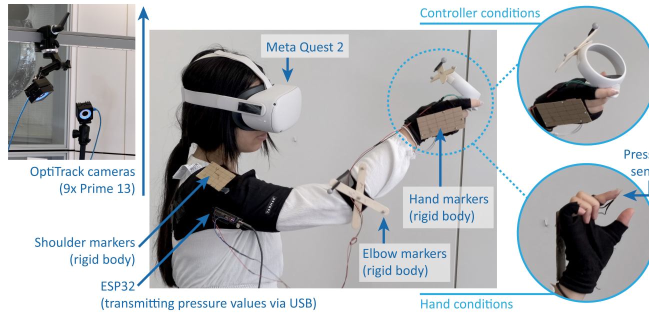

Fig. 2: Our study apparatus comprised a Meta Quest 2, nine OptiTrack Prime 13 cameras, four rigid bodies (16 markers) to track the participants' hand, controller, elbow, and shoulder, and a force sensitive resistance sensor on the index fingertip in the raycast setting.

3.1 Input Modalities and Apparatus

Figure 2 shows an overview of our setup. We conducted the experiment in a $3 \times 3m^2$ region inside an office space. Strict safety and sanitary protocols were followed to ensure COVID-19 compliance according to the local regulations, including wiping and sanitizing the VR equipment and sensors before and after each participant.

3.1.1 Tracking system

During the study, participants (all right-handed or ambidextrous) were equipped with a Meta Quest 2 headset and used in turn, the right **Meta Touch** (i.e., *VR controller*) and their right **bare hand** (see Fig. 1 and 2). We chose the Meta Quest 2 as it is one of the most popular VR headsets on the market [41], and the Meta Touch, as it is one of the lightest VR controllers (126 g without batteries).

To ensure accurate tracking of participants' motions throughout the experiment, we configured a 9-camera *OptiTrack Prime 13 system*, which has a 3D accuracy of $\pm .20$ mm [1], to continuously report the position and orientation of users' hands, the controller, and their arms using rigid bodies (see Fig. 2). We based participants' instrumentation on prior work [49, 65]. We designed these rigid bodies from sturdy cardboard and attached them to each participant's right shoulder, elbow, hand, and controller using a shoulder pad, Velcro straps, and a glove, respectively. Therefore, each participant had a total of 16 passive markers attached that reflected their motions (through 3D position and 3D orientation). We attached another 4 markers on a table outside of the users' motor space (but within the Meta and OptiTrack systems' view) to calibrate the OptiTrack system and the VR environment.

Participants performed all tasks while standing. A marker on the floor indicated where participants had to stand to ensure the best tracking quality. They were instructed to stay in place once a trial started while remaining comfortable.

We used a 5 m long Meta Link cable and made sure the cables and rigid bodies did not interfere with participants' arm and finger movements using the existing velcros and straps. The experiment ran on an Intel Core i7-9700K CPU 3.90 Hz computer with 32 GB of RAM, supported by an NVIDIA GeForce RTX 3070 GPU.

3.1.2 Selection activation

In the *raycast* condition, participants triggered selections by pulling the trigger button of the Meta Touch (*VR controller*), and by performing a pinch gesture using *free-hand interaction*. To avoid erroneous pinch detection, we affixed a 1 mm thick force-sensitive resistance (FSR) sensor under the participant's right index finger using double-sided tape. This allowed us to precisely detect participants' pinches and assess the pressure exerted on their index finger in the raycast setting. The pinching threshold was set manually from pilot studies. All pinch events were recorded no matter how subtle. The FSR sensor connected to an ESP32 board, which transmitted the pressure values directly to the computer via USB (see Fig. 2).

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3320211

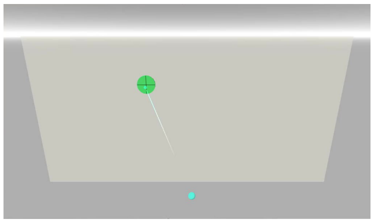

Fig. 3: View of the virtual environment in the *raycast* condition during the *selection task*. Participants' virtual input (hand or controller) was represented by a small semi-transparent white sphere, which turned blue when activating the selection. The target turned green for .1s when selected.

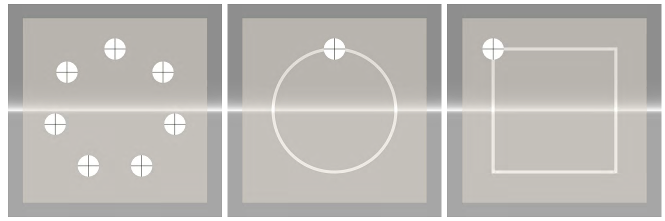

Fig. 4: (Left) Targets arrangement in the *selection task*. Only one target was displayed at a time. (Middle and right) In the *trajectory tracing task*, participants had to follow a target moving along the shown lines in one direction, then the other. Only the target was shown to the participants (not the traced trajectory).

3.1.3 Virtual Environment

The virtual environment was an empty gradient gray environment, developed in Unity 3D. The UIs were placed in the virtual environment based on the participants' arms length and shoulder position.

We chose to display a simplified representation of the participants' hands as previous work found it to result in a higher sense of agency [6]. Therefore, participants' right hands and controllers were represented by the same 5 cm diameter semi-transparent white virtual sphere. It was attached to either the participants' right palm or to the controller trigger button. The virtual sphere turned transparent blue when users activated the selection (see Fig. 3).

To make sure that the virtual environment was well calibrated, we asked participants to confirm that the virtual representation of their hands or controllers matched their physical motion before each trial. In our experiment, the environment and the objects participants saw in VR were solely driven by the output of the OptiTrack system, including the position and orientation of the virtual representation of participants' hands and controllers. Our apparatus did not rely on the inside-out tracking inside the Quest 2 to ensure consistency and sustained tracking accuracy across conditions.

3.2 Interaction Techniques

The experiment compared *VR controllers* and *free-hand interaction* in two INTERACTION TECHNIQUE conditions: *touch* and *raycast* (see Fig. 1). Stimuli was presented on a semi-transparent rectangular canvas, placed and scaled depending on the participants' arms length and conditions. Previous work found an increase in physical exertion in direct manipulation settings when the stimuli were placed at participants' shoulder level [17, 22, 55]. As we are interested in the difference in physical exertion elicited by the two INPUT MODALITIES, we chose a similar setting. The task plane was placed in front of the participant's shoulder at an $A_{touch} = armlength$ and $A_{raycast} = armlength + 1$ (in m) distance in the touch and raycast conditions, respectively.

Touch condition. The participant's task was to reach out and touch the stimuli with either INPUT MODALITY. Following the pilots, we fixed the target width to $W_{touch} = 7$ cm for all participants in the touch condition.

Raycast condition. In this condition, both hand and controller-based raycasting use the "raycasting from the eye" method described by Argelaguet and Andujar [4]: the ray originates from the Cyclops' eye [65] and points in a direction determined by the user's hand or controller. The direction was set using the OptiTrack rigidbody placed on the controller and the user's hand. The start of the visualized ray (see Fig. 3) corresponded to the center of the participant's right palm in the case of *free-hand interaction*, and on top of the trigger button for the *VR controller* input. A point marked the intersection point between the ray and the virtual environment. The intersection position was smoothed with a 1€ filter [16]. We determined the smoothing parameters empirically through pilots with four participants to ensure that there was no unintended movement from the Heisenberg effect—a change in the tracker position due to the confirmation action [12]. Participants triggered selection with the trigger button of the Meta Touch with the index finger for *VR controller*, and a pinch gesture for *free-hand interaction*.

Because the displayed elements were placed further away in the raycast condition, we determined the targets' size in the raycast condition $W_{raycast}$ by multiplying the original size of the targets in the touch condition W_{touch} by $f = (armlength + 1)/armlength$ to keep the Fitt [25]'s Law index of difficulty (ID) constant. We obtained f by solving the equation $ID_{raycast} = ID_{touch}$ with $ID = \log_2(A/W + 1)$; A is the movement amplitude (distance to the target) and W is the target width.

3.3 Tasks

Participants performed two tasks in the study: a *selection task* and a *trajectory tracing task* (see Fig. 4). The tasks are representative abstractions of tasks commonly performed in graphical interfaces [26]. We chose 2D tasks due to the pervasiveness of window-based interactions in 3D UIs [24, 44] and because they have been extensively used in hand-eye coordination and mid-air interactions studies [7, 8, 10, 17, 32, 35, 47, 68]. Participants performed each task twice in a row, taking 2 min per trial. Between trials, participants took a 3 min break.

Selection task. The experiment arrangement targets based on implementing the ISO 9241-9 standard [34], replicating the setup of previous experiments on mid-air selection performances [17, 35]. Our interface comprised seven circular targets arranged in a circle on a plane, with cross-hairs indicating the centers of each target (see Fig. 4). One target was displayed at a time and the sequence of selection was predefined as in prior work [17, 35].

Participants were instructed to touch as many targets as possible while ensuring good accuracy [35]. The selection was triggered when participants performed the appropriate action of their input (*touch* condition: by intersecting their hands or controllers with the task plane; *raycast* condition: by performing a pinching gesture or by pulling the trigger button of their controller). Selections counted as successful if the intersection point was within the target during input activation (i.e., first contact point; not when releasing). Audio-visual feedback accompanied successful and erroneous events; targets turned green or gray, respectively, for .1 s with adjusted sound effects before the task proceeded to the next target.

Trajectory tracing task. We implemented a pursuit task as conducted in previous work [8, 47]. Each trial consisted of tracing either a circle or a square (see Fig. 4) for 2 min and was divided into two parts. Participants were instructed to aim at the target, select it, then follow it as closely as possible while maintaining the selection action. First, participants traced each shape in one direction for 1 min. Then, they traced it in the other direction for the remaining minute. The order of the tracing direction was randomized per trial.

A text countdown in the center of the task plane informed participants 3 s before the direction changed. A circular target with a cross-hair indicated where participants needed to aim. The target started moving at a speed .25 m/s from the moment participants activated the selection for the first time. A gradient blue trail line of 1 cm indicated where participants were tracing in the touch condition (1 s of trail). The target turned green when the intersection point was within the target.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3320211

Table 1: Summary of the experimental design. There were two TRIALS per condition.

Independent Variables

PARTICIPANT	24	(random variable)	
INPUT MODALITY	2	*free-hand interaction*, *VR controller*	(within-subject)
INTERACTION TECHNIQUE	2	*touch*, *raycast*	(between-subject)
TASK	2	*selection*, *trajectory tracing*	(within-subject)

Dependent Variables

Self-reports	Physical exertion [0–10], sense of agency [1–7], input modality preference per task
Selection task perf.	Selection time, selection offset, error rate
Trajectory task perf.	Trajectory offset, number of strokes, tracing time
Behavioral metrics	Total movement, mean movement position, mean movement space, the pressure exerted on the index finger (mean, SD; only for *raycast*), target approach speed (mean, SD; only for *selection* × *touch*)
Ergonomic metrics	Consumed endurance, RULA

3.4 Measures

We analyzed the following metrics:

Self-reports. After each trial, we asked participants to evaluate their physical exertion level using the Borg's CR10 scale [11] (i.e., "How physically exerting was it to use?", from 0–not at all to 10–extremely hard) and their sense of agency (i.e., "From 1–7, how much did you feel in control?"). At the end of the study, participants selected their preferred input method (*free-hand interaction* or *VR controller*) for each task (*selection* and *trajectory tracing*).

Selection task performance. We recorded the average target selection time (excluding missed inputs), input accuracy (i.e., offset between participants' input and the target position on the task plane), error rate (number of errors divided by total number of activations), and throughput based on effective measures [9] for each trial.

Trajectory task performance. We assessed the mean input accuracy (i.e., mean offset between participants' input and the target on the task plane), sampled at 90 Hz during each trial, the number of strokes participants used, and the ratio of time during which they maintained the tracing interaction (each trial lasted 2 min, and participants were instructed to maintain the interaction as much/long as possible).

Behavioral metrics. To characterize users' behaviors, we logged all OptiTrack positions and orientations of the users' right shoulder, arm, wrist, knuckle and index finger at a 90 Hz sampling rate. From this, we quantified movement with the following metrics: *total movement*, *movement position*, and *movement space*.

Total movement refers to the accumulated angular movement of a given joint. We compute this for participants' *upper arm* (i.e. the accumulated angular difference between vectors pointing from the shoulder to the elbow through time), *forearm* (i.e. the accumulated angular difference between elbow to wrist vectors), and *hand* (i.e. the accumulated angular difference between the forward vectors of the hand, computed as the vector from the bottom of the wrist to the knuckle of the middle finger).

We additionally compute a mean *movement position* for participants' right *elbow* and *hand* for each trial (i.e. mean 3D position relative to their head), as well their *movement spaces* (i.e. amount of space occupied by movement). We compute movement position metrics always relative to participants' heads. To compute the *movement space* values, we firstly perform a principal component analysis of its 3D positions during the task. We then use the first two principal components to fit an ellipse. The *movement space* is set to the area of this ellipse.

Lastly, we measured the pressure exerted on the index finger in the *raycast* condition, which we standardized per participant. We also assessed the input *mean approach speed* to the targets in the *touch* × *selection* condition. We calculate this based on the input position 1 frame (≈11 ms) before the selection event and the position of the input registered during the selection.

Ergonomic metrics. From the right arm poses, we computed the mean consumed endurance [32] and Rapid Upper Limb Assessment (RULA) [50] scores. Consumed endurance uses shoulder torque as an index for muscle strain [32] and RULA assigns posture scores depending on joint angles to assess risk factors associated with upper-limb disorders [50]. For both metrics, low scores reflect low strains or minimal risk factors while higher numbers indicate high muscle strains or increasing risks.

3.5 Experimental Procedure

The experiment consisted of four parts totaling around 55 min:

1) Written consent and instructions. Participants filled out a consent form and signed a COVID-19 statement prior to the experiment. The experimenter then introduced them to the study, the equipment that was involved, and the data we recorded (which was anonymized). Afterwards, participants filled out the pre-questionnaire to assess their experience with VR, state of alertness, and demographic profile.

2) Training. The experimenter equipped participants with the OptiTrack markers, FSR sensor (for participants in the *raycast* group), and the VR headset. We asked participants to stand on the marker on the floor and to fixate in one direction. Then, the VR environment and OptiTrack system were calibrated. Participants stood still in different poses (T-pose, arm at 90°, and arm along the body) to register their shoulder position and arm's length and to calibrate the task placement. The participants tried each condition, once, in the same order as during the experiment (INPUT MODALITY and TASK). Once they felt comfortable with the inputs and tasks (at least 30 s per condition), they were asked to take a break for 3 min (no VR headset, seated). Participants were asked to report any questions they might have during this training phase.

3) Experiment. Before *each* TRIAL, we calibrated the VR environment and OptiTrack system. We asked participants to confirm that the INPUT MODALITY matched their movements. Then, we calibrated the task placement. After each TRIAL, participants reported their level of physical exertion and their sense of agency in VR. Between each TRIAL, participants took off the VR HMD and rested for 3 min to control for contamination effects. Participants were always able to request longer breaks if needed.

4) Debriefing. After the final condition, participants were asked to report which INPUT MODALITY they preferred for each task (i.e., for selection and trajectory tracing). They were then encouraged to report any thoughts and feedback they might have had.

3.6 Experimental Design

The experiment followed a 2 × 2 × 2 × 2 mixed experimental design with the following independent variables: INPUT MODALITY, INTERACTION TECHNIQUE, TASK, and TRIAL.

As summarized in Table 1, each participant solely experienced one of the INTERACTION TECHNIQUES (i.e., touch or raycast, balanced groups) to limit the length of the study. We assigned participants in each group based on their age, VR, and hand-tracking experience to obtain comparable groups. The other factors were within-subjects. The experiment was divided into two INPUT MODALITY blocks, in which participants completed both TASKS twice in a row. The order of INPUT MODALITY and TASK (in block) was counterbalanced. A TRIAL consisted of 2 min of performing the selection task (identical for the two trials) or 2 min of trajectory tracing (either circle or square; participants traced both shapes in a pseudo-randomized order in each INPUT MODALITY block).

3.7 Participants

The study was approved by the local ethics committee. Participants were 18–70 years old and right-handed or ambidextrous. They should neither have had any COVID-19 symptoms nor have been in contact with confirmed cases in the previous 14 days. They should not have known physical disabilities nor feel strain in their arms. Volunteered participants received chocolates for their time.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3320211

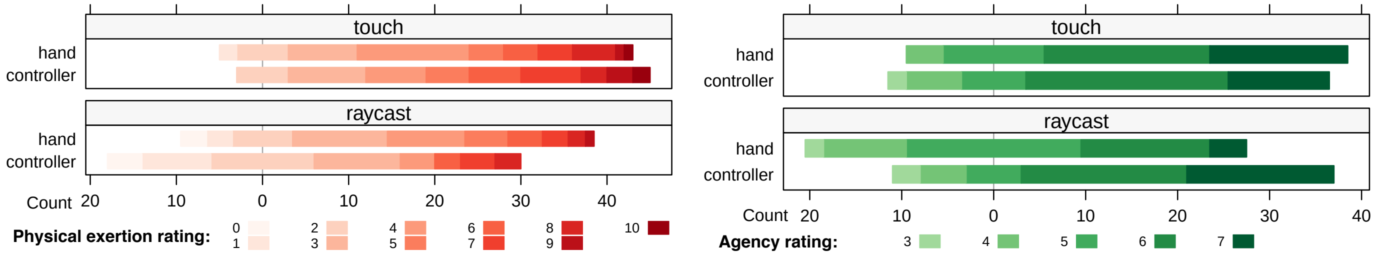

Fig. 5: (Left) Effect of INPUT MODALITY on physical exertion: scores are lower for *VR controller* in the *raycast* condition ($p < .01$). (Right) Effect of INPUT MODALITY on agency: scores are higher for *VR controller* in the *raycast* condition $p < .001$.

In a pre-questionnaire, participants reported their demographic information (age, gender, nationality), prior experience with VR technology, controller, and hand tracking on a 5-point Likert scale (from 1—never to 5—more than 20 times), frequency of playing games (from 1—never or occasionally to 5—at least once a week), their level of alertness using the Stanford Sleepiness Scale [33] (from 0—asleep to 7—active, vital, alert, or wide awake), and their level of physical activity using the International Physical Activity Questionnaire - Short Form [21] (1—low, 2—moderate, 3—high levels of physical activity, based on seven items related to walking, moderate, and vigorous activities in the last seven days).

We recruited 24 participants from places around ETH Zürich. Table 2 lists the key sample characteristics.

Table 2: Studied variables and characteristics of the analyzed study sample. Age is summarized as M (SD). Ratios are provided for gender and level of physical of activity. Following ordinal variables are summarized with a median value.

Variable	**Touch** ($N = 12$)	**Raycast** ($N = 12$)
Age [23–41] (in years)	28.58 (4.9)	28.58 (5)
Gender		
% Male	83.3 % ($n = 10$)	50 % ($n = 6$)
% Female	16.7 % ($n = 2$)	41.7 % ($n = 5$)
% Non-binary	0 %	8.3 % ($n = 1$)
Level of physical activity [21]		
% Low physical activity	0 %	16.7 % ($n = 2$)
% Moderate physical activity	66.6 % ($n = 8$)	50 % ($n = 6$)
% High physical activity	33.3 % ($n = 4$)	33.3 % ($n = 4$)
VR experience	2	3
VR controller experience	2	3
Hand tracking experience	1.5	2
Gaming frequency	2	2.5
Level of alertness [33]	6	6.5

4 Results

First, we compared INPUT MODALITY in terms of subjective ratings, task performance, and users' behavior. Then, we analyzed the relations between the dependent variables using Spearman correlations.

For the effect analysis, ordinal data (questionnaire ratings) was analyzed using an Aligned Rank Transform (ART) ANOVA [71]. Interval data was analyzed using a mixed ANOVA. When the assumptions about the normality of the residuals or homogeneity was violated (Shapiro-Wilk and Levene's test $p < .05$), the data was either transformed using a log or square root function or analyzed using ART analysis. For each variable, the PARTICIPANT was considered as a random factor, the INTERACTION TECHNIQUE as a between-subject factor, and all the other independent variables (INPUT MODALITY, TASK) as within-subject factors. In addition, we also tested eventual learning effects by aggregating objective data (Table 1) over non-overlapping time windows of 10 s and 30 s but did not find any significant effect of TIME on any of the objective metrics nor self-reports. When needed, pairwise post-hoc tests (Bonferroni adjusted p-values) were performed. For the sake of concision, we only report statistically significant main effects, and interaction effects that include INPUT MODALITY. The statistical analysis was performed using R.

Subjective Ratings

In summary, in the *raycast* condition, participants reported lower exertion and a higher sense of agency using *VR controllers* compared to *free-hand interaction*. We did not find any significant difference in the subjective ratings between the two input modalities in the *touch* condition. Figure 5 illustrates participants' self-reported measures.

Perceived exertion. The ART analysis showed an interaction effect between INPUT MODALITY and INTERACTION TECHNIQUE [$F_{1,22} = 14.38$, $p < .001$, $\eta_p^2 = .08$]. Post-hoc tests showed that *VR controller* was only perceived as significantly less exerting than *free-hand interaction* in the *raycast* condition ($p < .01$).

Sense of agency. Overall, participants felt more in control using the *VR controller* than using *free-hand interaction* [$F_{1,22} = 5.27$, $p = .02$, $\eta_p^2 = .03$]. The analysis also showed an interaction effect between INPUT MODALITY and INTERACTION TECHNIQUE [$F_{1,22} = 10$, $p < .01$, $\eta_p^2 = .06$]. Post-hoc tests suggest that the perceived difference between *VR controller* and *free-hand interaction* was only statistically significant in the *raycast* condition ($p < .001$).

Input Preference. All conditions considered, *VR controller* and *free-hand interaction* were preferred equally (i.e., participants ranked *VR controller* over *free-hand interaction* in 24 out of 48 instances, and vice versa in the remaining instances). However, in *touch* conditions, participants generally preferred using *free-hand interaction* ($N_{touch} = 19/24; N_{traj \times touch} = 10/12; N_{select \times touch} = 9/12$). In *raycast* conditions, they contrarily preferred *VR controller* ($N_{raycast} = 19/24; N_{traj \times raycast} = 11/12; N_{select \times raycast} = 8/12$).

Task Performances

In summary, in the *raycast* condition, participants were faster in the *selection* task, and more accurate in the *trajectory tracing* task using *VR controllers* compared to *free-hand interaction*. Figure 6 showcases the effect of INPUT MODALITY and INTERACTION TECHNIQUE on task performances.

Selection task performance. Overall, participants were on average 6.2% faster using *VR controllers* than *free-hand interaction* [$F_{1,22} = 4.68$, $p = .04$, $\eta_p^2 = .02$]. The analysis also showed an interaction effect between INPUT MODALITY and INTERACTION TECHNIQUE [$F_{1,22} = 7.59$, $p = .01$, $\eta_p^2 = .03$]. Post-hoc tests showed that participants were only significantly faster (by 11.8%) using *VR controllers* compared to *free-hand interaction* ($p < .01$) in the *raycast* condition.

We did not find any other significant result for selection performance but report the error rate and throughput for comparisons with past works. The mean error rate for the *touch* and *raycast* conditions is, respectively, $7.3\% (SD = 3.8\%)$ and $8.1\% (SD = 6.2\%)$ (aggregating *free-hand interaction* and *VR controller* data). The average effective throughput for *touch* and *raycast* is respectively $3.5(0.56)$ bps and $4.71(1.05)$ bps, which matches the range of past ISO 9241-9 throughput reported in the literature in VR [9].

Trajectory mean offset. We found an interaction effect between INPUT MODALITY and INTERACTION TECHNIQUE [$F_{1,22} = 10.90$, $p < .01$, $\eta_p^2 = .06$]. Post-hoc tests showed that in the *raycast* condition, participants were on average 18.9% more accurate using the *VR controller* than using *free-hand interaction* ($p < .001$).

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3320211

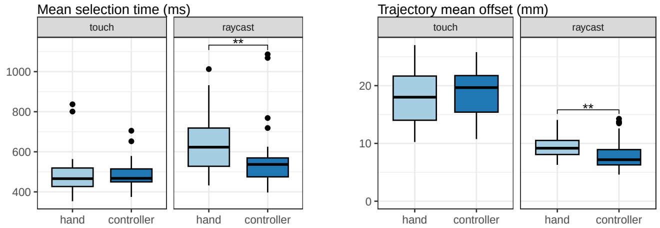

Fig. 6: Effect of input modality on task performance. The error bars correspond to the standard errors (participant × trial). The p-values were obtained via pairwise post-hoc test comparisons (Bonferroni correction) on transformed or aligned-and-ranked data for the variables that did not meet the ANOVA assumptions. Significances: $^{**}p < .01$.

Behavioral Metrics

Across all conditions, when participants used *free-hand interaction* as opposed to *VR controller*, they adopted a position with both their elbow and hand raised higher and further forward. They also moved their upper arm more and their elbow across a greater space. When specifically performing *raycast*, participants positioned their hand higher up, moved their hand across a greater space, and applied more pressure during activations when using *free-hand interaction*. When performing *touch*, they were faster at approaching targets and moved their hand across a greater space with a *VR controller*. Lastly, they moved their forearm more when performing selections with *free-hand interaction* than with *VR controllers*.

Total upper arm movement. Participants moved their upper arm on average 20% more when using *free-hand interaction* compared to with the *VR controller* [$F_{1,22} = 6.45$, $p = .02$, $\eta_p^2 = .04$].

Total forearm movement. We found an interaction effect between INPUT MODALITY and TASK [$F_{1,22} = 9.7$, $p < .01$, $\eta_p^2 = .02$]. Participants moved their forearm 21.9% more with *free-hand interaction* than with the *VR controller* during the *selection task* ($p = .02$).

Elbow position. Participants held their elbow 5.7% higher [$F_{1,22} = 38.15$, $p < .001$, $\eta_p^2 = .06$] and 14.1% further forward [$F_{1,22} = 11.87$, $p < .01$, $\eta_p^2 = .11$] when using *free-hand interaction* compared to *VR controller*.

Hand position. Participants generally positioned their hand 14.3% higher up (halving the vertical distance from the headset) [$F_{1,22} = 126.53$, $p < .001$, $\eta_p^2 = .30$] and 66.2% further forward [$F_{1,22} = 8.46$, $p < .01$, $\eta_p^2 = .06$] when using *free-hand interaction* compared to with *VR controllers*. For the vertical position of the hand, the analysis also revealed an interaction effect between INPUT MODALITY and INTERACTION TECHNIQUE [$F_{1,22} = 79.99$, $p < .001$, $\eta_p^2 = .21$]. Post-hoc tests suggest that participants only placed their hand higher using *free-hand interaction* in the *raycast* condition ($p < .0001$) (by 29.4%).

Elbow movement space. Participants moved their elbow around a greater space (26% more) when using *free-hand interaction* compared to with *VR controller* [$F_{1,22} = 8.18$, $p < .01$, $\eta_p^2 = .05$].

Hand movement space. Participants generally moved their hand around a greater space (62.4% more) when using *free-hand interaction* compared to with *VR controllers* [$F_{1,22} = 16.19$, $p < .001$, $\eta_p^2 = .12$]. The analysis also showed an interaction effect between INPUT MODALITY and INTERACTION TECHNIQUE [$F_{1,22} = 49.88$, $p < .001$, $\eta_p^2 = 29.3$]. Post-hoc tests showed that participants moved their hand around a greater space (239.3% more) when using *free-hand interaction* compared to with *VR controllers* in the *raycast* condition ($p < .0001$) while they moved around a greater space (41.4% more) with *VR controllers* compared to with *free-hand interaction* in the touch condition ($p = .043$).

Targets approach speed. In the *touch* condition, participants approached the selection targets 40% faster using the *VR controller* than when they used *free-hand interaction* [$F_{1,22} = 252.55$, $p < .001$, $\eta_p^2 = .85$].

Pressure exerted on the finger. We only assessed the pressure exerted on the index finger in the *raycast* condition. We found that participants exerted 85.9% more force on their index finger while pinching using *free-hand interaction* than while pressing the trigger button of the *VR controller* [$F_{1,22} = 23.06$, $p < .01$, $\eta_p^2 = .42$].

Ergonomic Assessments

In summary, for *raycast* conditions, *free-hand interaction* yielded a higher RULA score but *VR controller* yielded a higher CE score. For *touch* conditions, *free-hand interaction* yielded a higher CE score. Using *free-hand interaction* also yielded a higher CE score in *trajectory tracing* conditions.

Consumed endurance. We found an interaction effect between INPUT MODALITY and INTERACTION TECHNIQUE [$F_{1,22} = 26.39$, $p < .001$, $\eta_p^2 = .06$], and between INPUT MODALITY and TASK [$F_{1,22} = 12.84$, $p < .01$, $\eta_p^2 = .02$]. Participants' muscle strain in accordance with consumed endurance score was greater using *free-hand interaction* in the *touch* setting ($p < .01$) and it was greater using the *VR controller* in the *raycast* condition ($p < .001$). The consumed endurance score was significantly greater using *free-hand interaction* during *trajectory tracing* ($p < .001$).

RULA. Participants' RULA score was higher using *free-hand interaction* than using the *VR controller* [$F_{1,22} = 25.36$, $p < .001$, $\eta_p^2 = .13$]. We also found an interaction effect between INPUT MODALITY and INTERACTION TECHNIQUE [$F_{1,22} = 5.02$, $p = .04$, $\eta_p^2 = .03$]. In the *raycast* setting, participants' RULA score was significantly greater using *free-hand interaction* than using the *VR controller* ($p < .0001$).

4.1 Correlation Analysis

In order to understand subjective ratings and performance scores, we analyzed how physical exertion, sense of agency, and task performances correlate against each other, and against behavioral metrics using Spearman correlation.

Generally, the reported physical exertion level was negatively associated with the reported sense of agency ($\rho = -.18$, $p = .04$).

The correlations between subjective ratings and task performances are given in supplementary materials. Overall, higher level of exertion was associated with faster selections ($\rho = -.29$, $p = .03$), lower trajectory tracing accuracy (increase in offset ($\rho = .41$, $p < .001$), lower amount of strokes necessary to complete the trajectory tracing task ($\rho = .36$, $p < .01$)), and lower time during which participants maintained the tracing interaction ($\rho = -.24$, $p < .001$). We did not find any significant correlation between agency and task performance.

The correlations between subjective ratings, task performances, and behavioral metrics are detailed in supplementary materials.

In summary, exertion was positively associated with bad ergonomic posture (averaged consumed endurance score per trial), higher variation in the exerted pressure during pinching or trigger pulling, higher variation in the approach speed in the *touch* × *selection* condition with a controller, higher forward position values of the hands in the *touch* condition, higher hand movement space in the *raycast* condition, and higher vertical and lower forward position values of the participant's hand in the *raycast* condition.

Agency was positively associated with bad ergonomics posture with regards to consumed endurance and RULA scores, and higher variation in the controller approach speed to the selection targets. We also found a positive correlation between agency and the relative position of the elbow to the right in the *touch* condition.

Faster selection was associated with a greater amount of arm motion (upper arm, forearm, hand rotation), bad postures (consumed endurance, RULA), and lower pressure and variation in the pressure exerted on the index finger. Selection accuracy (lower offset) was positively associated with lower amount of hand rotation, and good ergonomic posture (consumed endurance).

Trajectory accuracy increased with lower amounts of arm motion (upper arm, forearm), and good postures (consumed endurance, RULA). The number of strokes and trajectory tracing time ratio followed the same trends as the trajectory accuracy. Tracing time was also negatively associated with the variation in pressure exerted on the finger.

We do not discuss the correlation between the variation in the pressure exerted on the finger, the selection speed, and the tracing time as these relations were likely caused by the increased amounts of INPUT MODALITY selection activation.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3320211

5 Discussion

In this paper, we conducted an empirical study with 24 participants to compare *VR controllers* vs. *free-hand interaction* for mid-air interactions across two settings (*touch*, *raycast*) and using two tasks (*selection*, *trajectory tracing*).

5.1 VR Controllers vs. Bare Hands

In our study, *VR controllers* overall outperformed *free-hand interaction* in terms of sense of agency, task speed, motor space, and ergonomic scores. In addition, *VR controllers* outperformed *free-hand interaction* in terms of physical exertion, preference, and task accuracy in the *raycast* setting. These results are in line with past work [15, 29, 42]. When using *raycast*, our participants had to use the controller button or a hand gesture to interact with the virtual content, similar to prior work. However, past studies all compared these input modalities for direct interaction with 3D objects, so we extend these results to raycast for selection and pursuit tasks on a plane in a 3D environment. We relied on robust and precise tracking systems in our study and obtained similar findings as past works.

Contrary to past studies [15, 29, 42], we did not find any significant difference between *VR controllers* and *free-hand interaction* subjective ratings and task performance in the direct interaction setting (i.e., *touch* condition). While past results balanced in favor of controllers for direct manipulation, some works explained the bad performances of controller-free input with the unreliability of the hand tracking system [28, 39]. In our experiment, we controlled for this confound with sub-millimeter tracking accuracy using a high-end motion capture system. We thus expected *free-hand interaction* to outperform the *VR controller* in the *touch* condition, as we relied on the same high-end external tracking system and abstract virtual hand representations for both input modalities. Moreover, this setting did not include any selection activation (i.e., no pinching) and controllers are heavier than bare hands, which should have fatigued participants more, particularly since they were in mid-air with no support [19]. Despite the weight of the controllers, we did not find any significant difference between both inputs in terms of task performance and subjective ratings in the touch setting. However, participants preferred free-hand input to the VR controller in the touch setting and our data slightly balances toward better task performance and self-reported measures for free-hand interaction (Figures 5 and 6).

Most of our observed quantitative difference lies in the behavioral metrics. In the *touch* condition, participants approached the selection targets with greater speed with the *VR controller* than the *free-hand interaction* and they moved their hands over a smaller space with *free-hand interaction*, suggesting a higher level of apprehension in approaching virtual interfaces with tracked fingers rather than with controllers. Overall, this might reflect a greater control of the input and a stronger sense of embodiment, respectively, which might be the origin of users' preference for free-hand manipulation in this setting.

In summary, while current developments in the consumer market indicate a trend toward free-hand interaction and a future without handheld controllers, motivated likely due to costs and system complexity, our findings suggest that this development is counterproductive for user interaction; in our experiment, free-hand interaction was less efficient, accurate, and led to more physical exertion, and produced a lower sense of agency than VR controllers. Our findings reveal that this difference mostly occurs during raycast—which is the dominant selection method in contemporary VR environments due to its limitless space availability [5]. To avoid additional information processing and friction in participants' handling of the task [13], we selected pinching as a straightforward, well-accepted, and unambiguous hand gesture for input, which is also commonly used in commodity VR systems. Still, however, we observed performance differences in favor of controllers.

Several interpretations may explain these results. Even with optimal tracking and gesture recognition, pinching still requires users to perform more finger movement than when pressing buttons, which can increase the task completion time. In our study, participants moved their hands and arms more and overall over a larger space during free-hand interaction than when they were using controllers, which can also explain the difference in physical exertion ratings. This increase in motion can stem from the raycast implementation for free-hand interaction, which we based on previous work [4]. We piloted the apparatus with novice VR participants to confirm that the ray angle would be natural. However, while less intuitive, other raycast implementations (e.g., other position and angle mapping) or indirect interaction techniques (e.g., relying on eye gaze and pinch [60]) could improve task performance and lower the users' arm strain. Beyond the additional movement required to perform a pinch, these task performance results can also be explained by the fact that people are not used to performing indirect interactions without any support and haptic feedback. Indeed, while humans are used to reaching out to touch physical objects with their hands, indirect interactions are more commonly performed through the use of devices in the real world (e.g., mouse, remote TV control, game controllers).

5.2 Understanding Physical Exertion and Task Performance

The second objective of this study was to understand the relationship between subjective ratings, task performance, and participants' behavior during VR interaction.

Unsurprisingly, we found that more arm motion and hand movement space were associated with more physical exertion, faster selection, lower task accuracy, and lower tracing time.

We also found a positive correlation between physical exertion and the amount of interaction (i.e., number of selections or task speed) and a negative correlation between physical exertion and task accuracy. While we cannot infer the causality between these dependent variables, a possible interpretation is that the more participants interacted, the more they were physically exerted, which in turn made them less accurate. Participants also felt less in control when they were more physically strained, which is in line with this interpretation. Further studies could focus on investigating these relations.

Interestingly, we found that an increase in physical exertion was associated with an increase in the variation of the participants' approach speed (i.e., when touching virtual targets). Future work could focus on this variation in the users' motion for the detection of physical exertion during direct mid-air interactions.

5.3 Other Interesting Results

While participants rated the *selection task* as more physically exerting than the *trajectory tracing task*, the ergonomic metrics reveal that participants' average postures were worst during the *trajectory tracing task* than during the *selection task*. We found that participants performed more motion during the *selection task* than during the *trajectory tracing task*, but they took more space during the *trajectory tracing task* than during the *selection task*. Therefore, beyond static considerations of arm and body postures and amplitude of arm movement, the dynamic and fast interaction of the *selection task* might have put a bigger toll on the users' muscle strains than continuously maintaining arm poses during the *trajectory tracing task*, which could not be revealed with the ergonomics assessment we chose. Future work could explore new ways to detect physical exertion using dynamic behavioral measurement by relying, e.g., on the variation in the participants' motion speed or input behavior.

5.4 Implications for Design

Summarizing our findings in short, we have gained insights into which INPUT MODALITY is more suitable in VR for which mid-air INTERACTION TECHNIQUE:

1. Use *VR controller* for *raycast* interactions (more speed, accuracy, control, less physical exertion).
2. Prefer *free-hand interaction* for direct *touch* interactions.
3. Use *VR controller* when limited interaction space is available.
4. Limit high hand positions, upper arm rotations, and fast interactions in case of high levels of physical exertion detected.

5.5 Benefits, Limitations, and Future Work

Above all, our results show that depending on the application context, free-hand manipulation may be a preferable input method compared to controllers, especially for direct interaction and in situations where high levels of body ownership are required (e.g., for therapy or psychological studies [3, 56, 69]). Alternatively, controllers highly benefit training and games for VR interactions, as users are often required to hold onto physical props in these applications and since the actuators can help immerse users into the virtual world through multi-sensory feedback [20, 70]. Future controllers may evolve and take other shapes, e.g., become lighter and less cumbersome while providing haptic feedback, which would make them suitable for VR direct interactions with mid-air interfaces. At the same time, future designs of VR interaction may leverage controller techniques to decrease the needed control space to prevent fatigue while maintaining input performance [40, 61]. Further investigations could extend this study to other controllers with different weights, weight distribution, shape factors, and grip styles.

Our findings result from an evaluation of human factors under submillimeter tracking apparatus, using a multi-camera OptiTrack system. In contrast, headset-based inside-out tracking faces several challenges in practice that affect tracking accuracy [66], including self-occlusion, camera resolution, and coverage, and may thus not deliver the same accuracy. Future work could focus on improving this technology by using other modalities, such as gaze tracking, inertial measurement units (e.g., inside wrist-worn smartwatches or bands [66]), or physiological signals to provide non-invasive inputs for efficient and accurate VR interactions.

Human-computer interaction typically seeks to adapt computers to human needs, but humans also adapt to the computer [54]. As mentioned earlier, users are currently not used to free-hand manipulation techniques, whereas they regularly use controllers for indirect interactions, such as in the aforementioned scenarios involving mice, remote controllers, or other tangibles. In the future, a majority of users might be more efficient at performing hand gestures and motions, which may improve their free-hand interaction task performance and make it a better-suited modality for at-a-distance interaction. To go toward that direction, future studies could investigate the influence of these interactions with exclusively expert VR or Mixed Reality users. Furthermore, while we did not observe any significant main or interacting learning effect in our study—possibly attributable to the mitigating influence of physical exertion, subsequent investigations could explore exposure over extended periods to model free-hand interaction learning curves.

We limited our work to 2D selection and tracking tasks in VR for different mid-air interaction techniques. Since the literature already covered direct manipulation with 3D objects, future work could focus on more complex 3D interactions (e.g., scaling or rotating 3D objects) at-a-distance to compare both input modalities and new gestures with non-linear interaction techniques such as Go-Go [61]. The effect of the selection method and the raycasting technique could also be separated to investigate eventual bottlenecks and optimum selection-raycast technique combinations.

Further studies could also extend this work to Mixed Reality interactions for which hand tracking is, by far, the most used input modality and might not be adapted for efficient mid-air interactions.

6 Conclusion

We have presented the results of a controlled user study with 24 participants that compared VR controllers and free-hand input during mid-air interaction using touch and raycast for input during selection and trajectory tracing tasks.

We found that controllers outperformed free-hand interaction in terms of sense of agency, task speed, and movement space, especially in the raycast environment where controllers were also less physically exerting and more accurate than free-hand interaction. Controllers were preferred in the raycast setting, and free-hand interaction was preferred in the touch setting. Our results suggest that controllers are overall more adapted for precise, accurate, and ergonomic interactions in a raycast setting, while free-hand input is more suitable for direct touch interaction.

We also showed that physical exertion was positively associated with task speed, movement space, quantity of motion, hand vertical motion, and variation in the users' motion speed. Our results also suggest that physical exertion lowers the sense of agency and decreases task accuracy.

Based on our results, we have provided a set of guidelines on the type of input recommended for use depending on the interaction setting. Overall, we expect that our findings can inform and contribute to the design and development of future interaction techniques and the interaction design of future immersive environments.

Acknowledgments

We thank the participants of our study for their time and feedback. We also thank the AIT lab at ETH Zürich for lending us their OptiTrack system.

References

[1] https://optitrack.com/cameras/primex-13. [Online; accessed 17-June-2023].

[2] D. Abdlkarim, M. Di Luca, P. Aves, S.-H. Yeo, R. C. Miall, P. Holland, and J. M. Galea. A methodological framework to assess the accuracy of virtual reality hand-tracking systems: A case study with the oculus quest 2. *bioRxiv*, 2022.

[3] A. Adkins, L. Lin, A. Normoyle, R. Canales, Y. Ye, and S. Jörg. Evaluating grasping visualizations and control modes in a vr game. *ACM Transactions on Applied Perception (TAP)*, 18(4):1–14, 2021.

[4] F. Argelaguet and C. Andujar. Efficient 3d pointing selection in cluttered virtual environments. *IEEE Computer Graphics and Applications*, 29(6):34–43, 2009.

[5] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013.

[6] F. Argelaguet, L. Hoyet, M. Trico, and A. Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE virtual reality (VR)*, pp. 3–10. IEEE, 2016.

[7] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. W. Fitzmaurice. Experimental evaluation of sketching on surfaces in vr. In *CHI*, vol. 17, pp. 5643–5654, 2017.

[8] A. Bansal, S. Weech, and M. Barnett-Cowan. Movement-contingent time flow in virtual reality causes temporal recalibration. *Scientific reports*, 9(1):1–13, 2019.

[9] A. U. Batmaz and W. Stuerzlinger. Effective throughput analysis of different task execution strategies for mid-air fitts' tasks in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3939–3947, 2022.

[10] A. U. Batmaz, X. Sun, D. Taskiran, and W. Stuerzlinger. Hitting the wall: Mid-air interaction for eye-hand coordination. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–5, 2019.

[11] G. Borg. *Borg's perceived exertion and pain scales.* Human kinetics, 1998.

[12] D. Bowman, C. Wingrave, J. Campbell, and V. Ly. Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments. 2001.

[13] D. A. Bowman and L. F. Hodges. Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages & Computing*, 10(1):37–53, 1999.

[14] D. A. Bowman, D. B. Johnson, and L. F. Hodges. Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 26–33, 1999.

[15] G. Caggianese, L. Gallo, and P. Neroni. The vive controllers vs. leap motion for interactions in virtual environments: a comparative evaluation. In *International Conference on Intelligent Interactive Multimedia Systems and Services*, pp. 24–33. Springer, 2018.

[16] G. Casiez, N. Roussel, and D. Vogel. 1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2527–2530, 2012.

[17] N. Cheema, L. A. Frey-Law, K. Naderi, J. Lehtinen, P. Slusallek, and P. Hämäläinen. Predicting mid-air interaction movements and fatigue using deep reinforcement learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.

[18] J. Chen, A. Moemeni, and P. Caleb-Solly. Comparing a graphical user interface, hand gestures and controller in virtual reality for robot teleoperation. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '23, p. 644–648. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3568294.3580165

[19] Y. F. Cheng, T. Luong, A. R. Fender, P. Streli, and C. Holz. ComforTable user interfaces: Surfaces reduce input error, time, and exertion for tabletop and mid-air user interfaces. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 150–159, 2022. doi: 10.1109/ISMAR55827.2022.00029

[20] I. Choi, E. Ofek, H. Benko, M. Sinclair, and C. Holz. CLAW: A multifunctional handheld haptic controller for grasping, touching, and triggering in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–13, 2018. doi: 10.1145/3173574.3174228

[21] C. Craig, A. Marshall, M. Sjostrom, A. Bauman, P. Lee, D. Macfarlane, T. Lam, and S. Stewart. International physical activity questionnaire-short form. *J Am Coll Health*, 65(7):492–501, 2017.

[22] J. M. Evangelista Belo, A. M. Feit, T. Feuchtner, and K. Grønbæk. Xrgonomics: Facilitating the creation of ergonomic 3d interfaces. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–11, 2021.

[23] F. Fahmi, K. Tanjung, F. Nainggolan, B. Siregar, N. Mubarakah, and M. Zarlis. Comparison study of user experience between virtual reality controllers, leap motion controllers, and senso glove for anatomy learning systems in a virtual reality environment. In *IOP Conference Series: Materials Science and Engineering*, vol. 851, p. 012024. IOP Publishing, 2020.

[24] S. Feiner, B. MacIntyre, M. Haupt, and E. Solomon. Windows on the world: 2d windows for 3d augmented reality. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, UIST '93, p. 145–155. ACM, USA, 1993. doi: 10.1145/168642.168657

[25] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.

[26] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan. Direct-touch vs. mouse input for tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, p. 647–656. ACM, USA, 2007. doi: 10.1145/1240624.1240726

[27] J. L. Gabbard, D. Hix, and J. E. Swan. User-centered design and evaluation of virtual environments. *IEEE computer Graphics and Applications*, 19(6):51–59, 1999.

[28] T. Galais, A. Delmas, and R. Alonso. Natural interaction in virtual reality: Impact on the cognitive load. In *Proceedings of the 31st Conference on l'Interaction Homme-Machine: Adjunct*, pp. 1–9, 2019.

[29] E. Gusai, C. Bassano, F. Solari, and M. Chessa. Interaction in an immersive collaborative virtual reality environment: a comparison between leap motion and htc controllers. In *International Conference on Image Analysis and Processing*, pp. 290–300. Springer, 2017.

[30] P. Habibi and D. Chattopadhyay. The impact of handedness on user performance in touchless input. *International Journal of Human-Computer Studies*, 149:102600, 2021. doi: 10.1016/j.ijhcs.2021.102600

[31] J. T. Hansberger, C. Peng, S. L. Mathis, V. Areyur Shanthakumar, S. C. Meacham, L. Cao, and V. R. Blakely. Dispelling the gorilla arm syndrome: the viability of prolonged gesture interactions. In *Virtual, Augmented and Mixed Reality: 9th International Conference, VAMR 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings 9*, pp. 505–520. Springer, 2017.

[32] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1063–1072, 2014.

[33] E. Hoddes, V. Zarcone, and W. Dement. Stanford sleepiness scale. *Enzyklopädie der Schlafmedizin*, 1184, 1972.

[34] I. ISO. 9241-9 ergonomic requirements for office work with visual display terminals (vdts)-part 9: Requirements for non-keyboard input devices (fdis-final draft international standard), 2000. *International Organization for Standardization*, 2000.

[35] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani. Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3328–3339, 2017.

[36] J. Jankowski and M. Hachet. A survey of interaction techniques for interactive 3d environments. In *Eurographics 2013-STAR*, 2013.

[37] R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan. Eye–hand coordination in object manipulation. *Journal of neuroscience*, 21(17):6917–6932, 2001.

[38] C. I. Johnson, D. E. Whitmer, J. Entinger, E. K. Peterson, and B. M. Sobel. Interacting with virtual reality with a controller instead of the body benefits performance and perceptions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 66, pp. 1294–1298. SAGE Publications Sage CA: Los Angeles, CA, 2022.

[39] J. Kangas, S. K. Kumar, H. Mehtonen, J. Järnstedt, and R. Raisamo. Trade-off between task accuracy, task completion time and naturalness for direct object manipulation in virtual reality. *Multimodal Technologies and Interaction*, 6(1):6, 2022.

[40] M. Kari and C. Holz. Handycast: Phone-based bimanual input for virtual reality in mobile and space-constrained settings via pose-and-touch transfer. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3580677

[41] J. Kelly, T. Doty, M. Ambourn, and L. Cherep. Distance perception in the oculus quest and oculus quest 2. 2022.

[42] C. Khundam, V. Vorachart, P. Preeyawongsakul, W. Hosap, and F. Noël. A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training. In *Informatics*, vol. 8, p. 60. MDPI, 2021.

[43] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[44] J. H. Lee, S.-G. An, Y. Kim, and S.-H. Bae. Projective windows: Bringing windows in space to the fingertip. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–8. ACM, USA, 2018. doi: 10.1145/3173574.3173792

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

[46] L. Lin, A. Normoyle, A. Adkins, Y. Sun, A. Robb, Y. Ye, M. Di Luca, and S. Jörg. The effect of hand size and interaction modality on the virtual hand illusion. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 510–518, 2019. doi: 10.1109/VR.2019.8797787

[47] L. Liu and R. Van Liere. Modeling object pursuit for desktop virtual reality. *IEEE transactions on visualization and computer graphics*, 18(7):1017–1026, 2012.

[48] A. Masurovsky, P. Chojecki, D. Runde, M. Lafci, D. Przewozny, and M. Gaebler. Controller-free hand tracking for grab-and-place tasks in immersive virtual reality: Design elements and their empirical study. *Multimodal Technologies and Interaction*, 4(4):91, 2020.

[49] S. Mayer, V. Schwind, R. Schweigert, and N. Henze. The effect of offset correction and cursor on mid-air pointing in real and virtual environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–13. ACM, USA, 2018.

[50] L. McAtamney and E. N. Corlett. Rula: a survey method for the investigation of work-related upper limb disorders. *Applied ergonomics*, 24(2):91–99, 1993.

[51] R. P. McMahan, A. J. D. Alon, S. Lazem, R. J. Beaton, D. Machaj, M. Schaefer, M. G. Silva, A. Leal, R. Hagan, and D. A. Bowman. Evaluating natural interaction techniques in video games. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 11–14, 2010. doi: 10.1109/3DUI.2010.5444727

[52] M. Meier, P. Streli, A. Fender, and C. Holz. TapID: Rapid Touch Interaction in Virtual Reality using Wearable Sensing. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 519–528. IEEE, 2021.

[53] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, vol. 38, pp. 21–45. Wiley Online Library, 2019.

[54] J. Mey. Adaptability in human-computer interaction. In K. Brown, ed., *Encyclopedia of Language & Linguistics (Second Edition)*, pp. 49–56. Elsevier, Oxford, second edition ed., 2006. doi: 10.1016/B0-08-044854-2/00394-1

[55] R. A. Montano Murillo, S. Subramanian, and D. Martinez Plasencia. Erg-o: Ergonomic optimization of immersive virtual environments. In *Proceedings of the 30th annual ACM symposium on user interface software*

and technology, pp. 759–771, 2017.

[56] A. Mottelson, C. Vandeweerdt, M. Atchapero, T. Luong, C. Holz, R. Böhm, and G. Makransky. A self-administered virtual reality intervention increases covid-19 vaccination intention. *Vaccine*, 2021. doi: 10.1016/j.vaccine.2021.10.004

[57] D. Navarro and V. Sundstedt. Evaluating player performance and experience in virtual reality game interactions using the htc vive controller and leap motion sensor. In *3rd International Conference on Human Computer Interaction Theory and Applications*, pp. 103–110. SciTePress, 2019.

[58] M. Oudah, A. Al-Naji, and J. Chahl. Hand gesture recognition based on computer vision: A review of techniques. *Journal of Imaging*, 6(8), 2020. doi: 10.3390/jimaging6080073

[59] S. Pei, A. Chen, J. Lee, and Y. Zhang. Hand interfaces: Using hands to imitate objects in ar/vr for expressive interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3501898

[60] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th symposium on spatial user interaction*, pp. 99–108, 2017.

[61] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pp. 79–80, 1996.

[62] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer graphics forum*, vol. 17, pp. 41–52. Wiley Online Library, 1998.

[63] H.-R. Rantamaa, J. Kangas, S. K. Kumar, H. Mehtonen, J. Järnstedt, and R. Raisamo. Comparison of a vr stylus with a controller, hand tracking, and a mouse for object manipulation and medical marking tasks in virtual reality. *Applied Sciences*, 13(4):2251, 2023.

[64] S. Rutkowski, M. Adamczyk, A. Pastuła, E. Gos, C. Luque-Moreno, and A. Rutkowska. Training using a commercial immersive virtual reality system on hand–eye coordination and reaction time in young musicians: A pilot study. *International journal of environmental research and public health*, 18(3):1297, 2021.

[65] V. Schwind, S. Mayer, A. Comeau-Vermeersch, R. Schweigert, and N. Henze. Up to the finger tip: The effect of avatars on mid-air pointing accuracy in virtual reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '18, p. 477–488. ACM, USA, 2018. doi: 10.1145/3242671.3242675

[66] P. Streli, R. Armani, Y. F. Cheng, and C. Holz. HOOV: Hand out-of-view tracking for proprioceptive interaction using inertial sensing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581468

[67] P. Streli, J. Jiang, A. R. Fender, M. Meier, H. Romat, and C. Holz. TapType: Ten-finger text entry on everyday surfaces via bayesian inference. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3501878

[68] R. J. Teather and W. Stuerzlinger. Pointing at 3d targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 87–94. IEEE, 2011.

[69] C. Vandeweerdt, T. Luong, M. Atchapero, A. Mottelson, C. Holz, G. Makransky, and R. Böhm. Virtual reality reduces covid-19 vaccine hesitancy in the wild: a randomized trial. *Scientific Reports*, 12(1):1–7, 2022.

[70] E. Whitmire, H. Benko, C. Holz, E. Ofek, and M. Sinclair. Haptic revolver: Touch, shear, texture, and shape rendering on a reconfigurable virtual reality controller. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. ACM, USA, 2018. doi: 10.1145/3173574.3173660

[71] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 143–146, 2011.

[72] L. Yang, J. Huang, T. Feng, W. Hong-An, and D. Guo-Zhong. Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware*, 1(1):84–112, 2019.

[73] X. Zhou, Y. Jin, L. Jia, and C. Xue. Study on hand–eye cordination area with bare-hand click interaction in virtual reality. *Applied Sciences*, 11(13):6146, 2021.