Accuracy Evaluation of Touch Tasks in Commodity Virtual and Augmented Reality Head-Mounted Displays

Daniel Schneider
daniel.schneider@hs-coburg.de
Coburg University of Applied
Sciences and Arts
Coburg, Germany

Verena Biener
verena.biener@hs-coburg.de
Coburg University of Applied
Sciences and Arts
Coburg, Germany

Alexander Otte
alexander.otte@hs-coburg.de
Coburg University of Applied
Sciences and Arts
Coburg, Germany

Travis Gesslein
travis.gesslein@hs-coburg.de
Coburg University of Applied
Sciences and Arts
Coburg, Germany

Philipp Gagel
philipp.gagel@stud.hs-coburg.de
Coburg University of Applied
Sciences and Arts
Coburg, Germany

Cuauhtli Campos
cc.mijangos@gmail.com
University of Primorska
Koper, Slovenia

Klen Čopič Pucihar
klen.copic@famnit.upr.si
University of Primorska
Koper, Slovenia

Matjaž Kljun
matjaz.kljun@famnit.upr.si
University of Primorska
Koper, Slovenia

Eyal Ofek
eyalofek@microsoft.com
Microsoft Research
Redmond, Washington, United States

Michel Pahud
mpahud@microsoft.com
Microsoft Research
Redmond, Washington, United States

Per Ola Kristensson
pok21@cam.ac.uk
Department of Engineering,
University of Cambridge
Cambridge, United Kingdom

Jens Grubert
jg@jensgrubert.de
Coburg University of Applied
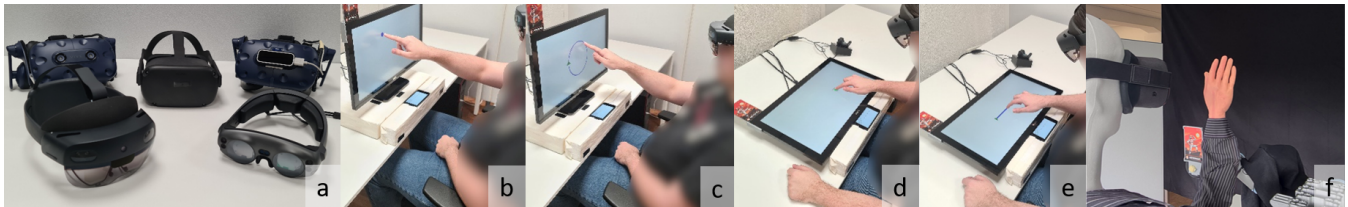Sciences and Arts
Coburg, Germany

Figure 1: a: HMDs used for the conditions: HTC Vive Pro (top left), Oculus Quest (top center), Leap Motion (top right), Microsoft HoloLens 2 (bottom left), MagicLeap 1 (bottom right). Human participants study participant conducting in Vertical orientation b: Target Acquisition task, c: Shape Tracing task and Horizontal orientation d: Target Acquisition task, e: Shape Tracing task, f: robot study setup.

ABSTRACT

An increasing number of consumer-oriented head-mounted displays (HMD) for augmented and virtual reality (AR/VR) are capable of finger and hand tracking. We report on the accuracy of off-the-shelf VR and AR HMDs when used for touch-based tasks such as pointing or drawing. Specifically, we report on the finger tracking accuracy of the VR head-mounted displays Oculus Quest, Vive Pro and the Leap Motion controller, when attached to a VR HMD, as well as the finger tracking accuracy of the AR head-mounted displays Microsoft HoloLens 2 and Magic Leap. We present the results of two experiments in which we compare the accuracy for absolute and relative pointing tasks using both human participants and a robot. The results suggest that HTC Vive has a lower spatial accuracy than the Oculus Quest and Leap Motion and that the Microsoft HoloLens 2 provides higher spatial accuracy than Magic Leap One. These findings can serve as decision support for researchers and practitioners in choosing which systems to use in the future.

CCS CONCEPTS

• **Human-centered computing** → **Gestural input**; **Mixed / augmented reality**; **Virtual reality**.

KEYWORDS

Finger tracking, Hand tracking, Accuracy evaluation, User study, Head-mounted displays.

1 INTRODUCTION

The ability to see one's hands and finger movement inside Virtual Reality (VR) opens up opportunities for natural interaction in VR. This raises the potential to bridge the virtual space with the limited input space available in today's touch surfaces such as laptops, desktops and touchscreen-based devices.

Lately, commercial VR head-mounted displays (HMDs) are progressing to 'inside-out' tracking using multiple built-in cameras [10]. Inside-out tracking allows a simple setup of the VR system, and the ability to work in uninstrumented environments. Tracking the user hands in real-time is a potentially useful capability, that is already implemented in a number of consumer-oriented AR and VR HMDs. Given the capability of spatial hand and finger sensing, it is possible to see VR extending the input space of existing computing devices, such as touchscreens and keyboards, to a space around the devices, thereby enhancing their usage. A knowledge worker can use existing 2D surfaces as well as using space around the devices and in front of the screen, reachable while sitting, to represent and manipulate additional information. In this scope, the applicability of hand and finger tracking has a crucial dependency on the accuracy of tracking the user's fingers: While detection of certain finger gestures can be achieved using only coarse positional accuracy, interaction with real physical objects requires accurate grounded positioning of the VR coordinate system relative to the relevant objects, and a good spatial positioning of the user's fingers. Further, when mixing the interaction with physical input devices and 3D hand tracking, there is special importance in the relationship between these. Different techniques of finger sensing, the ability to select an object, and the minimal distance between such an object and nearby ones to prevent selection errors, are elements that need to be considered. Also, hand tracking algorithms must cope with self-occlusions [10] and refer to the estimation of the complete hand pose, including all finger joints, the hand position itself, or the position of an end effector, such as the fingertip. Within this paper, we focus on touch-based interactions, that is, tasks in which the user interacts with physical or digital surfaces, for example, for the purpose of selecting objects (such as buttons) or for drawing on surfaces [19]. While the achievable accuracy is dependent on multiple factors (e.g., the accuracy and latency of actual finger and hand tracking sensors and algorithms as well as the localization system of the respective HMD), we deliberately focus on the overall accuracy achievable with commodity off-the-shelf HMDs as this is of major concern for many practitioners and researchers who want or need to use the HMDs without modifying individual subsystems. Specifically, we report results from a controlled study with human participants ($n = 20$), investigating the accuracy of VR headsets (HTC Vive, Oculus Quest) and a LeapMotion sensor, as well as AR headsets (Microsoft Hololens 2 and Magic Leap 1). Further, we complemented the study with human participants with a study using a robotic-arm, which allows for better repeatability of measurements, at the cost of ecological validity. Our findings suggest that for VR HMDs, the HTC Vive results in significantly lower spatial accuracy (mean distance between a target and the sensed fingertip location of around 37 mm with a standard deviation of 20 mm) compared to both Oculus Quest (mean = 16 mm, sd = 9 mm) and Leap Motion (mean = 13 mm, sd = 6 mm). For AR HMDs the Microsoft HoloLens 2 provides a significantly higher spatial accuracy (mean distance between a target and the sensed fingertip location of around 15 mm with a standard deviation of 9 mm) compared to the Magic Leap One (m = 40 mm, sd = 16 mm). We hope the reported results will support the design of 3D work spaces around touch devices in AR and VR.

2 RELATED WORK

Besides stylus-based [4, 18] and controller-based [2, 22] input, hand and finger input is important for 3D interaction techniques in XR for selection, spatial manipulation, navigation and system control [14]. There are recent surveys on these techniques [1, 17].

The performance of the Leap Motion controller was already studied in various context.Weichert et al. [26] evaluated accuracy and repeatability of the Leap Motion Controller using a pen (with variable diameters form 3 to 10mm) mounted on a robotic arm (position accuracy of 0.2mm). The authors indicated an accuracy of below 0.2 mm in a static measurement setup. However, the measured volume only encompassed 20 cm along each axis. This setup replicates one intended use case of the Leap Motion controller, namely lying flat on a desk with fingers being moved above the sensor. In contrast forward reach of arms can typically encompass 40-50 cm [20], which is a more common scenario for the Leap Motion controller mounted on a HMD. Brown et al. [6] compared the performance of the Leap Motion to mouse in a Fitts' Law task (with two confirmation methods) and found the throughput of the Leap Motion to be significantly lower than that of the mouse. Again, a table mounted setup was used for the Leap Motion. Tung et al. [24] indicated a mean pointing accuracy of the Leap Motion controller of 17.3 mm ($sd = 9.6$) in a study with human participants ($n = 9$). Participants were pointing at 15 targets at a vertical monitor at an approximate distance of 35 cm. Again the leap motion was mounted flat on the supporting desk. Valentini et al. [25] employed a similar test setup with the Leap Motion lying on a table and ten users touching a glass plate mounted above the device at distances between 20 and 60 cm. They report mean tracking errors between 4 and 7 mm. Marin et al. proposed to combine the data from an XBox Kinect 1 sensor and the Leap Motion (again in a table mounted setup) [16] but did not report positional accuracy. Lindsey [15] compared a HMD-mounted Leap Motion Controller with other commodity input devices (gamepad,

touchpad) in a human participants study (n = 23) but did not report accuracy measurements. Lindsey focused on time and errors as objective measures in a shape, color and texture matching task) and found the Leap Motion controller to be significantly slower to the other input devices and also to be the least preferred. Xiao et al. [27] evaluated a novel finger tracking algorithm for the Microsoft HoloLens2, but did not compare it against other devices. In a user study (n=17) they found the mean spatial accuracy of the system to be 5.4 mm (sd=3.2) with a systematic offset to the right of predefined touch targets. Han et al. [10] developed and evaluated a hand tracking algorithm which can be assumed to be a basis for the current Oculus Rift hand tracking system. They reported a mean positional error of 11mm relative to a ground truth optical outside-in tracking system for a number for in-air hand movements.

Complementary to these prior works, our study allows for direct comparison of the accuracy of multiple hand tracking systems specific to AR and VR HMDs in a joint experimental design.

Recently, Batmaz et al. [3] utilized a HMD-mounted Leap Motion to investigate human performance between a VR, AR and 2D touchscreen condition (with and without haptic feedback) in a eye-hand coordination reaction test. In a human participants study (n=15) they found the throughput of both the VR and AR condition to be significantly lower than in the 2d touchscreen condition and the throughput of AR to be also significantly lower compared to VR and hypothesized that the difference between AR and VR might be due to the display system (as the same hand tracking technology was used). The closest work to ours is by Schneider et al. [21], who compared the finger tracking accuracy of the HTC Vive Pro and the leap motion. Specifically, our work complements this prior work by including a larger set of both recent AR and VR displays as well as a more comprehensive evaluation using additional dependent variables such as workload, subjective feedback and usability ratings. Our work also includes a complementing robot study.

3 STUDY WITH HUMAN PARTICIPANTS

In a study with human participants ($n = 20$), we compared the overall spatial accuracy of different head-mounted displays. Specifically, we compared three solutions for hand tracking in VR HMDs and two AR systems.

3.1 Experimental Design

We conducted the an experiment consisting of two parts, one for VR devices and one for AR devices. The part comparing the VR systems was designed as a 3×2 within-subjects design. The part comparing AR systems used a 2×2 within-subjects design. The two independent variables in both parts were INTERFACE and ORIENTATION. INTERFACE represented tested platforms. For the VR experiment these were: VIVE: V, QUEST: Q and LEAPMOTION: LM For the AR experiment these were: HOLOLENS: HL and MAGICLEAP: ML. Each represents a commodity device or sensor listed in Section 3.4. Following prior work [27]), each participant performed the tasks in two different ORIENTATIONS: HORIZONTAL and VERTICAL. These two values refer to the orientation of the touchscreen, on which users had to perform the tasks and corresponds to typical interaction scenarios (e.g., wall interaction, desk interaction). HORIZONTAL describes a flat configuration, where the screen was parallel to the

Figure 2: Participants performing tasks (left) with respective VR view (right): a: TARGET ACQUISITION task in HORIZONTAL with LEAP-MOTION; b: SHAPE TRACING task in HORIZONTAL with QUEST; c: SHAPE TRACING in VERTICAL with HOLOLENS.

table (see Figure 2, a) and VERTICAL refers to a standing configuration, where the screen is perpendicular to the table (see Figure 2, c). To mitigate ordering effects, the sequence of INTERFACES and ORIENTATION were counterbalanced between participants.

Dependent variables for both experiments (VR and AR) included workload, as measured by NASA TLX [12], Usability as measured by the System Usability Scale (SUS) [5]) and Simulator Sickness as measured by the Simulator Sickness Questionnaire (SSQ) [13].

Based on prior work [9], we provided a three-item questionnaire (which we name Perceived Finger Assessment (PFA) within this paper) with a 7 point Likert scale (ranging from "totally disagree" to "totally agree"). The items were: "I felt that the fingers were my own." (PFA-F), "I felt that I could control the position of my fingers." (PFA-C) and "I felt that I hit the target I aimed for." (PFA-A).

Furthermore, the spatial accuracy was assessed using the following measures, see also Figure 3: DISTANCE FINGER-TARGET describes the Euclidean distance between the index finger position as tracked by the system and the target on the screen while touching and is an indicator for the spatial accuracy of the tracking system. DISTANCE TOUCH-TARGET describes the distance between the touch position on the screen and the target while touching. This metric was chosen as it describes the offset between the actual touch position and the position recognized by the target position. Z-DISTANCE is the distance between the finger as tracked by the respective system and the plane representing the screen while the participant is touching. ANGLE FINGER-TARGET describes the angle between the displayed target line and a line that is fitted through the tracked finger positions while the user was tracing the line. RADIUS DIFFERENCE FINGER-TARGET is the difference between the radius of the target circle and the circle that was fitted through the tracked finger positions while the user was tracing the circle. ANGLE FINGER-TARGET and RADIUS DIFFERENCE FINGER-TARGET were chosen to assess the relative accuracy of tracking a moving finger. Other possible measures like circle center or start and endpoint of the target line were
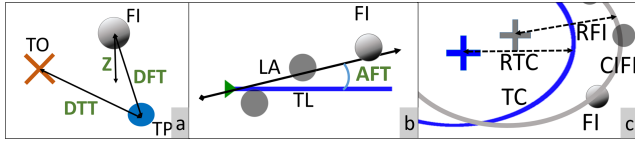
Figure 3: Measures used for assessing spatial accuracy (highlighted in green). a: Metrics used in the Target Acquisition task with Z: Z-Distance between the tracked finger (FI) and the display surface, DFT: Distance Finger-Target between FI and the target point (TP), DTT: Distance Touch-Target between the touch position (TO) and TP; b and c: Metrics used in the Shape Tracing Task with AFT: Angle Finger-Target being the angle between the line fitted through the tracked finger points (LA) and the target line (TL), Radius Difference Finger-Target is the difference between the radius (RFI) of the circle (CIFI) fitted through the FIs and the radius (RTC) of the target circle (TC).
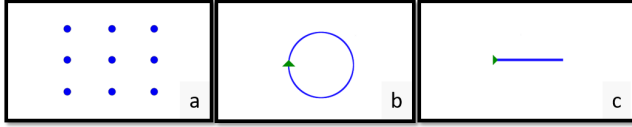

Figure 4: a: Target Acquisition task with 9 possible target locations displayed simultaneously for visualization. Shape Tracing task displays two shapes: b: a circle to be traced clockwise and c: a line to be traced from left to right.

not considered because the absolute accuracy was already assessed through Distance Finger-Target, Distance Touch-Target and Z-Distance.

3.2 Tasks

The participants were told to conduct two tasks in all conditions, inspired by prior work [11, 21, 27]. For each task, the participants started with their index finger on the smartphone in front of them, see Figure 2, left. The whole smartphone display served as home position. The home position in VR was represented as orange dot within the bounds of the smartphone (the smartphone was not visualized in VR), see Figure 2, right, i.e. participants could still touch slightly to the left or right of the visualized homing position in VR.

In the Target Acquisition task, participants needed to touch a blue round target (diameter: 18 mm) on a touchscreen in front of them as soon as it appeared and hold it with the tip of their dominant hand's index finger for about one second until its color changed from blue to green. The request for holding the fingertip for a short period enables the collection of multiple tracking samples at the target position. After touching the target, participants had to return their index finger to the resting position on the smartphone. The next target was shown as soon as the finger touched the smartphone surface. The return to the smartphone had two reasons. First, the touch on the smartphone was used to trigger the visualization of the next target, and, second, to ensure the movement of the hand for each position is similar. When using the VR-HMDs, the participants could not see the real world including the target touchscreen, the

phone or even their own hands. Therefore, the tip of their index finger was rendered as a sphere with a diameter of 16 mm (following recommendation by prior work [9]). Using a sphere with 16mm allowed to position it completely within the target circle having a diameter of 18 mm. Additionally, the participants saw 3D models of the table and the touchscreen. When using AR-HMDs there was no virtual rendering of the participant's hands or any other objects to avoid potential confusion between real and tracked fingertips when touching a target.

One trial in the Target Acquisition task consisted of hitting a total number of nine circular targets, which appeared in a fixed order left to right and top to bottom (see Figure 4, a). As we focused on pointing accuracy and not target acquisition time, randomizing the order of targets was not necessary. The distance between the center target and the starting position on the phone is 25.9 cm for the Vertical orientation and 20.1 cm for Horizontal. The other eight targets have an offset from the center target of 7.5 cm to the left (or right) and 7.5 cm to the top or bottom. Participants were asked to conduct five trials, with nine targets each, for a total of 5 trials × 9 targets × 3 interfaces 2 × orientations = 270 recorded target acquisitions per user for the VR experiment and 5 trials × 9 targets × 2 interfaces 2 × orientations = 180 recorded target acquisitions per user for the AR experiment.

The second task, Shape Tracing, also starts with placing the tip of the index finger on the phone's touchscreen. On the target touchscreen a blue shape was displayed, being either a circle with a radius of 7.5 cm or a straight 15 cm long line. A green triangle on the shape indicated the point that the users should touch with their index fingertip and follow the shape in the direction displayed by the triangle (see Figure 4, b and c). The circles had to be traced both clockwise and counterclockwise. The line segments had to be traced from left-to-right, right-to-left, top-to-bottom and bottom-to-top. Again, after the participants finished tracing one shape, they were asked to return their index finger to the phone screen. The Shape Tracing task also consisted of five trials. One trial consisted of two circles (clockwise and counterclockwise) and four lines (in all four directions). Again, the order of the target shapes was not randomized, as the experiment focused on accuracy rather than speed.

3.3 Procedure and Data Collection

After an introduction, participants were asked to fill out a demographic questionnaire. The study was divided into two parts. In one part, VR-HMDs were tested, and in another the AR-HMDs. The order of the parts as well as the order of the devices inside the parts were counter-balanced, as well as the order of Orientation. Before starting with the tasks in a specific orientation, a calibration step mapped the position and orientation of the touchscreen to the respective HMD coordinate system. Depending on the Interface, different techniques were used for the calibration process: LeapMotion and Vive where calibrated using a Vive Tracker that could be held to the corner of the screen and its position in 6DOF to the Vive coordinate system. Quest was calibrated with a modified Oculus controller (for pictures of the calibration tools see the appendix). Following the calibration process, participants were asked to conduct a training session before starting the data collection session.

With each touch on the touchscreen the touch data are sent via WiFi to the HMD. The HMD converted the 2D touch position of the monitor to 3D touch positions in the virtual world and saved with the target position and the index finger position in a file. At the beginning of each stage (being either Horizontal or Vertical Orientation) the participants were asked to sit in a comfortable position. After conducting all tasks (Target Acquisition and Shape Tracing) in both Horizontal and Vertical orientation, the participants filled out the SSQ, SUS and TLX questionnaires as well as the three questions for perceived finger assessment once for each HMD. Following the completion of all three VR-HMDs, the participants were asked which HMD they preferred. This was not done after AR conditions, because participants saw their real hand and the real touchscreen (no augmentations). Therefore, they were unlikely to have a preference related to the tracking performance. The duration of the study was about 110 minutes per participant. They were compensated with a voucher worth €10.

For the target acquisition task, we collected multiple touch samples (at minimum 20 samples, more if the users touched down longer) per target location, which were averaged into one. For the shape tracing task, we sampled touch (and the associated finger) positions at a distance of at least .01 mm to the previous sample. In this way, we ensure that the start (and end) positions of the drawn shapes only consist of single points, even if the participant's finger rests for a little while on that position. This is important for the fitting of the line and circle to not bias the start and end. For the line fitting we used linear optimization and the RANSAC algorithm [7] for robust line fitting. For fitting the circle, we utilized the singular value decomposition and the method of least-squares for the optimal circle fitting.

3.4 Apparatus

The following HMDs were employed: **HoloLens**: A Microsoft HoloLens 2 AR-HMD (see Figure 1 a. bottom-left) running Windows 19041.1136. **MagicLeap**: A Magic Leap 1 AR-HMD (see Figure 1 a. bottom-right) running LuminOS 0.98.20. **Quest**: An Oculus Quest 1 VR-HMD (see Figure 1 a. top-center) running Quest OS 25.0.0.77. **Vive**: An HTC VIVE Pro VR-HMD (see Figure 1 a. top-left) using VIVE Hand Tracking SDK 0.9.4. **LeapMotion**: An Ultraleap LeapMotion sensor attached to a VIVE Pro VR-HMD (see Figure 1 a. top-right). It uses LeapMotion Software 4.1.0. The VIVE Pro HMD is similar to the one used in Vive, without using its built-in hand tracking capabilities.

Both AR-HMDs used the image marker tracking provided by Vuforia 9.7.5 using a default image, to allow a better comparison between the devices. Also both HTC VIVE Pro devices were used with Steam VR 1.16.8 and Steam VR Unity Plugin 2.7.3 (SDK version 1.14.15). The system for all VR Interfaces was implemented in Unity 2020.2 and deployed on a PC (Intel Xeon E5-2620 v 4 processor, 64 GB RAM, Nvidia GTX 1060 graphics card) running Windows 10. The touchscreen was a Dell S2340T monitor with a screen width of 56.2 cm and a height of 34 cm. The smartphone for the resting position was a Fire Phone with a screen width of 10.4 cm and a height of 5.8 cm.

Participants were seated on a standard office chair (seat height between a minimum of 46.5 cm and a maximum of 57 cm), adjusted to a comfortable height for each user. They were asked to initially place their index fingertip of their dominant hand on a touchscreen of a phone that lay on the table in front of them (See Figure 2). For the communication between touchscreen, smartphone and the HMD-device, we implemented an application to send the touch information to the respective device via WiFi and UDP protocol.

3.5 Participants

We recruited 20 participants (11 female, 9 male, mean age 33.4 years, sd = 8.9) with diverse backgrounds. All participants were familiar with touch sensitive screens. Skin types (which might influence tracking capabilities) were ranging from I to IV according to Fitzpatrick [8]. Three of the participants indicated no prior VR experience, 11 participants rarely used VR, yet more than once, two participants occasionally, two often and two participant very frequently. Four participants indicated no prior AR experience, 11 had rarely used AR yet more than once, two participants occasionally, one often and two participant very frequently. Eight participants wore contact lenses or glasses which corrected to normal vision. 19 participants were right handed while one was left handed. All participants used the index finger of their dominant hand to conduct the tasks.

3.6 Results

Unless otherwise specified, statistical significance tests for performance data (task completion time, distance to target) were carried out using general linear model repeated measures analysis of variance (RM-ANOVA) with Holm-Bonferroni adjustments for multiple comparisons at an initial significance level $\alpha = 0.05$. We indicate effect sizes whenever feasible (η_p^2). For subjective feedback, or for data that did not follow a normal distribution or could not be transformed to a normal distribution using a log-transform, we used Friedman's test with Holm-Bonferroni adjustments for multiple comparisons using Wilcoxon signed-rank tests. The anonymized raw and aggregated data of the study are available under https://gitlab.com/mixedrealitylab/finger-tracking-accuracy.

Selected additional results are depicted in the appendix.

3.6.1 AR. For the Target Acquisition Task we analyzed the distance between the tracked finger position and the displayed target (Distance Finger-Target: DFT) and the distance between the tracked finger position and the display-area (Z-Distance: Z). We did not analyze the distance between the touch position on the display and the displayed target (Distance Touch-Target: DTT), because this measure does not tell us anything about the tracking accuracy of the AR-devices (the user aligned their real fingertip with the target, not the virtual fingertip as in the VR HMDs). The descriptive statistics and the results from the null hypothesis significance tests (NHST) are presented in Table 1 and Figure 5. For the significance tests, the data was log-transformed to ensure a normal-distribution. The results indicate that the Interface significantly influenced both the Distance Finger-Target and the Z-Distance, in such a way, that the HoloLens was more accurate than the MagicLeap. The results show no significant influence of Orientation on the accuracy.

The Shape Tracing Task consisted of two types of tasks, tracing a line and tracing a circle. For the line tracing we analyzed the Angle Finger-Target: AFT which describes the angle between

Table 1: RM-ANOVA results for the human participants study and mean and standard-deviation values for AR-devices (regardless of Orientation). Gray rows show significant findings. I = Interface, O = Orientation.

AR - Target Acquisition Task

		Distance Finger-Target					Z-Distance			
	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2
I	1	19	91.3542	< .001	.828	1	19	64.612	< .001	.764
O	1	19	.0281	.869	.001	1	19	.134	.719	.007
I × O	1	19	1.4861	.238	.073	1	19	.357	.557	.018

AR - Shape Tracing Task

		Angle Finger-Target					Radius Difference Finger-Target			
	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2
I	1	19	23.46	< .001	.55	1	19	11.846	.003	.384
O	1	19	.0	.975	0.00	1	19	.0483	.82	.003
I × O	1	19	4.20	.054	.18	1	19	11.126	.003	.363

AR - Target Acquisition Task / AR - Shape Tracing Task

Device	Distance Finger-Target		Z-Distance		Angle Finger-Target		Radius Difference Finger-Target	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
HoloLens	14.54 mm	9.22	11.8 mm	19.15	3.29°	1.31	1.64 mm	1.47
MagicLeap	39.84 mm	16.11	29.85 mm	19.99	9.55°	7.4	8.62 mm	20.39
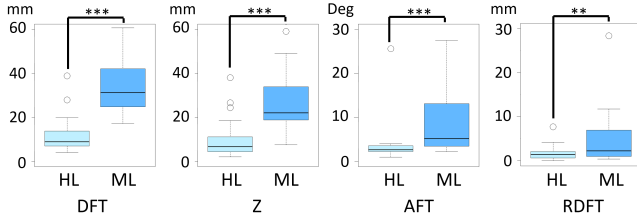
Figure 5: Accuracy measures from the human participants study for the HoloLens (HL) and MagicLeap (ML); DFT = Distance Finger-Target in millimeter; Z = Z-Distance in millimeter; AFT = Angle Finger-Target in degrees; RDFT = Radius Difference Finger-Target in millimeter; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

Figure 6: Accuracy measures from human participants study for the Vive (V), the LeapMotion (LM) and Quest (Q). DTT = Distance Touch-Target in millimeter; DFT = Distance Finger-Target in millimeter; Z = Z-Distance in millimeter; AFT = Angle Finger-Target in degrees; RDFT = Radius Difference Finger-Target in millimeter; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

the target line and the line fitted through the tracked finger positions. For the circle tracing we analyzed the Radius Difference Finger-Target, which describes the difference between the radius of the target circle and the radius of the circle fitted through the tracked finger points. These metrics are depicted in Figure 3, b and c. The descriptive statistics and the NHST results are presented in Table 1. For the significance tests, the data was log-transformed to ensure a normal-distribution. For both metrics the results indicate a significant influence of interface on the measures. Post-hoc tests indicate that for both metrics the HoloLens is significantly more accurate than the MagicLeap, which means that the direction of the fitted line and the radius of the fitted circle are more similar to the target shapes. For Radius Difference Finger-Target the analysis also indicates interaction effects, because the HoloLens performs slightly better in Horizontal orientation and MagicLeap slightly better in Vertical orientation. However, post-hoc test did not reveal significant differences. Workload and Usability: The results of the four questionnaires showed no significant differences between the two AR-HMDs regarding usability, simulator sickness, task load or perceived finger assessment (see also appendix).

3.6.2 VR. For the Target Acquisition Task we analyzed the distance between the tracked finger position and the displayed target (Distance Finger-Target), between the touch position on the display and the displayed target (Distance Touch-Target) and the distance between the tracked finger position and the display-area (Z-Distance). These metrics are depicted in Figure 6. For the significance tests, the data was log-transformed to ensure a normal-distribution. The descriptive statistics and the NHST results are presented in Table 2 and Figure 6. The results indicate that
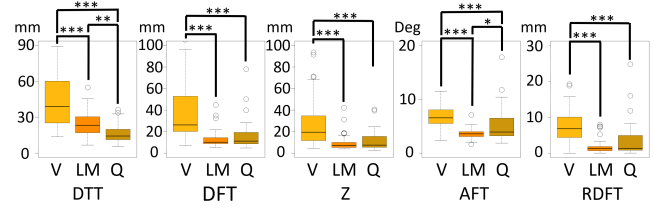
the Interface has a significant influence on Distance Touch-Target and Distance Finger-Target. Post-hoc tests show that for both metrics Vive is significantly less accurate than both Quest and LeapMotion. In addition, LeapMotion is significantly less accurate than Quest for the Distance Touch-Target measure. For both measures an interaction effect is detected. The results indicate that the LeapMotion and Quest are more accurate in the Horizontal orientation, while Vive is more accurate in the Vertical orientation. Analyzing the Z-Distance results indicates a significant influence of Interface and Orientation as well as interaction effects. Post-hoc tests indicate that, again, Vive is less accurate than LeapMotion and Quest. They also indicate that the horizontal orientation ($M = 10.87$ mm, $SD = 8.07$) is more accurate than the Vertical orientation ($M = 21.38$ mm, $SD = 22.08$).

For the Shape Tracing Task, the same measures as for the AR HMDs were employed. These metrics are depicted in Figure 3, a. For the significance tests, the data was log-transformed to ensure a normal-distribution. The mean values and results from analyzing these metrics are presented in Table 2. Analyzing the results indicates that Interface significantly influences the two metrics. Post-hoc tests showed that for both the Angle Finger-Target and Radius difference Finger-Target the results from the Vive differed significantly more from the target shapes than LeapMotion and Quest. The analysis also indicated that Orientation significantly influences the Radius difference Finger-Target. Post-hoc tests indicate that the Horizontal orientation ($M = 3.75$ mmm $SD = 3.94$) matches the radius closer than the Vertical orientation ($M = 4.94$ mmm $SD = 5.47$). Additionally, regarding Angle Finger-Target, LeapMotion differed significantly less from the

Table 2: RM-ANOVA results for the human participants study and mean and standard-deviation values for VR-devices (regardless of Orientation). Gray rows show significant findings. I = Interface, O = Orientation.

| | VR - Target Acquisition Task | | | | | | | | | | | | | | | VR - Shape Tracing Task | | | | | | | | | |
| | Distance Touch-Target | | | | | Distance Finger-Target | | | | | Z-Distance | | | | | Angle Finger-Target | | | | | Radius Diff. Finger-Target | | | | |
	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2
I	2	38	42.81	<.001	.69	2	38	50.8	<.001	.73	2	38	25.83	<.001	.58	2	38	27.73	<.001	.593	2	38	21.411	<.001	0.53
O	1	19	1.76	.2	.09	1	19	.0	.99	.0	1	19	9.11	.007	.32	1	19	0.002	0.96	0	1	19	5.406	0.031	0.222
I × O	2	38	25.13	<.001	.57	2	38	11.8	<.001	.38	1	38	10.24	<.001	.38	2	38	10.83	<.001	.36	2	38	0.781	0.465	0.039

| | VR - Target Acquisition Task | | | | | | VR - Shape Tracing Task | | | |
| | Distance Touch-Target | | Distance Finger-Target | | Z-Distance | | Angle Finger-Target | | Radius Difference Finger-Target | |
Device	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Vive	45.04 mm	19.51	37.22 mm	20.73	27.67 mm	23.98	6.76°	1.67	7.76 mm	4.80
LeapMotion	24.05 mm	8.13	13.15 mm	5.45	9.56 mm	7.57	3.61°	0.65	1.94 mm	2.14
Quest	16.14 mm	5.81	16.22 mm	8.76	11.14 mm	9.1	5.17°	2.44	3.34 mm	4.81
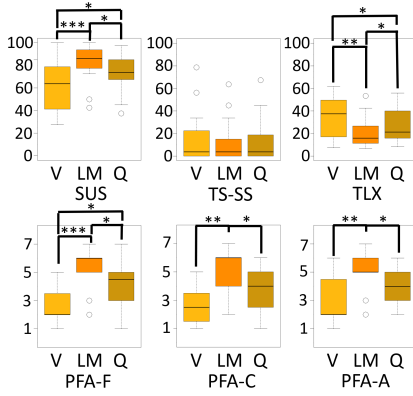
Figure 7: Results from Questionnaires for Vive (V), LeapMotion (LM) and Quest (Q). SUS = System Usability Scale; TS-SS = Total Severity Simulator Sickness; TLX = Overall Taskload; PFA-F = Perceived Finger; PFA-C = Perceived Control; PFA-A = Perceived Accuracy; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

target shapes than Quest. Analyzing Angle Finger-Target also indicated interaction effects. Post-hoc tests showed that the Quest differed significantly less from the target line in the Horizontal compared to the Vertical condition.

Workload and Usability: The analyzed questionnaire results for the three VR-HMDs and the mean and standard-deviation values are displayed in Table 3 and Figure 7. No significant influence of Interface could be found on the simulator sickness measurement. Regarding usability, LeapMotion received the highest scores, followed by Quest and then Vive. This aligns with the perceived overall taskload, where Vive resulted in the highest taskload, followed by Quest and the least taskload was recorded for LeapMotion. The results from the perceived finger assessment also indicate the highest scores for LeapMotion, second highest for Quest and least for Vive.

3.6.3 Preferences and Open Comments. We solely asked participants for preferences regarding the VR hand tracking systems, as no finger visualization was employed in the AR HMDs, and, hence we did not expect feedback specific to their hand tracking capabilities. From the 20 participants, 17 preferred the LeapMotion. Ten participants said "It was accurate." (P2, P4, P7, P10, P13, P15, P16, P17, P18, P19) and 8 said "It followed my finger movements" (P3, P7, P8, P11, P12, P17, P18, P19). This is in line with the usability

results. Three participants preferred the Quest. Twelve participants stated the Vive to be the worst HMD in regards of hand tracking accuracy. This is supported by all measures (DTT, DFT, Z, AFT and RDFT), as the VIVE was significantly less accurate than both Quest and LeapMotion. Six participants said about the Vive that "The tracking indicator was not on my finger" (P5, P11, P15, P16, P17, P18). Five participants said that the Quest was the worst HMD. Three of the participants using the Quest said "The tacking indicator jitters too much" (P8, P12, P13). Three participants could not decide which HMD was the best or worst. Participants were asked to make open comments during the experiment. On the LeapMotion, participants commented: "It feels very responsive."(P03), "much better than before [Quest and Vive] and less frustrating" (P08), "The sphere [fingertip visualization] does not want to go where I want it to" (P06), "It works good while hovering. But on contact with the screen it moves behind it" (P09). On the Vive participants commented: "It feels very accurate but slow" (P04), "I can get used to this" (P09), "The sphere [fingertip visualization] dips too far in the surface of the screen." (P14) and "I hope this won't be used in medical applications" (P12). Comments on the Quest were the following: "It feels better at 0 degrees [Horizontal orientation] " (P13), "Drawing the line with this one is much better" (P19), "Counterclockwise is worse than clockwise [tracing]" (P10) as well as "Oh god, the sphere [fingertip visualization] jitters" (P18).

4 ROBOT STUDY

While the study with human participants was a controlled experiment and HMDs are likely designed for coping with the hand movements of real users, we also collected measurements using a robotic arm following prior work [26]. This allows predefined spatial positions to be repeatedly selected with higher accuracy, thereby facilitating reproducibility of results. Please note the trade-off between reproducibility and ecological validity (as no actual human hand but an artificial hand was used). Hence, the results reported in this section should be seen as complementary to the results from the study with human participants.

4.1 Apparatus

The setup is shown in Figure 1(f) (for a schematic view see the appendix) and consists of the robot arm with an artificial hand model (made of silicon) and calibration tool attached and a Styrofoam dummy-head for mounting the HMDs.

Table 3: RM-ANOVA results of the questionnaire data for the VR-HMDs. Gray rows show significant findings. SUS: System Usability Scale. SS: Simulator Sickness Questionnaire. PFA: Perceived Finger Assessment. $df_1 = df_{effect}$ and $df_2 = df_{error}$.

	\multicolumn{5}{c}{RM-Anova}	\multicolumn{2}{c}{VIVE}	\multicolumn{2}{c}{LEAPMOTION}	\multicolumn{2}{c}{QUEST}
	df_1	df_2	F	p	η_p^2	Mean	SD	Mean	SD	Mean	SD
SUS	2	38	21.389	< 0.001	0.53	60.63	22.08	83.63	15.36	73.63	16.03
Total-severity-SS	2	38	0.31	0.74	0.02	20.01	35.71	11.22	17.58	17.2	32.7
Overall Taskload	2	38	15.53	< 0.001	0.45	34.75	18.16	20.79	13.33	26.38	14.28
PFA-Finger	2	38	29.01	< 0.001	0.6	2.65	1.14	5.3	1.38	4.25	1.55
PFA-Control	2	38	17.13	< 0.001	0.47	2.6	1.31	5.1	1.59	3.8	1.51
PFA-Accuracy	2	38	14.18	< 0.001	0.43	2.9	1.62	5.05	1.47	3.85	1.31

The devices and the versions of the software used in the robot study are the same as in the human participants study (see section 3.4 for more details). The robot used in the study is the Universal Robot UR3 and was mounted on a aluminum frame with its base at a height of 124 cm. The UR3 has a reach of 50 cm and it can move the arm to a specific position with an accuracy of ±0.1 mm. The positioning of the robot arm is done with the software of the robot to ensure the arm is always on the correct position. The HMD was positioned around 55 cm away from the middle position of the hand and looked directly at mimicking interaction with vertically mounted surfaces. The offset of the index fingertip relative to the top left corner of marker was measured with an Optitrack digitizing probe. The index fingertip was used as the reference point since it is most similar with the touch point in the study with human participants. To avoid potential interference with the tracking systems, the arm of the robot was covered in black fabric. The background was also covered with black fabric. To mimic a user's right arm we used a shirt, with the end of the sleeve attached to the hand model and the rest of the shirt attached on the dummy head. In contrast to the study with human participants, the hand model did not show a dedicated pointing pose using the index finger (as such a hand model was not available). However, the operator made sure that the index finger was used for data collection by monitoring the virtual hand model on an external screen. Similar to the study with human participants, a smartphone was used to trigger the next target.

4.2 Procedure

Following the procedure of the study with human participants, the robot fulfilled the same tasks, (see Section 3.2), namely the TARGET ACQUISITION task with the nine targets repeated five times and the SHAPE TRACING task with tracing the line and circle again repeated five times, but only in the VERTICAL orientation and without an actual touch by the silicon hand. For the TARGET ACQUISITION TASK the robot arm was moved to the specific position and at that position the data was collected. For the SHAPE TRACING TASK the robot arm was moved to the starting position, the data collection was started and then the robot arm moved along the corresponding path. In the robot study we only conducted the task in the VERTICAL orientation as it was not possible for us to mount the arm in a horizontal fashion. We did not use a touchscreen because the touch of the silicon hand could be registered by the touch monitor.

For each device, the robot study started with a calibration sequence. For this, the robot was moved to the top left position and a calibration tool was used to position the target finger positions (similar to calibrating the monitor in the study with human participants). For the AR devices, the image marker was used and after calibration the marker was covered with black fabric. For the VIVE and the LEAP MOTION, the Vive Tracker was used for calibrating the target positions. After calibration the tracking of the Vive Tracker was disabled in software. For the QUEST, the calibration was conducted via the corners of the image marker using the Quest calibration tool. For the TARGET ACQUISITION task the robot arm was positioned for the specific target. Then the operator touched on the tablet and collected 3000 samples per target location. After completing the data collection part, the operator triggered the transition to the next target. The operator then again moved the robot arm to the next target position and repeated the process.For the SHAPE TRACING task the operator positioned the robot arm at the starting position of the shape and then triggered a prerecorded motion pattern (line or circle).

The HMDs were not moved on the Styrofoam dummy-head during the data collection process to create a stable environment in contrast to the human participants study, where the user had the freedom to move their head. The preparation took ten minutes, calibration lasted five minutes and the main study was conducted within 45 minutes. The overall procedure for each HMD lasted 60 minutes.

4.3 Results

As for the study with human participants, additional results can be found in the appendix.

4.3.1 AR. Similar to the human participants study, for the TARGET ACQUISITION TASK with AR devices, we analyzed the DISTANCE FINGER-TARGET and the Z-DISTANCE. In contrast to the study with human participants, we used paired samples t-test to analyze the data (normality assumptions were met), as there was only one independent variable (INTERFACE with two levels HOLOLENS and MAGICLEAP. The results from the analysis and the mean and standard-deviation values are shown in Table 4 and Figure 8. For the significance tests, the data was log-transformed to ensure a normal-distribution. The analysis indicates that HOLOLENS is significantly more accurate than MAGICLEAP regarding both DISTANCE FINGER-TARGET and Z-DISTANCE.

Table 4: RM-ANOVA results for the robot study and mean and standard-deviation values for AR-devices (regardless of ORIENTATION). Gray rows show significant findings. I = INTERFACE, O = ORIENTATION.
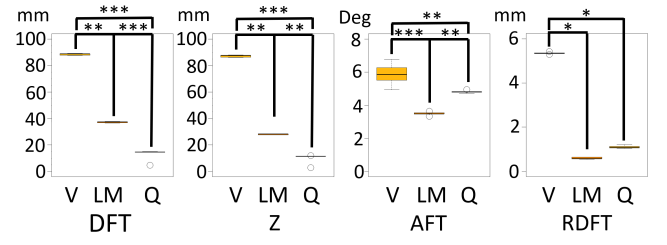
	AR - Target Acquisition Task			AR - Shape Tracing Task
	Distance Finger-Target	Z-Distance	Angle Finger-Target	Radius Difference Finger-Target
df	4.0	4.0	4.0	4.0
p	$< .001$	$< .001$	$< .001$	.398
Effect Size	-6.42	-7.34	-9.45	-0.945

	AR - Target Acquisition Task				AR - Shape Tracing Task	
	Distance Finger-Target		Z-Distance		Angle Finger-Target	Radius Difference Finger-Target
Device	Mean	SD	Mean	SD	Mean	SD	Mean	SD
HOLOLENS	17.65 mm	0.24	13.89 mm	0.45	1.26°	0.11	2.4 mm	0.14
MAGICLEAP	31.79 mm	2.75	30.39 mm	2.66	1.75°	0.08	4.08 mm	3.86
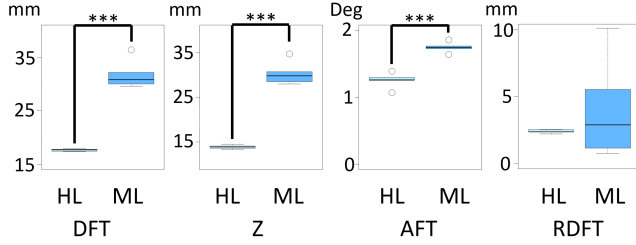
Figure 8: Accuracy measures from the robot study for the HOLOLENS (HL) and MAGICLEAP (ML); DFT = Distance Finger-Target in millimeters; Z = Z-Distance in millimeters; AFT = Angle Finger-Target in degrees; RDFT = Radius Difference Finger-Target in millimeters; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

Figure 9: Accuracy measures from robot study for the VIVE (V), the LEAPMOTION (LM) and QUEST (Q). DFT = Distance Finger-Target in millimeter; Z = Z-Distance in millimeter; AFT = Angle Finger-Target in degrees; RDFT = Radius Difference Finger-Target in millimeter; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$

Similar to the human participants study, for the SHAPE TRACING TASK with AR devices we analyzed the ANGLE FINGER-TARGET and the RADIUS DIFFERENCE FINGER-TARGET (again using a paired samples t-test). The results from the analysis and the mean and standard-deviation values are shown in Table 4. For the significance tests, the ANGLE FINGER-TARGET-data was log-transformed to ensure a normal-distribution. The analysis again indicates that HOLOLENS is significantly more accurate than MAGICLEAP regarding ANGLE FINGER-TARGET but not regarding RADIUS DIFFERENCE FINGER-TARGET.

4.3.2 VR. Similar to the study with human participants, for the TARGET ACQUISITION TASK with VR devices we analyzed the DISTANCE FINGER-TARGET and the Z-DISTANCE, but not the DISTANCE TOUCH-TARGET, because the touchscreen was not used in the robot study. Because we had one independent variable with three levels, we used a repeated measures ANOVA to analyze the data. The results from the analysis and the mean and standard-deviation values are shown in Table 5. For the significance tests, the data was log-transformed to ensure a normal-distribution. The analysis indicates that the INTERFACE significantly influences the DISTANCE FINGER-TARGET and the Z-DISTANCE measures. Post-hoc tests revealed that QUEST was significantly more accurate regarding both measures than LEAPMOTION and VIVE. Furthermore, LEAPMOTION was significantly more accurate than VIVE.

Similar to the study with human participants, for the SHAPE TRACING TASK with VR devices we analyzed the ANGLE FINGER-TARGET and the RADIUS DIFFERENCE FINGER-TARGET. As with the TARGET ACQUISITION TASK we used a repeated measures ANOVA to analyze the data. The results from the analysis and the mean and standard-deviation values are shown in Table 5. For the significance tests, the RADIUS DIFFERENCE FINGER-TARGET-data was log-transformed to ensure a normal-distribution. The analysis indicates that the INTERFACE significantly influences the ANGLE FINGER-TARGET and the RADIUS DIFFERENCE FINGER-TARGET. Post-hoc tests indicate that the angle between the target line and the line through the tracked finger-points was significantly smaller for LEAPMOTION compared to VIVE and QUEST and that it was significantly smaller for QUEST compared to VIVE. The VIVE differed significantly more from the target shapes than LEAPMOTION and QUEST.

5 DISCUSSION

Our study reported on the accuracy of AR and VR commodity HMDs for finger tracking when interacting with surfaces. Our findings suggest that for VR HMDS, the HTC Vive results in significantly lower spatial accuracy compared to both Oculus Quest and Leap Motion. For AR HMDs the Magic Leap One resulted in significantly lower spatial accuracy compared to the Microsoft HoloLens 2. Those results were indicated both in the study with human participants and in the robot study. They are also supported by the subjective

Table 5: RM-ANOVA results for the robot study and mean and standard-deviation values for VR-devices (regardless of Orientation). Gray rows show significant findings. I = Interface, O = Orientation.

			VR - Target Acquisition Task								VR - Shape Tracing Task								
		Distance Finger-Target					Z-Distance					Angle Finger-Target					Radius Difference Finger-Target			
	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2
I	2	8	57.0	< .001	0.93	2	8	48.1	< .001	0.92	2	8	43.7	< .001	0.92	2	8	6.46	0.021	0.618

	VR - Target Acquisition Task				VR - Shape Tracing Task			
	Distance Finger-Target		Z-Distance		Angle Finger-Target		Radius Difference Finger-Target	
Device	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Vive	88.54 mm	0.69	87.2 mm	0.81	5.87°	0.70	5.35 mm	0.05
LeapMotion	37.25 mm	0.52	28.07 mm	0.41	3.5°	0.11	3.16 mm	5.72
Quest	12.75 mm	4.52	9.87 mm	3.95	4.82°	0.09	1.1 mm	0.07

feedback of the participants, as most participants (17 out of 20) preferred the Leap Motion and more than half (12/20) liked the HTC Vive least.

The indicated spatial accuracy results have to be understood as a cumulative measure, integrating multiple tracking systems. Specifically, the accuracy of the respective HMD tracking system is added to the accuracy of the actual sensors employed for hand and finger tracking.

In contrast to prior work, the mean spatial accuracy results reported in this paper are typically lower (specifically for the Leap Motion). For example while Weichert et al. [26] reported spatial accuracy of below 0.2 mm for the Leap Motion controller, in our test setup the mean spatial accuracy in a pointing task was 24 mm (sd = 8) for the study with human participants. The difference can potentially be attributed to different sensor-hand configurations (e.g., sensor lying on a table with the hand interacting up close vs. head-mounted sensor with hand interacting at arm's reach). Also, we witnessed a high inter-subject variability of the hand tracking accuracy.

For the AR HMDs and the QUEST, the robot study mostly replicated the results from the human participants study. However, for the Vive and LeapMotion, the mean spatial accuracy for the target acquisition task resulted in approximately three times higher mean distances in the robot study compared to the human participants study. We see multiple possible explanations for this diverging behaviour: First, in the VR conditions, users typically aligned their virtual finger tip with the respective target, potentially lowering the accuracy error compared to when they would align their physical fingertip with the target. In contrast, in the robot study no such compensating movement was made, but the physical fingertip of the artificial hand was aligned with the target. Second, in the study with human participants, the participants moved their hand back and forth in space when switching between the homing position and the touchscreen. This potentially allowed the tracking systems to get a better initial estimate of the finger position which could then be updated as the finger moved towards the touch surface. Also, in the study with human participants, participants were free to move their head back and forth to facilitate tracking. In contrast, in the robot study, the artificial hand was moved in a fixed plane (mostly) parallel to the HMD. Also, the HMD was fixed. This resulted in substantially less variance in the Z-distance between the hand and the HMD. Third, compared to the Oculus Quest, Vive and

LeapMotion have substantially smaller baselines for their sensors (Vive uses two cameras with a baseline of ca. 6.5 cm, LeapMotion also uses two cameras with a baseline of ca. 4 cm, Quest uses four cameras with horizontal baselines between 12 and 15 cm and vertical baselines of approximately 7.5 cm). This also likely contributes to a worse accuracy at larger distances for Vive and LeapMotion compared to Quest.

The paper is looking only at tracking a single fingertip, visible to the tracking cameras. However, there are additional factors that may reduce the quality of tracking. For example, occlusion of the fingers by the hand or another hand or object may increase the error of fingertips locations. While the general shape of the hand, as rendered by the HMD, may look plausible, location-critical rendering, such as use of world-grounded haptics may be sensitive to such errors. Also, fast movements of the hands or the head may reduce the quality of tracking due to motion blur. In this paper we focused on moderate hand motions and head rotations. Finally, there are hand trackers, such as the Leap Motion, that may be positioned on the screen, or on the table. Due to being separated from the HMD, this generates problems of portability, power, and calibration, but these locations may give the sensors better view of the fingers, unclouded by the back of the hand, and less dependency on the accuracy of the head tracking by the HMD.

5.1 Limitations

Similar to prior work [27], our study focused on accuracy as the primary objective measure for characterising performance for touch-based interaction. We deliberately chose this measure to be able to compare measures from the study with human participants and the robot study. However, prior work [23] has indicated that latency can have a stronger effect on human performance in 2D pointing and constrained 3D object movement tasks. Hence future work may consider exploring throughput as an additional measure for jointly capturing movement speed and accuracy in pointing tasks. Also, our work focused on the combination of a VR HMD with the Leap Motion Sensor, but we are aware that the Leap Motion is also used in AR HMDs (e.g., [3]). However, prior work also suggests that differences in performance between AR and VR HMDS when using the same sensing system, such as Leap Motion, is likely due to the display system [3]. For the AR HMDs we deliberately did not show a virtual finger tip to avoid confusion on whether to align

the physical or virtual finger with the target. However, showing a virtual finger tip could allow users to better react to potential tracking issues, and, therefore lead to a higher spatial accuracy. In the study with human participants we used both VERTICAL and HORIZONTAL orientations while in the robot study we only used the VERTICAL orientation, as it was not possible for us to mount the arm in a horizontal fashion. We also did not use a touchscreen in the robot study since the robot could position the hand accurately in a predefined plane. Hence, technically, no touch interactions were performed in the robot study. Also, the lack of a touchscreen could potentially be considered a confounding factor as potentially distracting reflections of the touch display were not present in the robot study. We also considered including further orientations (e.g., 45°), but did not do so due to the already long duration of the study with human participants. Finally, the study solely reported results from the overall spatial accuracy. Theoretically, individual components of the HMDs overall tracking architecture could be exchanged, for example, the hand tracking systems could be combined with outside-in tracking systems for tracking the HMD position.

6 CONCLUSIONS

We have evaluated commodity hand tracking systems for AR and VR HMDs including the Oculus Quest, Vive Pro, Microsoft HoloLens 2 and Magic Leap One, as well as the Leap Motion controller. We reported and compared the accuracy for absolute and relative pointing tasks in order to inform researchers' and designers' decisions when building new systems or experiences using finger tracking. Future work includes user experience research with touch displays and finger tracking to allow supporting touch and hover tracking at the same time to create experiences that uses the approach trajectory of the fingers to seamlessly adapt the user interface accordingly. The accuracy of hand tracking when interacting in mid-air is also a fruitful avenue of future work. To this end, controlled experiments could be conducted investigating performance in larger interaction volumes.

REFERENCES

[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] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George W Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR.. In *CHI*, Vol. 17. 5643–5654.
[3] Anil Ufuk Batmaz, Aunnoy K Mutasim, Morteza Malekmakan, Elham Sadr, and Wolfgang Stuerzlinger. 2020. Touch the wall: Comparison of virtual and augmented reality with conventional 2D screen eye-hand coordination training systems. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 184–193.
[4] Anil Ufuk Batmaz, Aunnoy K Mutasim, and Wolfgang Stuerzlinger. 2020. Precision vs. Power Grip: A Comparison of Pen Grip Styles for Selection in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 23–28.
[5] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
[6] Michelle A Brown, Wolfgang Stuerzlinger, and EJ Mendonça Filho. 2014. The performance of un-instrumented in-air pointing. In *Graphics Interface 2014*. AK Peters/CRC Press, 59–66.
[7] Martin A Fischler and Robert C Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (1981), 381–395.
[8] Thomas B Fitzpatrick. 1975. Soleil et peau. *J Med Esthet* 2 (1975), 33–34.
[9] Jens Grubert, Lukas Witzani, Eyal Ofek, Michel Pahud, Matthias Kranz, and Per Ola Kristensson. 2018. Effects of hand representations for typing in virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 151–158.
[10] Shangchen Han, Beibei Liu, Randi Cabezas, Christopher D Twigg, Peizhao Zhang, Jeff Petkau, Tsz-Ho Yu, Chun-Jung Tai, Muzaffer Akbay, Zheng Wang, et al. 2020. MEgATrack: monochrome egocentric articulated hand-tracking for virtual reality. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 87–1.
[11] Chris Harrison, Hrvoje Benko, and Andrew D Wilson. 2011. OmniTouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 441–450.
[12] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
[13] 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 int. journal of aviation psychology* 3, 3 (1993), 203–220.
[14] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.
[15] Summer Lindsey. 2017. *Evaluation of Low Cost Controllers for Mobile Based Virtual Reality Headsets*. Ph.D. Dissertation.
[16] Giulio Marin, Fabio Dominio, and Pietro Zanuttigh. 2016. Hand gesture recognition with jointly calibrated leap motion and depth sensor. *Multimedia Tools and Applications* 75, 22 (2016), 14991–15015.
[17] 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.
[18] 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. In *25th ACM Symposium on Virtual Reality Software and Technology*. 1–11.
[19] Hugo Romat, Andreas Fender, Manuel Meier, and Christian Holz. 2021. Flashpen: A High-Fidelity and High-Precision Multi-Surface Pen for Virtual Reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 306–315.
[20] MS Sanders and EJ McCormick. 1993. Applied anthropometry, work-space design and seating. *Human factors in engineering and design* 7 (1993).
[21] Daniel Schneider, Alexander Otte, Axel Simon Kublin, Alexander Martschenko, Per Ola Kristensson, Eyal Ofek, Michel Pahud, and Jens Grubert. 2020. Accuracy of commodity finger tracking systems for virtual reality head-mounted displays. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 805–806.
[22] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based text entry in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
[23] Robert J Teather, Andriy Pavlovych, Wolfgang Stuerzlinger, and I Scott MacKenzie. 2009. Effects of tracking technology, latency, and spatial jitter on object movement. In *2009 IEEE symposium on 3D user interfaces*. IEEE, 43–50.
[24] James J Tung, Tea Lulic, Dave A Gonzalez, Johnathan Tran, Clark R Dickerson, and Eric A Roy. 2015. Evaluation of a portable markerless finger position capture device: accuracy of the Leap Motion controller in healthy adults. *Physiological measurement* 36, 5 (2015), 1025.
[25] Pier Paolo Valentini and Eugenio Pezzuti. 2017. Accuracy in fingertip tracking using leap motion controller for interactive virtual applications. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 11, 3 (2017), 641–650.
[26] Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. 2013. Analysis of the accuracy and robustness of the leap motion controller. *Sensors* 13, 5 (2013), 6380–6393.
[27] Robert Xiao, Julia Schwarz, Nick Throm, Andrew D Wilson, and Hrvoje Benko. 2018. MRTouch: adding touch input to head-mounted mixed reality. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1653–1660.