Usability test of a modified Leap Motion hand-sensing virtual reality interface in a virtual grasping-and-placing task

Master Thesis
submitted in fulfilment of the requirements for the degree
Master of Arts (M.A.)
in the master's program „Mind and Brain"

Humboldt-Universität zu Berlin
Berlin School of Mind and Brain

Handed in by Alexander Masurovsky
date of birth 06.10.1987
in Washington, DC, USA
Address Theodor-Heuss-Platz 5, 14052 Berlin, Germany

1. Supervisor: Prof. Dr. Carsten Finke
2. Supervisor Dr. Michael Gaebler

Berlin, den ……….. 5 March, 2020

Table of contents

Index of figures

Index of tables

Abstract

Hand-sensing technology such as the Leap Motion has enabled the creation of controller-free virtual reality (VR) interfaces which could, compared to handheld controllers, increase ecological validity in research and transfer effects in rehabilitation and training. For the use case of grasping and placing virtual objects on a table, we modified the basic Leap Motion VR interface, adapting visual indicators and physical rules for object interaction. We systematically compared the modified Leap Motion, the basic Leap Motion, and the Oculus Touch controller along performance and subjective experience measures. Modifications reduced errors without affecting accuracy or total time per trial, though time to grab the cube was increased and time to place the cube was decreased. The Oculus controller outperformed the modified Leap Motion on all performance metrics and was rated higher on ease of use, precision, gripping and releasing; we did not see expected higher ratings of naturalness, intuitiveness and agency for hand sensing. For this use case, our results show that modifying grab and release parameters and adding visual feedback improves the basic Leap Motion when accuracy and error rate are prioritized over speed of interaction; and, that hand sensing on its own does not offer any user experience advantage over the Oculus controller. Future iterations or combinations with other technologies (i.e., a haptic feedback device) may change this outlook.

1 Introduction

1.1 Object interaction in virtual reality

The capability for object interaction in immersive virtual reality has increased its applicability in various fields, including research, training, rehabilitation, and gaming. Immersive, interactive VR[1] creates a sense of presence, the sense of being in the virtual world (Slater, 2014), and immersion, the level to which one experiences VR is if it were real (Kim et al., 2017; Slater, 2014), by blocking out vision of the outside world using a headset and presenting a constructed 3D world that updates according to the user's head movements (Rizzo & Koenig, 2017). Handheld controllers such as the Oculus Touch and HTC Vive enable hand interaction with objects presented within the virtual environment. Object interaction in VR allows not only the development of games, which comprise of 43% of the VR market (Dujmovic, 2019), but apps for more "serious" topics

[1] In this paper, "VR" from this point refers to immersive, interactive virtual reality.

as well. Healthcare and education were two areas expected by VR industry professionals to receive high investment for the development of VR solutions in 2018 (Don et al., 2018). Recent studies have tested VR for safety and equipment training in mining scenarios (e.g., Kim & Choi, 2019) and grip training for manufacturing scenarios (e.g., Geiger et al., 2018). In neuroscience research, VR can increase ecological validity – the degree to which study findings apply to the real world – by allowing for increased control over the subject's environment, the presentation of dynamic, multisensory stimuli, and the ability to make precise, automatic measurements (Park et al., 2018; Hofmann et al., 2018; Tromp et al., 2017; Parsons, 2015). This makes it a promising tool for neurorehabilitative therapy as well, where it could improve diagnostic capabilities and increase transfer effects, bringing benefits to the the patient's life outside the clinic (Belger et al., 2019; Massetti et al., 2018; Pedroli et al., 2018).

1.2 Controller-free object interaction

Currently, the standard way to achieve object interaction in VR is with a handheld controller such as the Oculus Touch or HTC Vive, but advancements in the quality and affordability of hand-sensing technology have made it a candidate for improving this function. Hand-sensing technology enables users to interact with virtual objects using their bare hands, potentially increasing the naturalness and intuitiveness of object interaction within VR. This could make VR an even more effective tool for the uses mentioned in 1.1 and increase adoption by untapped user markets, such as children, older adults and neurological patients, for whom the lack of intuitiveness might be a barrier. By removing the need for the controller, the equipment burden is also reduced, making the VR system overall more portable and compact. The Leap Motion system provides an effective option for controller-free interaction within VR environments via a small, lightweight, combined infrared and normal spectra-based camera. It can attach to the front of a VR headset to scan the user's hands during a VR session. Leap Motion provides a modifiable application program interface (API) and software development kit (SDK) for customization of the interface. Researchers have begun testing the feasibility and usability of Leap Motion interfaces in a variety of contexts (Belger et al, 2019; Geiger et al., 2018; Vosinakis & Koutsabasis, 2017; Argelaguet et al., 2016; Wozniak et al., 2016) and indicated a number of challenges and potential workarounds, detailed below.

1.2.1 Design challenges

There are challenges to implementing a controller-free interface for virtual object interaction. The somatosensory system (sense of touch) provides information about an object: texture, hardness, shape and *if* object interaction is occurring. When using hand-sensing technology in VR (without additional haptic feedback technology), the sense of touch provides no information about virtual objects (Wozniak et al., 2016; Slater, 2014). Handheld controllers, while also unable to provide the information afforded by the somatosensory system, have vibration capabilities, which can be leveraged by designers to indicate object interaction, and buttons with discrete states that indicate unambiguously when the user is trying to make an action such as grabbing. Lacking buttons is also an advantage for hand sensing, as it enables a more complex array of possible gestures.

Detection with hand sensing encounters different challenges than with a handheld controller. The Oculus Touch uses LED sensors, which rarely become occluded by the user's hands, to detect the position of the controller. With hand sensing, machine vision algorithms determine the positions of the user's fingers, which often occlude each other, leaving potential for error. The Leap Motion solution also requires the user to keep their head directed towards their hands, while the Oculus controller does not have this restriction. Other technologies can potentially be combined with the Leap Motion to compensate for this gap in coverage, but our aim was to develop and test a prototype based in the Leap Motion with minimal additional hardware.

The novelty of hand sensing creates another challenge. Handheld VR controllers are a small step from video game controllers. Hand-sensing interfaces, however, have few commonplace technological precursors. Only one, the Xbox Kinect sensor from Microsoft, ever approached widespread use. The Kinect enables gesture recognition as a supplement to a handheld controller for gaming and has been tested for rehabilitative uses (e.g., Proffitt et al., 2015) and other niche markets. Microsoft ended production of the Kinect in 2017; while a reconceptualization of it now exists as the Microsoft Azure Kinect, marketed as a versatile sensor kit (Azure Kinect DK, 2020), the fate of the original Kinect shows that a controller-free interface has not yet reached sustained, widespread user adoption. Nintendo and Sony, Microsoft's larger competitors, do not support mainstream controller-free interfaces for use with their gaming systems. People are therefore unlikely to have had much experience with a hand-sensing interface in the past.

1.2.2 Addressing design challenges

These challenges can be addressed: Good design can make unfamiliar technology more intuitive (Norman, 2013). To compensate for the lack of somatosensory feedback, visual and auditory feedback can be added to indicate stages of object interaction. Color changes to both the virtual representation of the hand and to the object to indicate interaction has shown success in improving time on task, accuracy of placement and the subjective user experience (Geiger et al., 2018; Vosinakis & Koutsabasis, 2017; Argelaguet et al., 2016). The customizable Leap Motion SDK also allows for the rules of object interaction be altered to offset common errors made by users. One way to do this is to change the definition of the gripping hand shape (i.e., which fingers can be involved, how close together they must be to form a grip), or requiring the user to hover over the object with their hand before being able to grab it. Here we present a prototype hand-sensing interface based on modifications to the Leap Motion system (referred to in this paper as the HHI_Leap). A summary of our modifications can be found in Table 1.

1.2.3 Natural interaction: hands-free vs. handheld controller

While it seems logical that transposing hand movements directly into VR creates a more natural user experience than a handheld controller, this has not yet been supported by research. Naturalness within VR can be construed as how similar an action feels to performing it in real life. It is an important factor in two anticipated benefits of immersive VR: ecological validity in research and transfer in rehabilitative therapies. Two related studies assumed that a Leap Motion-based hand-sensing interface for object-interaction in VR would enable a more natural user experience (Vosinakis & Koutsabasis, 2018; Argelaguet et al., 2016), and another study used a version of our interface (the HHI_Leap) to enable what they refer to as "natural (i.e., controller-free) hand interaction," aiming to maximize ecological validity and transferability in a rehabilitation task (Belger et al., 2019). Immersion is assumed by some researchers (i.e., Belger et al., 2019; Rizzo et al., 2018) to be related to ecological validity; Kim and Choi (2019) found greater immersion with a Kinect-based hand-tracking interface combined with a haptic glove than with a handheld controller. We did not find any previous studies that directly compared naturalness between a hand-sensing interface on its own and a controller.

Naturalness is a broad construct that deserves more attention than the scope of our study allowed. We included one questionnaire item addressing naturalness and two questions addressing related constructs: intuitiveness and agency (see Appendix C: Questionnaires for specific question wording). The sense of agency, a feeling of control over events, body parts or tools (Krugwasser et al., 2019; Caspar et al., 2015; Haggard

& Tsakaris, 2009), may be important to the naturalness of the experience. In VR, agency has been found to be related to the amount of control the user feels over a virtual hand avatar (Argelaguet et al., 2016). Based on its affordance for bare-hand object interaction (Vosinakis & Koutsabasis, 2018), and the assumption in the aforementioned studies of its higher naturalness, we hypothesized that our hand-sensing interface would be rated higher on naturalness, intuitiveness and agency.

1.3 Usability testing

Interface design can be biased by the intuitions of the designer, to the detriment of the user (Norman, 2013). Usability testing gathers data from users to determine what does and does not work about an interface (Lazar et al., 2017). Incorporating cycles of usability testing improves the design process by confronting designer intuitions with data (Spurlock & Ravasz, 2019). Typical measures include accuracy, time and error rate to evaluate performance and questionnaire data to understand the user's subjective experience (Tullis and Albert, 2013). While usability studies of controller free interfaces exist (e.g., Vosinakis & Koutsabasis, 2018; Geiger et al., 2018), we found few that systematically measured the usability of a hand-sensing prototype, such as one based in the Leap Motion, alone (Kim et al., 2017 present a hand-sensing prototype that is combined with a haptic glove). Furthermore, we found no studies that systematically compared the usability of a hand-sensing interface alone to a controller for object interaction within VR. Our research here aims to fill this gap in the research by presenting a hand-sensing interface and an accompanying usability study, in which we include a systematic comparison to a handheld controller.[2]

1.4 Present study

Here we present a prototype for a hand-sensing interface based on Leap Motion technology for use in VR, as well as the results of a usability test that systematically compared our prototype (HHI_Leap) to the basic Leap Motion system (B_Leap) and to an Oculus Touch (Oculus) handheld controller. The Oculus Touch, already available to consumers, was included to compare our prototype to an accepted standard for object interaction in VR. The initial impression by our designers was that the Oculus controller was generally higher performing and easier to use than our prototype.

Our research questions were: 1) Did modifications to the Leap Motion interface improve its usability for the use case of grabbing and placing objects in VR? We predicted that the HHI_Leap would have have better accuracy (lower distance displacement from target), lower times on all time metrics and fewer accidental drops than the B_Leap, and receive higher ratings on all measures of subjective experience; and 2) How did the modified hand-sensing interface (HHI_Leap) compare to the Oculus controller on measures of usability for this use case? Due to challenges presented in 1.2.1 and initial impressions of the designers, we predicted that the HHI_Leap would have lower performance than the Oculus controller on all performance measures and receive lower ratings on all measures of subjective experience *except* for naturalness, intuitiveness and agency, on which we predicted the HHI_Leap would receive higher ratings than the Oculus controller.

---

[2] According to guidelines by Wobbrock and Kientz (2016), our work here serves as an artifact contribution, by presenting an interface prototype, and an empirical contribution, by presenting usability data about the three interfaces described here.

Table 1 Grab rules and visual specifications for interfaces in the present study

	Basic Leap (B_Leap)	HHI-modified Leap (HHI_Leap)	Oculus Touch controller
Grab rules	- no restrictions on grab	- must hover over object with open hand before grab is possible - if grab is attempted before hover, must open hand to non-grab state and wait 0.35 seconds before attempting hover again - if hand is moving too fast, must wait 0.25 seconds before hover is possible	- push trigger buttons on controller to grab
Visual specifications	- hand representation is solid grey - no additional visual feedback	- hand is semi-transparent grey when empty - green spotlight projects from hand when grab is possible - hand turns green when object is grabbed/held	- hand representation is solid grey

2 Methods

2.1 Study design

To address our research questions (section 1.3), we used a within-subjects design with interface (B_Leap, HHI_Leap, Oculus controller) as our independent variable. All subjects completed the experimental task using all interfaces and answered questions about their experience after each interface. Performance measures were recorded automatically during the task; subjective measures were derived from the questionnaire data (Fig. 2). We also conducted an exploratory analysis to determine if interfaces differed significantly at different cube sizes, with interface (B_Leap, HHI_Leap, Oculus controller) and cube size (small, medium large) as within-subjects factors.

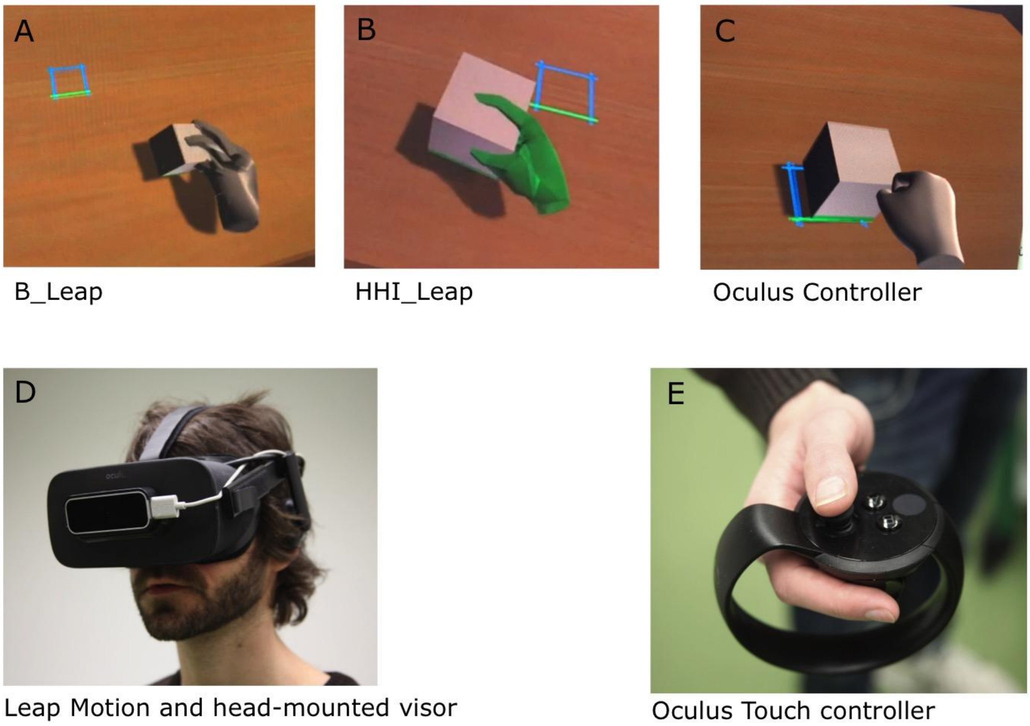


Fig. 1 Virtual representations of the hands for (A) B_Leap, (B) HHI_Leap, (C) Oculus controller, the (D) mounted Leap Motion sensor and (E) Oculus Touch controller

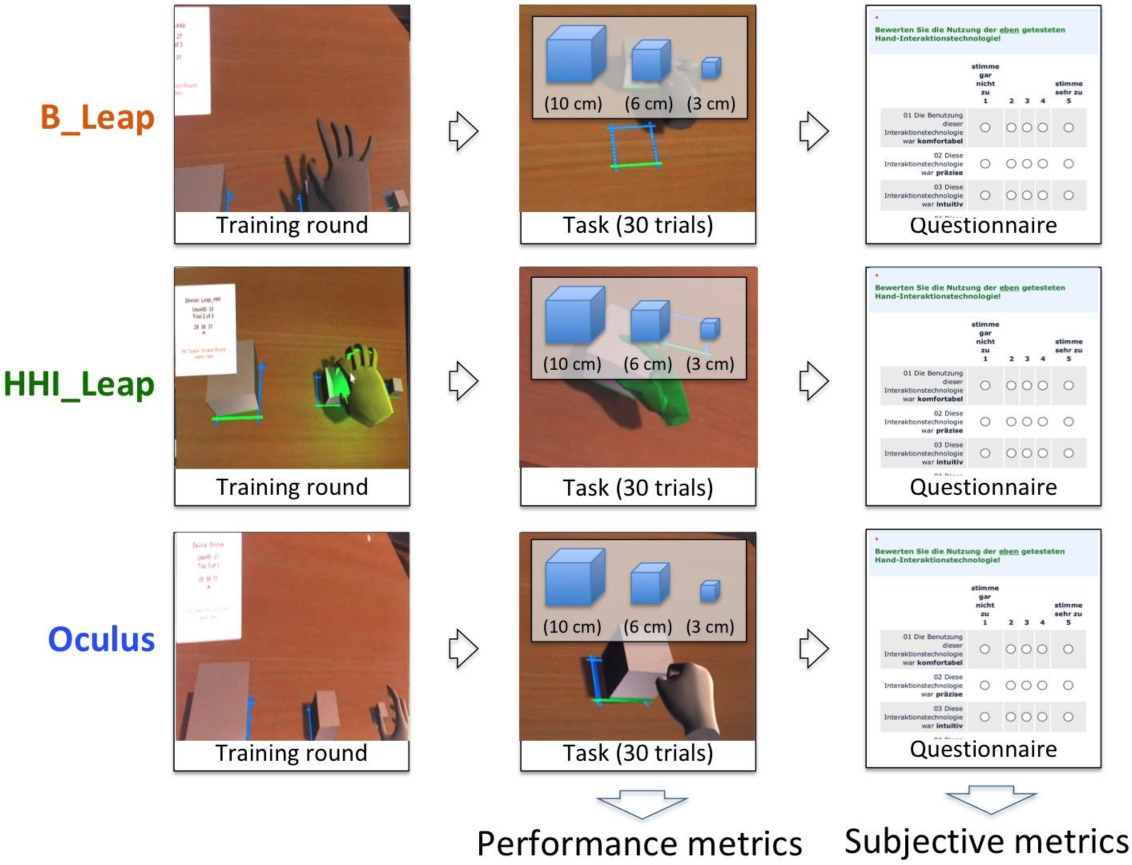


Fig. 2 Task procedure: practice round (left panel), task (middle panel) and questionnaire session (right panel)

2.2 Sample

Subjects were recruited from employees, interns and their associates at the Fraunhofer Heinrich-Hertz-Institut in Berlin. The only inclusion criteria were that they had no injuries or vision impairments that would prevent them from completing the task. All 32 subjects (23 males, 8 females, 1 undisclosed) were included and completed the study. Ages ranged from 22 to 36 years old (M = 27.8, SD = 3.34) and heights ranged from 159 cm to 194 cm (M = 177 cm, SD = 9.13) (see Table 9 for all demographics). The majority of subjects (20) reported no (10) or some (10) prior experience with virtual reality (Fig. 25). All subjects reported at least some prior experience with video game controllers (Fig. 24) and most subjects (16) reported frequent current use with electronic games (including mobile phone games) (Fig. 26). Table 2 (below) shows the number of subjects who completed the study in each combination of interface orders.

Table 2 Interface order group counts

Note: The first row shows the number of subjects who completed the study for each interface order (B=B_Leap, H=HHI_Leap, O=Oculus). The second row shows interface orders grouped into whether the subject used the B_Leap before the HHI_Leap (B_Leap first) or the HHI_Leap before the B_Leap (HHI_Leap first).

BHO	BOH	HBO	HOB	OBH	OHB
5	6	7	3	6	5

B_Leap first (BHO, BOH, OBH)	HHI_Leap first (HBO, HOB, OHB)
17	15

2.3 Task

Participants completed a virtual pick-and-place task using all three interfaces in randomized order. The task was to place virtual cubes, one at a time, onto a target area on a virtual table (Fig. 3, Fig. 4). There were three different cube sizes: small (3x3 cm), medium (6x6 cm) and large (10x10 cm). 10 cubes of each size were presented (with size order randomized) for a total of 30 trials for each interface. At the beginning of each trial, the cube would appear at one of 10 positions (with position order randomized) at a distance of 30 cm from the target (Fig. 5). The subject was asked to use their dominant hand to place each cube as quickly and accurately as possible onto a 2-dimensional square target, the size of one face of the cube drawn onto the table (screenshots of the cube and target can be seen in Fig. 1). After grabbing the cube, the subject had one chance to place it on the target; they could not adjust the cube once it made contact with the table.

Fig. 3 Subject engaging in task (front view)

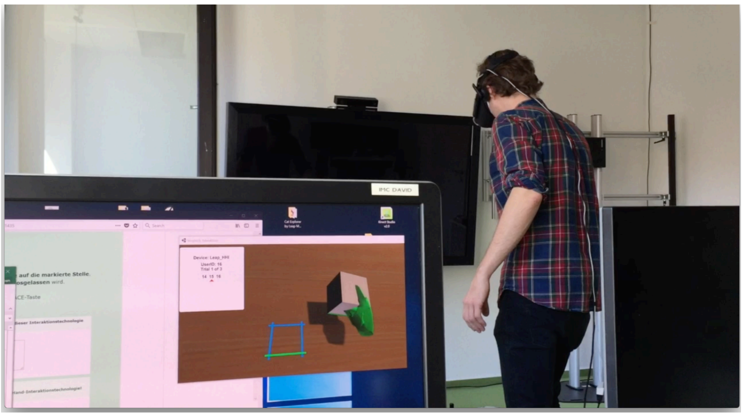


Fig. 4 Subject engaging in task (back view) and computer setup

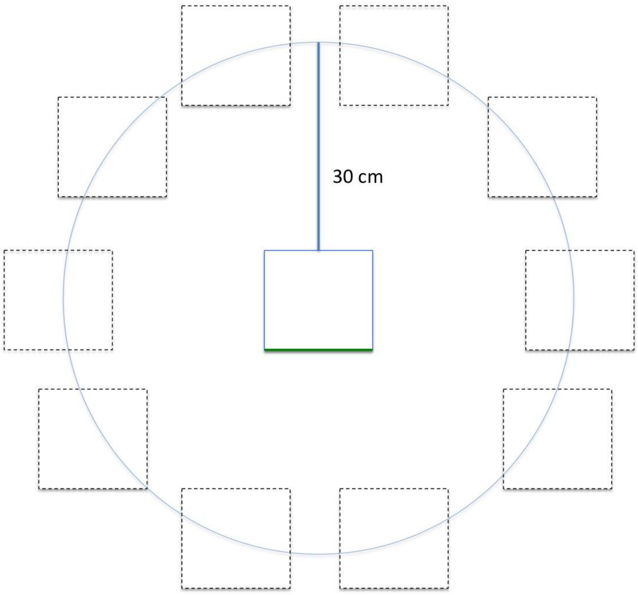


Fig. 5 Diagram of target and cube spawn positions: cubes spawned (appeared) at one of 10 positions located 30 cm from the target

2.4 Measures

2.4.1 Performance Measures

Accuracy: Accuracy was measured as the Euclidean distance, in meters, from the 2D center of the bottom face of the cube to the center of the target square.

Total time per trial: Total time per trial was the time from cube spawn (appearance on the table) until the time the cube once again made contact with the table after having been picked up, equal to the sum of the following two time measures.

Grab time (time from spawn to grab): Time from when the cube appeared on the table to the time it was grabbed by the subject.

Release time (time from grab to placement): Time from when the object was grabbed to the time the cube made contact with the table after being released by the subject.

Accidental drops: Accidental drops were recorded as a binary error metric. The experimenter noted in which trials the subject appeared to accidentally drop the cube before getting a chance to try to place it on the target. During analysis, these and additional trials that met certain criteria (see 2.5 below) were removed from the calculation of the other performance metrics (listed above) and counted as errors.

2.4.2 Subjective Experience Measures

System Usability Scale (SUS): The SUS is a "quick and dirty" (Brooke, 1996) measurement tool to assess the usability (primarily ease of use) of any sort of product, from websites to cell phones to kitchen appliances. It contains 10 5-point Likert scale questionnaire items with response options from 1 (strongly disagree) to 5 (strongly agree). Responses are transformed by a scoring rubric, resulting in a score out of 100.

Single Likert-scale subjective questions (5-point): Subjects responded to 8 Likert-scale questions assessing user experience (i.e., comfort, ease of gripping, likelihood to recommend to friends) with responses ranging from 1 (strongly disagree) to 5 (strongly agree) (Appendix A).

Agency (7-point Likert): Agency describes a feeling of control and connectedness to a part of one's own body or a representation of the body (Caspar et al., 2015), "when oneself is the agent of one's own actions" (Argelaguet et al., 2016). The item "I felt like I controlled the virtual representation of the hand as if it was part of my own body," on which participants rated the degree of agency on a scale from 1 (strongly disagree) to 7 (strongly agree) was selected from an agency questionnaire created by Argelaguet et al. (2016).

Overall satisfaction (7-point Likert): Subjects rated overall satisfaction on a Likert scale from 1 (least satisfied) to 7 (most satisfied) with smiley faces representing degree of overall satisfaction (Fig. 23).

Overall preference: One question was given after the subject completed the task with all 3 interfaces: "Of the three interfaces you used, which did you like best?" / "Welche der kennengelernten Möglichkeiten hat Ihnen am besten gefallen?"

See Appendix C: Questionnaires for questionnaire text.

2.5 Data cleaning and accidental drop counting

No trials were removed entirely. 246 of the original 2970 trials were these trials were removed from calculation of the other four performance metrics (accuracy, total time, grab time and release time) and added to accidental drop (error) counts. 2724 trials were included in the final calculation of accuracy, total time, grab time and release time.

Accidental drops were calculated using a combination of visual and metric-based methods. The visual method was for the researcher to watch each trial during testing for what looked like an accidental drop. The trial number was constantly displayed so that the researcher could record those trials in which an accidental drop occurred. For metric-based detection, we plotted grab time by release time for each interface (Fig. 6), binning each trial into an accuracy range (<0.1 m, 0.2 - 0.3 m, >0.3 m).

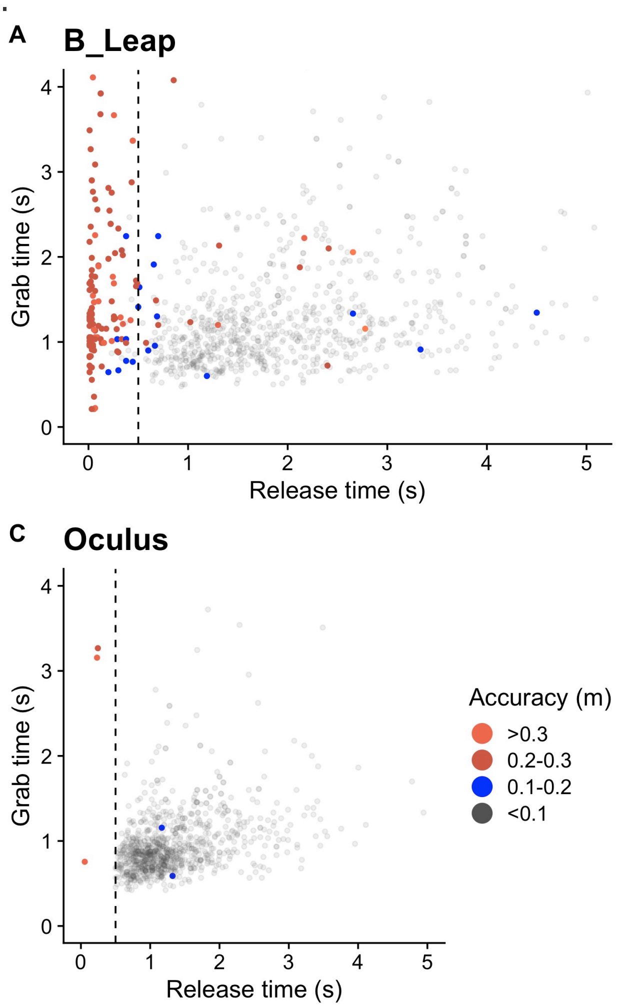


Fig. 6 Grab time by release time for each interface, with color coding for accuracy and dashed line indicating release time of 0.5 s

Using the visually-detected accidental drops as a guide (as we did not have prior research on which to base our criteria), we decided that trials with a release time of less than 0.5 seconds (left of dashed line) and an accuracy of greater than 10 cm (blue dots), or with an accuracy of greater than 20 cm (all shades of red dot) could reasonably be considered accidental drops. Of the 246 trials flagged as accidental drops (combination of both methods), 209 were detected visually during data collection and 37 were detected using the metric-based method. 181 of 246 error trials (74%) were flagged by both the visual and metric-based methods.

Two subjects, S12 and S30, were considered for removal due to extreme mean scores (> 3.5 SD's from the grand mean) on some performance metrics. We decided they should not be removed because we could not determine that there were extenuating circumstances (i.e., improperly functioning equipment or confusion about the task on the part of the subject) to warrant removal.

2.6 Analysis

For each metric, we conducted a one-way within-subjects ANOVA with three levels for the interface factor (B_Leap, HHI_Leap and Oculus controller) and followed up with paired-samples t-tests. For performance metrics, we also conducted an exploratory analysis of cube size, for which we used a 3x3 two-way within-subjects ANOVA with interface (B_Leap, HHI_Leap, Oculus controller) and cube size (small, medium, large) as factors, also followed up with paired-samples t-tests. For all statistical tests, a two-sided alpha of .05 was used to determine significance. Holm adjustment (Holm, 1979) was applied to p-values to correct for multiple comparisons. Paired-samples t-tests were conducted only between the HHI_Leap and the other two interfaces (we were not interested in the comparison between the B_Leap and Oculus). For measures of central tendency, we used means and 95% confidence intervals. Despite some controversy, parametric tests are a viable method for analyzing Likert-scale data, even when they violate assumptions (i.e., normally distributed data) for those analyses (Norman, 2010; Carifio & Perla, 2008; Meek et al., 2007). Effect sizes were calculated using Cohen's d (Cohen, 1988), with magnitude interpretations from Cohen (1992) applied via the rstatix package in R. The analysis was conducted in R Studio using the following packages: readr v. 1.3.1, readxl v. 1.3.1, tidyverse v. 1.3.0, dplyr v. 0.8.3, ggplot2 v. 3.2.1, see v. 0.3.0, cowplot v. 1.0.0, lsr v. 0.5, ggpubr v. 0.2.4, car v. 3.0-84 and rstatix v. 0.3.0 (the R code can be found in the accompanying R Markdown HTML file). In plots, asterisks are used to indicate Holm-adjusted p-values, using the following demarcations: No asterisks ($p > .05$), * ($p < .05$), ** ($p < .01$), *** ($p < .001$).

3 Results

3.1 Performance measures

3.1.1 Main results

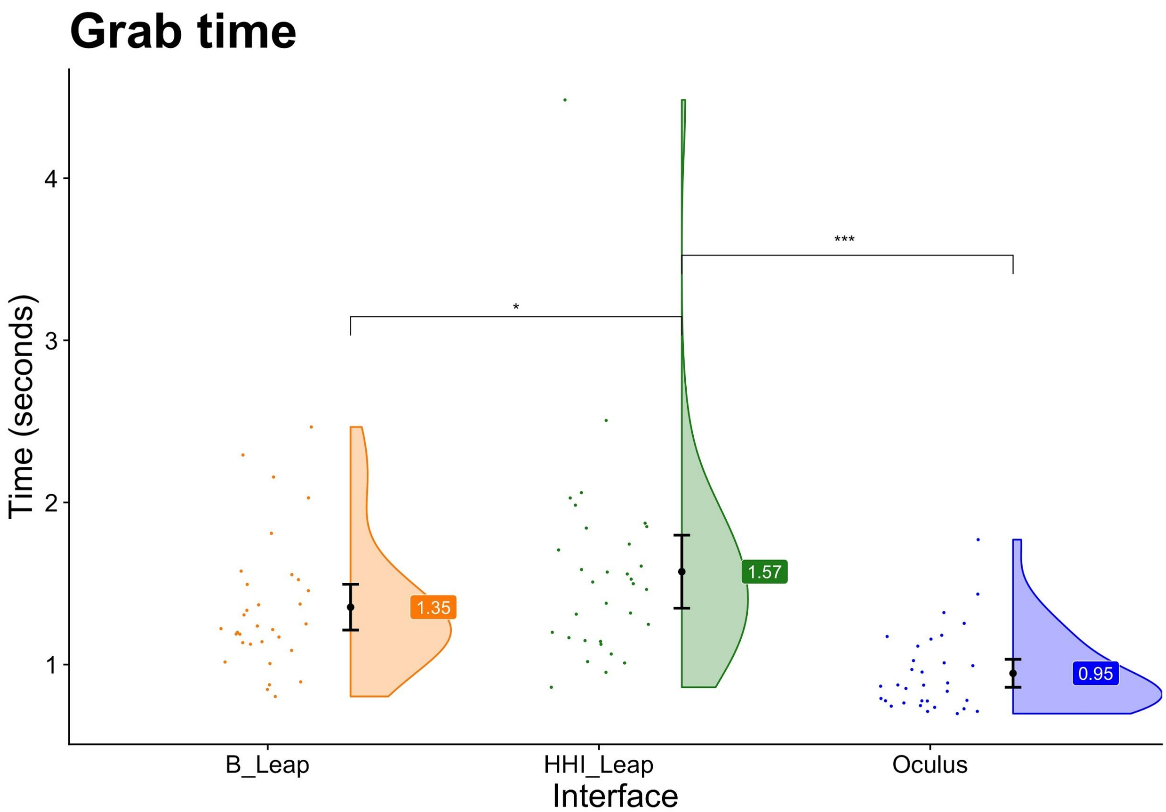


Fig. 7 Grab time by interface, with means and 95% confidence intervals, individual subject means (colored dots), smoothed density distributions and results of paired-samples t-tests

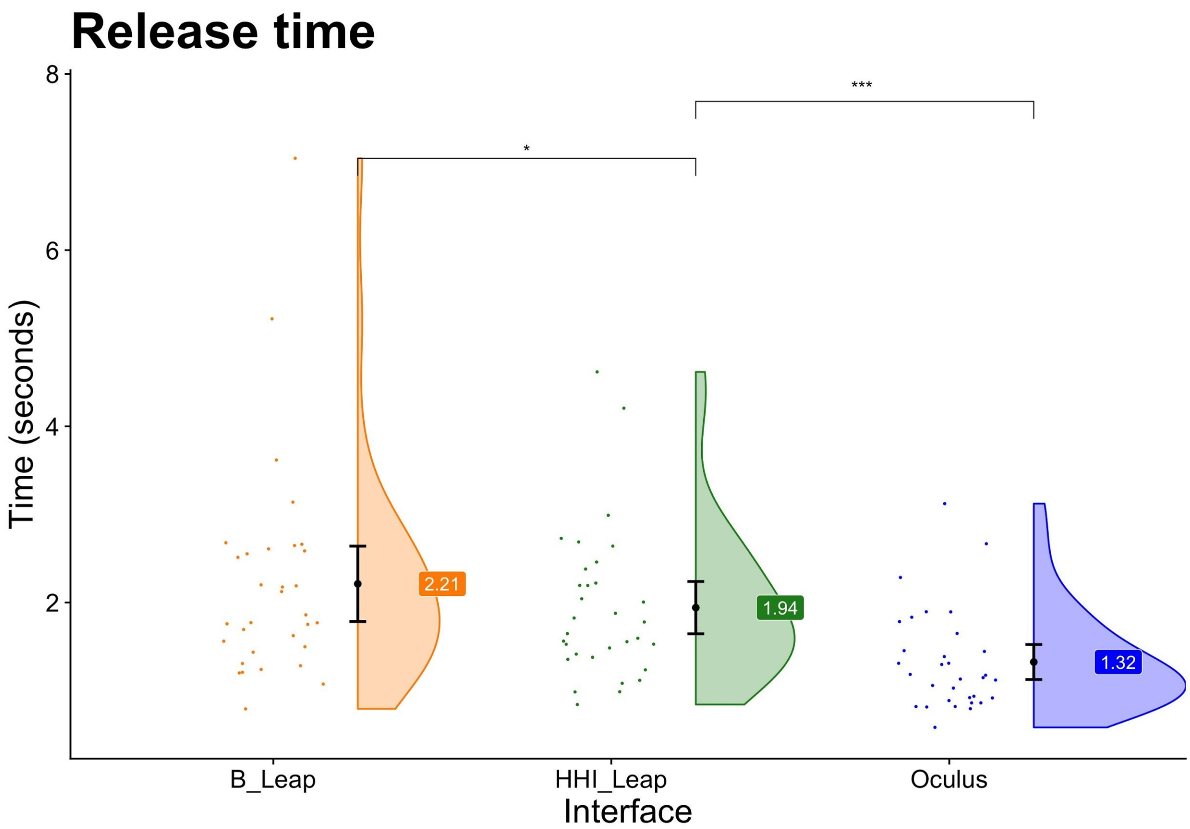


Fig. 8 Release time by interface, with means and 95% confidence intervals, individual subject means (colored dots), smoothed density distributions and results of paired-samples t-tests

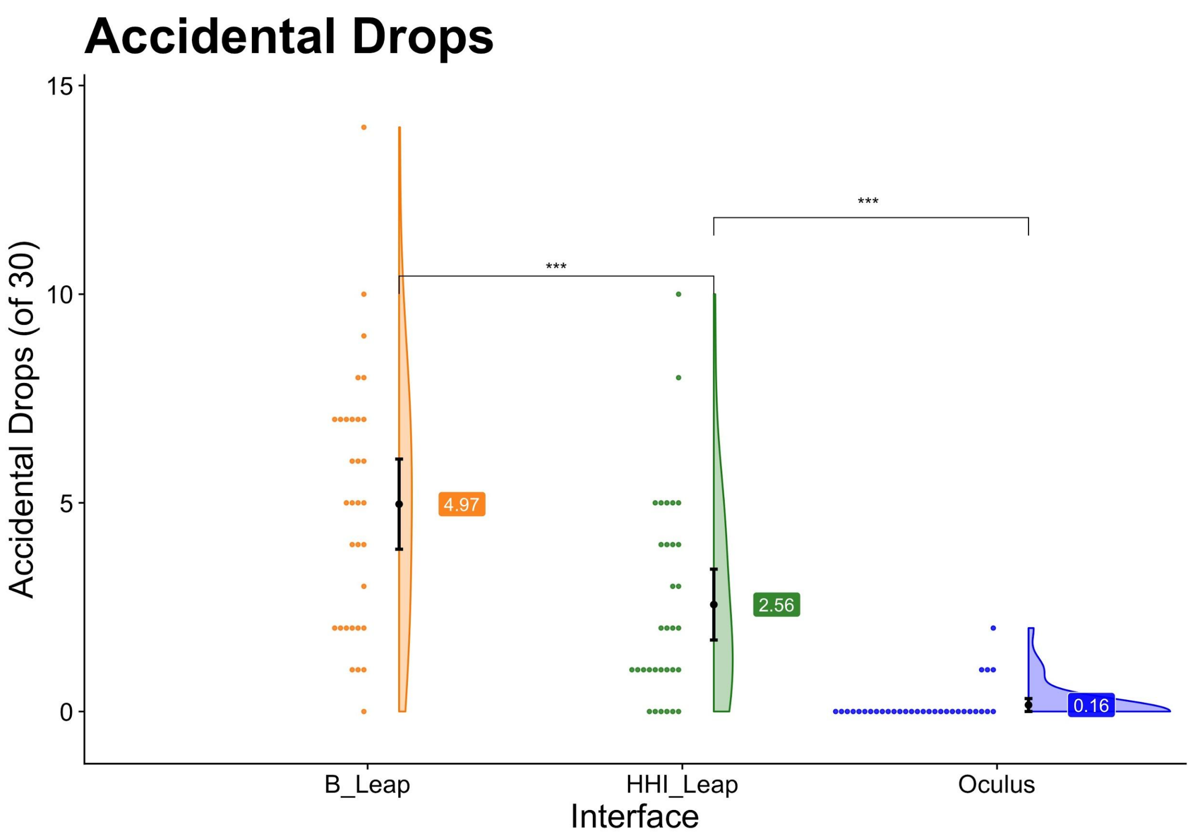


Fig. 9 Accidental drop errors by interface, with means and 95% confidence intervals, individual subject totals (colored dots), smoothed density distributions and results of paired-samples t-tests

One-way within-subjects ANOVAs for all performance metrics showed significant effects ($p < .001$; see Table 3). Pairwise t-tests showed no significant differences between the HHI_Leap and B_Leap for accuracy, $t(31) = -.029$, $p = .77$ (Fig. 17), or total time per trial, $t(31) = 0.29$, $p = .77$ (Fig. 18). For grab time, the HHI_Leap ($M = 1.58$, $SD = 0.65$) was significantly slower than the B_Leap ($M = 1.35$, $SD = 0.41$), $t(31) = 6.94$, $p = .032$, $d = 0.40$ (small) (Fig. 7), an average difference of 0.23 seconds; but for release time, the HHI_Leap ($M = 1.94$, $SD = 0.86$) was significantly faster than the B_Leap ($M = 2.21$, $SD = 1.24$), $t(31) = -2.34$, $p = .027$, $d = 0.41$ (small) (Fig. 8), an average difference of 0.27 seconds. For accidental drop errors, the HHI_Leap ($M = 2.56$, $SD = 2.45$) had significantly fewer errors than the B_Leap ($M = 4.97$, $SD = 3.12$), $t(31) = -3.95$, $p < .001$, $d = 0.70$ (moderate) (Fig. 9), an average difference of 2.41 drops. The Oculus controller outperformed the HHI_Leap on all performance metrics. See Table 3 for ANOVA results and Table 4 for paired-samples t-test results.

3.1.2 Exploratory results: Cube size

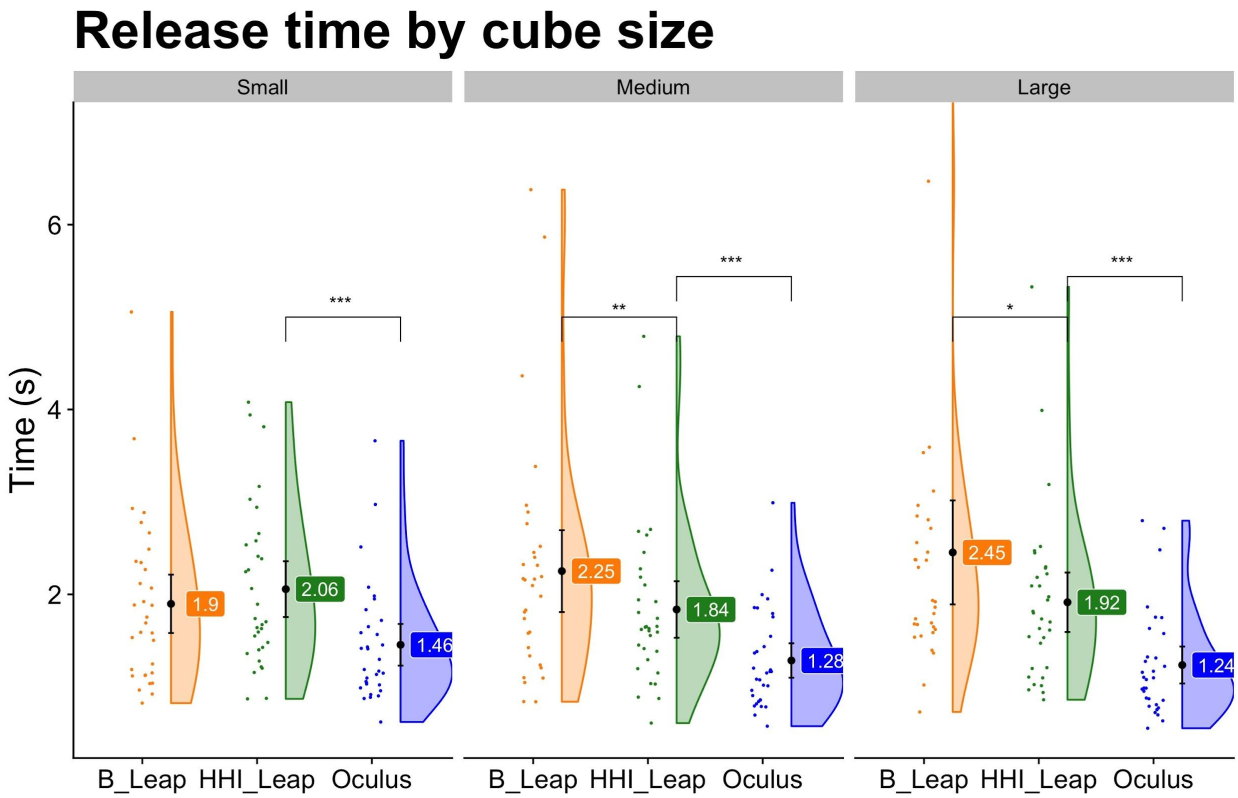


Fig. 10 Release time by cube size and interface, with means and 95% confidence intervals, individual subject means (colored dots), smoothed density distributions and results of paired-samples t-tests

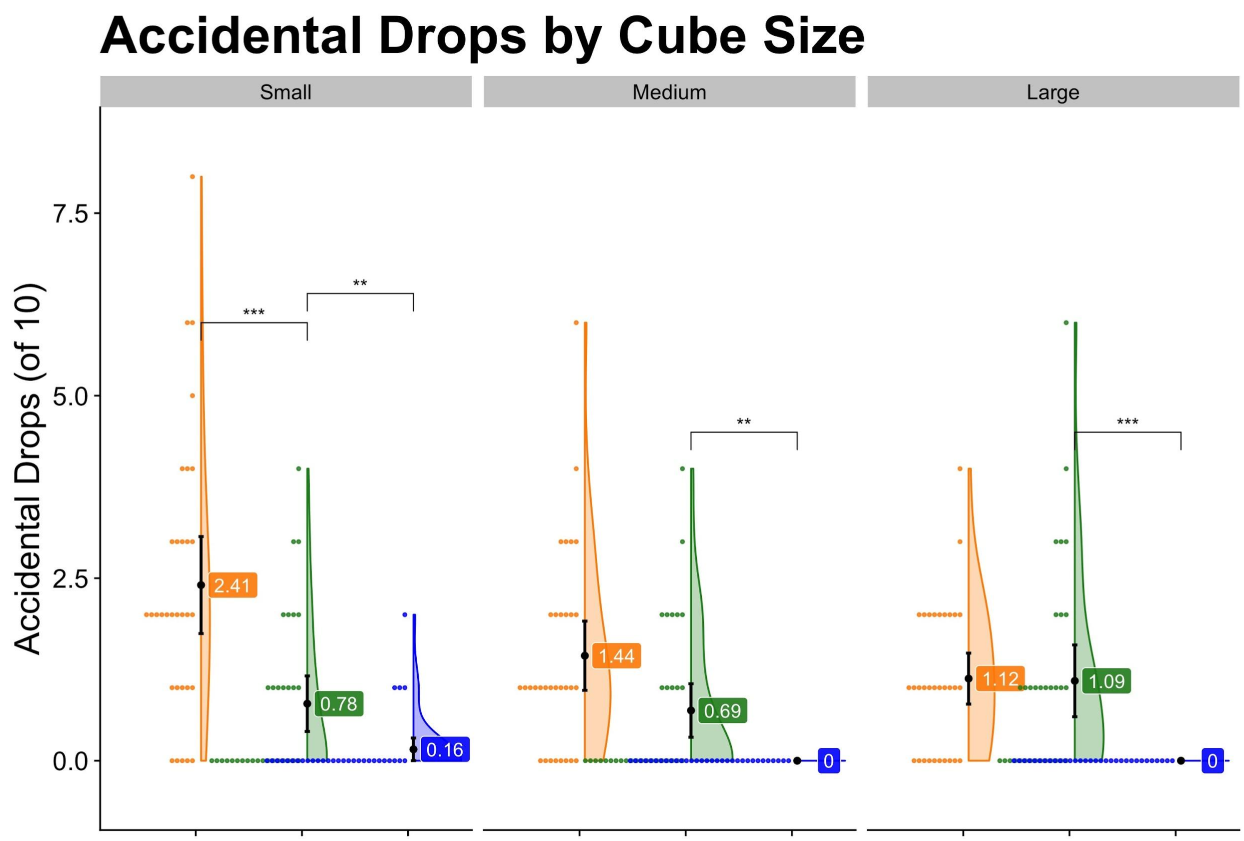


Fig. 11 Accidental drop errors by cube size and interface, with means and 95% confidence intervals, individual subject totals (colored dots), smoothed density distributions and results of paired-samples t-tests

Significant differences were found between the HHI_Leap and B_Leap for two metrics: release time (Fig. 10) and accidental drops (Fig. 11). The HHI_Leap had significantly lower release times than the B_Leap at the medium cube size (HHI_Leap: $M = 1.84$, $SD = 0.88$; B_Leap: $M = 2.25$, $SD = 1.28$), $t(31) = -3.42$, $p = .006$, $d = 0.31$ (small), and the large cube size (HHI_Leap: $M = 1.92$, $SD = 0.86$; B_Leap: $M = 2.45$, $SD = 1.62$), $t(31) = -2.85$, $p = .016$, $d = 0.31$ (small). For the accidental drop metric at the small cube size, the HHI_Leap ($M = 0.78$, $SD = 1.10$) had significantly fewer accidental drops than the B_Leap ($M = 2.41$, $SD = 1.92$), $t(31) = -4.40$, $p < .001$, $d= 0.40$ (small). At the medium cube size there was a non-significant trend between for fewer errors for the HHI_Leap ($M = 0.69$, $SD = 1.06$) than the B_Leap ($M = 1.44$, $SD = 1.37$), $t(31) = -2.25$, $p = .064$, $d = 0.40$ (small). No significant differences were found between the HHI_Leap and B_Leap for accuracy (Fig. 19), total time (Fig. 20), or grab time (Fig. 21). Performance metrics for the Oculus controller did not differ from main results when including cube size as a factor (see Table 5 for all paired-samples t-test results).

3.2 Subjective measures

3.2.1 System Usability Scale (SUS)

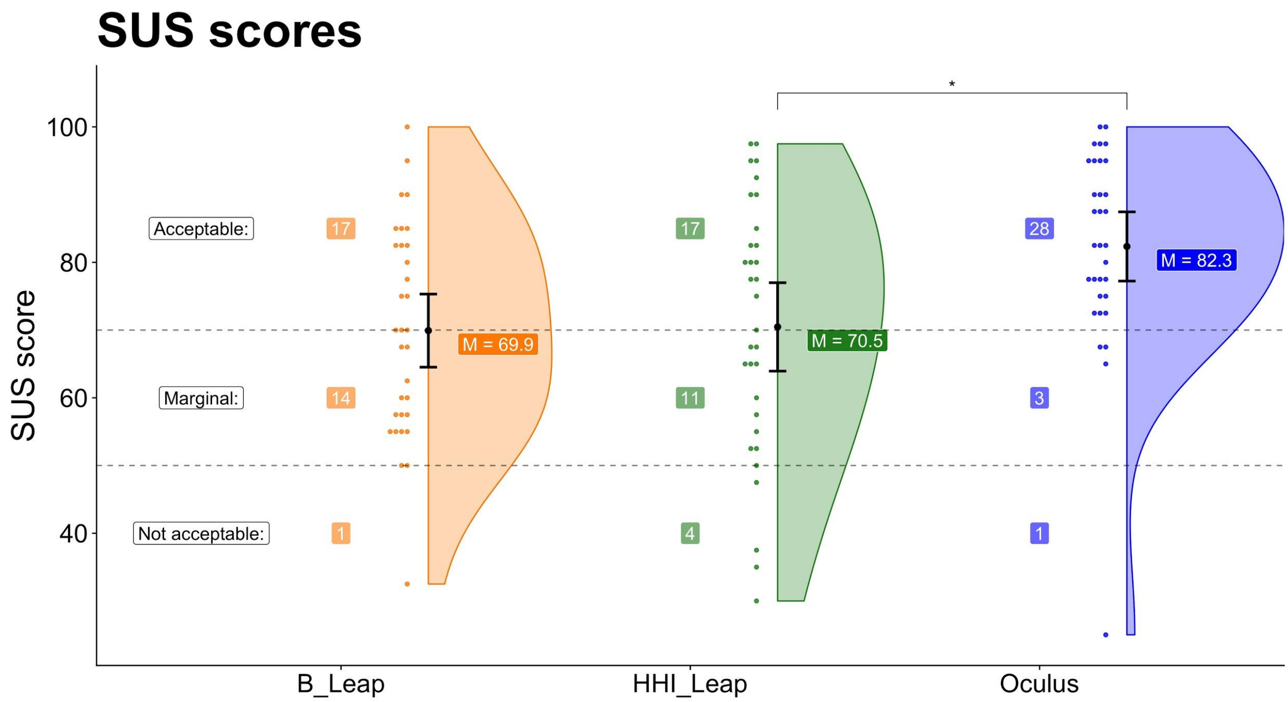


Fig. 12 SUS scores by interface, with means and 95% confidence intervals, individual subject scores (colored dots), smoothed density distributions, results of paired-samples t-tests and scoring categories as determined by Banger, Kortum and Miller (2009)

ANOVA showed a main effect of interface on SUS score, $F(2, 62) = 7.13$, $p = .002$, $\eta_p^2 = 0.19$. Paired-samples t-tests showed that the difference between the HHI_Leap ($M = 70.5$, $SD = 18.8$) and the B_Leap ($M = 69.9$, $SD = 16.6$) was not significant $t(31) = 0.19$, $p = .853$. Scoring guidelines based on SUS score data found by Bangor, Kortum and Miller (2009) put both Leap interfaces at the border (score = 70) of "acceptable" and "marginal." SUS scores for the HHI_Leap were significantly lower than for the Oculus controller ($M = 83.3$, $SD = 14.7$), $t(31) = -3.39$, $p = .006$, $d = 0.60$ (moderate) (Fig. 12).

3.2.2 Individual subjective questions

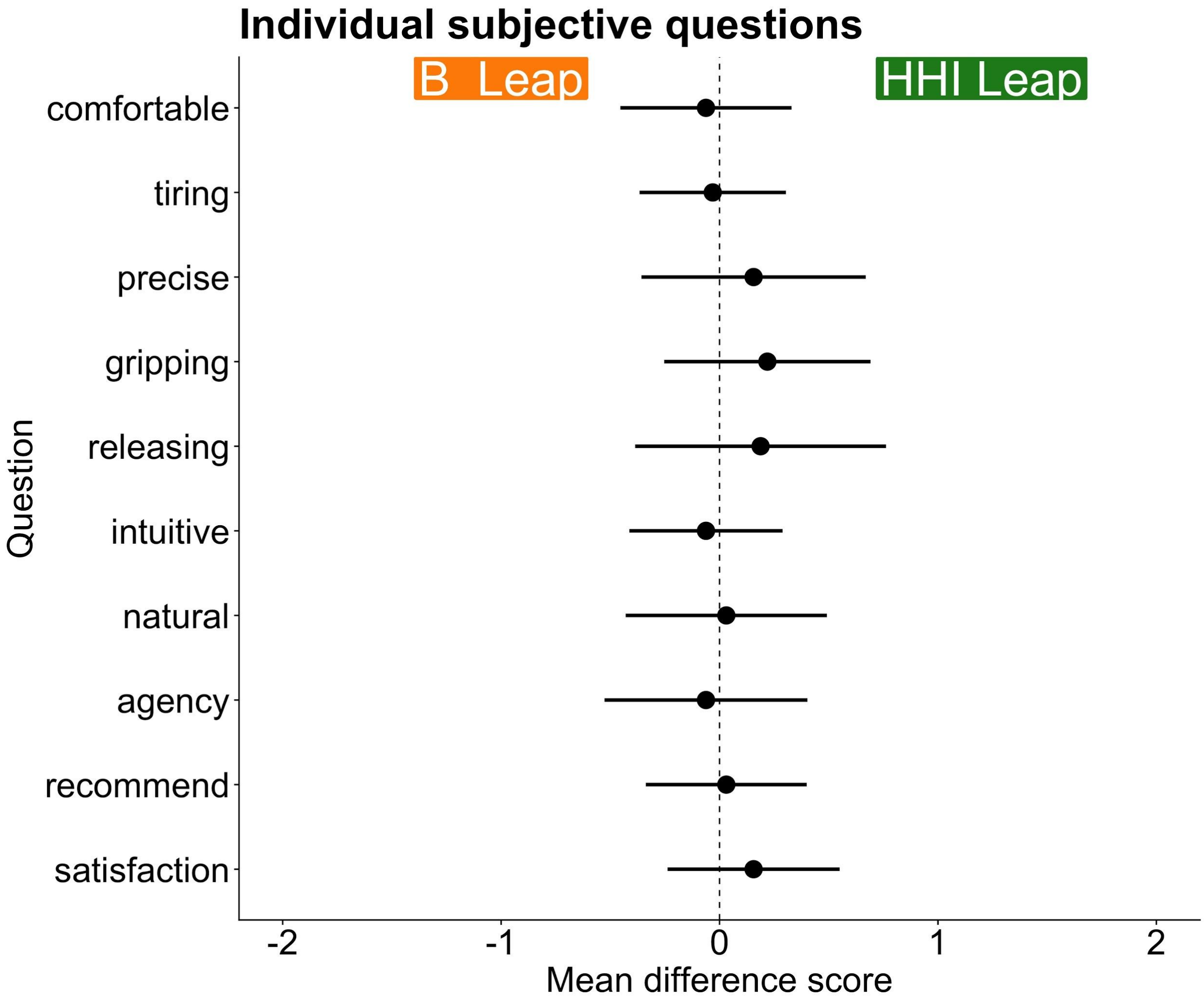


Fig. 13 Difference scores (HHI_Leap – B_Leap) for individual subjective questions, with means and 95% confidence intervals

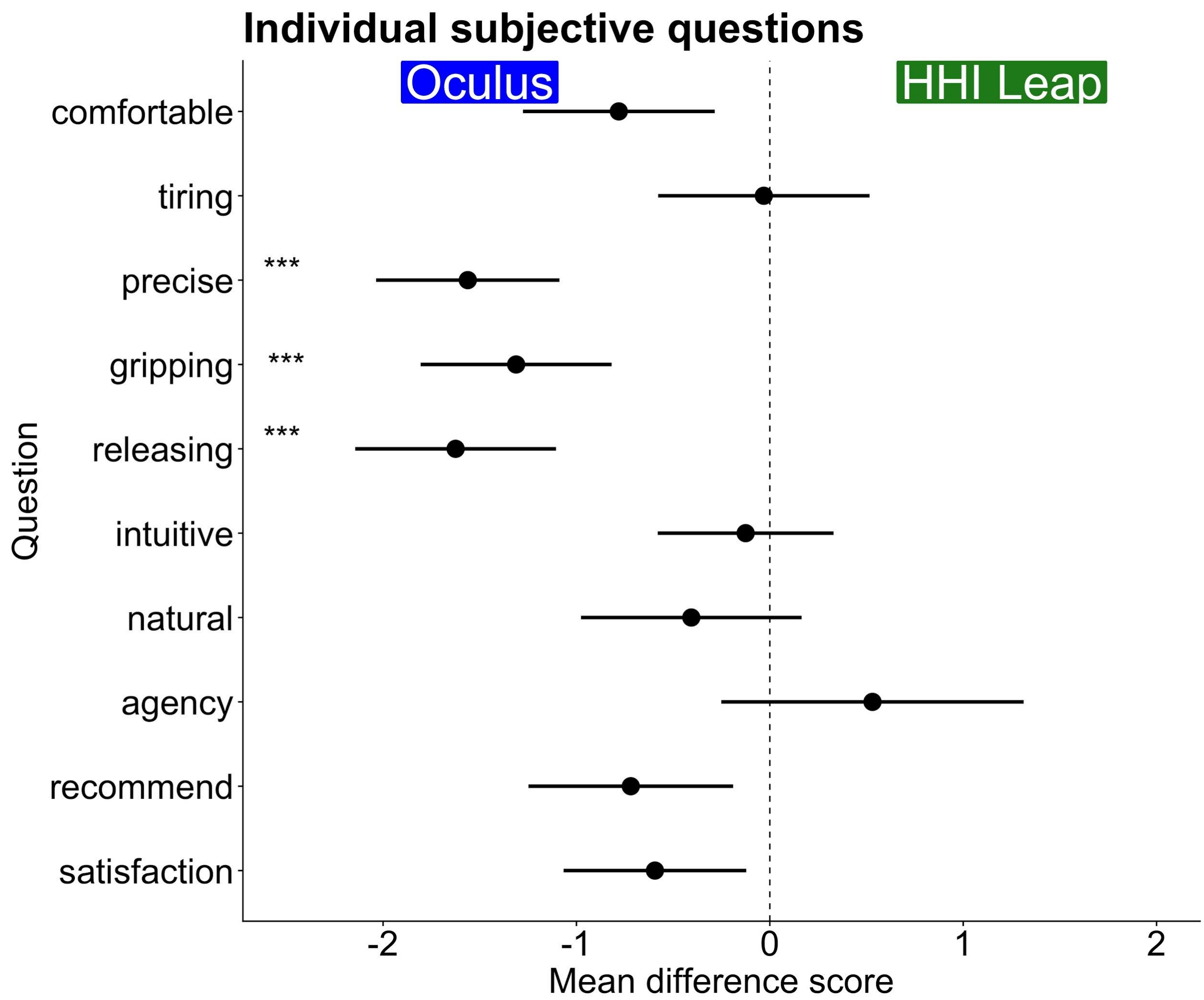


Fig. 14 Difference scores (HHI_Leap – Oculus) for individual subjective questions, with means and 95% confidence intervals; results of paired-samples t-test are indicated next to question labels

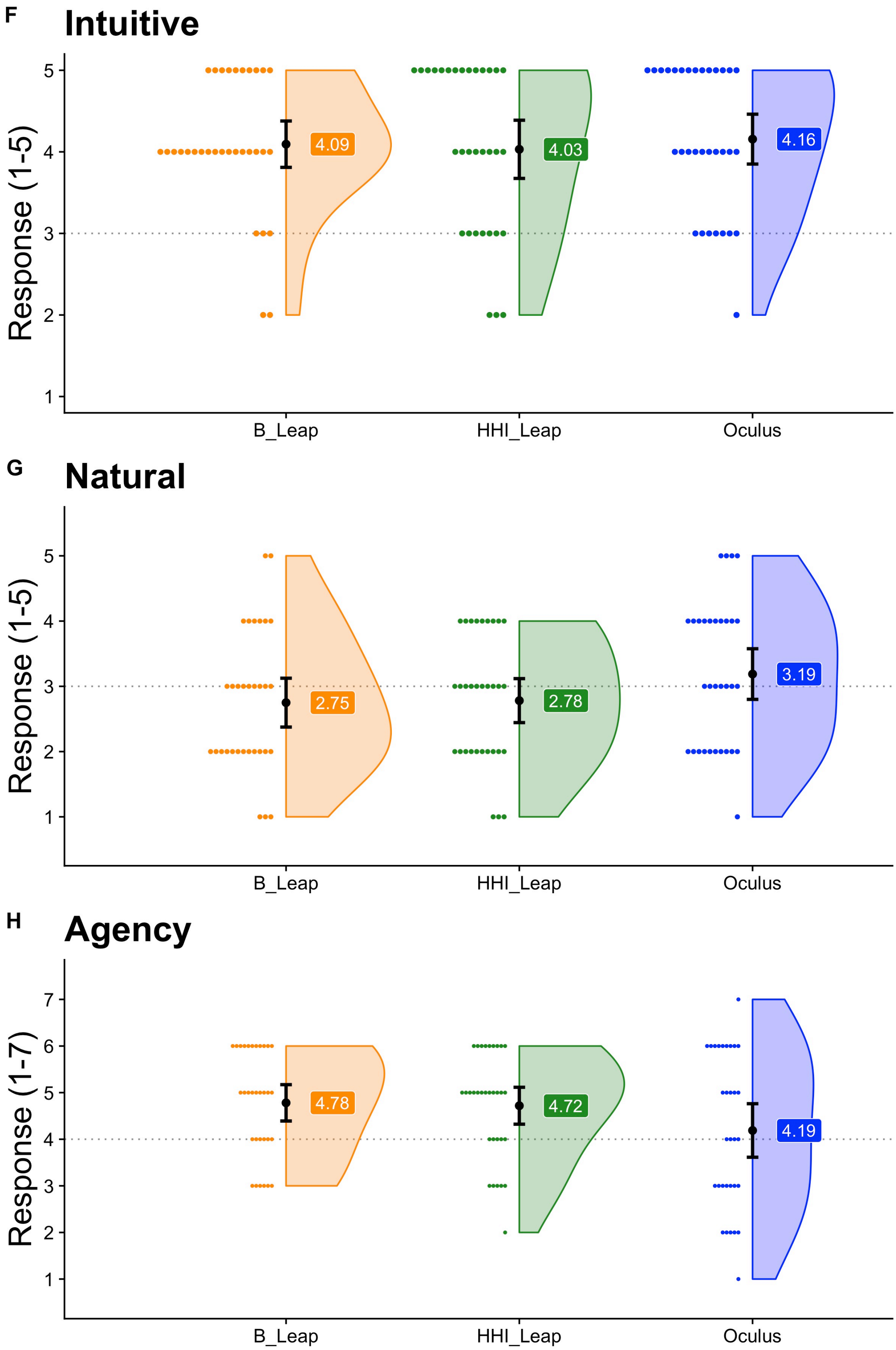


Fig. 15 Subjective ratings by interface for (A) naturalness, (B) intuitiveness and (C) agency, with means and 95% confidence intervals, individual subject ratings (colored dots) and smoothed density distributions; all t-tests indicated that differences were non-significant

Paired-samples t-tests showed no significant difference between the HHI_Leap (grand M = 3.24, SD = 1.18) and the B_Leap (grand M = 3.18, SD = 1.19) on grand means of 5-point Likert question scores, $t(31) = 0.44$, $p = .66$, or for any individual questions (Fig.

13). Grand means of 5-point Likert question ratings were significantly lower for the HHI_Leap than for the Oculus controller (grand M = 4.06, SD = 1.07), $t(31) = -4.13$, $p < .001$, $d = 0.95$ (large). t-tests showed that Oculus ratings were significantly higher than for the HHI_Leap on questions assessing gripping, $t(31) = -5.21$, $p < .001$, $d = 0.40$ (small), releasing, $t(31) = -6.14$, $p < .001$, $d = 0.40$ (small), and precision, $t(31) = -6.47$, $p < .001$, $d = 0.40$ (small) (Fig. 14). See Table 6 and Table 7 for descriptive statistics, Table 8 for all individual subjective question t-test results and Fig. 22 for plots of all individual subjective question data.

3.2.3 Overall preference

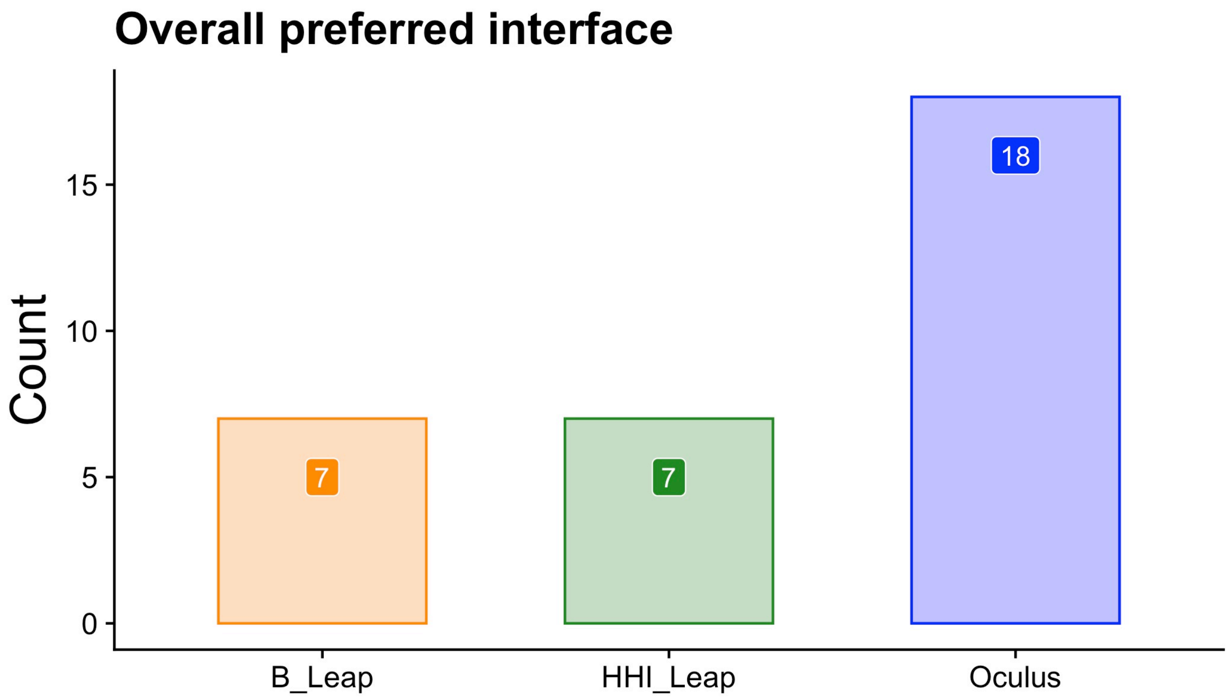


Fig. 16 Counts for overall preferred interface

18 subjects selected the Oculus controller as their overall preferred interface; 7 selected the HHI_Leap and 7 selected the B_Leap (Fig. 16).

4 Discussion

We assessed the usability of a prototype hand-sensing interface (HHI_Leap) for object interaction in VR by systematically comparing it to the basic Leap Motion (B_Leap) and the Oculus Touch controller. Performance was measured by accuracy, time and errors, and subjective experience was assessed using questionnaires.

Our first research question was: Did modifications to the basic Leap Motion hand-sensing interface improve it for the use case of grabbing and placing virtual objects in VR? Performance metrics showed that the HHI_Leap improved upon the B_Leap by reducing errors (accidental drops) during our pick-and-place task. The HHI_Leap was slightly slower at grabbing (grab time), slightly faster at placing (release time), and not significantly different in accuracy or total time per trial. The modification requiring the HHI_Leap to hover with the hand over the object before grabbing is possible was likely responsible for the increased grab time. Added to make the grabbing motion more purposeful (see Table 1), this rule may also explain the lower error rate; we see the tradeoff between grab time and error rate as acceptable for this use case. Related use cases, such as that of the Immersive Virtual Memory Task (Belger et al., 2019), benefit more from the user being able to pick up and place the desired object accurately and intentionally, rather than quickly. The use cases of training for mining (Kim & Choi, 2019) and manufacturing (Geiger et al., 2018) also do not require quick object interaction and would benefit more by using the HHI_Leap over the B_Leap. For use cases in which the object needs to be grabbed quickly (faster than the average HHI_Leap grab time of 1.57 seconds), the HHI_Leap would be less suitable. Despite the improved performance, we found no significant difference between the HHI_Leap or the B_Leap on any subjective experience measure. Subjective scores were slightly higher overall for the HHI_Leap on ease of use (as measured by the SUS) and individual questions, but none of these differences were statistically significant (see Appendix B). Ultimately, the HHI_Leap and B_Leap were too similar to create observable differences in subjective response.

Subjects completed our experimental task with three different sized cubes: small, medium and large. Analysis at this level provides some insight into why we saw the main performance results that we did when comparing the HHI_Leap to the B_Leap. Differences between the B_Leap and HHI_Leap for release time were more pronounced at the medium and large cube sizes than at the small cube size (Fig. 10). This may be explainable by challenges in detection of the hand shape at different cube sizes: As the cube size grows larger, the difference between the gripping and non-gripping[3] hand shape becomes more subtle. Our results suggest some aspect of our modifications either improved detection of the user's hand shapes for the releasing action, or encouraged the user to form more easily-detectable hand shapes. Another finding, that differences in accidental drops between the HHI_Leap and B_Leap were more pronounced at the small cube size than at the medium and large cube sizes (Fig. 11), also suggests that hand shape was crucial for the main results seen for this metric. For small cubes, the user has to keep the hand in a particular grabbing shape, with no feedback (tactile or visual) to warn when they are in danger of widening their grip too much and letting the cube fall out of their hands. If the HHI_Leap detected hand shapes better, or encouraged more purposeful user behavior, this may explain the cube size results we observed for release time and accidental drops. It also suggests that providing feedback to indicate an impending drop (i.e., with color hue or warning signal) could improve the accidental drop rate even more.

Our second research question was: How did the HHI_Leap prototype compare to the Oculus controller? As we expected, the Oculus controller outperformed the HHI_Leap on all performance metrics. Subjective response scores were generally higher for the Oculus controller and were statistically significant for ease of use (as measured by the SUS) and for individual questions assessing performance (gripping, releasing and precision); there was also a non-significant trend for higher comfort for the Oculus controller ($p = .07$; see Table 8). The higher subjective ratings for the Oculus controller may have been due to its superior performance, though as a product already brought to market, it may have also had more stages of usability testing and redesign. We thought that because the HHI_Leap enabled bare-hand interaction, subjects would find the experience similar to using their hands in real life and rate the HHI_Leap higher on naturalness, intuitiveness and agency. However, we found no significant differences on any of these ratings between the HHI_Leap and Oculus controller (Fig. 15).

There are several possible explanations why we did not see significantly higher ratings of naturalness, intuitiveness and agency for the HHI_Leap than for the Oculus controller. One is expectations: Users may have had unrealistic preconceptions of the capabilities of

[3] Resting hand positions often involve some curvature, which makes detection the difference between gripping and resting more difficult.

hand sensing, setting them up for disappointment (Spurlock & Ravasz, 2019; Myers et al., 2019, discuss how expectations similarly shape the user experience with voice interfaces). Another possible factor is that haptic feedback was more important to naturalness than we initially thought: Users may have found it highly unnatural to manipulate objects with their hands, but without the sense of touch; combining the Leap Motion with a haptic feedback device has been found to create stronger feelings of presence and immersion than the Leap Motion on its own (Kim et al., 2017). Though the Oculus controller did not map haptic feedback to interaction with virtual objects, it was itself a physical object, with buttons to press to engage in interaction. Subjects may have responded positively to even this minimal amount of tactile sensation when interacting with virtual objects. Additionally, technological limitations with the mapping of the virtual hand to the real hand and the responsiveness of the Leap Motion system may have lowered feelings of naturalness (Wozniak et al., 2016) as well as agency: even small incongruences between the user's intended hand gesture and how the virtual hand actually behaves can lower the sense of agency (Krugwasser et al., 2019; Caspar et al., 2015; Haggard & Tsakaris, 2009). Finally if, for some, the HHI_Leap felt too realistic, though not quite real enough, it may have created an uncanny valley effect (Mathur & Reichling, 2016; Mori, 2012), in which the nearness to reality results in the VR experience feeling highly unsettling. Argelaguet et al. (2016) found that less realistic virtual hand avatars created higher feelings of agency.

Considering the Oculus controller in the context of familiarity (sameness or similarity to a previously used tool) and tool use may also explain the results we observed for naturalness, intuitiveness and agency. Humans and their evolutionary ancestors have been using tools for thousands of years (Ambrose, 2001). Modern humans have demonstrated proficiency with simple tools such as hammers to more complex tools such as smartphones. That humans are able to achieve a fluid proficiency with tools is fundamental to philosophers Andy Clark and David Chalmers' (2002) proposition that certain tools, such as a calendar or walking stick, can in some circumstances be considered part of the mind itself. There is evidence that a user's familiarity with a tool is important for the user experience. Familiar tools activate different neural circuits than unfamiliar tools (Vingerhoets, 2008). Familiarity with an interface can increase user satisfaction and performance over one that is less familiar (Myers et al., 2019). Even when the design of a new technological tool is more efficient than its predecessor, users are hesitant to adopt it due to initial learning difficulties (Tomasi et al., 2018). In the absence of sufficient training, users report lower satisfaction with new software (Tomasi et al., 2018) and can end up abandoning it entirely (Jasperson et al., 2005). A well-known example is that the QWERTY keyboard – initially designed to slow human typing so as not to overburden early typewriters – continues to be the standard English language keyboard, despite attempts to replace it with a more efficient design (Norman, 2013). User experience researchers theorize that users create mental models for how to use an interface, and that familiar interfaces will be more usable due to previously existing mental models (Norman, 2013; Corry, 2002). Our subjects likely had more experience with game controllers like the Oculus controller than with the hand-sensing interfaces. All subjects in our study reported having used a video game controller at least once before, and many fell into the category of "frequent past use" (see Appendix D). Hand sensing is an emerging interface technology that our subjects, even as employees of a technology research institute, had fewer opportunities to use. Any existing mental model would likely have come from the use of their own hands in real life, or from depictions of such interfaces in science-fiction (Spurlock & Ravisz, 2019). We encourage future usability studies to include questions about familiarity of all interfaces tested to support such assumptions.

Our results on naturalness, intuitiveness and agency, while limited, may have implications for researchers aiming to maximize ecological validity and for neurorehabilitation practitioners aiming to maximize transfer to outside of the clinic. Naturalness is important to both ecological validity and transfer, which are theorized to be highest by approaching the real world as much as possible within the lab (Parsons, 2015) or clinic (Belger et al., 2019). Our results indicate that with our design and in this use case, hand sensing does not afford higher ecological validity or transfer; therefore, other factors should be considered when choosing an interface for a VR study or therapy, such as time-to-learn, equipment burden or frequency of technical malfunction. However, as we sought to form a general picture of our interface's usability rather than specifically investigate ecological validity or transfer, our conclusions on this matter are limited.

There are other limitations to our study that should be considered for the interpretation of our results: As our sample of convenience was pulled from a population of young, healthy adults, many of whom were male engineers in their 20's (M = 27.8 years old), generalization of the results is unclear – particularly to older adults, a population of interest for researchers working on VR as a tool for clinical testing and rehabilitation. This is true not only for performance measures, but also subjective response: for instance, older adults may find hand sensing more or less natural than what we observed here. Future usability studies could target older adults and people with disabilities to determine usability in these populations.

Controllers and hand-sensing interfaces differ on many dimensions, making them difficult to compare directly. For instance, it was difficult to control across interface types for familiarity and expectations. Also, the physical aspects of the controller, such as buttons with discrete interaction states and LED-based tracking, may give it inherent performance advantages over hand sensing in its current form. Furthermore, a parameter in the implementation of the Oculus controller interface, undiscovered by us until after beginning main trial testing, might have inflated differences in performance: Users could sink the cube nearly all the way through the table while holding it, allowing them to align it nearly perfectly with the target before releasing (this could not be done with the Leap Motion interfaces). Some users discovered this "hack" and some did not; we did not record which subjects discovered it.

Video capture of the screen could have helped resolve questions that arose during analysis, including which users discovered the "hack" with the Oculus controller mentioned above, confirming the number of accidental drops, and justifying the removal of subjects with outlier performance metrics by confirming extraneous circumstances. We unfortunately did not record video of testing and recommend doing so for future research.

There is always more data that could be collected, but we particularly recommend that future studies measure positive emotion (i.e., joy, pleasure, excitement, novelty). If hand-sensing interfaces are higher in these areas than a handheld controller, it could justify accepting deficits in performance in exchange for a more positive user experience. Collecting qualitative data (i.e., "What did you like/dislike about this interface?") could also have helped us understand our subjects' ratings of the interfaces.

A within-subjects design is useful: it allows for a larger sample and standardizes the rating scale for each subject (within-subject variability is likely to be lower than between-subject variability). However, it can also create order effects that systematically affect observed outcomes, such as practice effects, in which the user's performance improves over time, or fatigue effects, in which the user's performance decreases over time (Shaughnessy et al., 2012). Randomized, counterbalanced conditions can minimize practice and fatigue effects, but does not account for intrinsic, unequal effects between interfaces (i.e., if using the B_Leap first improves performance on the HHI_Leap more than the reverse). The Oculus controller, which differs from hand sensing on many dimensions, could have created an order effect by influencing ratings of the hand-sensing interfaces if used first. Future studies should consider a between-subjects design if order effects are anticipated between the interfaces to be tested.

5 Conclusions

Our usability metrics indicate that modifications made to the pre-packaged Leap Motion hand-sensing VR interface had small effects on performance and none on subjective experience. The improved accidental drop rate and slightly slower grab speed (in the absence of any loss to accuracy or total time) optimize our prototype for VR applications that prioritize accuracy over speed. Analysis by cube size suggests that the HHI_Leap either detected hand shape better than the B_Leap, or encouraged the user to form more easily-detectable hand shapes, and that performance could be improved further by an indicator of the user's grip tightness on the object (i.e., color hue or warning signal). The Oculus Touch controller performed higher and provided a better overall user experience than our prototype. However, the lack of significant differences between the HHI_Leap and Oculus controller on many of our individual subjective measures suggests that, despite the novelty, people respond as positively to our hand-sensing prototype as they do to the more familiar, market-ready Oculus controller. Our results from questions assessing naturalness, intuitiveness and agency do not support higher ecological validity in research or transfer in rehabilitative therapy for hand sensing over an Oculus controller. However, combining with a haptic feedback device, improving the system's capability to detect hand shapes and innovative future designs could increase performance, and with it the subjective experience. Oculus, for example, currently has its own hand-sensing interface in development, capitalizing on self-haptic gestures such as pinching to promising results (Spurlock & Ravasz, 2019). Leap Motion has merged with Ultrahaptics to form a new venture, UltraLeap, aiming to leverage ultrasound technology as a haptic interface for use with the Leap Motion. Hand sensing as a VR interface is a promising emerging technology, with much to gain from further usability research, design and development.

6 References

Ambrose, S. H. (2001) Paleolithic technology and human evolution. *Science*, *291*(5509), 1748–1753. https://doi.org/10.1126/science.1059487

Azure Kinect DK. (2020, January 11) https://azure.microsoft.com/en-us/services/kinect-dk/

Argelaguet, F., Hoyet, L., Trico, M., & Lécuyer, A. (2016) The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *Proceedings - IEEE Virtual Reality* (Vol. 2016-July, pp. 3–10). IEEE Computer Society. https://doi.org/10.1109/VR.2016.7504682

Bangor, A., Kortum, P., & Miller, J. (2009) Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*, *4*(3), 114–123. Retrieved from http://uxpajournal.org/wp-content/uploads/sites/8/pdf/JUS_Bangor_May2009.pdf

Belger, J., Krohn, S., Finke, C., Tromp, J., Klotzche, F., Villringer, A., Gaebler, M., Chojecki, P., Quinque, E. and Thöne-Otto, A. (2019) Immersive Virtual Reality for the Assessment and Training of Spatial Memory: Feasibility in Neurological Patients. In: WG Wright, S Subramanian, G Fluet, M Agmon, RM Proffitt, M Roberts (Eds), Proc. 13th Int'l Conf. on Virtual Rehab., Tel Aviv, Israel, 21-24 July 2019.

Brooke, J. SUS-A quick and dirty usability scale. (1996) In Jordan, P.W., Thomas, B., McLelland, I. and Weerdmeester, B.A. (eds) Usability Evaluation in Industry. Taylor and Francis, London: 189-194

Carifio, J. and Perla, R. (2008) Resolving the 50-Year Debate around Using and Misusing Likert Scales. Medical Education, 42(12):1150–52. https://doi.org/10.1111/j.1365-2923.2008.03172.x.

Caspar EA, Cleeremans A, Haggard P. (2015) The relationship between human agency and embodiment. Conscious Cogn 33:226-236. doi:10.1016/j.concog.2015.01.007

Clark, A., & Chalmers, D. (2002). The Extended Mind. In: Chalmers, D. (ed) *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press.

Cohen, J. (1988) Statistical Power Analysis for the Behavioral Sciences. 2nd Edition, Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

Don, K., Farhi, M., Pekarek Krohn, D., Schneiderman, J., Soderquist, K., Grant, A., Straughan, B. et al. (2018). *Industry insights into the future of AR/VR*. Retrieved from https://dpntax5jbd3l.cloudfront.net/images/content/1/8/v2/187785/2018-VR-AR-Survey-Digital.pdf

Dujmovic, J. (2019, July 15). Opinion: Here's why you will be hearing more about virtual reality. Retrieved from: https://www.marketwatch.com/story/heres-why-you-will-be-hearing-more-about-virtual-reality-2019-07-15

Geiger A, Bewersdorf I, Brandenburg E, Stark R. (2018) Visual feedback for grasping in virtual reality environments for an interface to instruct digital human models. In: Advances in Intelligent Systems and Computing. Springer Verlag, Vol 607:228-239. doi:10.1007/978-3-319-60492-3_22

Haggard P, Tsakiris M. (2009) The experience of agency: Feelings, judgments, and responsibility. Curr Dir Psychol Sci, 18(4):242-246. doi:10.1111/j.1467-8721.2009.01644.x

Hofmann, S., Klotzsche, F., Mariola, A., Nikulin, V., Villringer, A. and Gaebler, M. (2018) Decoding Subjective Emotional Arousal during a Naturalistic VR Experience from EEG Using LSTMs. 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR):128-131. 10.1109/AIVR.2018.00026.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6 (2): 65–70.

Jasperson, J., Carter, P. E., & Zmud, R. W. (2005). A comprehensive conceptualization of post-adoptive behaviors associated with information technology enabled work systems. *MIS Quarterly: Management Information Systems*, *29*(3), 525–557. https://doi.org/10.2307/25148694

Kim, H., & Choi, Y. (2019). Performance comparison of user interface devices for controlling mining software in virtual reality environments. Applied Sciences (Switzerland), 9(13). https://doi.org/10.3390/app9132584

Kim, M., Jeon, C., & Kim, J. (2017). A Study on Immersion and Presence of a Portable Hand Haptic System for Immersive Virtual Reality. Sensors (Basel, Switzerland), 17(5). https://doi.org/10.3390/s17051141

Krugwasser R, Harel E V., Salomon R. (2019) The boundaries of the self: The sense of agency across different sensorimotor aspects. J Vis. 19(4):1-11. doi:10.1167/19.4.14

Lazar J, Feng JH, Hochheiser H. Usability testing. (2017) In: Research Methods in Human Computer Interaction. 2nd ed. Morgan Kaufmann, Cambridge, MA, USA, pp 263-298. doi:10.1016/b978-0-12-805390-4.00010-8

Mathur, Maya B.; Reichling, David B. (2016). "Navigating a social world with robot partners: a quantitative cartography of the Uncanny Valley". Cognition. 146: 22–32. doi:10.1016/j.cognition.2015.09.008. PMID 26402646.

Massetti, T., da Silva, T. D., Crocetta, T. B., Guarnieri, R., de Freitas, B. L., Bianchi Lopes, P., et al. (2018). The Clinical Utility of Virtual Reality in Neurorehabilitation: A Systematic Review. Journal of Central Nervous System Disease, 10, 117957351881354. https://doi.org/10.1177/1179573518813541

Meek G, Ozgur C, Dunning K. (2007) Comparison of the t vs. Wilcoxon Signed-Rank test for likert scale data and small samples. J of Applied Stat Methods, 6(1), 91-106.

Mori, M. (2012). Translated by MacDorman, K. F.; Kageki, Norri. "The uncanny valley". IEEE Robotics and Automation. 19 (2): 98–100. doi:10.1109/MRA.2012.2192811

Myers, C. M., Furqan, A., & Zhu, J. (2019). The impact of user characteristics and preferences on performance with an unfamiliar voice user interface. Conference on Human Factors in Computing Systems - Proceedings, (Paper 47), 1–9. https://doi.org/10.1145/3290605.3300277

Norman, D. The Design of Everyday Things. Basic Books, 2013.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Park, J. L., Dudchenko, P. A., & Donaldson, D. I. (2018). Navigation in real-world environments: New opportunities afforded by advances in mobile brain imaging. Frontiers in Human Neuroscience, 12(Sept), 1–12. https://doi.org/10.3389/fnhum.2018.00361

Parsons, Thomas D. (2015) Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. Frontiers in Human Neuroscience 9 (DEC): 1–19. https://doi.org/10.3389/fnhum.2015.00660.

Pedroli, E., Greci, L., Colombo, D., Serino, S., Cipresso, P., Arlati, S., et al. (2018). Characteristics, Usability, and Users Experience of a System Combining Cognitive and Physical Therapy in a Virtual Environment: Positive Bike. Sensors, 18(2343). https://doi.org/10.3390/s18072343

Proffitt, R., Sevick, M., Chang, C.-Y., & Lange, B. (2015). User-Centered Design of a Controller-Free Game for Hand Rehabilitation. Games for Health Journal, 4(4), 259–264. https://doi.org/10.1089/g4h.2014.0122

Rizzo, A. S., & Koenig, S. T. (2017) Is clinical virtual reality ready for primetime? *Neuropsychology*, *31*(8), 877–899. https://doi.org/10.1037/neu0000405

Rizzo, A. S., Koenig, S. T., & Talbot, T. Clinical Virtual Reality: Emerging Opportunities for Psychiatry. (2018) Clinical Virtual Reality: Emerging Opportunities for Psychiatry. *FOCUS*, *16*(3), 266–278. https://doi.org/10.1176/appi.focus.20180011

Shaughnessy, John J, Eugene B Zechmeister, and Jeanne S Zechmeister. (2012) Research Methods in Psychology. (Ninth Edit.) McGraw-Hill, New York, NY, pp 230-239.

Slater, M. Grand Challenges in Virtual Environments. (2014) Frontiers in Robotic and AI. Vol. 1 (May): pp. 1-4.

Spurlock, J & Ravasz, J. (September 2019). *Hand Tracking: Designing a New Input Modality*. Presented at the Oculus Connect conference in San Jose, California, USA. Available online: https://www.youtube.com/watch?v=or5M01Pcy5U. Accessed: 13 Jan., 2020.

Tomasi, S, Schuff, D and Turetken, O. (2018) Understanding novelty: how task structure and tool familiarity moderate performance. Behav Inf Technol., 37(4):406-418. doi:10.1080/0144929X.2018.1441325

Tullis, T and Albert, B. (2013) Ch.1: Introduction. In: Measuring the User Experience, 2nd ed. Morgan Kaufman, pp 1-14.

Vingerhoets, G. (2008). Knowing about tools: Neural correlates of tool familiarity and experience. NeuroImage, 40(3), 1380–1391. https://doi.org/10.1016/j.neuroimage.2007.12.058

Vosinakis, S., & Koutsabasis, P. (2018). Evaluation of visual feedback techniques for virtual grasping with bare hands using Leap Motion and Oculus Rift. Virtual Reality, 22(1), 47–62. https://doi.org/10.1007/s10055-017-0313-4

Wozniak, P., Vauderwange, O., Mandal, A., Javahiraly, N., & Curticapean, D. (2016). Possible applications of the LEAP motion controller for more interactive simulated experiments in augmented or virtual reality. Optics Education and Outreach IV, 9946(September 2016), 99460P. https://doi.org/10.1117/12.2237673

Wobbrock, J., Kientz, J., (2016) Research contributions in human-computer interaction. Interactions 23 (3), 38–44.

Appendix A: Performance Data

A.1 Main results

A.1.1 Main result Plots

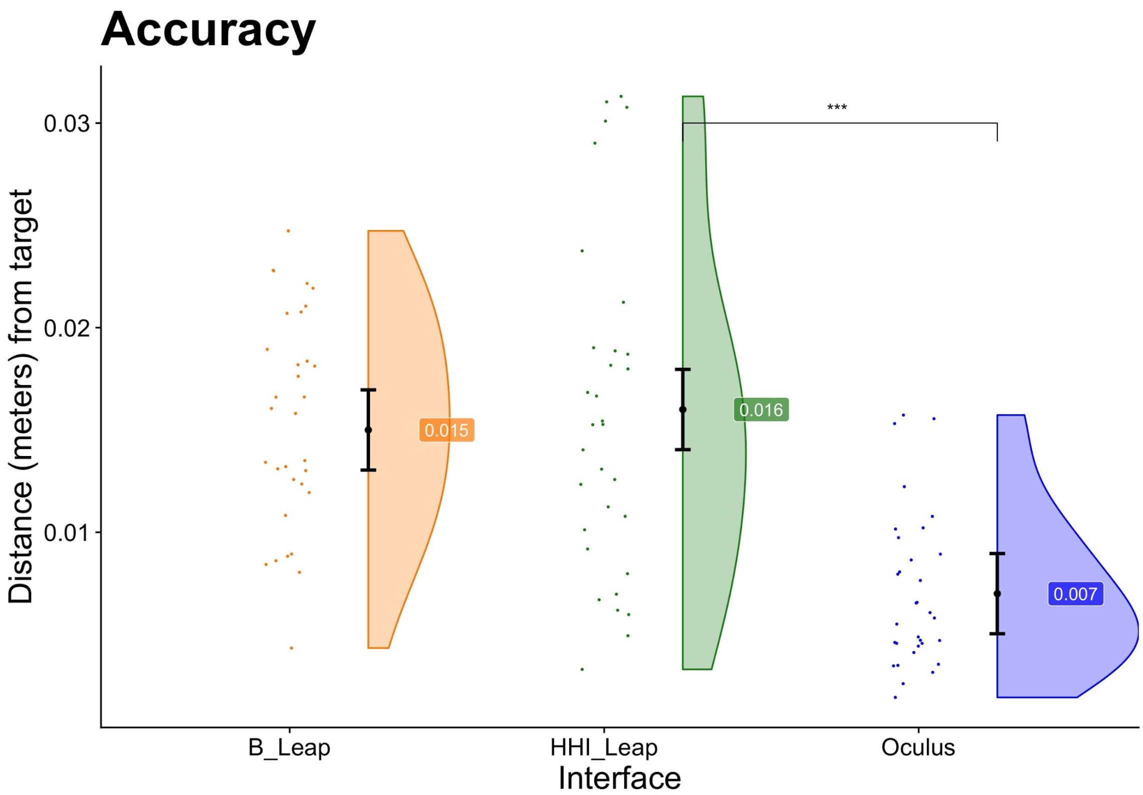


Fig. 17 Accuracy by interface, with means and 95% confidence intervals, individual subject means (colored dots), smoothed density distributions and results of paired-samples t-tests

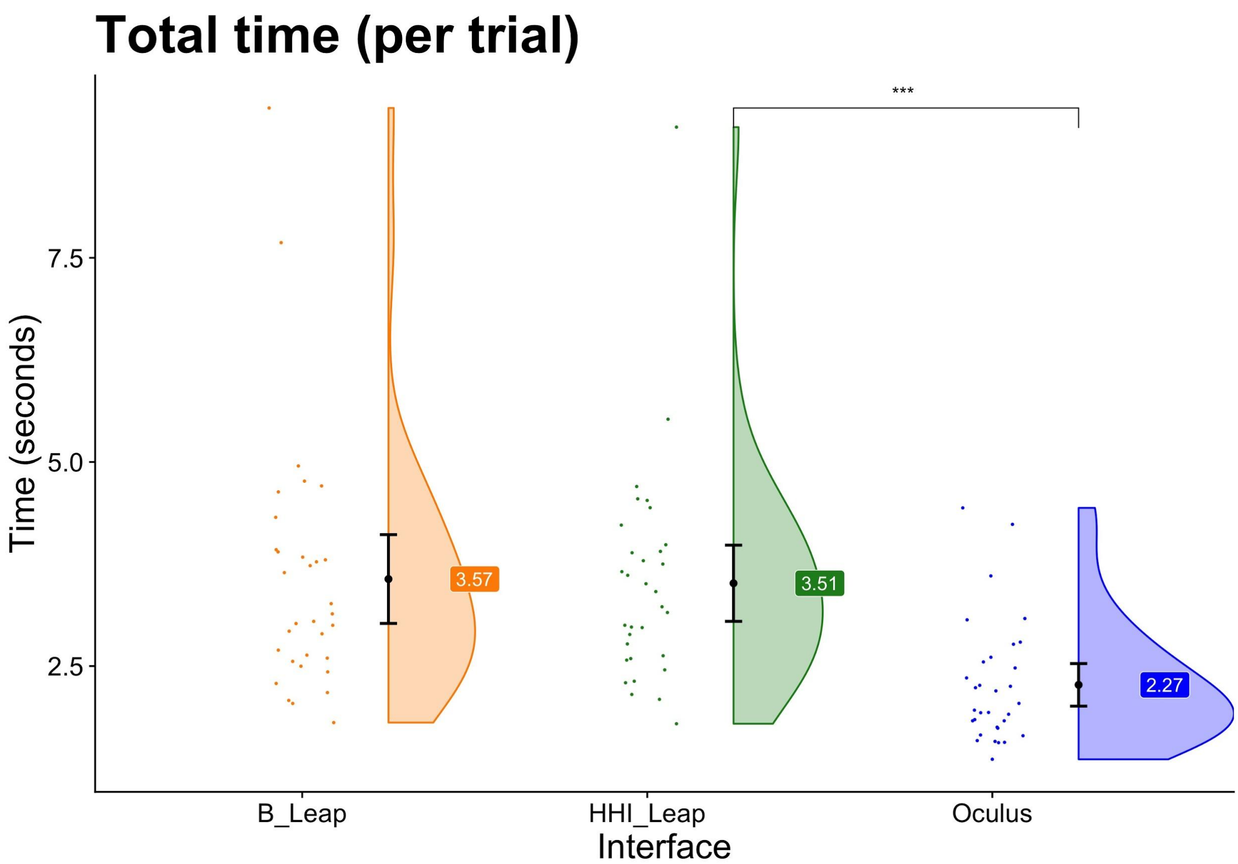


Fig. 18 Total time by interface, with means and 95% confidence intervals, individual subject means (colored dots), smoothed density distributions and results of paired-samples t-tests

A.1.2 Main result tables

Table 3 Performance metrics by interface: one-way within-subjects ANOVA results

	DFn	DFd	F	p	η_p^2
Accuracy	2	62	41.71	< .001	0.574
Total time	2	62	36.62	< .001	0.542
Grab time	2	62	29.06	< .001	0.484
Release time	2	62	30.34	< .001	0.495
Accidental drops	2	62	43.06	< .001	0.581

Table 4 Performance metrics by interface: paired-samples t-test results

	Group1	Group2	t	df	p^a	d	Magnitude
Accuracy	B_Leap	HHI_Leap	-0.29	31	.771	0.05	negligible
	HHI_Leap	Oculus	7.13	31	< .0001	1.26	large
Total time	B_Leap	HHI_Leap	0.29	31	.772	0.05	negligible
	HHI_Leap	Oculus	7.88	31	< .0001	1.39	large
Grab time	B_Leap	HHI_Leap	-2.25	31	.032	0.40	small
	HHI_Leap	Oculus	6.94	31	< .0001	1.23	large
Release time	B_Leap	HHI_Leap	2.32	31	.027	0.41	small
	HHI_Leap	Oculus	6.70	31	< .0001	1.18	large
Accidental	B_Leap	HHI_Leap	3.95	31	.0004	0.70	moderate

Drops	HHI_Leap	Oculus	5.96	31	< .0001	1.05	large

[a]Holm-adjusted for multiple comparisons

A.2 Cube Size

A.2.1 Cube Size Plots

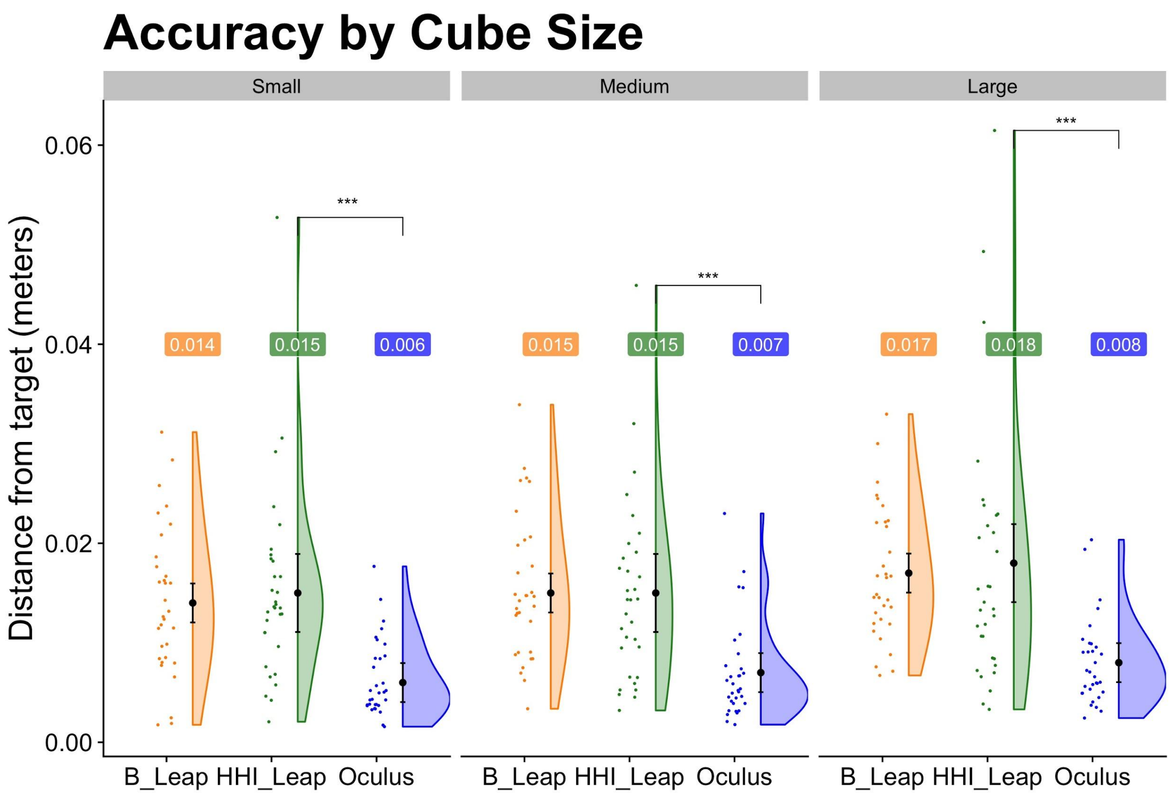


Fig. 19 Accuracy by cube size and interface, with means and 95% confidence intervals, individual subject means (colored dots), smoothed density distributions and results of paired-samples t-tests

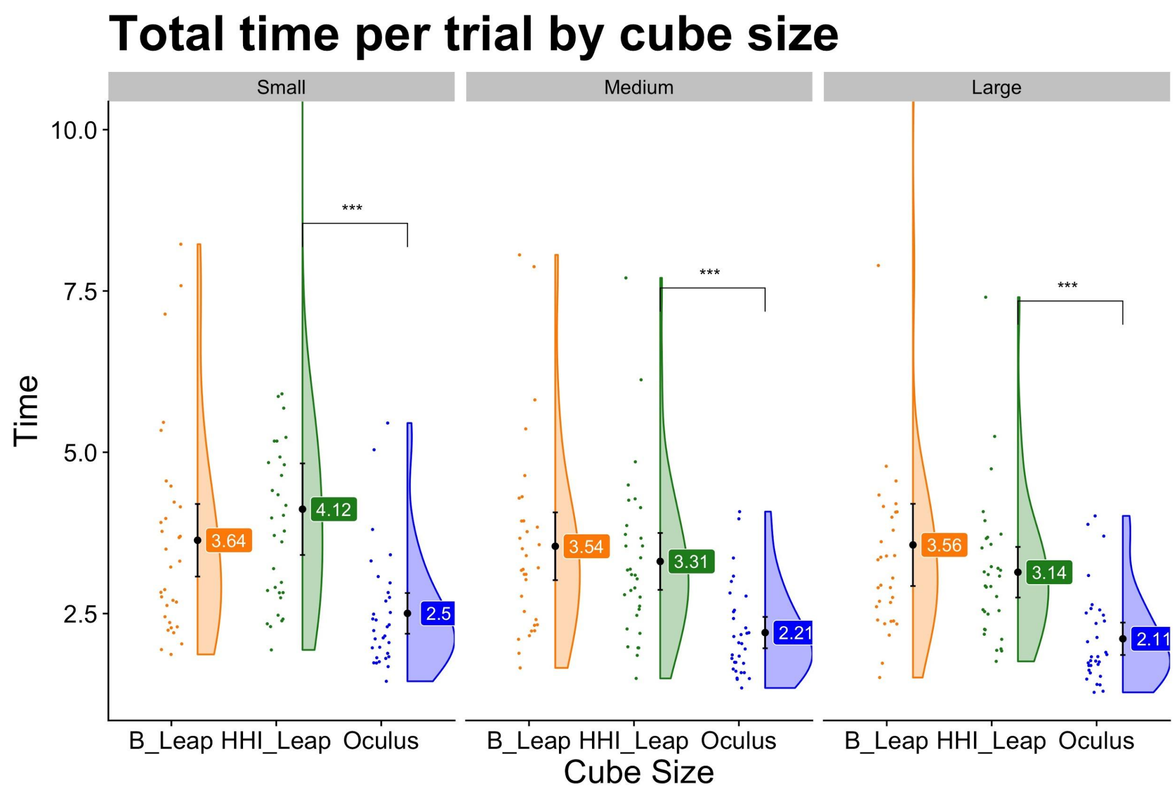


Fig. 20 Total time by cube size and interface, with means and 95% confidence intervals, individual subject means (colored dots), smoothed density distributions and results of paired-samples t-tests

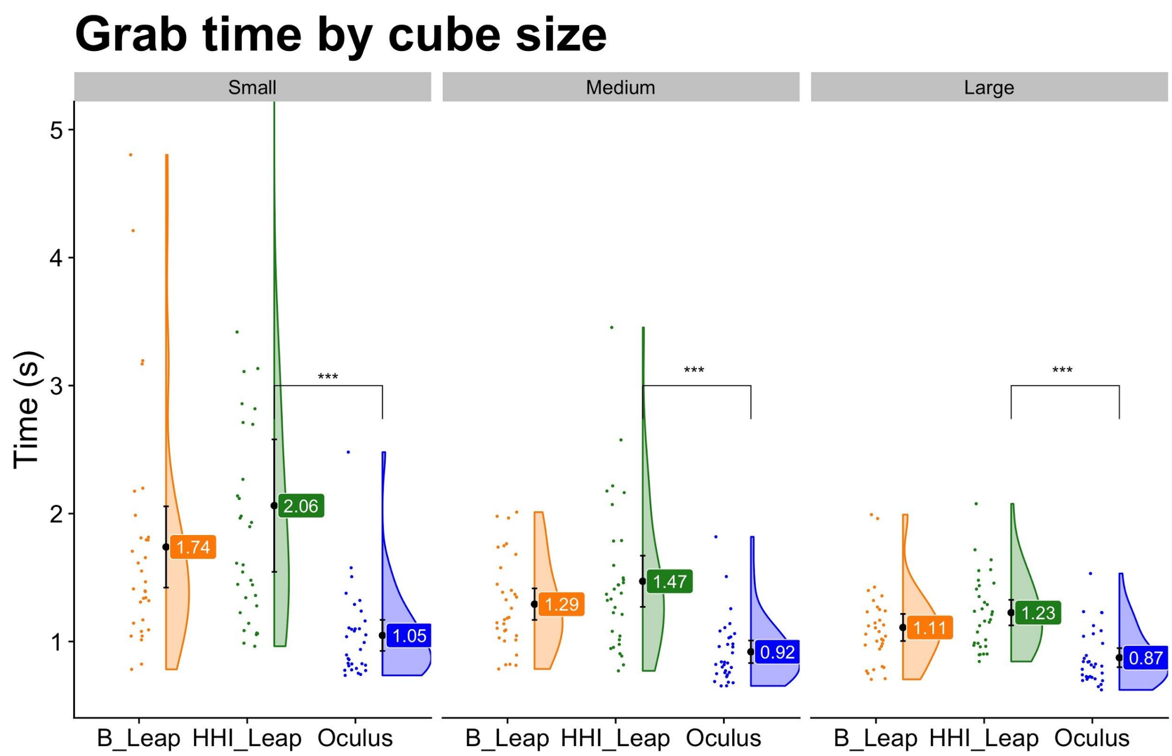


Fig. 21 Grab time by cube size and interface, with means and 95% confidence intervals, individual subject means (colored dots), smoothed density distributions and results of paired-samples t-tests

A.2.2 Cube Size Tables

Table 5 Performance metrics by cube size and interface: paired-samples t-test results

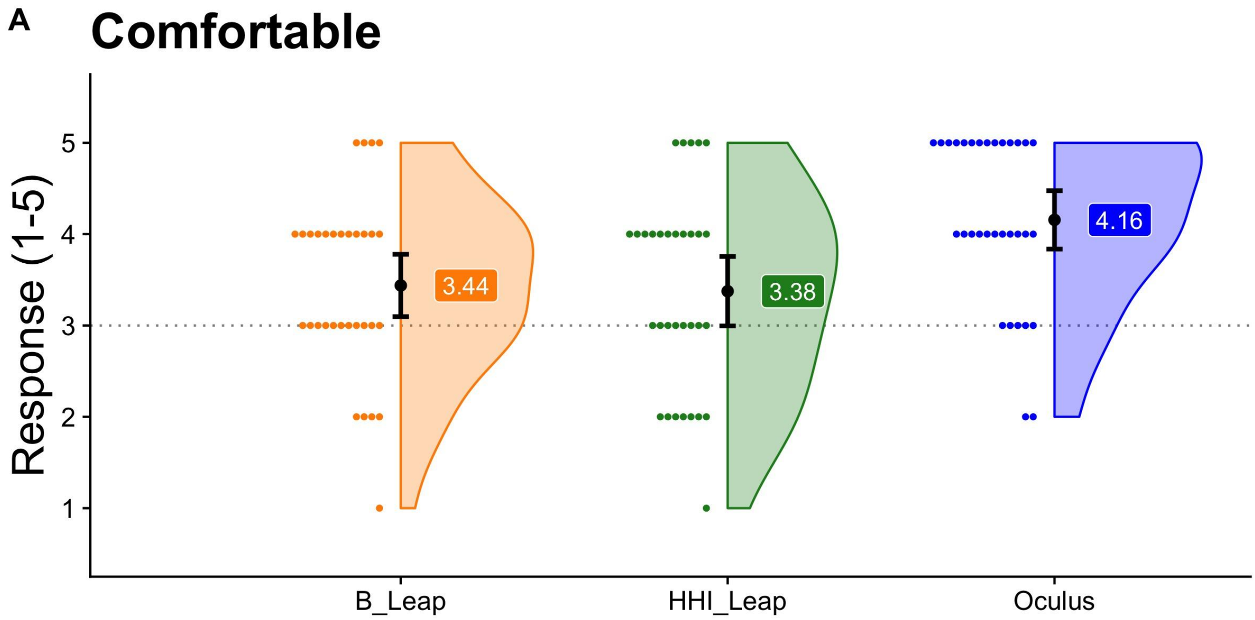
	Cube_Size	Group1	Group2	t	df	p[a]	d	Magnitude
Accuracy	Small	B_Leap	HHI_Leap	-0.64	31	> .99	0.04	negligible
	Medium	B_Leap	HHI_Leap	0.36	31	> .99	0.04	negligible
	Large	B_Leap	HHI_Leap	-0.28	31	> .99	0.04	negligible
	Small	HHI_Leap	Oculus	5.54	31	< .001	0.88	large
	Medium	HHI_Leap	Oculus	4.76	31	< .001	0.88	large
	Large	HHI_Leap	Oculus	4.66	31	< .001	0.88	large
Total time	Small	B_Leap	HHI_Leap	-1.67	31	.21	0.04	negligible
	Medium	B_Leap	HHI_Leap	1.50	31	.21	0.04	negligible
	Large	B_Leap	HHI_Leap	1.88	31	.21	0.04	negligible
	Small	HHI_Leap	Oculus	5.89	31	< .001	1.10	large
	Medium	HHI_Leap	Oculus	7.68	31	< .001	1.10	large
	Large	HHI_Leap	Oculus	7.21	31	< .001	1.10	large
Grab time	Small	B_Leap	HHI_Leap	-1.43	31	.18	0.25	small
	Medium	B_Leap	HHI_Leap	-1.74	31	.18	0.25	small
	Large	B_Leap	HHI_Leap	-2.01	31	.16	0.25	small
	Small	HHI_Leap	Oculus	4.58	31	< .001	0.77	moderate
	Medium	HHI_Leap	Oculus	7.09	31	< .001	0.77	moderate

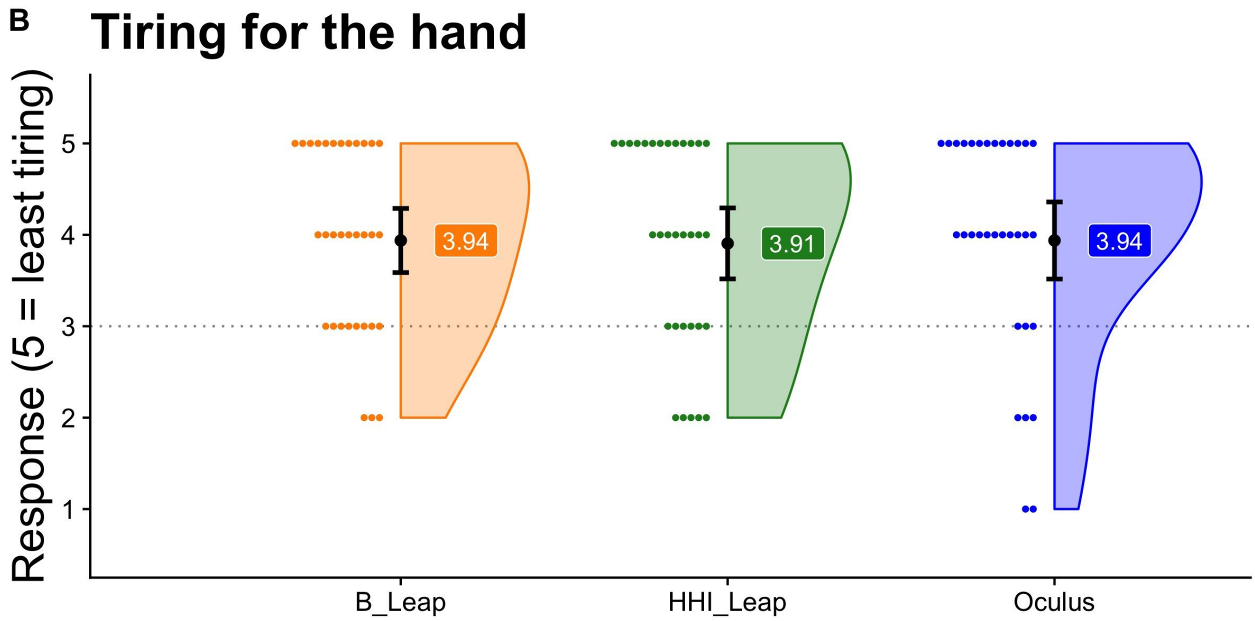
	Large	HHI_Leap	Oculus	6.73	31	< .001	0.77	moderate
Release time	Small	B_Leap	HHI_Leap	-1.51	31	.14	0.31	small
	Medium	B_Leap	HHI_Leap	3.42	31	.01	0.31	small
	Large	B_Leap	HHI_Leap	2.85	31	.02	0.31	small
	Small	HHI_Leap	Oculus	5.50	31	< .001	1.03	large
	Medium	HHI_Leap	Oculus	5.57	31	< .001	1.03	large
	Large	HHI_Leap	Oculus	6.26	31	< .001	1.03	large
Accidental drops	Small	B_Leap	HHI_Leap	4.40	31	< .001	0.40	small
	Medium	B_Leap	HHI_Leap	2.25	31	.06	0.40	small
	Large	B_Leap	HHI_Leap	0.10	31	.92	0.40	small
	Small	HHI_Leap	Oculus	3.75	31	< .001	0.69	moderate
	Medium	HHI_Leap	Oculus	3.67	31	< .001	0.69	moderate
	Large	HHI_Leap	Oculus	4.35	31	< .001	0.69	moderate

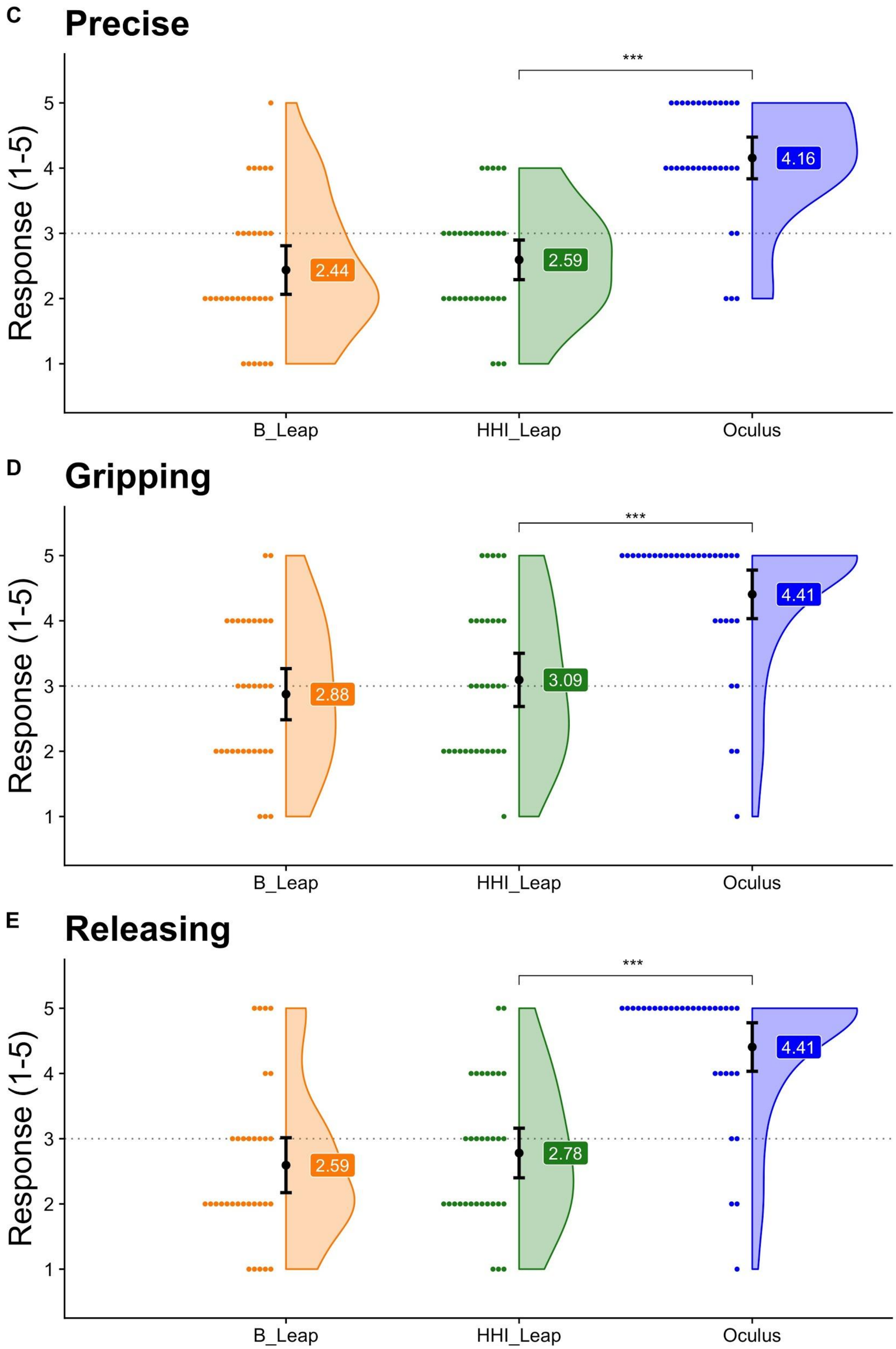
[a]Holm-adjusted for multiple comparisons

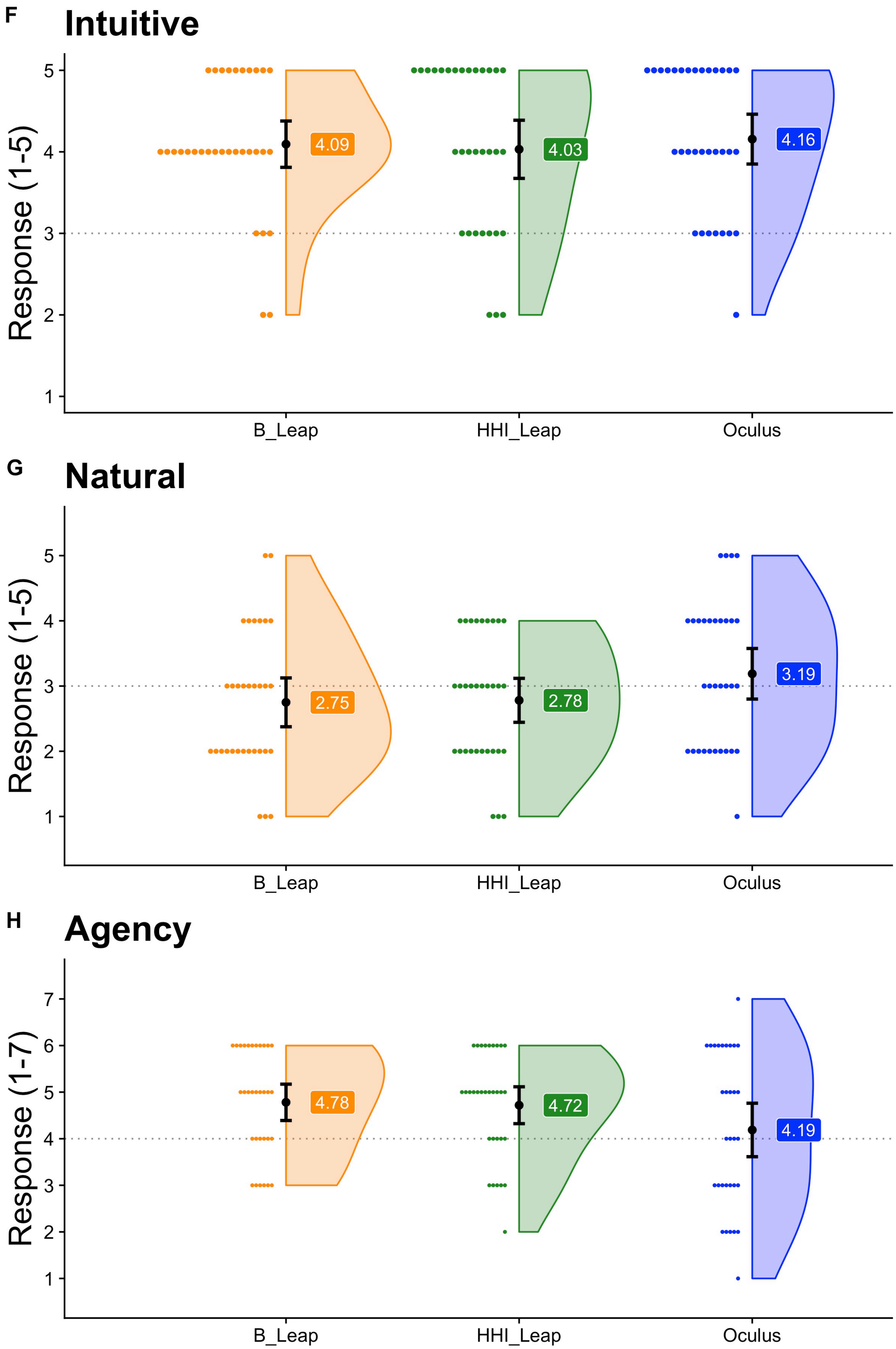
Appendix B: Subjective Data

B.1 Plots





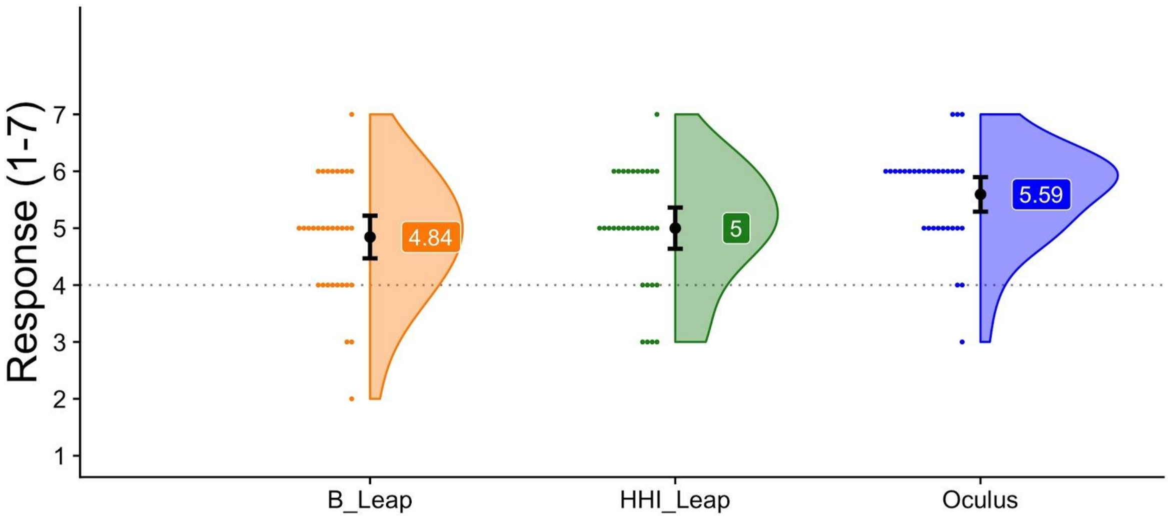

C
Precise
***
Response (1-5)
1
2
3
4
5
2.44
2.59
4.16
B_Leap
HHI_Leap
Oculus
D
Gripping
***
Response (1-5)
1
2
3
4
5
2.88
3.09
4.41
B_Leap
HHI_Leap
Oculus
E
Releasing
***
Response (1-5)
1
2
3
4
5
2.59
2.78
4.41
B_Leap
HHI_Leap
Oculus



F
Intuitive
Response (1-5)
4.09
4.03
4.16
B_Leap
HHI_Leap
Oculus
G
Natural
Response (1-5)
2.75
2.78
3.19
B_Leap
HHI_Leap
Oculus
H
Agency
Response (1-7)
4.78
4.72
4.19
B_Leap
HHI_Leap
Oculus

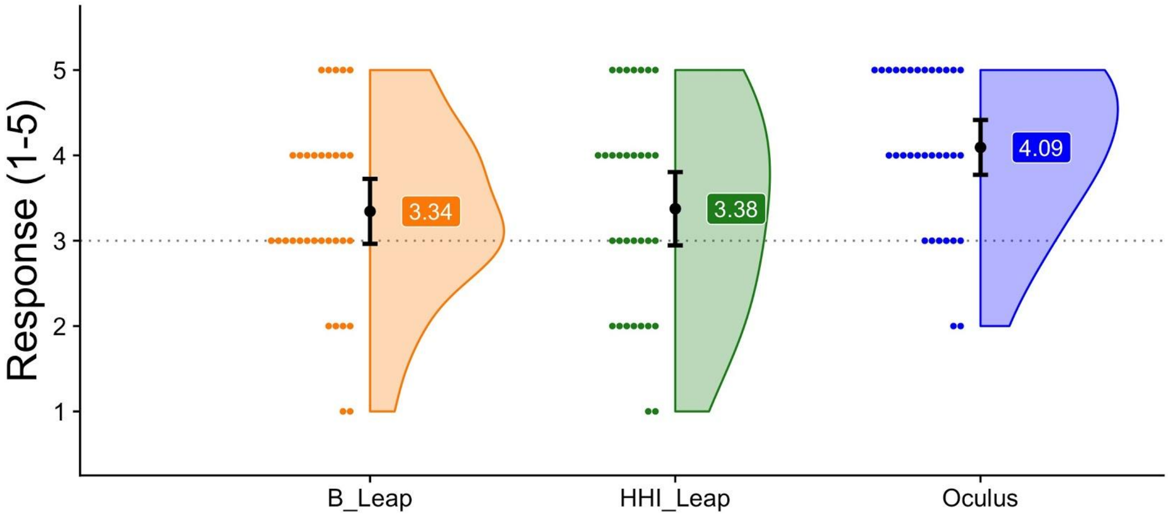
I

Overall satisfaction



J

Recommend



K

Overall preferred interface

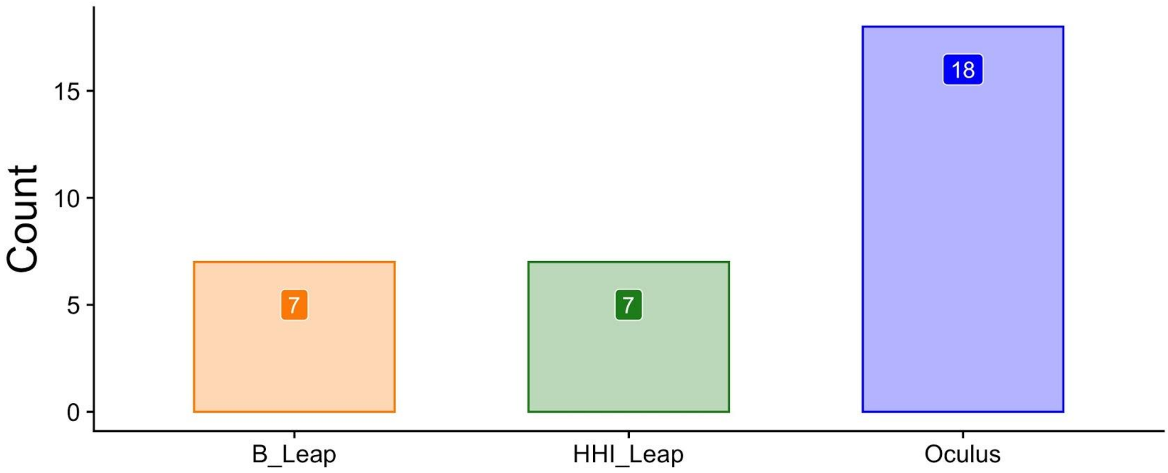


Fig. 22 (A-J) Subjective ratings by interface for all individual subjective questions, with means and 95% confidence intervals, individual subject ratings (colored dots), smoothed density distributions and results of paired-samples t-tests and (K) counts for overall preferred interface

B.2 Tables

Table 6 Individual subjective questions (5-point Likert scale): descriptive statistics

Question	Interface	Mean	SD	Median	IQR	Min	Max
Comfort	B_Leap	3.438	0.982	3.5	1	1	5
	HHI_Leap	3.375	1.1	3.5	1.25	1	5
	Oculus	4.156	0.92	4	1	2	5
Gripping	B_Leap	2.875	1.129	3	2	1	5
	HHI_Leap	3.094	1.174	3	2	1	5
	Oculus	4.406	1.073	5	1	1	5
Intuitive	B_Leap	4.094	0.818	4	1	2	5
	HHI_Leap	4.031	1.031	4	2	2	5
	Oculus	4.156	0.884	4	1.25	2	5
Natural	B_Leap	2.75	1.078	3	1.25	1	5
	HHI_Leap	2.781	0.975	3	2	1	4
	Oculus	3.188	1.12	3	2	1	5
Precise	B_Leap	2.438	1.076	2	1	1	5
	HHI_Leap	2.594	0.875	3	1	1	4
	Oculus	4.156	0.92	4	1	2	5

Recommend	B_Leap	3.344	1.096	3	1	1	5
	HHI_Leap	3.375	1.238	3.5	2	1	5
	Oculus	4.094	0.928	4	1.25	2	5
Releasing	B_Leap	2.594	1.214	2	1	1	5
	HHI_Leap	2.781	1.099	3	2	1	5
	Oculus	4.406	1.073	5	1	1	5
Tiring	B_Leap	3.938	1.014	4	2	2	5
	HHI_Leap	3.906	1.118	4	2	2	5
	Oculus	3.938	1.216	4	1.25	1	5
Grand (all 5-point Likert questions)	B_Leap	3.184	1.192	3	2	1	5
	HHI_Leap	3.242	1.177	3	2	1	5
	Oculus	4.062	1.072	4	1.25	1	5

Table 7 Individual subjective questions (7-point Likert scale): descriptive statistics

Question	Interface	Mean	SD	Median	IQR	Min	Max
Agency	B_Leap	4.781	1.128	5	2	3	6
	HHI_Leap	4.719	1.143	5	2	2	6
	Oculus	4.188	1.655	4	3	1	7
Satisfaction	B_Leap	4.844	1.081	5	2	2	7
	HHI_Leap	5	1.047	5	1.25	3	7

	Oculus	5.594	0.875	6	1	3	7
Grand (all 7-point Likert questions)	B_Leap	4.812	1.097	5	2	2	7
	HHI_Leap	4.859	1.096	5	2	2	7
	Oculus	4.891	1.492	5	2	1	7

Table 8 Individual subjective questions (all): paired-samples t-test results

Question	Group1	Group2	t	df	p[a]	d	Magnitude
Agency[b]	HHI_Leap	B_Leap	-0.26	31	> .99	0.04	negligible
	HHI_Leap	Oculus	1.33	31	> .99	0.40	small
Comfort	HHI_Leap	B_Leap	-0.31	31	> .99	0.04	negligible
	HHI_Leap	Oculus	-3.09	31	.07	0.40	small
Gripping	HHI_Leap	B_Leap	0.91	31	> .99	0.04	negligible
	HHI_Leap	Oculus	-5.21	31	< .0001	0.40	small
Intuitive	HHI_Leap	B_Leap	-0.35	31	> .99	0.04	negligible
	HHI_Leap	Oculus	-0.54	31	> .99	0.40	small
Natural	HHI_Leap	B_Leap	0.13	31	> .99	0.04	negligible
	HHI_Leap	Oculus	-1.40	31	> .99	0.40	small
Precise	HHI_Leap	B_Leap	0.60	31	> .99	0.04	negligible
	HHI_Leap	Oculus	-6.47	31	< .0001	0.40	small
Recommend	HHI_Leap	B_Leap	0.17	31	> .99	0.04	negligible

	HHI_Leap	Oculus	-2.66	31	.19	0.40	small
Releasing	HHI_Leap	B_Leap	0.64	31	> .99	0.04	negligible
	HHI_Leap	Oculus	-6.14	31	< .0001	0.40	small
Satisfaction[b]	HHI_Leap	B_Leap	0.78	31	> .99	0.04	negligible
	HHI_Leap	Oculus	-2.46	31	.30	0.40	small
Tiring	HHI_Leap	B_Leap	-0.18	31	> .99	0.04	negligible
	HHI_Leap	Oculus	-0.11	31	> .99	0.40	small
Grand means (5-point Likert questions)	HHI_Leap	B_Leap	0.44	31	.66	0.08	negligible
	HHI_Leap	Oculus	-4.13	31	< .001	0.73	moderate
Grand means (7-point Likert questions)	HHI_Leap	B_Leap	-0.26	31	> .99	0.05	negligible
	HHI_Leap	Oculus	-2.46	31	.30	0.40	small

[a]Adjustment for multiple comparisons: Holm
[b]7-point Likert scale question (all others are 5-point)

Appendix C: Questionnaires

5-point Likert questions (English)

Rate the following questions:
1 (strongly disagree) 2 3 4 5 (strongly agree):

1. Using this interface was comfortable
2. This interface was precise
3. This interface was intuitive
4. This interface was tiring for the hand *(reverse scored)*
5. The gripping of objects gave me a lot of trouble *(reverse scored)*
6. The releasing of objects gave me a lot of trouble *(reverse scored)*
7. The gripping and releasing of objects was very natural
8. I would recommend this interface to friends

7-point Likert questions (English)

Agency: I felt like I controlled the virtual representation of the hand as if it was part of my own body. Please choose one of the following answers:
1 (strongly disagree) 2 3 4 5 6 7 (strongly agree)

Overall satisfaction: Which of the following faces best represents your overall satisfaction with using these interfaces?

Fig. 23 Smiley faces accompanying overall satisfaction question

5-point Likert questions (German)

Bewerten Sie die Nutzung der eben getesteten Hand-Interaktionstechnologie!
1 (stimme gar nicht zu) bis 5 (stimme sehr zu).

1. Die Benutzung dieser Interaktionstechnologie war komfortabel
2. Diese Interaktionstechnologie war präzise
3. Die Interaktionstechnologie war intuitiv
4. Diese Interaktionstechnologie war für die Hand ermüdend
5. Das Greifen der Objekte bereitete mir große Mühe
6. Das Loslassen der Objekte bereitete mir große Mühe
7. Das Greifen und Loslassen der Objekte war sehr natürlich
8. Ich würde diese Interaktionstechnologie Freunde empfehlen

7-point Likert questions (German)

Agency: Ich hatte das Gefühl, die virtuelle Darstellung der Hand so zu steuern, als ob sie Teil meines eigenen Körpers wäre. Bitte wählen Sie eine der folgenden Antworten:
1 (stimme gar nicht zu) 2 3 4 5 6 7 (stimme sehr zu)

Overall satisfaction: Welches Gesicht entspricht am ehesten Ihre Gesamtzufriedenheit mit der Nutzung dieser Interaktionstechnologie?
1 2 3 4 5 6 7

Appendix D: Demographics and Experience

Table 9 Demographics and experience

	N	Min.	Max	Median	Mean	SD
Age	32	22	36	27.5	27.812	3.37
Arm length*	31	63	89	79	78.194	5.74
Height	32	159	194	179	177.750	8.88
Exp. w/ controllers	32	2	5	3	3.000	1.05
Exp. w/ VR	32	1	5	2	2.250	1.14

Exp. w/ all games	32	1	5	4	3.719	1.35

Gender	Male	Female	N/A
	23	8	1

*missing arm length measurement from one subject

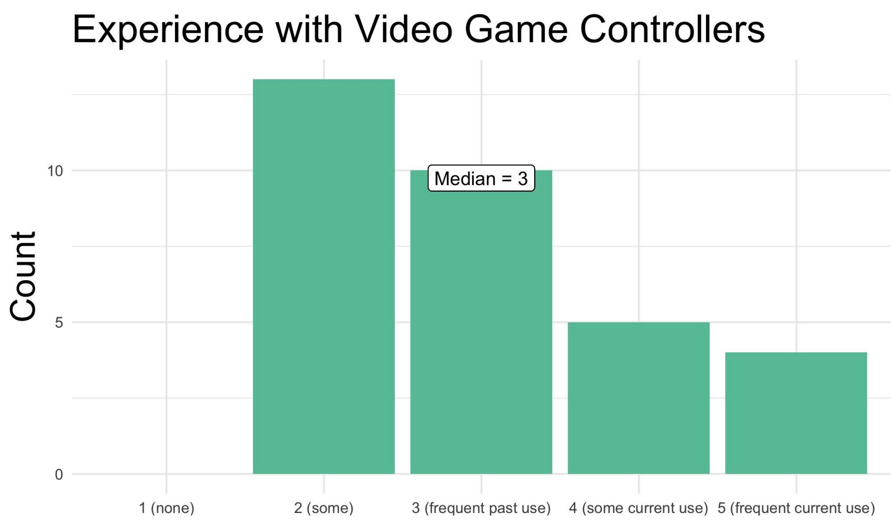


Fig. 24 Experience with video game controllers

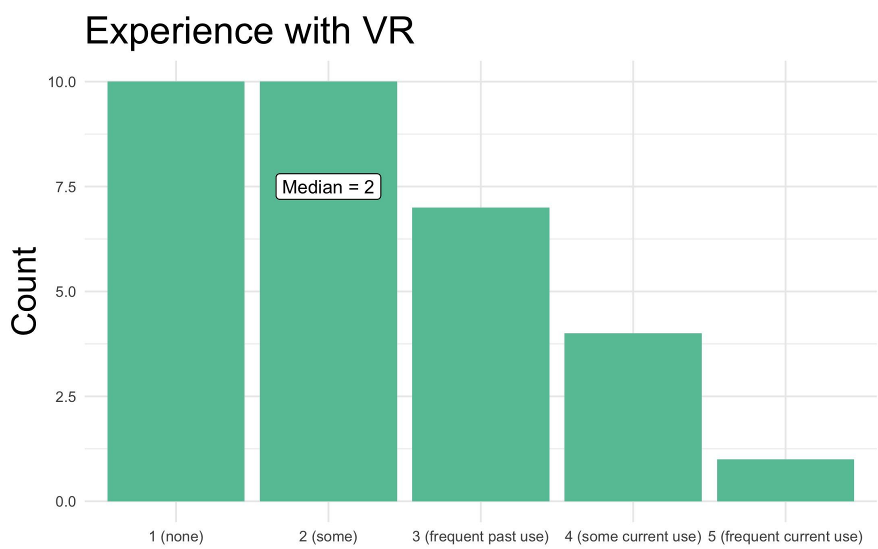


Fig. 25 Experience with virtual reality

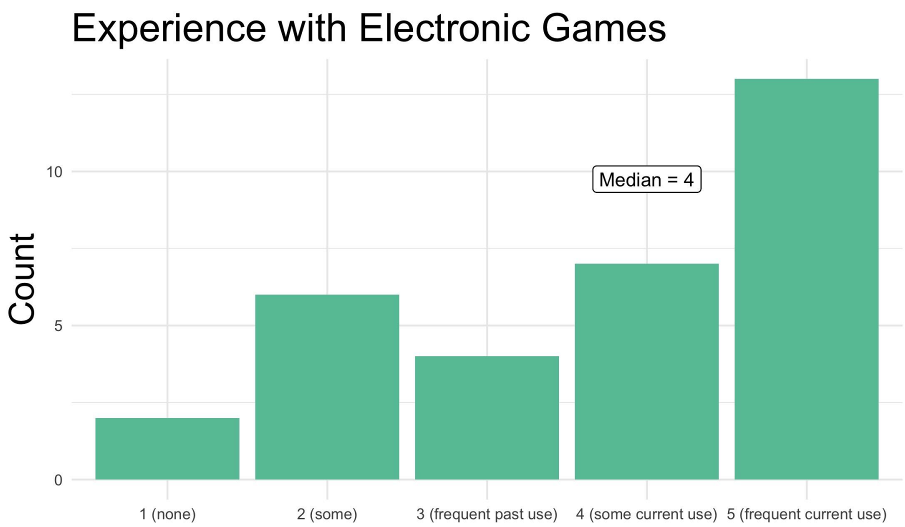


Fig. 26 Experience with electronic games

Declaration of Authorship

I hereby declare that my thesis is the result of my own work, and has been written only with the help of the indicated sources.

In particular, I confirm that I have cited the source for all instances, without exception, in which I have reproduced the work of another author, including material from the internet. This includes both direct reproduction, such as direct quotation or the use of unmodified tables, graphs, images and similar, and indirect reproduction, such as statements of another author's opinion in my own words or the use of modified tables, graphs or other images.

I am aware that any breach of these rules is considered to be plagiarism and will be punished according to the examination regulations and/or the general university regulations (Fächerübergreifenden Satzung zur Regelung von Zulassung, Studium und Prüfung der Humboldt-Universität zu Berlin (§ 111 ZSP-HU)).

This thesis, in the same or similar form, has not previously been submitted to any other institution.

Berlin, 04.03.2020 (Signature)