Cognitive and Touch Performance Effects of Mismatched 3D Physical and Visual Perceptions

Jason Hochreiter* Salam Daher† Gerd Bruder‡ Greg Welch§

University of Central Florida

ABSTRACT

While research in the field of augmented reality (AR) has produced many innovative human-computer interaction techniques, some may produce physical and visual perceptions with unforeseen negative impacts on user performance. In a controlled human-subject study we investigated the effects of mismatched physical and visual perception on cognitive load and performance in an AR touching task by varying the *physical fidelity* (matching vs. non-matching physical shape) and *visual mechanism* (projector-based vs. HMD-based AR) of the representation. Participants touched visual targets on four corresponding physical-visual representations of a human head. We evaluated their performance in terms of touch accuracy, response time, and a cognitive load task requiring target size estimations during a concurrent (secondary) counting task. After each condition, participants completed questionnaires concerning mental, physical, and temporal demands; stress; frustration; and usability. Results indicated higher performance, lower cognitive load, and increased usability when participants touched a matching physical head-shaped surface and when visuals were provided by a projector from underneath.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed/augmented reality; Human-centered computing—Human computer interaction (HCI)—Interaction techniques

1 INTRODUCTION

Touching physical objects is considered one of the most natural and common activities performed by humans since early childhood. Being able to touch 3D objects is important for many application fields in the domains of simulation and training. For instance, training of 3D manufacturing tasks requires trainees to reach towards and touch the object at different points on the surface to learn the physical movements involved and associated motor control perception-action loops. Similarly, training healthcare providers typically involves touching physical patient simulators at precise locations while the trainee experiences high cognitive load due to performing multiple tasks simultaneously. There is a strong desire in these application fields to leverage technologies from the fields of virtual reality (VR) and augmented reality (AR) to improve the fidelity of training.

In the field of AR, multiple approaches have been introduced to improve the display fidelity in these applications, with monoscopic or stereoscopic computer-generated imagery provided by projector-based spatial augmented reality (SAR) [6] or head-mounted displays (HMDs). Orthogonally, several methods have been presented to improve haptic touch feedback of virtual imagery, ranging from haptic gloves to static or dynamic 3D physical surfaces [1, 31, 37]. While there are many differences in these approaches and limitations in the current state of technology, it is important for the aforementioned application fields to understand which ecologically valid approaches are available, their limitations, and how state-of-the-art technologies compare in terms of user performance, cognitive load, and usability.

We present a controlled human-subject study concerning the task of accurately touching different points on a human head, which is a common occurrence in medical and non-medical training fields. We evaluated and compared four experimental conditions that match state-of-the-art prototypes of ecologically valid training systems. We formalized the considered AR approaches by considering two technical dimensions: *physical fidelity* and *visual mechanism*, with two levels each. The experiment comprised four corresponding conditions: rear-projection SAR imagery on a matching head-shaped surface, rear-projection SAR imagery on a flat surface, HMD-based virtual imagery registered to a matching head-shaped surface, and HMD-based visual imagery with no physical surface (free space). These conditions reflect possible realizations of such a touch task, depending on the available technology (e.g. whether one has access to a rear-projection surface, a touchscreen, an object not suitable for rear-projection, or an HMD). We compared touch performance in terms of accuracy, response time, usability, and subjective ratings. To measure differences in cognitive load, we incorporated a decision-making task involving size estimation on the surface, which commonly occurs in fields such as healthcare when a nurse sees a mole on a human head and has to make a decision based on its size, alongside a concurrent verbal counting task.

A primary finding of our study is that a mismatch between the visual and tactile perception of an object increases cognitive load in AR touch tasks. Our results demonstrate that participants were more accurate when touching in the SAR conditions and subjectively found them easier to use. There was a strong subjective preference for the SAR condition with geometry matching the virtual content, and participants experienced lower cognitive load in this case. We discuss the results and implications for practitioners aiming to use such state-of-the-art approaches for comparable touch applications.

2 RELATED WORK

2.1 Cognitive Load

Human working memory draws from finite cognitive resources, for which several theoretical models have been proposed [4, 15]. A model of cognition and working memory was proposed by Baddeley and Hitch [3, 4], which considers manipulation and storage of visual and spatial information in a speech-based loop. According to this model, access to verbal and spatial working memory and general attention is handled by a central executive.

Cognitive load describes the resource demands in working memory when learning, training, or performing a task [40]. Task workload varies as a function of the perceptual, cognitive, and motor requirements imposed on the person using a system. There are three types of cognitive load: intrinsic, extraneous, and germane. *Intrinsic* cognitive load refers to the essential load that cannot be reduced. The

*email: jhochrei@cs.ucf.edu
†email: salam@knights.ucf.edu
‡email: Gerd.Bruder@ucf.edu
§email: welch@ucf.edu

IEEE Virtual Reality 2018
18-22 March 2018, Reutlingen, Germany
© IEEE 2018

load introduced by the means of instruction presented to the learner is called *extraneous* cognitive load. *Germane* cognitive load leads to permanent storage of knowledge. Since working memory has a fixed capacity, an increase in the extraneous cognitive load of a task can reduce the remaining capacity for germane cognitive load. Cognitive overload can happen when the means of instruction interfere with the user's ability to process information, which can lead to decreased learning and reduced performance [41]. Ideally, user interfaces wish to reduce extraneous load to free up resources for germane load. For instance, an HMD that provides instructional information to the user performing a physical task might allow him to better focus his mental efforts on the task itself [42]. In VR, researchers have demonstrated that users with a self-avatar whose movements matched their own can perform better in cognitive tasks [38].

The structure of a task can significantly affect mental workload and performance. When a person's task workload exceeds their mental capacity, they will start making mistakes in that task. If two tasks are completed by the user simultaneously, such as a primary task and a secondary task, then these tasks may compete for the same finite mental resources [24, 33]. If an increased workload in one task results in a decrease in performance in the other task, then it can be inferred that these tasks use resources from the same pool. In this case, mental workload can be assessed by the effect on either the primary task or the secondary task, whichever does not receive priority. Such dual-task designs are sensitive measures that can be used to assess changes in spare capacity of cognitive resources.

Researchers have employed different strategies to induce cognitive load using secondary tasks, such as Stroop effects [19], repetitive stimulus comparison tasks [8], or counting tasks with varying difficulty [27]; for example, counting down from 1000 in decrements of 7 leads to a higher cognitive load than counting up in increments of 2 [39]. Similar dual-task paradigms have been used successfully in previous research in the field of VR (e.g. to assess cognitive demands of locomotion [28] and redirected walking [8]). In this paper, we leverage a secondary counting task to assess the performance and cognitive load of users who interact with physical and virtual representations of a 3D model via a primary touch task. Based on internal pilot testing, we opted for subtraction by 7 in response to visual prompts from the 3D model.

2.2 Touch Performance

User performance in interactive tasks such as touching 2D/3D objects is affected by several factors, including aspects of the system such as tracking performance, the type of display, interaction techniques, and the amount of cognitive load induced. Assessing systematic differences in performance in interactive tasks induced by system characteristics is important for a wide range of application fields, requiring effective and efficient user interfaces. Touch performance can be evaluated using various measures, including touch accuracy, response time, and subjective feedback, depending on which dimensions of performance are the most important in the considered application field. For common tasks, such as touching a sequence of objects in 2D/3D user interfaces, multiple standardized tests have been proposed to assess task performance, and several models of human task performance have been established in the field of human-computer interaction (e.g., outlined in the ISO 9241-9 [22]). For instance, Fitts' Law [13] describes the trade-off between touch accuracy and movement time in such interfaces. However, it should be noted that many user interfaces in the fields of VR and AR do not impose similar stress on users to complete tasks as quickly as possible as in studies involving classical Fitts' Law tasks, and the focus shifts more towards the accuracy of touches and cognitive effects than the time it takes to complete that action.

Over the last few years, multiple studies have focused on understanding touch performance in interactive 3D tasks in head-mounted or projector-based display environments [9, 10, 25, 43–45]. In particular, the addition of tactile feedback to touch tasks has been shown to reduce task time and improve accuracy [2, 32]. Moreover, responses to tactile stimuli were found to be significantly less delayed than responses to visual stimuli, which might point towards benefits in decreased cognitive load [11]. While many previous studies focused on semi-natural methods to provide tactile feedback (e.g. using vibro-tactile gloves or finger cups [1, 31, 37]), we are not aware of previous studies investigating the benefits or limitations that may be induced by leveraging 3D surface shapes as commonly used in state-of-the-art SAR displays.

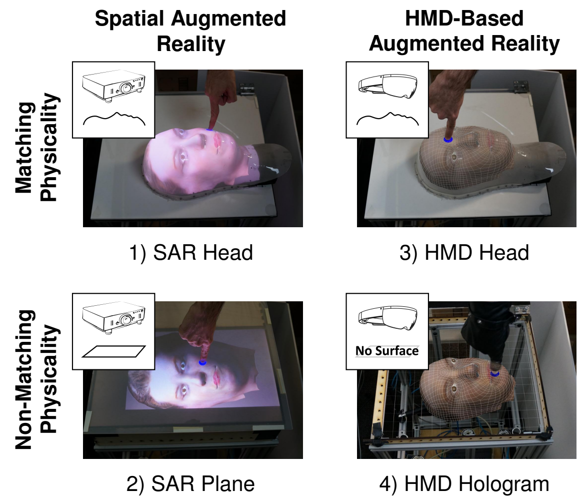


Figure 1: Participants experienced four study conditions, each differing in physical and virtual representation of a 3D human head model. The physical object with which they interacted either matched or did not match the virtual object; the virtual object was displayed either via a projector or through an HMD. For the HMD conditions, the imagery shown above is simulated.

3 Experimental Setup

In this section we describe the four specific realizations of hardware prototypes supporting a controlled touch task on physical-virtual 3D content (namely a human head model), which we termed *SAR Head*, *SAR Plane*, *HMD Head*, and *HMD Hologram* (Fig. 1). Virtual imagery was provided by a projector in the SAR conditions and by a Microsoft HoloLens, a state-of-the-art optical see-through head-mounted display, in the HMD conditions. Participants interacted by touching a matching head-shaped surface (SAR Head and HMD Head), touching a flat surface (SAR Plane), or placing their finger in free space (HMD Hologram) and submitting via a separate device.

Our four conditions were intended to be reasonable representations of how these technologies are used in practice. Many current HMDs, including the HoloLens, provide a controller or other device to indicate input to the system, which induces a baseline cognitive load inherent to such paradigms. As an alternative, requiring users to leave their finger in place for several seconds to submit a touch would make time metrics inconsistent between conditions and might make it difficult for participants to prioritize in a dual-task scenario. Furthermore, we designed the study so that participants would interact solely according to their visual and physical perception of the virtual content as dictated by the capabilities of each condition. For instance, users interacting with 3D content may need to view it from multiple viewpoints. In the HMD Hologram condition, perception is limited to the hologram and the user's proprioception. While we could have added an audio or visual cue to assist the participant in estimating target positions, this would override his or her own perception of the content and likely lead to further cognitive load.

We developed a unified study platform (Fig. 2), used in all four conditions, that consisted of an aluminum frame housing four calibrated Point Grey Blackfly monochrome cameras (resolution

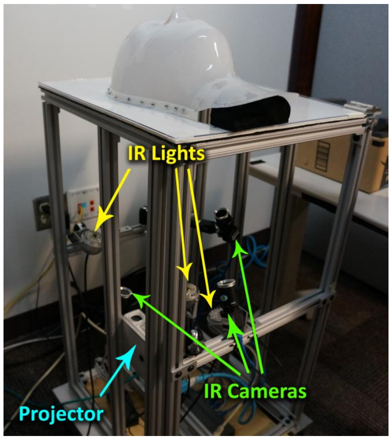


Figure 2: Study platform: the physical head surface, IR cameras and rear-diffused light used for touch sensing, and projector for the SAR conditions. The head surface is attached to the aluminum frame with hinges and can be replaced with a flat surface or removed for the relevant study conditions.

640×512 at 30 frames per second) with removable 780 nm IR filters, one calibrated AAXA P300 pico projector (1920×1080), and three IR illuminators (850 nm). We are able to mount different 3D surfaces onto the frame that match specific application contexts. Our 3D head surface attached directly to the frame with hinges, so it could be removed and replaced in the same position for the SAR Head and HMD Head conditions, and we built a wooden support to provide a consistent position for the flat surface in the SAR Plane condition. For the HMD Hologram condition, we placed a wooden guard with a wire grid to prevent participants from inadvertently reaching too far and touching sensors or other equipment.

3.1 Touch

To support touch interaction for our four conditions, we developed two related optical touch detection methods: one for physical rear-projection surfaces and one for free-space interaction. Both employ rear-diffused IR light, which is largely invisible to the human eye and does not interfere with projected imagery [5, 29, 48]. When a user touches the surface, IR light is reflected and captured by several cameras. Typical rear-diffused illumination setups use simple geometric objects (e.g. flat tabletops and spheres) that can be easily parameterized, so it is straightforward to relate touches detected in 2D camera imagery to 3D locations on the object. Some approaches use frustrated total internal reflection (FTIR) [16], which is applicable to objects with simple curvature (e.g. [47]); however, complex curvature can lead to light bleeding or otherwise disrupt FTIR [36]. Our method is inspired by work by Hochreiter et al. [20], which builds a lookup table of camera, projector, and 3D correspondences for touch sensing on surfaces with more complex geometry that cannot be parameterized and for which FTIR is not suitable.

SAR Head, SAR Plane, and HMD Head

We project a series of structured coded light patterns onto our 3D head-shaped and flat surfaces, localize features in the camera imagery to obtain camera-projector correspondences, and triangulate features to create 3D surface meshes. Using the projector-3D correspondences, we calibrate the projector by computing the maximum likelihood estimate of its projection matrix [18].

The 3D head mesh is likely not appropriate as a 3D model, as it may be of insufficient resolution or have a topology unsuitable for texturing and animating. Thus, we clean up and smooth the mesh and create an appropriate topology before texturing, which may require small adjustments to vertex coordinates. We update the 3D coordinates in the lookup table accordingly so that detected touches are converted only to 3D model coordinates. By back-projecting all camera and projector pixels to 3D rays and finding their points of intersection on the smoothed model, we create lookup tables of higher resolution. For the SAR Plane condition, we analytically project the smoothed model onto the flat mesh.

We rely on simple and fast image processing techniques to detect touches at run time. Incoming camera imagery is background-subtracted and thresholded. Intense white contours that survive the thresholding may represent touches, but they must be distinguished from non-touch events (e.g. fingers or palms hovering over the surface). All potential touch contours from all cameras are converted to projector space via the lookup table and combined. As a finger approaches and ultimately touches the surface, the contributions from each camera converge in projector space. Thus, only contours with sufficient overlap among camera contributions in projector space are accepted as touches. Each detected touch in projector space is converted to 3D coordinates on the graphical model using the lookup table. Thus, touch sensing resolution is affected by how projector pixels map to the physical object, and in practice this can mean slightly less resolution in areas with high curvature for which a single projector pixel may cover more physical space.

HMD Hologram

We developed a related touch detection method for virtual model interactions without a physical surface present by localizing a finger in free space. A user "touches" the model by placing his or her finger at the desired location and indicating the touch through a separate device (e.g. the HoloLens clicker). As in the rear-projection touch sensing system, we perform background subtraction and intensity thresholding on incoming camera imagery. Potential fingertip contours are triangulated, and only contours with consistent triangulations across all pairs of stereo cameras are accepted. To simplify the process of detecting fingers, we ask users to wear a non-IR-reflective glove with a hole cut through the index fingertip.

HoloLens IR Considerations

The HoloLens uses an IR light-based time-of-flight sensor to map the user's environment. It occasionally projects IR signals, appearing as short, drastic intensity spikes that interfere with the above touch sensing techniques. We discard frames in which a substantial change has occurred compared to the background model. Touches that are detected prior to and persist after an IR spike are not interrupted. Using this approach, we found it possible to use IR-based touch sensing together with the HoloLens and similar sensors.

3.2 Visuals

Virtual imagery was provided by a calibrated projector for the SAR conditions and a HoloLens for the HMD conditions. Though this HMD was designed for content 2 m away from the user, Microsoft states that content can be comfortably viewed at distances of 50 cm if the content and user remain generally stationary [21], which our study setup maintains. Furthermore, while the HoloLens has a limited field of view, the dimensions of our 3D head model are such that the virtual content fits almost entirely in it. As with many HMDs, the user's hand is occluded by virtual content with the HoloLens. As we were interested in comparing perceptual effects on touch performance, we asked users to place their finger based solely on their own perception of the virtual content, and they received no feedback regarding their touch accuracy.

Virtual Alignment

The HMD conditions require a mechanism for aligning virtual imagery to a physical object. Some approaches detect physical markers in the user's environment to place virtual content relative to them [30, 34], but this assumes the physical relationship between the markers and virtual content is accurately modeled. Our registration method uses head-tracking position and gaze information from the

HoloLens, inspired by previous work such as the single point active alignment method [46].

Let V be the 3D vertices of the virtual model and P be the corresponding 3D coordinates on the physical object in the coordinate space defined by the HMD. The goal is to compute a transform that moves V to P, which will align the virtual model to the physical object: $RV + T = P$. The coordinates P are not known; if they were, we could compute the transform directly. Instead, we collect gaze vectors passing through known 3D points on the virtual model to approximate points in P, which can be used to estimate the desired transform. Let $\{C_i^V\} \subseteq V$ be a set of control points on the virtual model with known 3D positions. These control points have corresponding coordinates $\{C_i^P\} \subseteq P$ on the physical object. Each control point C_i^V is displayed on the physical object using the projector by retrieving projector-3D correspondences from the lookup table. Through the HMD, an operator carefully aligns his or her gaze with the projected control point 2 or more times. The resulting gaze vectors originate at the operator's head position as reported by the HMD and pass through the current control point. For a given control point C_i^V, let G_i be the set of gaze vectors. These gaze vectors in general will not intersect; let X_i be the closest point to the set of rays G_i. Each point X_i is an approximation of the control point C_i^P on the physical surface. From here, the virtual and physical models can be aligned by computing a rotation matrix R and translation matrix T that together transform the virtual control points $\{C_i^V\}$ to the physical control point approximations $\{X_i\}$. The transformation can be found by minimizing the expression $\sum_i (RC_i^V + T) - X_i$. In practice, we found that using roughly 60 control points with 2 gaze vectors each produced reasonable results. The resulting alignments generally require a manual adjustment of only about 1–2 cm.

While the above approach assumes the availability of a projector to display the virtual control points on the physical surface, this is not a strict requirement. One simply needs a way to align gaze vectors with a set of known 3D points in a specific order on the desired physical object. Moreover, degenerate cases such as planar or symmetric control points will only allow for alignments up to a rotation or prevent a successful alignment.

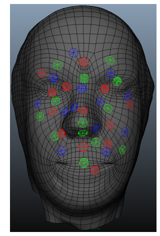
3.3 Computing

A single Unity server controlled the study state for all four conditions. We created a virtual camera in Unity and applied the physical projector's intrinsic and extrinsic calibration data to it; the image rendered by this camera was sent to the projector to display the visual stimulus to the participant. For the HMD conditions, the server interfaced with a HoloLens client with a stored spatial anchor that rendered the virtual model at a consistent position over time. To assist the HoloLens with maintaining the anchor, we placed a variety of color images with asymmetric features in the immediate environment. We projected roughly 60 control points and collected 2 gaze vectors each to compute a transform to align the virtual head imagery to the physical head surface (Sect. 3.2). As a one-time fine-tuning of the alignment, an experimenter touched the head surface at a few locations and applied small adjustments until the detected touches rendered by the HMD visually matched the actual touch positions. This alignment persisted across all participants.

A separate computer handled touch sensing, processing incoming camera imagery using the appropriate mechanism—rear-projection or midair touch. Accepted touches were sent to the server over the network. The server logged information regarding the current displayed target or targets along with the positions and response times of each participant's touch.

4 Experiment

In this section we present the experiment which we conducted to investigate the differences between four state-of-the-art physical-virtual representations of a touch-sensitive 3D human head model.



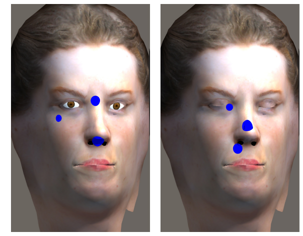


Figure 3: Visual targets. (a) The 39 touch targets across the 3D head. These targets were further divided into small (5 mm, shown in red), medium (7.5 mm, green) and large (10 mm, blue) groups for the *Touch Accuracy Phase*. All targets are shown above as small-sized targets for visualization purposes. (b) Example cognitive load trials. In all trials, one small, one medium, and one large target are shown to participants. They are tasked with touching the medium-sized target. When the virtual human's eyes are closed (right), participants must also provide a verbal response related to a secondary counting task (subtraction by 7 from a given starting number around 500).

We analyze the effects of display type and touch input approach on interaction performance, cognitive load, and subjective preference.

4.1 Participants

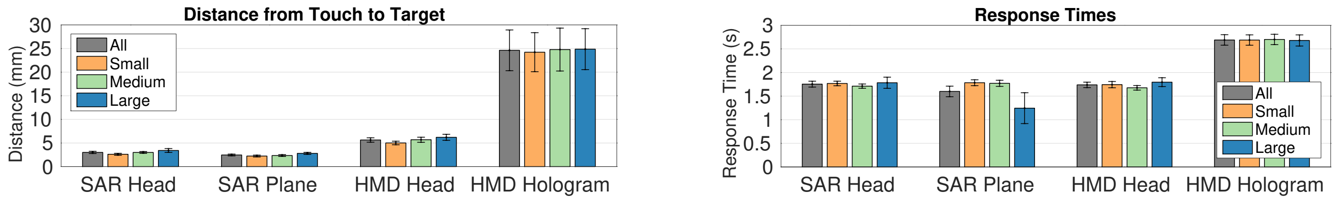
We recruited 24 students or professionals (14 males) from the local university community to participate in our experiment, with ages ranging from 18 to above 50. All had correct or corrected vision, with no reported known visual or motor disorders; eight wore glasses during the experiment. Twenty-two were right-handed, though one of the left-handed participants opted to use their right hand due to a medical condition. Fourteen had experience interacting with 3D content, with 3 having previously worn a HoloLens. We measured the interpupillary distance (IPD) of each participant before the experiment and applied it to render the virtual content on the HoloLens ($M = 6.12$ cm, $SD = 0.3$ cm).

We used a within-subject design due to the expected interpersonal differences in touch behavior and performance. Each participant completed all four study conditions (Fig. 1), which were presented in counterbalanced order using a Latin square design. Each condition included two phases: *Touch Accuracy* and *Cognitive Load*.

4.2 Phases and Tasks

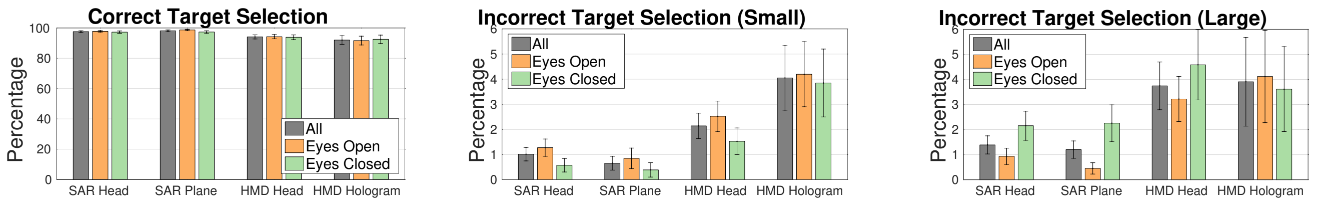
In both phases, one or more visual targets were displayed to participants. These targets came from a set of 39 vertices chosen uniformly across the 3D model (Fig. 3a). At each vertex, we created three spheres of different radii—small (5 mm), medium (7.5 mm), and large (10 mm)—and intersected them with the head geometry, leaving roughly circular targets. This provided consistent targets across all conditions: had we simply used 3D spheres, targets in the HoloLens conditions would have appeared as true 3D spheres, while targets in the SAR conditions would have appeared "flattened" onto the respective surface. For all touch tasks, a touch was accepted as the first point of contact when participants interacted with a physical surface and as their finger location at the moment they pressed the HoloLens clicker in the HMD Hologram condition.

Touch Accuracy Phase We divided the 39 targets into 3 groups chosen uniformly across the 3D model: 13 each of small, medium, and large. For each condition, we created a predetermined, randomized ordering of the targets, with each target appearing exactly twice. Targets were shown sequentially, yielding 78 touch trials. Each participant experienced the same ordering for a given condition. Participants were asked to carefully touch the center of each target and then press the spacebar key of a provided keyboard. Pressing spacebar triggered the next target and also maintained a consistent starting position for all participants and touches across all conditions; in particular, it was not used to confirm a touch.



(a) (b)

Figure 4: Results for the *Touch Accuracy Phase*, separated by study condition. The error bars show the standard error. (a) Average distance between displayed target and user touch. (b) Average time between spacebar press and the next touch. We found significant main effects of display condition on both touch-target distance and response time. Participants also subjectively felt that the SAR Head and Plane conditions were easier than the others in terms of accurately touching and visually locating the targets (Fig. 7), matching their objective results.



(a) (b) (c)

Figure 5: Results for the target selection task in the *Cognitive Load Phase*. (a) The percentage of trials for which participants correctly touched the medium-sized targets. (b) and (c) The percentage of trials for which participants incorrectly touched either the small- or large-sized targets, respectively. We found significant main effects of display condition on medium (correct) and small (incorrect) selections. Questionnaire responses indicated significantly higher perceived task load and significantly lower usability for the HMD Hologram compared to the SAR Head (Fig. 6).

Cognitive Load Phase This phase featured a dual-task paradigm: participants were given two concurrent tasks that competed for their finite mental resources. For the primary task, a set of three targets was displayed—one small, one medium, and one large target from the set of 39 (Fig. 3b)—and the participant needed to touch the medium-sized one. Each of the 39 possible targets appeared as the medium target (i.e. the correct answer) exactly twice, while the remaining two distractor targets (small and large) were drawn from the set of 39 targets. Thus, there were 78 cognitive load trials in total. As before, each condition had an associated predetermined, randomized ordering of displayed targets that was consistent for all participants.

As a secondary task, participants were given a starting number around 500. For each trial, if the virtual human's eyes were closed, the participant had to subtract 7 from this number and state the result verbally to the experimenter; if the eyes were open, the participant had to provide no verbal response and not alter the count. The virtual human's eyes could only change status when a new set of targets was shown, and if the eyes were closed for two consecutive trials, the participant had to update the count twice. After a short period of time (4 seconds), the next set of targets was automatically displayed; no spacebar press was required. Participants were instructed to prioritize the main target size estimation and touching task over the secondary counting task. The starting number and virtual human's eye behavior were predetermined and specific to each condition.

4.3 Study Procedure

Participants first read an informed consent form and completed a demographic questionnaire. They then watched a 2-minute video of sample cognitive load tasks, allowing them to practice counting backward by 7 in response to the virtual human's eyes closing. The video also showed sets of three targets, but they were all the same size; participants were instructed to not estimate target size or touch anything and to focus only on practicing the counting task. The sets of targets advanced at the same speed as in the actual *Cognitive Load Phase* (4 seconds). Participants then took part in a short training phase consisting of 6 and 25 examples of the touch accuracy and cognitive load tasks, respectively. After completing this training, they had a final chance to ask questions, if desired.

Upon completing the *Touch Accuracy Phase*, participants were given a 60-second break during which no imagery was displayed; they could use this time to rest their arms, neck, and eyes. They were informed when 10 seconds remained in the break and were reminded of their starting number for the secondary counting task of the upcoming *Cognitive Load Phase*. After finishing each condition, participants completed subjective questionnaires relating to cognitive load and usability. Finally, once a participant completed all four conditions and questionnaires, we asked them to rank the conditions in terms of difficulty, and we collected additional informal qualitative feedback by asking them a few brief followup questions.

4.4 Measures

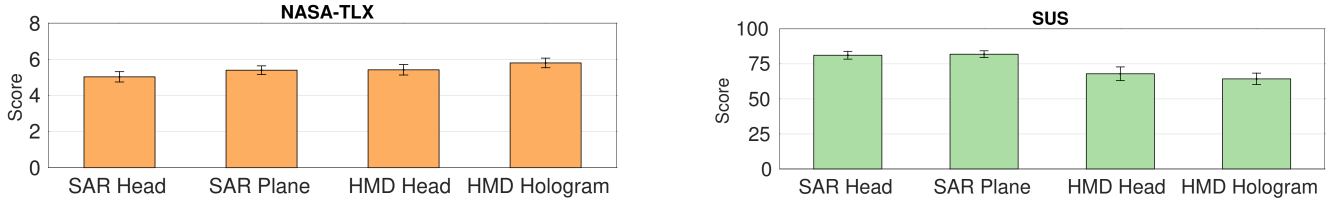
The *Touch Accuracy Phase* included two dependent variables. We measured the Euclidean distance between a displayed target and a participant's touch location (*distance from touch to target*). After touching a target, participants advanced to the next by pressing the spacebar key on a provided keyboard; we measured the time between each spacebar press and the participant's next touch (*response time*). By requiring participants to press the spacebar key, we enforced a consistent starting location for all touches across all conditions.

The two *Cognitive Load Phase* tasks each had an associated dependent variable. Participants were presented with three targets (small, medium, large). We computed the percentage of time that they correctly touched the medium target (*selection of correct target—primary task*); the target with the minimum Euclidean distance to the touch location was used as the participant's selection. Additionally, participants had to give a verbal response after updating a count if and only if the virtual human's eyes were closed for a trial. We computed the percentage of correct responses (*verbal counting task response—secondary task*). Two experimenters logged each verbal response in real time and verified them using recorded videos.

Furthermore, participants completed two questionnaires following each condition: a NASA Task Load Index (TLX) questionnaire [17] and a Simple Usability Scale (SUS) questionnaire [7]. These questionnaires provide information about perceived usability and mental, physical, and temporal demands in the conditions.

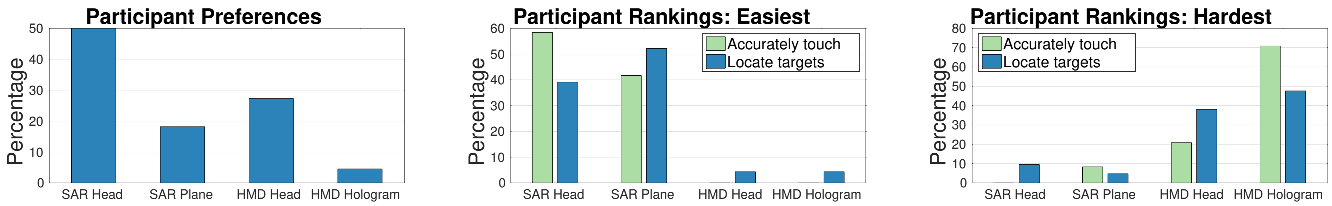
5 Results

We present the descriptive and inferential statistical analysis of the *Touch Accuracy* and *Cognitive Load* phases, along with the subjective questionnaire responses. We underlined the key findings



(a) (b)

Figure 6: Results of the subjective questionnaires for the four experimental conditions: (a) NASA Task Load Index (TLX) and (b) Simple Usability Scale (SUS). Note that lower is better for NASA-TLX, whereas higher is better for SUS. We found significant main effects of display condition on both scores. These results align with objective user performance in terms of touch-target accuracy and response time (Fig. 4).



(a) (b) (c)

Figure 7: Subjective participant rankings for the four experimental conditions. (a) Overall preferred condition. (b) and (c) Subjectively ranked easiest and hardest conditions, respectively, for accurately touching and locating the visual targets. Participants preferred interacting with the rear-projection head, since it affords physical feedback on touch, and they found the HoloLens uncomfortable and disliked the field of view limitations. In general, participants did objectively perform the best on the conditions they subjectively found easiest (Fig. 4 and Fig. 5).

among the results. We analyzed the results with repeated-measures ANOVAs and Tukey multiple comparisons with Bonferroni correction at the 5% significance level. We confirmed the normality with Shapiro-Wilk tests at the 5% level and QQ plots. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated. We used parametric statistical tests to analyze the questionnaire responses in line with the ongoing discussion in the field of psychology indicating that parametric statistics can be a valid and often more expressive method for the analysis of ordinal data measured by the experimental questionnaires [23, 26].

5.1 Touch Performance

In the *Touch Accuracy Phase*, participants touched individually displayed targets. We found significant main effects of display condition on touch-target distance and on response time.

Fig. 4a shows the average distance between the target touched by a participant and the location of the touch, separated by the four study conditions. We found a significant main effect of display condition on touch distance, $F(1.02, 23.50) = 25.16, p < .001, \eta_p^2 = .52$. Pairwise comparisons revealed significant differences between each two of the conditions (all $p < .05$, including $p = .022$ between SAR Head and SAR Plane).

Fig. 4b shows the mean response time between target display and participant's touch. We found a significant main effect of display condition on response time, $F(2.08, 47.90) = 32.45, p < .001, \eta_p^2 = .59$. Pairwise comparisons revealed that the HMD Hologram had significantly longer response times than each of the other three conditions (all $p < .001$). We found no other significant pairwise differences.

5.2 Cognitive Load

In the *Cognitive Load Phase*, participants had to determine which target to touch under increased cognitive load. We found significant main effects of display condition on medium (correct) selections and on small (incorrect) selections.

Fig. 5 shows the results for the selection choice task, including the average percentage of time that participants correctly selected the medium target (Fig. 5a) or incorrectly selected the small or large target (Fig. 5b and c, respectively).

We found a significant main effect of display condition on medium (correct) selections, $F(1.40, 32.10) = 3.87, p = .045, \eta_p^2 = .14$. Pairwise comparisons revealed a significant lower percentage of correct selections for the HMD Head than for the SAR Plane ($p = .019$). We found no other pairwise effects.

We found a significant main effect of display condition on small (incorrect) selections, $F(1.39, 31.99) = 4.83, p = .025, \eta_p^2 = .17$. Pairwise comparisons revealed a trend for a lower percentage of small selections for the SAR Plane than for both the HMD Head ($p = .066$) and HMD Hologram ($p = .062$). We found no other pairwise effects. We found no significant main effect of display condition on large (incorrect) selections, $F(1.39, 31.88) = 2.52, p = .113, \eta_p^2 = .10$.

We found no significant main effect of display condition on correct or incorrect responses to the secondary counting task. However, we observed an error rate of roughly 20% in the participants' verbal responses, pointing to increased cognitive load compared to the *Touch Accuracy Phase*.

5.3 Subjective Responses

We found significant main effects of display condition on the NASA-TLX task load scores, on SUS usability scores, and on subjective participant rankings of the easiest and hardest conditions to accurately touch and visually locate targets.

Task Load The results of the NASA-TLX questionnaire are shown in Fig. 6a. We found a significant main effect of the display conditions on the NASA-TLX task load scores, $F(3, 69) = 3.40, p = .023, \eta_p^2 = .129$. Pairwise comparisons revealed a significantly ($p = .033$) higher task load for the HMD Hologram compared to the SAR Head. We found no other significant pairwise effects.

Usability The results of the SUS questionnaire are shown in Fig. 6b. We found a significant main effect of the display conditions on the SUS usability scores, $F(1.75, 40.22) = 9.95, p = .001, \eta_p^2 = .302$. Pairwise comparisons revealed a significantly higher usability for the SAR Head compared to the HMD Hologram ($p = .003$), and for the SAR Plane compared to the HMD Hologram ($p = .002$). Moreover, we observed a trend suggesting a higher usability for the SAR Head compared to the HMD Head ($p = .074$), and for the SAR Plane compared to the HMD Head ($p = .067$).

Preference During the informal post-experiment interview, participants were asked which of the four conditions they would prefer to use and why. There was a strong preference for the SAR Head, with 11 participants preferring it over the others ($z = 2.357$, $p = 0.018$) (Fig. 7a). Many participants mentioned that they found the SAR Head easier than the other conditions, that they liked the physicality afforded by the head surface, and that they did not enjoy the complexity of wearing an HMD. While some were excited to try the HoloLens, others complained that it was heavy, bothered their neck and nose, and was uncomfortable. All participants were aware of the narrow field of view of the HoloLens, and several limited their head movement while wearing the HMD; we noticed that many tried to touch the spacebar key without turning their heads in the *Touch Accuracy Phase*. These preferences were expressed together with recurring qualitative comments, such as, *The SAR Head was easier to use "because [it was] physically there," The HoloLens was "uncomfortable" and "very heavy,"* and, *The SAR Plane was easier to use because "[I am] used to seeing [imagery] on [a] screen."*

Rankings We asked participants to rank the four conditions in terms of accurately touching and visually locating the targets from easiest (Fig. 7b) to hardest (Fig. 7c). To analyze the rankings, we calculated the exact Clopper-Pearson confidence interval [12, 14]. For accurately touching targets, 14 participants felt it was easiest on the SAR Head ($z = 3.771$, $p < 0.001$), and 17 ranked the HMD Hologram as the hardest ($z = 5.185$, $p < 0.001$). Furthermore, 21 expressed that it was easier to visually locate targets in the SAR conditions ($z = 3.674$, $p < 0.001$), and 18 felt that the HMD conditions were harder than the SAR conditions ($z = 2.449$, $p = 0.014$).

6 Discussion

Our results suggest guidelines for practitioners interested in such AR touch tasks. When high user accuracy, decreased cognitive load, and increased usability are important factors, training systems should use projectors to display visual stimuli to users as opposed to HMDs, if possible. In the case when the desired interactive object is not amenable to projected imagery, an HMD that augments the object is preferable to interaction in free space. Below, we discuss both objective and subjective measures that support these findings.

Most participants preferred the SAR conditions and found them easier and more intuitive. In fact, there was a strong preference for the SAR Head condition. We think this is due to the simplicity and user-friendliness of this paradigm compared to the HMD conditions. Many participants ranked the SAR Plane as easiest in terms of locating the targets, perhaps because it is the most similar to devices they use in their everyday lives (e.g. smart phones and touchscreens). Furthermore, all targets are visible at once on the SAR Plane, while participants often had to look at the SAR Head from multiple vantage points. However, participants generally ranked the SAR Head as being easier to accurately touch than the SAR Plane, likely because the physical head surface more closely matches the virtual imagery.

For the *Touch Accuracy Phase*, the average distance between displayed and touched targets was smallest for the SAR Plane and higher for the SAR Head, suggesting that the tasks of visually locating the center of a projected target and precisely touching it are more challenging on a curved surface. Despite interacting with the same physical surface as in the SAR Head condition, participants were less accurate when touching targets on the HMD Head, likely due to the differences in visual display. Distances were significantly higher for the HMD Hologram condition—nearly 2.5 cm on average. These results comport with post-experiment interviews, where participants rated the SAR conditions as easier to accurately touch than the HMD conditions. Participants were generally most accurate when touching small targets and least accurate when touching large ones, though the differences were slight, pointing to difficulties in visually locating the centers of larger targets—especially on the physical head on which targets may slightly deform due to surface geometry.

Despite the differences in touch accuracy, response times were very similar for the SAR Head and HMD Head. Participants had slightly shorter response times when using the SAR Plane, perhaps because targets cannot be occluded by surface geometry. The longest response times appeared in the HMD Hologram condition. Target size did not have an impact on response time.

During the *Cognitive Load Phase*, participants were more likely to select the wrong targets in the two HMD conditions, pointing to an increase in cognitive load over the SAR conditions. Many participants felt the SAR Head was the easiest and the HMD Hologram the hardest in terms of remembering to observe the virtual human's eyes and update their counts; however, participants on average performed the counting task nearly as well in all conditions.

The majority of participants preferred interacting with a physical object with geometry matching the virtual content (SAR Head and HMD Head). Wearing the HoloLens adds physical and mental demands on the user, as evident in participants' performance, cognitive load and usability questionnaire answers, and post-experiment interview responses. Also, the HoloLens displays imagery directly in front of a user's eyes, overpowering non-HoloLens imagery such as his or her hands and the surrounding environment. In the HMD Hologram condition, participants mentioned that being required to use the clicker to submit a touch was both less enjoyable and less user-friendly. Receiving physical feedback upon touching a physical surface (with any geometry) leads to an improved experience.

We note that results and preferences for the HMD conditions are impacted by our choice of the HoloLens. A lighter HMD or one with a larger field of view may have produced more enjoyable interactions. However, we expect that issues common to current HMDs (e.g. imagery occlusion and the need to wear a device at all) would likely have comparable effects on similar AR touch tasks.

7 Conclusion and Future Work

We explored the differences between four physical-visual representations of a touch-sensitive 3D human head model and their impacts on touch performance, cognitive load, and subjective preferences. Overall, participants expressed a preference for interaction with a physical surface with geometry matching the 3D model; furthermore, they found it easier when the imagery was projected directly onto the surface (from underneath) as opposed to provided by an HMD. Supporting these subjective responses, we found significant benefits in touch accuracy, response time, and size estimation for the spatial augmented reality display conditions. We believe this quantified advantage of SAR for touch sensing is new, compared for example to the often-cited SAR advantage whereby multiple users can see and interact with the surfaces as naturally as they would with real objects, while looking at each other's faces (no HMD), etc. [6, 35].

On the one hand, we are particularly interested in the impacts of physicality and display type for medical applications, where such virtual 3D head and body models are used as a state-of-the-art complement for traditional mannequins in healthcare provider training. However, we are more generally interested in user interface paradigms for free-space "touching" with visual AR via HMDs. We hope that our results here with physical surfaces can serve as a goal or gold standard for future free-space touch mechanisms and that our choice of measures encompassing performance, cognitive load, and subjective preferences can serve as a basis for future assessment of user interface paradigms in general.

References

[1] O. Ariza, P. Lubos, F. Steinicke, and G. Bruder. Ring-shaped haptic device with vibrotactile feedback patterns to support natural spatial interaction. In *Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments, ICAT-EGVE*, vol. 15, pp. 175–181, 2015.

[2] R. Arsenault and C. Ware. Eye-hand co-ordination with force feedback. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 408–414. ACM, 2000.

[3] A. Baddeley. Working memory: theories, models, and controversies. *Annual review of psychology*, 63:1–29, 2012.

[4] A. D. Baddeley and G. Hitch. Working memory. *Psychology of learning and motivation*, 8:47–89, 1974.

[5] H. Benko, A. D. Wilson, and R. Balakrishnan. Sphere: multi-touch interactions on a spherical display. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 77–86, 2008.

[6] O. Bimber and R. Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A. K. Peters, Ltd., Natick, MA, USA, 2005.

[7] J. Brooke. SUS–A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194):4–7, 1996.

[8] G. Bruder, P. Lubas, and F. Steinicke. Cognitive resource demands of redirected walking. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):539–544, 2015.

[9] G. Bruder, F. Steinicke, and W. Stuerzlinger. Touching the void revisited: Analyses of touch behavior on and above tabletop surfaces. In *IFIP Conference on Human-Computer Interaction*, pp. 278–296. Springer, 2013.

[10] G. Bruder, F. Steinicke, and W. Sturzlinger. To touch or not to touch?: Comparing 2D touch and 3D mid-air interaction on stereoscopic tabletop surfaces. In *Proceedings of the 1st Symposium on Spatial User Interaction*, SUI '13, pp. 9–16. ACM, New York, NY, USA, 2013. doi: 10.1145/2491367.2491369

[11] A. Chan and A. Ng. Finger response times to visual, auditory and tactile modality stimuli. 2196:1449–1454, 03 2012.

[12] C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, pp. 404–413, 1934.

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

[14] J. L. Fleiss, B. Levin, and M. C. Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, 2013.

[15] A. S. Gevins and B. C. Cutillo. Neuroelectric evidence for distributed processing in human working memory. *Electroencephalography and Clinical Neurophysiology*, 87:128–143, 1993.

[16] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 115–118, 2005.

[17] S. G. Hart and L. E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[18] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[19] D. Hecht and M. Reiner. Stroop interference and facilitation effects in kinesthetic and haptic tasks. *Advances in Human-Computer Interaction*, 2010:1, 2010.

[20] J. Hochreiter, S. Daher, A. Nagendran, L. Gonzalez, and G. Welch. Optical touch sensing on nonparametric rear-projection surfaces for interactive physical-virtual experiences. *Presence: Teleoperators and Virtual Environments*, 25(1):33–46, 2016.

[21] Hologram stability. `https://developer.microsoft.com/en-us/windows/mixed-reality/hologram_stability`. Accessed: 2017-11-21.

[22] ISO. Ergonomic requirements for office work with visual display terminals (VDTs)—Part 9: Requirements for non-keyboard input devices. ISO, International Organization for Standardization, Geneva, Switzerland, 2007.

[23] T. R. Knapp. Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing Research*, 39(2):121–123, 1990.

[24] W. B. Knowles. Operator loading tasks. *Human factors*, 5(2):155–161, 1963.

[25] L. Kohli, M. C. Whitton, and F. P. Brooks. Redirected touching: Training and adaptation in warped virtual spaces. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pp. 79–86. IEEE, 2013.

[26] W. M. Kuzon Jr, M. G. Urbanchek, and S. McCabe. The seven deadly sins of statistical analysis. *Annals of Plastic Surgery*, 37(3):265–272, 1996.

[27] B. P. Lewis and D. E. Linder. Thinking about choking? Attentional processes and paradoxical performance. *Personality and Social Psychology Bulletin*, 23(9):937–944, 1997.

[28] W. E. Marsh, J. W. Kelly, V. J. Dark, and J. H. Oliver. Cognitive demands of semi-natural virtual locomotion. *Presence*, 22(3):216–234, 2013.

[29] N. Matsushita and J. Rekimoto. HoloWall: designing a finger, hand, body, and object sensitive wall. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 209–210, 1997.

[30] Mixed reality companion kit. `https://github.com/Microsoft/MixedRealityCompanionKit`. Accessed: 2017-09-07.

[31] A. Nguyen and A. Banic. *3DTouch: A wearable 3D input device for 3D applications*. IEEE, 2015.

[32] I. Poupyrev, S. Maruyama, and J. Rekimoto. Ambient touch: Designing tactile interfaces for handheld devices. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology*, UIST '02, pp. 51–60. ACM, New York, NY, USA, 2002. doi: 10.1145/571985.571993

[33] R. W. Proctor and T. Van Zandt. *Human factors in simple and complex systems*. CRC press, 2008.

[34] L. Qian, E. Azimi, P. Kazanzides, and N. Navab. Comprehensive tracker based display calibration for holographic optical see-through head-mounted display. *arXiv preprint arXiv:1703.05834*, 2017.

[35] R. Raskar, G. Welch, and H. Fuchs. Spatially augmented reality. In *In First IEEE Workshop on Augmented Reality (IWAR '98)*, pp. 11–20, 1998.

[36] A. Roudaut, H. Pohl, and P. Baudisch. Touch input on curved surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1011–1020. ACM, 2011.

[37] S. Schätzle, T. Ende, T. Wüsthoff, and C. Preusche. Vibrotac: An ergonomic and versatile usable vibrotactile feedback device. In *RO-MAN, 2010 IEEE*, pp. 670–675. IEEE, 2010.

[38] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *Virtual Reality (VR), 2016 IEEE*, pp. 67–76. IEEE, 2016.

[39] S. R. Steinhauer, R. Condray, and A. Kasparek. Cognitive modulation of midbrain function: task-induced reduction of the pupillary light reflex. *International Journal of Psychophysiology*, 39(1):21–30, 2000.

[40] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988.

[41] J. Sweller, J. J. G. van Merrienboer, and F. G. W. C. Paas. Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251–296, Sep 1998. doi: 10.1023/A:1022193728205

[42] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pp. 73–80. ACM, New York, NY, USA, 2003. doi: 10.1145/642611.642626

[43] R. J. Teather, D. Natapov, and M. Jenkin. Evaluating haptic feedback in virtual environments using ISO 9241–9. In *Virtual Reality Conference (VR), 2010 IEEE*, pp. 307–308. IEEE, 2010.

[44] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie. Effects of tracking technology, latency, and spatial jitter on object movement. In *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on*, pp. 43–50. IEEE, 2009.

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

[46] M. Tuceryan, Y. Genc, and N. Navab. Single-point active alignment method (SPAAM) for optical see-through HMD calibration for augmented reality. *Presence: Teleoperators and Virtual Environments*, 11(3):259–276, 2002.

[47] S. Voelker, C. Sutter, L. Wang, and J. Borchers. Understanding flicking on curved surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pp. 189–198. ACM, New York, NY, USA, 2012. doi: 10.1145/2207676.2207703

[48] A. D. Wilson. TouchLight: an imaging touch screen and display for gesture-based interaction. In *Proceedings of the ACM International Conference on Multimodal Interfaces*, pp. 69–76, 2004.