Selection at a Distance Through a Large Transparent Touch Screen

Sebastian Rigling* University of Stuttgart
Steffen Koch† University of Stuttgart
Dieter Schmalstieg‡ University of Stuttgart
Bruce H. Thomas§ Univ. of South Australia
Michael Sedlmair¶ University of Stuttgart

Abstract

Large transparent touch screens (LTTS) have recently become commercially available. These displays have the potential for engaging Augmented Reality (AR) applications, especially in public and shared spaces. However, the interaction with objects in the real environment behind the display remains challenging: Users must combine pointing and touch input if they want to select objects at varying distances. There is a lot of work on wearable or mobile AR displays, but little on how users interact with LTTS. Our goal is to contribute to a better understanding of natural user interaction for these AR displays. To this end, we developed a prototype and evaluated different pointing techniques for selecting 12 physical targets behind an LTTS, with distances ranging from 6 to 401 cm. We conducted a user study with 16 participants and measured user preferences, performance, and behavior. We analyzed the change in accuracy depending on the target position and the selection technique used. Our findings include: (a) Users naturally align the touch point with their line of sight for targets farther than 36 cm behind the LTTS. (b) This technique provides the lowest angular deviation compared to other techniques. (c) Some user close one eye to improve their performance. Our results help to improve future AR scenarios using LTTS systems.

Index Terms: touch interaction, augmented reality, transparent display, target selection, pointing

1 Introduction

In recent years, large transparent touch screens (LTTS) have become commercially available for digital signage. We are motivated by the idea of using an LTTS as an Augmented Reality (AR) display (Fig. 1), as the LTTS combines a transparent OLED display with touch surface input. In our ongoing project, we are developing what we refer to as "Touch-And-Point" (TAP) selection, an AR-assisted interaction technique for the selection of physical objects behind an LTTS. In pursuit of this goal, we lay the foundation for TAP interaction. We want to gain a better understanding of the challenges and limitations, and therefore focus on the perception and natural user interaction of the selection task. In the following, we embed our work into the larger context of its AR application.

LTTS as AR displays. AR technology is anticipated to become ubiquitous within the next decades. However, there are several technical and non-technical challenges along the way [3]. Current state-of-the-art AR systems often use a stereoscopic head-mounted display (HMD). An HMD has many advantages, but the form factor can be obtrusive, the battery life limited, eye contact obstructed, and the price too high for the average consumer. In public and shared spaces, wearing an HMD affects both users and bystanders: It raises privacy concerns [16], affects social interaction [19, 33], and increases perceived isolation and safety concerns [7]. For these reasons, the use of an HMD is currently primarily restricted to the completion of a specific, time-limited task. We believe there is room for other forms of AR which rely on ambient displays in the environment instead of personal, mobile devices. Especially in public and shared spaces, where no training is possible, AR displays should allow one to simply "walk-up and use". This unobtrusive mode of operation could accelerate the widespread adoption of AR.

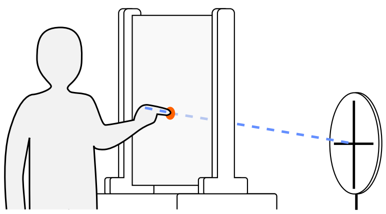
*e-mail: sebastian.rigling@visus.uni-stuttgart.de
†e-mail: steffen.koch@vis.uni-stuttgart.de
‡e-mail: dieter.schmalstieg@visus.uni-stuttgart.de
§e-mail: bruce.thomas@unisa.edu.au
¶e-mail: michael.sedlmair@visus.uni-stuttgart.de

Figure 1: Our prototype of a TAP system for AR. The LTTS displays view-dependent information. The user can select objects behind the screen via touch input.

An LTTS can seamlessly blend into human-made environments where glass surfaces are already used to separate users and physical objects of interest. It can turn the separation into an interaction-and-display layer for objects that may not be touched or manipulated directly. Research examples with interactive transparent surfaces include museum showcases [5, 24], aquariums [28], cheese counters in supermarkets [17], or even the side windows in cars [39]. Their common use case is the selection of an object of interest for information or purchase. Other potential applications are public info screens, shop windows, and vending machines. With TAP selection, there is no context switch or disconnect between selection target and user input.

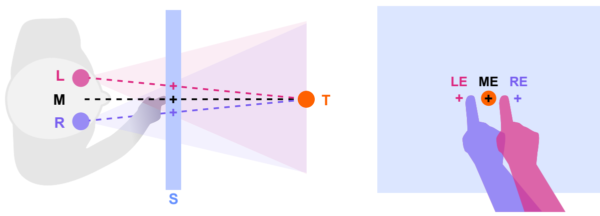
In these examples, an LTTS offers several advantages over an HMD: (1) Users do not need to wear, handle or manage the device. (2) The LTTS does not affect social interaction of users in front of the LTTS. (3) When viewed at arm's length, an LTTS can cover a larger area of the user's field of view than is currently possible with an optical see-through HMD. (4) Touch input is a familiar concept and, for large screens, has been proven to be faster and more accurate than mid-air gestures [31] and laser pointers [44]. For these reasons, we believe in the potential of the LTTS as an AR display variant. However, because an LTTS is not mobile and versatile like an HMD, its use is limited to environments that allow for meaningful integration, as in the examples above.

LTTS interaction at a distance. Information displayed in 2D on the LTTS can be determined based on the real-world context of the objects behind the surface. However, if we want to pursue full three-dimensional interaction with real objects at a distance via the touch screen, the problem of selection and targeting emerges [22]. Since the object viewed at a distance cannot be touched directly, it



(a) **ORI** Oriented (b) **EF** Eye-Finger (c) **ORT** Orthogonal

Figure 2: The three main types of ray casting as defined by Cabric et al. [11]. The ray passes through the touch point on the screen. (a) In Oriented mode, the ray has the same direction as the touching finger's orientation. (b) In Eye-Finger mode, the ray connects the eye and the finger tip. (c) In Orthogonal mode, the ray is perpendicular to the screen surface.



(a) Top View (b) First Person View

Figure 3: In the top view (a) there are EF rays for the left eye **L**, middle eye **M** and right eye **R** towards the target **T**. This situation corresponds to the eye directions when the user focuses on the target, converging in the focus point. The three rays pass the screen **S** in the points **LE**, **ME**, and **RE**, as can be seen in the first person view (b). Due to the binocular parallax, the user perceives a double image of their finger, which does not appear to be on the target.

must be selected using a pointing or ray casting method in combination with touch. For a monoscopic LTTS, this TAP interaction becomes more challenging with increasing distance to the target, as the user's eyes cannot focus on the touch surface (or touching hand) and the pointing target at the same time [4]: When one is focused, the other appears blurred (*accommodation*) and double (*vergence*, Fig. 3). If the user sees the touching finger as a double image, it is unclear which point is used as a reference for the input.

Cabric et al. [11] developed *TouchGlass* and explored how visitors would interact with a touch-enabled museum showcase when asked to select the exhibits behind it. The selection targets were arranged in symmetric patterns on a plate that was parallel to the LTTS and up to 55 cm behind it, creating a spatial frame of reference. They found that users would apply one of three TAP modes: *oriented* mode (ORI), shooting a ray from the touch point in the direction indicated by the pointing finger, *eye-finger* mode (EF), shooting a ray from the eye through the touch point, and *orthogonal* mode (ORT), casting a ray orthogonally from the touch surface to the target (Fig. 2). Their results suggest that ORT is the most common TAP mode and accurate for objects up to 35 cm behind the glass. The authors did not measure the accuracy of the other TAP modes or target distances above 55 cm.

We wonder if their findings apply to a scenario with a wider range of object positions and distances, as may occur in many realistic settings. We assume that participants might choose a different TAP mode if there are no clear spatial references or patterns. A potential TAP mode could be EF, but binocular parallax poses a challenge. While other LTTS systems used ocular dominance to circumvent this issue [27], we believe that a system in public and shared spaces should—realistically—be agnostic of the user's ocular dominance. Inspired by shooting sports, we suspect that users could naturally mitigate the parallax effect by closing one eye.

For this reason, we extend on the work of Cabric et al.: (1) We test a wider range of object positions and distances. (2) We do not arrange targets in a visible pattern or spatial frame of reference. (3) We test EF with both eyes open and with one eye closed. (4) We use sensors to measure and compare the user behavior and accuracy of each TAP mode.

This work. To the best of our knowledge, there is no empirical investigation on the natural selection of objects behind an LTTS that compares measured errors of the three different TAP modes over a wide range of distances. We believe that this can contribute to a more comprehensive or holistic understanding of the problem. To that end, we want to collect subjective user feedback and measure a variety of variables to determine how user preference, behavior, and selection accuracy change with the object position and the TAP mode used. Based on these preliminary thoughts, we formulated the following research questions:

RQ1 How does natural interaction change with target position?

RQ2 How accurate are the different TAP modes?

RQ3 Does secondary input improve accuracy?

RQ4 What are users' thoughts on each TAP mode?

We conducted a comparative user study and observed natural user interaction for the selection of 12 targets in a range between 6 cm and 401 cm behind the LTTS. We intentionally did not provide clear spatial references, patterns, visual aids, or feedback. We explored conditions with and without pointing instructions.

We found that around 36 cm target distance, a switch in natural user interaction occured from ORT to EF (RQ1). The angular deviation of EF was more than five times smaller than that of ORT and ten times smaller than that of ORI (RQ2). Gaze input proved to be a viable alternative or addition to the TAP modes (RQ3). Participants agreed that EF felt the most natural (RQ4). We address the research questions in detail in section 6.1.

Our results suggest that EF is the natural TAP mode, even when users are instructed to use ORT. Interestingly, EF (with both eyes open) performs much better than users think. Some users close one eye to further improve accuracy. ORI and the trajectory of the hand did not show promise for accurate selection. In summary, we contribute the results from a comparative user study and novel insights on natural user behavior to improve future AR LTTS systems.

2 Background and Related Work

Our work builds on findings from various research studies on the interaction with LTTS and similar display technologies and devices.

2.1 LTTS

There has been little related work on LTTS in recent years while the commercial availability and popularity of the HMD increased a lot. Cabric et al. [11] built an LTTS prototype *TouchGlass* for a museum use case. They tested three distances (15 cm, 35 cm, 55 cm) and found that 75% of the participants naturally picked the ORT mode. In a second experiment, they examined the success rate of ORT selection. It is worth noting that the experimental setup provided a natural spatial frame of reference. This grounding is important, as ORT relies on the user's spatial understanding and mapping. While it was designed only as an input device, Hirakawa et al. [27] developed an LTTS to use as an AR display. In a user study, they instructed participants to use the EF mode and measured the dominant eye and touch position to calculate the error for three target distances from 0.5 m to 2 m. Our research takes a very similar direction. We extend the range of target positions and use sensors to measure the pointing error for all mentioned TAP modes.

2.2 Other Transparent Screens

There are more systems that use interactive transparent screens in a variety of ways. Stationary transparent screens have been used as AR displays for museum showcases [8] and for human collaboration, in which users on both sides of the screen interact with the user interface while maintaining eye contact [29]. A large body of work has focused on using transparent screens for AR output on desktop computers. This research dates back to the early 1980s [48] and continued until the early 2010s [37, 23], before newer HMD models became commercially available for the same purpose. These systems used transparent screens to render a three-dimensional computer workspace between the user's head and hands. Direct touch input was not possible. Instead, the research investigated stylus input, hand gestures, and typical desktop computer input modalities, i.e. mouse and keyboard. In our research, we explore touch and do not limit the interaction space to the desktop, but include the environment behind the screen.

Research has also been conducted on transparent mobile devices, such as transparent pads [47] and AR tablets, which can be used as a transparent overlay for annotations and interaction [25, 26] when placed on physical surfaces with images and text. In more recent research, Krug et al. [35] created *CleAR Sight*, which uses a transparent handheld touch display as a tangible six degrees of freedom (DOF) input device for an AR HMD. In their work, they explored the combination of 2D (tablet) and 3D (HMD) visualizations and both 2D (touch) and 3D (6 DOF) input. The handheld device was used as an empty picture frame to select the target, which could be seen through it. In their work, the authors also describe binocular parallax as one of the challenges for the selection tasks.

2.3 Selection Tasks

There is a lot of research on a variety of interaction techniques for AR (and virtual reality) selection at a distance. For decades, ray casting has been the most popular technique [9]. Related works propose different solutions to improve selection techniques and visual feedback depending on the task and data sets used [2, 6]. These solutions are especially necessary for dense or cluttered environments and complex data, such as volumetric data sets and point clouds. Examples include progressive refinement [13, 34], bubble mechanisms and depth selection along a ray [41, 50, 52], the combination of 2D touch and 3D spatial input by the use of additional tangible input devices [6, 35], or the addition of hand-tracking [30].

Further improvements can be achieved through multi-modal input techniques. A potentially useful addition to touch interaction is gaze input. The use of eye and gaze as input modality for AR is well studied [46]. Eye-based interaction was found to be a reliable and fast input modality for selection tasks [36, 42, 49, 54, 56]. For this reason, we wanted to include gaze in our investigation.

2.4 Binocular Parallax

Binocular parallax has already been investigated in the context of selection tasks on transparent and stereoscopic touch screens. Both systems suffer equally from the effect that users see a double image of the respective other when they focus on either their finger or the target. Obviously, selection becomes ambiguous if the target is smaller than the distance between the two finger images [38], and no further instructions are given. If instructions are given, the effect can even be used to an advantage: Users can be asked to consciously use the space between the double image of the finger as an area to select everything in between [57]. A different approach is to introduce visual aids. Lee et al. [38] developed a novel "binocular cursor" for transparent displays: The cursor consists of two parts that overlap and appear as one whole when the user focuses the target in the distance. For stereoscopic displays, there are even more approaches to mitigate binocular parallax, e.g., rendering a cursor only for one eye [55], flattening the image [1], a virtual shadow [18], or using a depth-adaptive cursor [58]. In our work, we wanted to investigate perception and natural user interaction to better understand the problem instead of finding a technical solution for binocular parallax.

The role of ocular dominance has been investigated for touch selection tasks on stereoscopic displays for targets -5 cm behind and up to 20 cm in front of the display. Valkov et al. [51] found that most of the participants would touch close to the middle between the left and the right eye projections with a slight tendency towards the dominant eye. Contrary to these results, Bruder et al. [10] reported that large groups of their participants consistently touched either the projection of their dominant eye, non-dominant eye, or the point between both. A user may prefer any of these three points. Therefore, knowing one's dominant eye would not improve the accuracy of EF selection. We are curious to find out the effect of ocular dominance on the touch position for our system.

2.5 Finger Pointing

Pointing gestures can be similar to EF or ORI [40] and can cause confusion if it is not communicated which of the two gestures is being used. In addition, the precise pointing direction can be unclear [21]: For EF pointing, the direction looks correct from the user's perspective, but the direction the arm or finger is pointing looks to bystanders as if they were pointing above the target. For ORI, there is no common understanding of how to translate the arm, hand, and finger posture into a single ray cast [17]. We simulate different rays for the joints of the index finger to see which approximation comes closest to the intended target.

3 System Overview

We built a system that resembles previous LTTS research [11, 27], the main difference of our approach being that we tested a wider range of target positions without any clear spatial references to their placement relative to the screen. We therefore anticipate that our results are generalizable to a large number of realistic scenarios. Another important difference is that we implemented real-time interaction of the three mentioned TAP modes and are able to compare measured pointing errors.

3.1 Main Components

Our experimental setup consisted of two main components, the LTTS prototype and the selection targets behind it. References to the appendix of this paper are labeled with the letter 'A'.

LTTS prototype. For our experiment, we built a functional LTTS prototype (Fig. 4) with AR display capabilities. It used a 55-inch transparent OLED touch screen (LG 55EW5TK) attached to a height-adjustable table. It was connected to a computer (Intel HM570, 16 GB RAM, NVIDIA GeForce RTX 3070 laptop) which

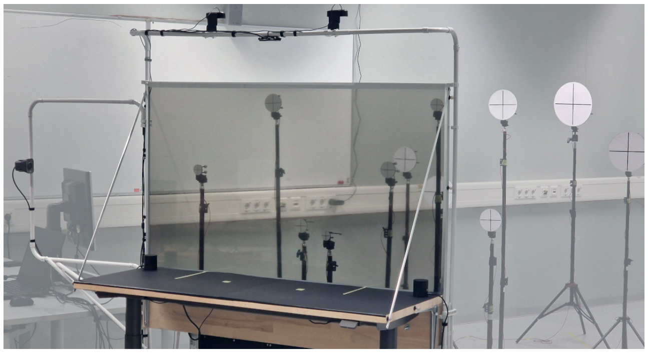


Figure 4: The physical prototype used for our user study.

ran our software to process the input and output data streams. These included the user interface, sound, and target LED lights (output), as well as gaze, touch, and tracked 3D poses (input). The software simulated the 3D ray casts in real-time and saved the log files for each selection task. It used the *Unity* 3D game engine (v. 2020.3).

Targets. We used 12 targets on free-standing tripods with height adjustment behind the screen (Fig. 5a). Their design consisted of a black cross on a circular white background with a red LED light in the center (Fig. A2.1). We chose this design because we wanted to measure the accuracy of pointing regardless of the target area or volume. The targets were arranged into groups of three with vertical and lateral variation (Fig. A2.2a) for an overall even distribution. We chose the distance between the targets so that it was almost linear at the beginning and increased exponentially with distance (Fig. A2.2b) in order to cover a larger area with a reasonable number of targets. This configuration was possible because the change in the angle of vergence and thus the change in the effects of binocular parallax is less pronounced with increasing distance. We set the targets at the following distances behind the screen in cm: 6, 14, 23, 36, 52, 73, 100, 134, 178, 238, 307, and 401. We believe that this range of distances reflects many realistic indoor scenarios.

We arranged the targets in such a way that none of the targets was occluded in the starting head position. In addition, the targets were not at eye level, but either above or below. We chose the starting head position to be a comfortable position for LTTS interaction. To avoid an effect of distance on the saliency of the targets, their size increased with distance to the screen as can be seen in Fig. 5a. At a distance of 50 cm between the participant's eye and the screen, the subjective perception of the optical size appeared to be approximately the same for all targets. In addition, the target size was such that the black cross could not be completely occluded by the participant's finger at a typical viewing distance during touch input. Except for saliency and occlusion, it could be expected that the target size does not influence the accuracy, i.e. the distribution of touch positions on the screen [45].

3.2 Sensors

Our system was equipped with various sensors to record user input and behavior. To simulate the ray casts of each TAP mode, we needed the user's eye, touch and index finger joint positions, and the user's left and right eye states (open or closed). Moreover, we needed target positions to calculate the pointing error. Extending on what is known from previous AR research, we assumed that users would focus on the targets during the selection task [43]. Consequently, gaze could prove to be a viable alternative or addition to TAP. In addition to eye tracking hardware, it usually requires another user action to confirm the selection, e.g., a hand gesture, a button click, or a voice command [22]. In our system, this confirmation was the touch. Therefore, we installed sensors that recorded spatial positions (eyes, targets, touch, hand, index finger joints), 6DOF poses (screen, headset), gaze direction, and eye states.

Touch. The LTTS has a built-in capacitive touch sensor that provides pointer events and coordinates via the input interface of Microsoft Windows.

Spatial tracking. We used the *VICON* room-scale optical tracking system. The tracked space included the area in front of the screen and the targets behind it. Passive infrared markers were fixed to the screen, targets, eye tracking headset, and hand tracking camera.

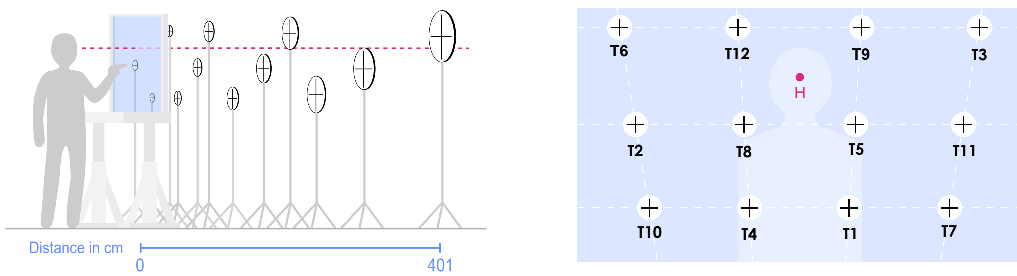
Hand tracking. A *Leap Motion Controller 2* was located above the interaction space in front of the screen. We encountered some challenges that we addressed in our technical implementation. We share our learnings in the appendix (A13). Initial testing showed that the hand tracking sensor still provided inaccurate and inconsistent angles, especially for the posture of the index finger. For this reason, we installed cameras above and to the side of the interaction space. These cameras took photos when the user touched the screen. We created a tool to manually adjust the tracked 3D finger joint positions for every registered touch from the aggregation of both the top- and the side-view photos (Fig. A3.2). This procedure gave us a more accurate position for the index finger joints at the time of touch input.

Eye tracking. Participants wore a *Pupil Labs Neon* eye tracking headset. Gaze data was transferred to our software via a local Python server and websocket connection. We used the headset pose and 2D gaze data to calculate gaze direction and simulate the ray cast in real-time. To increase gaze accuracy, we created an eye calibration procedure for our system: Different targets lit up; we collected a number of gaze positions on the screen and calculated the spread relative to the projected target position. This spread translated into an angular offset and cone-shaped tolerance. It should be considered that this eye tracking system is highly inaccurate when one eye is closed. In addition, its real-time interface (API v. 1.2.1) does not provide the open or closed state of the eyes. We found that comparing the pupil size values of the left and right eyes provided a good estimate of whether one eye was closed and which. With this approach, we measured the total time each eye was closed during the selection task. At the moment of touch, we also saved the photos of the left and right eye from the eye tracking camera stream. We used these photos to manually correct falsely detected open or closed eye states at the moment of touch.

4 Experiment

Previous work on LTTS does not provide data on natural user interaction and selection of objects at a wide range of positions behind an LTTS including a quantitative comparison of the different TAP modes. More specifically, we wanted to explore how user preference and behavior change with respect to the object position (RQ1) and the accuracy of the different TAP modes (RQ2). Furthermore, we wanted to know if secondary input (i.e., finger trajectory and gaze) can improve accuracy (RQ3) and subjective feedback (RQ4). We believe that this data will be relevant to improve TAP input for future LTTS as AR displays.

To this end, we used our LTTS prototpye with 12 targets behind the display and various sensors to record user input and behavior. Participants were instructed to try and select the center of the active LED (active target) as accurately as possible and repeated this task for each target. There was a condition for natural user interaction without any instructions on the TAP modes. For other conditions, participants were instructed to specifically use the TAP modes. The study was approved by the university's ethics committee.



(a) Study Setup

(b) Orthogonal Projection

Figure 5: The spatial distribution of the selection targets (a). Their 2D projections **T1–12** and the default head position **H** on the screen (b).

Supplemental Material. The pre-registration and all supplemental materials are available on OSF[1]. Supplemental materials include (1) study materials, (2) a CSV file containing the data used in our analysis, and (3) an appendix with detailed figures and images.

4.1 Method

Our study followed a within-participant design and a similar user-centered approach as in the work of Cabric et al. [11]. Apart from our setup, our work differs from previous work by replacing self-reports and visual observations with recorded user behavior and simulated pointing errors for each of the known TAP modes. When we repeated the experiment for the different conditions, we did not change the system parameters to obtain comparative data.

Participants. We recruited 19 paid participants via email announcements. Participants were required to be over 18 years of age and have good vision without glasses (at a distance of up to 6 m). The wearing of corrective contact lenses was allowed. After the first two participants, we made changes to our system. The data set of one other participant was incomplete. For these reasons, we excluded these three data sets. We analyzed the data of the remaining 16 participants (P1–P16).

TAP Modes. We introduced three TAP modes: ORT (orthogonal to the touch point), EF (from the eye through the touch point), and ORI (from the touch point in the direction the finger is pointing). We split EF into two different sub-modes: *eye-finger-stereo* **EFS** for both eyes open and *eye-finger-mono* **EFM** for one eye closed. When we refer to EF, we include both EFS and EFM modes. One additional ray cast that was tested in our system was based on the recorded trajectory **TRAJ** of the index finger (Fig. 6b).

Selection Task. One set of selection tasks consisted of 12 unique tasks (one for each target). For each set, the order of the targets was randomized using a Fisher–Yates shuffle. Before each task, participants were instructed to return to their initial position, standing upright in the center in front of the screen. They had to place their index fingers on two marks on the table in front of them. Thus, the distance between the fingers and the touch positions on the screen was the same for all participants. They were instructed to move freely during the task and choose one of the index fingers for the touch input. For each task, one of the LED lights in the targets lit up to indicate which one should be selected. We asked participants to try and select the LED at the center of the target as accurately as possible. To increase the saliency, the LED flashed for one second, and a short stereo audio signal indicated the general direction of the active target. We used Unity's spatial sound capability to create the effect that the sound was coming from the target. The purpose of this signal was to prevent participants from overlooking the most distant or lateral targets and minimize the effect on task completion time. The task started when the LED lit up and ended when the participant touched the screen.

Conditions. We repeated the experiment for different conditions:

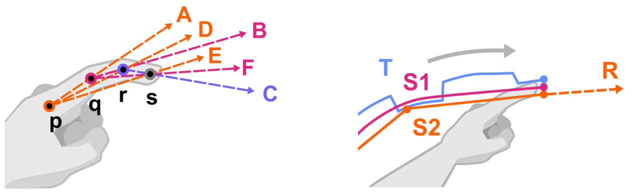
- **s2DT** The *2D target* condition was to collect data as a baseline for touch selection "at 0 cm distance." Instead of physical targets, the LTTS displayed 12 image targets. Their design was the same as the physical targets.
- **sFRS** Participants would complete a set without instructions on how to make TAP selections. During initial testing, users reported that they would think about the technical capabilities or limitations of the system. Thus, we instructed them to "not think about what input this or other systems might expect from them. They should think of this as *freestyle*. Their way of doing it is the right way to do it. They should do what feels natural."
- **sORT** Participants were instructed to complete a full set using only *ORT* mode for selection.
- **sEFS** Repeated for *EFS* mode.
- **sEFM** Repeated for *EFM* mode.
- **sORI** Repeated for *ORI* mode.
- **sQAA** We told participants that they could use and combine any TAP mode. Their goal was only to complete the set *as quickly and accurately as possible.*

4.2 Procedure

Preparation. Participants got general information about the purpose of the study and signed a consent form. We measured the interpupillary distance of the participants before they put on the eye tracking headset. We used the tracked position of the headset to manually adjust the height of the screen and targets for each participant to match the positions shown in Fig. 5b. From an upright standing position, the targets had the same relative position and distance to the participants' eyes, so that every participant had the same perspective on all targets. In this way, we created equal conditions in the event that perspective would have an effect on the relative touch position. In the second step, the LTTS displayed six buttons for the participants to touch. This step was added so that the participants could assume a distance from the touch screen that felt natural for touch input. When the preparations were completed, we started the video and audio recording of the experiment.

Baseline. First, participants would complete a *baseline set* (s2DT) of selecting 12 image targets displayed by the LTTS. This was also a way to prepare the participants for the following conditions, as the overall procedure—a target lit up, a sound played, the participant selected a target, and repeat—was the same for all conditions.

[1] Pre-registration and supplemental materials at https://osf.io/fuzhy/?view_only=ed2c4a1425144ad9be6173e9431c73e1



(a) ORI rays over the finger joints

(b) Trajectory ray cast

Figure 6: The image on the left (a) shows how the finger joints **p–s** can be used to define rays **A–F** for: A = $\overrightarrow{pq}$, B = $\overrightarrow{qr}$, C = $\overrightarrow{rs}$, D = $\overrightarrow{pr}$, E = $\overrightarrow{ps}$, and F = $\overrightarrow{qs}$. On the right (b) is how the finger's trajectory is used for casting a ray. It shows the raw trajectory data (T) smoothed (S1) and simplified (S2). The trajectory ray (R) origin and direction is defined by the last two points on the simplified trajectory.

Freestyle. Next, participants would complete four *freestyle sets* (sFRS). On completion of all four sets, we asked the participants three questions: **Q1** on their strategy used, **Q2** if they noticed any differences between close and distant targets, and **Q3** whether their eyes focused on the finger, the screen or the target.

Explanation. Afterwards, we provided a detailed explanation on the ORT, EFS, EFM, and ORI modes. Participants received two or more unrecorded training sets to try out each TAP mode.

Reference. When participants felt confident that they could use the four TAP modes, they completed one set for each. These are our four *reference sets* (i.e., sORT, sEFS, sEFM, and sORI). The order of these four conditions was counterbalanced across P1–P16 using a balanced Latin square. The reference sets were different from the freestyle set as between each task, we asked the participants **Q4** how they would rate the technique on a scale of 1 (bad) to 5 (great).

On completion of a reference set, we asked participants to answer six Likert scale questions on a scale of 1 (strongly disagree) to 5 (strongly agree):

- **Q5** I think this technique is easy to use.
- **Q6** I think this technique is accurate.
- **Q7** I think this technique is adequate for close targets.
- **Q8** I think this technique is adequate for distant targets.
- **Q9** I think I performed well on near targets.
- **Q10** I think I performed well on distant targets.

Quick and Accurate. When the participants finished all four reference sets, they did one final set *as quickly and accurately as possible* (sQAA). This condition was similar to sFRS, with the difference that participants were exposed to the different TAP modes, they had the opportunity to practice each, and the focus was on performance.

Post Study. After the participant completed the experiment, we asked six open interview questions on their general feedback and experience, we used the Dolman method (hole-in-the-card test) to determine the participant's ocular dominance, and they completed the demographic questionnaire.

4.3 Design of Data Collection

We used the sensors described above to record data for the duration of the selection task at a rate of 30 Hz. In our analysis dataset, we limited this log to the last three seconds of each task before touch. We also recorded the video from the eye-tracking headset's point-of-view camera and the video from an external camera positioned to the side of the experimental setup. Quantitative user feedback and demographic information was collected on paper forms.

Simulation. Because the user's strategy to the EFS and ORI modes is unknown, we simulated all possible rays cast for the target selection. For EFS, we simulated three rays casts: One for each starting at the left eye and right eye position, and one for the point in-between—the middle eye position. For EFM, it was only necessary to simulate one ray cast from the eye open at that time. Likewise, we simulated ORI ray casts for all possible combinations of starting points and directions of finger joints as seen in Fig. 6a. In our analysis, we determined which finger-pointing ray cast (A–F) provided the best estimate. For the TRAJ mode, we had to convert the finger movement into a ray cast. Due to tracking errors in the hardware used, we performed smoothing (polynomial regression) and used a Ramer-Douglas-Peucker algorithm for simplification. The last two points on the simplified trajectory corresponded to the final—and presumably most accurate—phase before the touch. These points therefore defined the starting point and the direction of our ray cast.

Measures There are two spatial frames of references to our setup, the screen space with 2D positions on the screen surface, and the 3D space. On the one hand, we measured the *touch error* in the screen space as the absolute distance of the touch point **P** to a reference point on the screen. The relevant reference points are the 2D projection of the gaze **G**, left eye **LE**, right eye **RE**, middle eye **ME**, head position **H**, and every target's orthogonal projection **T1–T12** (Fig. A3.3). The EFS touch error is the absolute distance from P to ME, while the EFM touch error is the distance from P to LE or RE, depending on which eye was closed. On the other hand, we measured the *pointing error* as the angle between the vector from P to the 3D target position and the vector from P to the closest point to the target in the ray cast. The coordinates are in the local space of the LTTS, with the origin in its bottom left corner. X is the horizontal, Y is the vertical, and Z is the depth axis (target distance).

5 Results

Null hypothesis significance testing (NHST) has drawn criticism in the natural sciences [20]. We report our results using effect size and 95% confidence intervals (CI) as recommended by the APA [53]. This approach is also considered good scientific practice in computer science and human–computer interaction [12, 14] for the improved characterization and statistical credibility of the reported results. Due to the uneven distribution of the touch samples and the relatively small sample size, we reported pointing error with 95% bias-corrected and accelerated bootstrap CI. Unless stated otherwise, the error bars show 95% CI.

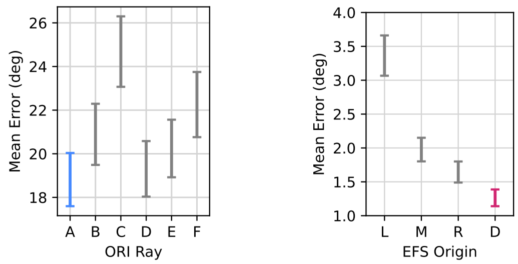
5.1 Participants and Behavior

The following results are based on the demographic questionnaires of the participants and the behavior that could be observed during the analysis of the quantified data.

Participants had diverse backgrounds. The 16 participants included in our analysis were between 23 and 33 years of age (m=26.6 years, SD=2.83). We had eight female and eight male participants. They indicated seven countries as their cultural background (Bosnia, China, Canada, Egypt, Germany, India, Turkey) and six professional backgrounds (architecture, chemistry, computer science, education, engineering, and medicine).

Right eye dominance and handedness were predominant. In the Dolman test, 13 participants showed right-eye dominance compared to three participants with left-eye dominance. Only one was left-handed. The remaining participants were right-handed.

Task completion times varied depending on the condition. The mean values and CI (Fig. A4.1) showed that the task completion time for sORT (m=2.77 s, 95% CI [2.63, 2.92]) was higher than for sFRS (m=2.35 s, 95% CI [2.28, 2.42]) and the other reference sets, which all performed similarly to sFRS. As expected, sQAA showed



(a) ORI mean pointing error

(b) EFS mean pointing error

Figure 7: Mean pointing error for ORI rays **A–F** (a) and EFS rays of different origins (b): the user's left eye **L**, right eye **R**, middle eye **M**, and the dynamic origin **D**. For the latter, we chose L, R, and M as ray origin depending on the proximity of P to the respective eye projection.

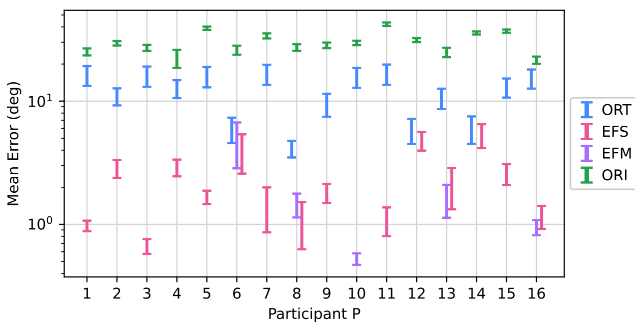


Figure 8: Mean pointing error for P1–16 during sFRS.

clear evidence that it is faster than the other conditions (m=1.64 s, 95% CI [1.56, 1.71]), but equally accurate (Fig. 9a).

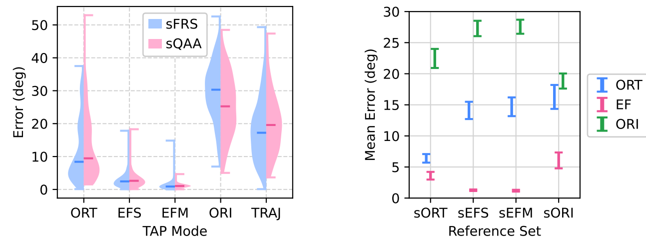
Participants moved the most during sORT. The CI analysis showed that the mean distance from H at the beginning of the task to H after the task is almost twice as large for sORT (m=22.8 cm, 95% CI [21.1, 24.6]) as for the other conditions (Fig. A4.2).

Closing one eye was used naturally and strategically. We counted the number of participants who had one eye closed during and before touching (Fig. A4.3). In the sFRS condition, 3–4 participants would close one eye for targets in >23 cm distance. This behavior changed in the sQAA condition, where around half of the participants closed one eye for targets in >100 cm distance. Interestingly, during the sORT condition, 1–2 participants would have one eye closed during touch, and for >100 cm distance, 3–4 would close one eye for >0.3 s and open it again before touching.

5.2 Pointing Error

Of the ORI rays, A has the lowest average pointing error. Fig. 7a shows the pointing errors of all ORI ray casts A–F. Ray A (m=18.8°, 95% CI [17.6, 20.0]) and Ray D (m=19.3°, 95% CI [18.0, 20.6]) outperformed rays C (m=24.7°, 95% CI [23.0, 26.3]) and F (m=22.3°, 95% CI [20.8, 23.8]). For the sake of simplicity, we only report the results of ray A when discussing further ORI results. For pointing error over distance, see Fig. A5.1.

Picking EFS ray origin by touch position improves its accuracy. As seen in Fig. 7b, evidence suggests that the results for EFS pointing error improve when we choose the ray origin (the user's left eye **L**, right eye **R**, or the point between both eyes **M**) dynamically depending on whether the touch point P is closer to the projected points LE, ME, or RE. The pointing error is smaller



(a) Freestyle pointing error

(b) Reference mean pointing error

Figure 9: A comparison of the TAP modes and their pointing error distribution for sFRS and sQAA (a), and mean pointing error across the reference sets (b). Violin plots show the median.

for this dynamic ray origin (m=1.26°, 95% CI [1.14, 1.38]) compared to using either the left eye ((m=3.36°, 95% CI [3.07, 3.66]), middle eye (m=1.98°, 95% CI [1.80, 2.16]), or right eye (m=1.64°, 95% CI [1.49, 1.80]). We used this approach to calculate the following EFS pointing error.

EF provides the best estimate for natural user interaction. For 13 out of 16 participants, EFS and EFM mean pointing error (Fig. 8) shows strong evidence that it is the smallest. ORT performed just as well as EF for the other three participants, while both modes performed slightly worse overall. This observation indicates that the participants combined these TAP modes and that ORT has a negative effect on the pointing error.

The dataset can be used to simulate target sizes. We used the samples collected during the sFRS conditions and calculated the size that a round target would have needed to be selected by at least 80% of the simulated TAP input (including outliers). The resulting diameters at a distance of 100 cm would be: 6 cm for EFM, 12 cm for EFS (15 cm without dynamic ray origin), 75 cm for ORT, 97 cm for TRAJ, and 153 cm for ORI. Smaller targets are possible if we consider asymmetrical shapes. The appendix (A6) contains pointing error ellipses for all TAP modes and conditions.

EF has the smallest error between conditions. In the sFRS condition, the mean values and the CI analysis showed evidence that the pointing error is different for all simulated TAP modes. From lowest to highest, the modes are EF, ORT, TRAJ, and ORI. This order is also reflected in Fig. 9a. This counterintuitive evidence suggests that EF outperforms the other TAP modes, even when participants were instructed not to use EFS or EFM (Fig. 9b). Comparing the mean pointing error of EFM m=1.19° (95% CI [1.04, 1.33]) and EFS m=1.26° (95% CI [1.14, 1.38]) does not provide evidence that one is consistently more accurate than the other.

5.3 Touch Positions

Touch error indicates three types of natural user strategy. We examined the mean touch error and CI for each target and participant of the sFRS condition (see A7). We found three patterns: Seven participants had a consistently low EF touch error. It appears that they did not actively influence the ORT error. Six participants showed similar touch error for EF and ORT up to a distance of 23 cm or 36 cm. One participant was even consistently lower at 23 cm distance. In this group, the evidence is very strong that the EF error is smaller at distances greater than 36 cm. Finally, three participants showed a more dynamic ORT error, which was closer to or similar to EF across all distances.

When instructed to use the ORT mode, it is accurate up to 36 cm. Fig. 10 shows the mean touch error over distance during the sORT condition. Up to a target distance of around 36 cm, the

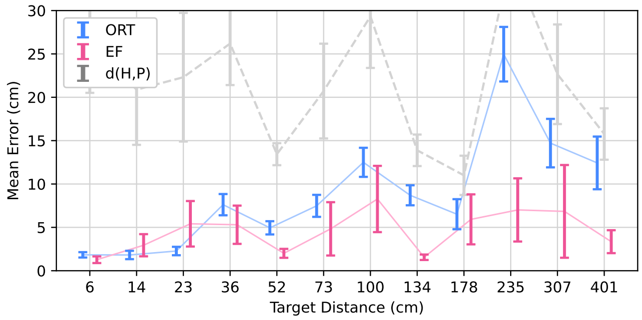
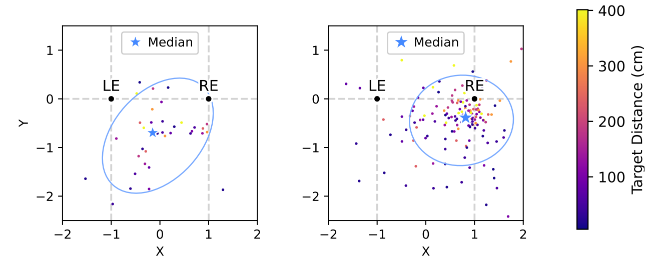
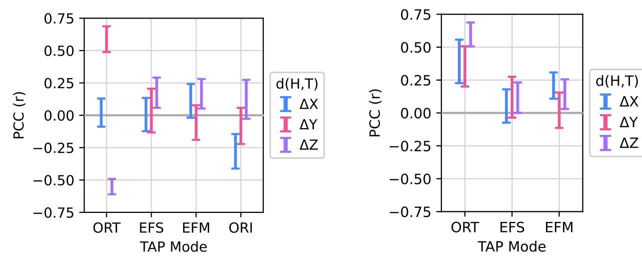


Figure 10: The comparison of EF and ORT mean touch error over distance during the sORT condition. For reference, we show the mean distance between the touch point P and the participant's projected head position H.



(a) Left-Eye Dominance (b) Right-Eye Dominance

Figure 11: The position of P in the sFRS condition normalized for the distance between **LE** and **RE** for participants with left-eye dominance (a) and right-eye dominance (b). The ellipses show the 95% confidence region.



(a) Pointing error and target position (b) Touch error and target position

Figure 12: These graphs show the Pearson correlation coefficient (PCC) for the correlation between the reference sets' head-to-target position and both pointing error (a) and touch error (b).

simulated ORT touch error is similar to or smaller than EF. This result could indicate that ORT was one of the strategies used for very close targets. The ORT error increases slightly with distance from the target. The mean distance between H and P indicates that there is likely a growing correlation between perspective and ORT touch error.

EFM and EFS touch show a vertical offset. We calculated the mean, median, and 95% confidence ellipse of all sEFS and sEFM touch errors (Fig. A8.1) and found a vertical offset for touch error in the sEFS (median y=-8.1 mm) and the sEFM (median y=-8.2 mm) conditions. The sEFS condition also shows a lean toward the right on the horizontal axis, which can be attributed to ocular dominance.

EFS touch leans towards the dominant eye. We normalized the position of P from the sEFS condition between the eye projections LE (at x=-1) and RE (at x=1). Fig. 11 shows that right-eye dominance had a strong influence: The median is at x=0.82 towards RE. The limited data on left-eye dominant participants does not show a tendency. For participants with a dominant right eye that used their left hand, the median shifted towards LE (median x=-0.2) with more touch points near LE and ME (x=0) than RE. Due to the small sample size, the data does not provide evidence in support of or against an influence of handedness.

ORT accuracy strongly correlates with target position. We calculated the Pearson correlation coefficient (PCC) for the correlation between touch or pointing error with the position from head to target at the moment of touch (Fig. 12). We found evidence for a strong positive and negative correlation for the ORT mode, as well as a weak correlation for the ORI pointing error on the horizontal axis. EFS, EFM, and ORI tend to have a weak positive correlation with the target distance. A scatter plot with linear regression analysis is included in the appendix (A9).

Gaze is on the target. The evidence follows our observation during the experiment and the analysis of the video footage (Fig. A10.1) that participants look at the target before and during touch. On average, the touch error was very close to ME (m=6 mm, 95% CI [2, 10]) and P (m=14 mm, 95% CI [7, 21]), with an incline towards ME (the direction of the target). Depending on the distance to the target and the current perspective, these points are very close on the screen. When we compare CI over the target distance, it is noticeable that the participants look at ME (Fig. A10.2). Participants do not look at the orthogonal projection of the target T (m=136 mm, 95% CI [126, 148]). Three participants reported that their focus is either on the target or alternates between the target and their finger, since both are in the same line of sight.

5.4 Subjective Feedback

Participants show a tendency towards EFM at greater distances. The average results (Fig. 13) indicate that other TAP modes received slightly lower scores than EFM, the further away the distance. During the sORT condition, participants also stated that their ability or effort needed to reach a certain point on the screen (ORT) or to orient their finger in a certain direction (ORI) influenced their rating. In the appendix (A11), we created a detailed overview of the 2D target positions and scores for all reference sets.

Participants agree that EF feels natural. 14 out of 16 participants stated that they would prefer to use EFS or EFM alone or in combination with other TAP modes, if they were to use such a system regularly. Approximately half of the participants indicated that they prefer ORT for close targets because they have to concentrate less. This finding confirms our previous results from the quantitative analysis.

EFM is accurate and provides natural feedback. Fig. 14 shows the mean values of the responses to the questions Q5–Q10 defined in Section 4.2. On average, participants agreed that all TAP modes were easy to use, with a tendency towards EFS and EFM. Participants would strongly agree that they found EFM to be accurate and adequate for all distances, while they would still agree that ORT and EFS are accurate, but less so for distant targets. The comments of the participants indicate that EFM performs better because it works well without additional visual feedback. They strongly felt they performed well for close targets, independent of the TAP mode. Regarding accuracy, adequacy, and subjective performance, there is a tendency to agree less for ORI than for the other modes.

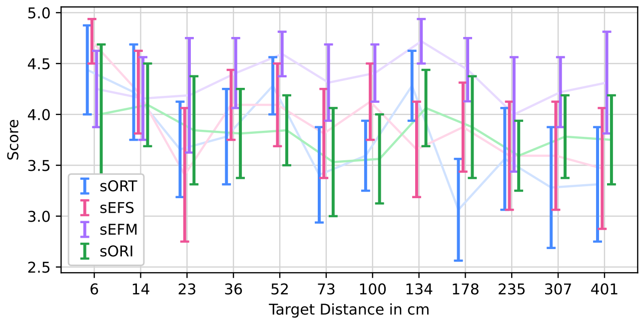


Figure 13: Participants' rating (where 1 means "bad" and 5 means "great") of each TAP mode over target distance.

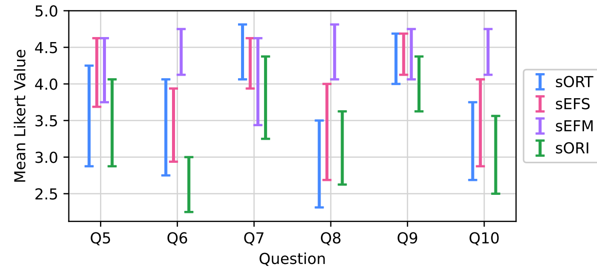


Figure 14: The mean Likert values for the questions **Q5–10** where 1 means "strongly disagree" and 5 means "strongly agree".

Most participants still prefer not to close one eye. Nine participants indicated that closing one eye felt unnatural or uncomfortable. Two participants gave the reason that they practice archery as a hobby and train to aim with both eyes open. Only four participants stated that they did not mind closing one eye, while none had a hobby that required them to close one eye to aim at a target. In general, participants agreed that they would like to limit the time they close one eye to a minimum.

6 Discussion

We wanted to gain a better understanding of the selection of objects behind an LTTS. In the following, we reflect on our study results.

6.1 Research Questions

RQ1: How does natural interaction change with target position? We found evidence suggesting that a large portion of users naturally attempt to touch the closest point on the screen when the target is closer than 36 cm. This finding confirms the results of Cabric et al. [11] who found ORT to be a natural and effective TAP mode for distances up to about 35 cm. For other users and targets above that distance, we found that a switch from ORT to EF mode occurs. For small distances, touch error strongly affects ORT pointing error, which explains why the EF mode is consistently more accurate in terms of pointing error.

RQ2: How accurate are the different TAP modes? Our results suggest that EFS and EFM are the most accurate TAP modes. If users are instructed to use either EFM or EFS, EFM performs slightly better. Our freestyle dataset can be used to simulate the minimum target size for different TAP modes and parameters. Our example calculation from the results section can be further improved, e.g., by taking into account the offsets, correlations and asymmetric distribution found.

We found evidence that the touch error of the ORT mode is very sensitive to the direction and distance of the target. Although the ORT pointing error correlates negatively with target distance, this result is misleading at best. The orthogonal target position is a constant point regardless of the target distance. The angular deviation of any point on the screen decreases and would eventually reach zero at an infinite distance. However, since we already know that the touch error increases, it would be larger than the dimensions of the screen much sooner. At this point, it is already impossible to make a meaningful selection. Because we found that EF provides a more accurate estimate of natural user interaction under all conditions, ORT mode should only be considered for systems that cannot implement EF. Given the large spread of ORI samples, we do not believe this mode to be an accurate TAP mode.

RQ3: Does secondary input improve accuracy? We found that participants look reliably at the target during the task. If a system's gaze tracking capability is accurate enough, it could even be used as the leading input modality. The hand trajectory did not show evidence that it can be used effectively or reliably for selection tasks.

RQ3: What are users' thoughts on each TAP mode? EFM was generally considered an accurate and adequate technique. Participants noted that the fact that they could accurately aim with the finger gave them an estimate of their own performance, which increased their confidence in the mode. EFS was not much less accurate, but participants did not feel confident. They tended to underestimate accuracy and their own performance. In contrast, they overestimated accuracy and their own performance on ORT and ORI. When asked which TAP modes they would like to use in a future system, participants confirmed our previous findings.

6.2 Other Interesting Findings

Closing one eye as a user interaction. Many systems do not consider closing one eye a form of user interaction. We found that it might be worthwhile to consider in an LTTS. Around 25% of the users closed one eye naturally. And after they learned about the technique, 50% would employ closing one eye to their advantage for targets at greater distances. A system that would detect and react to users closing one eye could further improve accuracy and user experience.

Insights on Improving the Accuracy. We found that the accuracy of EFS improves when the ray origin is selected as the left eye, right eye, or middle eye point, depending on which of their projections is closer to the point of touch. This finding could indicate that eye dominance alone is not decisive and that participants could even change their strategy depending on the hand or target position. Investigating this area in the future would be useful.

This finding helps us to better understand EFS touch and improve accuracy, regardless of the user's ocular dominance. Because the left eye, right eye, and middle eye points are projected for the presumed target, this technique is only possible if the potential selection targets are known to the system and all possible projections can be calculated. For targets with closely spaced projections, this approach reveals ambiguities that could be resolved through visual feedback and further user interaction.

6.3 Limitations and Future Work

We have made every effort to obtain meaningful results. It should be emphasized that our study had only a small sample size and is not representative of the entire population. Our sample group showed a strong bias towards right-eye dominance and handedness. In our questionnaire, we measured participants' agreement with given statements on the Likert scale. This could lead to a bias towards the original wording of the statement (acquiescence bias) [32]. Therefore, the results should be viewed with caution. We want to run further tests to confirm or refine the previous results, e.g, through a Fitt's Law [15] test on our LTTS prototype. We will use an improved questionnaire design for more conclusive subjective results. Another open question for future work is how the user's accuracy and behavior change when selection feedback is provided.

7 Conclusion

We built a functional LTTS prototype for selection tasks within a range of 6 cm to 401 cm behind the screen. We conducted an exploratory user study with 16 participants and a quantitative analysis. Our results suggest that most users would align their fingers with the line of sight as a natural user interaction, and some would close one eye. This technique turned out to be more accurate than others. We contribute a data set from our analysis which can be used to simulate target sizes. Our results help to better understand and outline the possibilities and limitations of selecting objects behind an LTTS. We believe that this result is of great relevance for improving the design of future LTTS systems.

Acknowledgments

This work was supported by the Alexander von Humboldt Foundation funded by the German Federal Ministry of Education and Research, the German Research Foundation *DFG* (grants 251654672, 390831618, 390740016), and a joint Weave project between the *DFG* and the Austrian Science Fund *FWF* (DFG 495135767, FWF I5912).

References

[1] F. Argelaguet and C. Andujar. Visual feedback techniques for virtual pointing on stereoscopic displays. In *Proc. VRST*, pp. 163–170. ACM, New York, USA, 2009. doi: 10.1145/1643928.1643966 3

[2] F. Argelaguet and C. Andujar. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, May 2013. doi: 10.1016/j.cag.2012.12.003 3

[3] R. T. Azuma. The road to ubiquitous consumer augmented reality systems. *Human Behavior and Emerging Technologies*, 1(1):26–32, Feb. 2019. doi: 10.1002/hbe2.113 1

[4] N. Barbotin, J. Baumeister, A. Cunningham, T. Duval, O. Grisvard, and B. H. Thomas. Evaluating visual cues for future airborne surveillance using simulated augmented reality displays. In *Proc. VR*, pp. 213–221. IEEE, New York, USA, 2022. doi: 10.1109/VR51125.2022.00040 2

[5] A. Bellucci, P. Diaz, and I. Aedo. A see-through display for interactive museum showcases. In *Proc. ITS*, pp. 301–306. ACM, New York, USA, 2015. doi: 10.1145/2817721.2823497 1

[6] L. Besançon, M. Sereno, L. Yu, M. Ammi, and T. Isenberg. Hybrid touch/tangible spatial 3D data selection. *Computer Graphics Forum*, 38(3):553–567, Jul. 2019. doi: 10.1111/cgf.13710 3

[7] V. Biener, S. Kalamkar, J. J. Dudley, J. Hu, P. O. Kristensson, J. Müller, and J. Grubert. Working with XR in public: Effects on users and bystanders. In *Proc. VRW*, pp. 779–780. IEEE, New York, USA, 2024. doi: 10.1109/VRW62533.2024.00186 1

[8] O. Bimber, L. M. Encarnação, and D. Schmalstieg. The virtual showcase as a new platform for augmented reality digital storytelling. In *Proc. EGVE*, p. 87–95. ACM, New York, USA, 2003. doi: 10.1145/769953.769964 3

[9] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proc. I3D*, pp. 35–ff. ACM, New York, USA, 1997. doi: 10.1016/j.cag.2012.12.003 3

[10] G. Bruder, F. Steinicke, and W. Stuerzlinger. Touching the void revisited: Analyses of touch behavior on and above tabletop surfaces. In *Proc. INTERACT*, pp. 278–296. Springer, Berlin/Heidelberg, Germany, 2013. doi: 10.1007/978-3-642-40483-2_19 3

[11] F. Cabric, E. Dubois, P. Irani, and M. Serrano. TouchGlass: Raycasting from a glass surface to point at physical objects in public exhibits. In *Proc. INTERACT*, pp. 249–269. Springer, Cham, Switzerland, 2019. doi: 10.1007/978-3-030-29387-1_15 2, 3, 5, 9

[12] A. Cockburn, P. Dragicevic, L. Besançon, and C. Gutwin. Threats of a replication crisis in empirical computer science. *Commun. ACM*, 63(8):70–79, Aug. 2020. doi: 10.1145/3360311 6

[13] H. G. Debarba, J. G. Grandi, A. Maciel, L. Nedel, and R. Boulic. Disambiguation canvas: A precise selection technique for virtual environments. In *Proc. INTERACT*, pp. 388–405. Springer, Berlin/Heidelberg, Germany, 2013. doi: 10.1007/978-3-642-40477-1_24 3

[14] P. Dragicevic. Fair statistical communication in HCI. In J. Robertson and M. Kaptein, eds., *Modern Statistical Methods for HCI*, pp. 291–330. Springer, Cham, Switzerland, Mar. 2016. doi: 10.1007/978-3-319-26633-6_13 6

[15] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381, Jun. 1954. doi: 10.1037/h0055392 10

[16] A. Gallardo, C. Choy, J. Juneja, E. Bozkir, C. Cobb, L. Bauer, and L. Cranor. Speculative privacy concerns about AR glasses data collection. In *Proc. PoPETs*, vol. 4, pp. 416–435. Self-published, 2023. doi: 10.56553/popets-2023-0117 1

[17] S. Gehring, M. Löchtefeld, F. Daiber, M. Böhmer, and A. Krüger. Using intelligent natural user interfaces to support sales conversations. In *Proc. IUI*, pp. 97–100. ACM, New York, USA, 2012. doi: 10.1145/2166966.2166985 1, 3

[18] A. Giesler, D. Valkov, and K. Hinrichs. Void shadows: multi-touch interaction with stereoscopic objects on the tabletop. In *Proc. SUI*, p. 104–112. ACM, New York, USA, 2014. doi: 10.1145/2659766.2659779 3

[19] J. Gugenheimer, C. Mai, M. McGill, J. Williamson, F. Steinicke, and K. Perlin. Challenges using head-mounted displays in shared and social spaces. In *Proc. CHI*, pp. 1–8. ACM, New York, USA, 2019. doi: 10.1145/3290607.3299028 1

[20] L. G. Halsey. The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? *Biology Letters*, 15(5):20190174, May 2019. doi: 10.1098/rsbl.2019.0174 6

[21] O. Herbort and W. Kunde. Spatial (mis-)interpretation of pointing gestures to distal referents. *Journal of Experimental Psychology: Human Perception and Performance*, 42(1):78, Jan. 2016. doi: 10.1037/xhp0000126 3

[22] J. Hertel, S. Karaosmanoglu, S. Schmidt, J. Bräker, M. Semmann, and F. Steinicke. A taxonomy of interaction techniques for immersive augmented reality based on an iterative literature review. In *Proc. ISMAR*, pp. 431–440. IEEE, New York, USA, 2021. doi: 10.1109/ISMAR52148.2021.00060 1, 4

[23] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson. HoloDesk: direct 3d interactions with a situated see-through display. In *Proc. CHI*, pp. 2421–2430. ACM, New York, USA, 2012. doi: 10.1145/2207676.2208405 3

[24] J. D. Hincapié-Ramos, X. Guo, and P. Irani. Designing interactive transparent exhibition cases. In *Proc. PATCH*. Self-published, 2014. 1

[25] J. D. Hincapié-Ramos, S. Roscher, W. Büschel, U. Kister, R. Dachselt, and P. Irani. cAR: Contact augmented reality with transparent-display mobile devices. In *Proc. PerDis*, pp. 80–85. ACM, New York, USA, 2014. doi: 10.1145/2611009.2611014 3

[26] J. D. Hincapié-Ramos, S. Roscher, W. Büschel, U. Kister, R. Dachselt, and P. Irani. tPad: designing transparent-display mobile interactions. In *Proc. DIS*, pp. 161–170. ACM, New York, USA, 2014. doi: 10.1145/2598510.2598578 3

[27] M. Hirakawa, Y. Kojima, and A. Yoshitaka. Transparent interface: a seamless media space integrating the real and virtual worlds. In *Proc. HCC*, pp. 120–122. IEEE, New York, USA, 2003. doi: 10.1109/HCC.2003.1260214 2, 3

[28] C.-Y. Huang, L.-H. Wang, W.-L. Hsu, K.-P. Chang, F.-R. Lin, H.-Y. Chen, K.-T. Chen, and J.-C. Ho. A design of augmented-reality smart window using directive information fusion technology for exhibitions. In *Proc. HCII*, pp. 447–457. Springer, Cham, Switzerland, 2019. doi: 10.1007/978-3-030-22643-5_35 1

[29] H. Ishii and M. Kobayashi. ClearBoard: A seamless medium for shared drawing and conversation with eye contact. In *Proc. CHI*, pp. 525–532. ACM, New York, USA, 1992. doi: 10.1145/142750.142977 3

[30] A. W. Ismail, N. A. A. Halim, R. Talib, and A. J. Sihes. Target selection method on the occluded and distant object in handheld augmented reality. *Indonesian Journal of Electrical Engineering and Computer Science*, 29(2):1157–1165, Feb. 2023. doi: /10.11591/ijeecs.v29.i2.pp1157-1165 3

[31] M. R. Jakobsen, Y. Jansen, S. Boring, and K. Hornbæk. Should i stay or should i go? Selecting between touch and mid-air gestures for large-display interaction. In *Proc. INTERACT*, pp. 455–473. Springer, Cham, Switzerland, 2015. doi: 10.1007/978-3-319-22698-9_31 1

[32] G. Kalton and H. Schuman. The effect of the question on survey responses: A review. *Journal of the Royal Statistical Society. Series A (General)*, 145(1):42–57, Jan. 1982. doi: 10.2307/2981421 9

[33] M. Koelle, M. Kranz, and A. Möller. Don't look at me that way! Understanding user attitudes towards data glasses usage. In *Proc. MobileHCI*, pp. 362–372. ACM, New York, USA, 2015. doi: 10.1145/2785830.2785842 1

[34] R. Kopper, F. Bacim, and D. A. Bowman. Rapid and accurate 3D selection by progressive refinement. In *Proc. 3DUI*, pp. 67–74. IEEE, New York, USA, 2011. doi: 10.1109/3DUI.2011.5759219 3

[35] K. Krug, W. Büschel, K. Klamka, and R. Dachselt. CleAR sight: Exploring the potential of interacting with transparent tablets in augmented reality. In *Proc. ISMAR*, pp. 196–205. IEEE, New York, USA, 2022. 3

[36] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proc. CHI*, pp. 1–14. ACM, New York, USA, 2018. doi: 10.1145/3173574.3173655 3

[37] J. Lee and C. Boulanger. Direct, spatial, and dexterous interaction with see-through 3D desktop. In *Proc. SIGGRAPH*. ACM, New York, USA, 2012. doi: 10.1145/2342896.2342980 3

[38] J. H. Lee and S.-H. Bae. Binocular cursor: enabling selection on transparent displays troubled by binocular parallax. In *Proc. CHI*, p. 3169–3172. ACM, New York, USA, 2013. doi: 10.1145/2470654.2466433 3

[39] J. Li, E. Sharlin, S. Greenberg, and M. Rounding. Designing the car iWindow: exploring interaction through vehicle side windows. In *Proc. CHI*, p. 1665–1670. ACM, New York, USA, 2013. doi: 10.1145/2468356.2468654 1

[40] C. J. Lin, S.-H. Ho, and Y.-J. Chen. An investigation of pointing postures in a 3D stereoscopic environment. *Applied Ergonomics*, 48:154–163, May 2015. doi: 10.1016/j.apergo.2014.12.001 3

[41] Y. Lu, C. Yu, and Y. Shi. Investigating bubble mechanism for ray-casting to improve 3D target acquisition in virtual reality. In *Proc. VR*, pp. 35–43. IEEE, New York, USA, 2020. doi: 10.1109/VR46266.2020.00021 3

[42] F. L. Luro and V. Sundstedt. A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. In *Proc. ETRA*, pp. 1–9. ACM, New York, USA, 2019. doi: 10.1145/3317956.3318153 3

[43] M. N. Lystbæk, P. Rosenberg, K. Pfeuffer, J. E. Grønbæk, and H. Gellersen. Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality. *Proc. ACM Hum.-Comput. Interact.*, 6(ETRA):1–18, May 2022. doi: 10.1145/3530886 4

[44] B. A. Myers, R. Bhatnagar, J. Nichols, C. H. Peck, D. Kong, R. Miller, and A. C. Long. Interacting at a distance: measuring the performance of laser pointers and other devices. In *Proc. CHI*, pp. 33–40. ACM, New York, USA, 2002. doi: 10.1145/503376.503383 1

[45] K. Pietroszek and E. Lank. Clicking blindly: using spatial correspondence to select targets in multi-device environments. In *Proc. MobileHCI*, pp. 331–334. ACM, New York, USA, 2012. doi: 10.1145/2371574.2371625 4

[46] A. Plopski, T. Hirzle, N. Norouzi, L. Qian, G. Bruder, and T. Langlotz. The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Computing Surveys*, 55(3):1–39, Mar. 2022. doi: 10.1145/3491207 3

[47] D. Schmalstieg, L. M. Encarnação, and Z. Szalavári. Using transparent props for interaction with the virtual table. In *Proc. I3D*, pp. 147–153. ACM, New York, USA, 1999. doi: 10.1145/300523.300542 3

[48] C. Schmandt. Spatial input/display correspondence in a stereoscopic computer graphic work station. *SIGGRAPH Computer Graphics*, 17(3):253–261, Jul. 1983. 3

[49] R. Shi, Y. Wei, X. Qin, P. Hui, and H.-N. Liang. Exploring gaze-assisted and hand-based region selection in augmented reality. *Proc. ACM Hum.-Comput. Interact.*, 7(ETRA):1–19, May 2023. doi: 10.1145/3591129 3

[50] R. Shi, J. Zhang, Y. Yue, L. Yu, and H.-N. Liang. Exploration of bare-hand mid-air pointing selection techniques for dense virtual reality environments. In *Proc. CHI*, pp. 1–7. ACM, New York, USA, 2023. doi: 10.1145/3591129 3

[51] D. Valkov, F. Steinicke, G. Bruder, and K. Hinrichs. 2D touching of 3D stereoscopic objects. In *Proc. CHI*, pp. 1353–1362. ACM, New York, USA, 2011. doi: 10.1145/1978942.1979142 3

[52] L. Vanacken, T. Grossman, and K. Coninx. Exploring the effects of environment density and target visibility on object selection in 3D virtual environments. In *Proc. 3DUI*. IEEE, New York, USA, 2007. doi: 10.1109/3DUI.2007.340783 3

[53] G. R. VandenBos. *Publication manual of the American Psychological Association*. APA, Washington, USA, 7 ed., 2020. doi: 10.1037/0000165-000 6

[54] R. Vertegaal. A Fitts Law comparison of eye tracking and manual input in the selection of visual targets. In *Proc. ICMI*, pp. 241–248. ACM, New York, USA, 2008. doi: 10.1145/1452392.1452443 3

[55] C. Ware and K. Lowther. Selection using a one-eyed cursor in a fish tank VR environment. *ACM Trans. Comput.-Hum. Interact.*, 4(4):309–322, Dec. 1997. doi: 10.1145/267135.267136 3

[56] Y. Wei, R. Shi, D. Yu, Y. Wang, Y. Li, L. Yu, and H.-N. Liang. Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions. In *Proc. CHI*. ACM, New York, USA, 2023. doi: 10.1145/3544548.3581042 3

[57] K. Yoshimura and T. Ogawa. Binocular interface: Interaction techniques considering binocular parallax for a large display. In *Proc. VR*, pp. 315–316. IEEE, New York, USA, 2015. doi: 10.1109/VR.2015.7223422 3

[58] Q. Zhou, G. Fitzmaurice, and F. Anderson. In-depth mouse: Integrating desktop mouse into virtual reality. In *Proc. CHI*, pp. 1–17. ACM, New York, USA, 2022. doi: 10.1145/3491102.3501884 3