A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces

Uta Wagner
uta.wagner@cs.au.dk
Aarhus University
Denmark

Mathias N. Lystbæk
mathiasl@cs.au.dk
Aarhus University
Denmark

Pavel Manakhov
pmanakhov@cs.au.dk
Aarhus University
Denmark

Jens Emil Grønbæk
jensemil@cs.au.dk
Aarhus University
Denmark

Ken Pfeuffer
ken@cs.au.dk
Aarhus University
Denmark

Hans Gellersen
h.gellersen@lancaster.ac.uk
Lancaster University
United Kingdom
Aarhus University
Denmark
hwg@cs.au.dk

ABSTRACT

Gaze-Hand Alignment has recently been proposed for multimodal selection in 3D. The technique takes advantage of gaze for target pre-selection, as it naturally precedes manual input. Selection is then completed when manual input aligns with gaze on the target, without need for an additional click method. In this work we evaluate two alignment techniques, Gaze&Finger and Gaze&Handray, combining gaze with image plane pointing versus raycasting, in comparison with hands-only baselines and Gaze&Pinch as established multimodal technique. We used Fitts' Law study design with targets presented at different depths in the visual scene, to assess effect of parallax on performance. The alignment techniques outperformed their respective hands-only baselines. Gaze&Finger is efficient when targets are close to the image plane but less performant with increasing target depth due to parallax.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Pointing**; **Interaction design theory, concepts and paradigms**.

KEYWORDS

eye-tracking, gaze interaction, pointing, mid-air gestures, augmented reality, menu selection

ACM Reference Format:
Uta Wagner, Mathias N. Lystbæk, Pavel Manakhov, Jens Emil Grønbæk, Ken Pfeuffer, and Hans Gellersen. 2023. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3581423

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581423

1 INTRODUCTION

Object selection at-a-distance is a fundamental task for interaction with 3D environments in Augmented (AR), Virtual (VR) and Mixed Reality (MR). Researchers have long investigated the design of remote pointing techniques to extend users' manual reach across the vast space afforded by virtual environments [26]. This is typically solved by handing a tracked controller to the user, but it is highly challenging with a freehand gestural interface. Hand gestures are appropriate for close-up, reachable user interfaces that afford direct grasping and manipulation of objects – but they do not intuitively scale into distance. We therefore investigate how gaze, as input that naturally extends from near to far, can enhance the user's manual pointing.

As such, recent work has proposed Gaze-Hand Alignment (GHA) as a novel mechanism for selection of objects by spatial alignment of the gaze ray with a manual pointer on the target [21]. The technique can be viewed as a gaze pointing technique where moving the manual pointer into the line of sight to the target acts as a selection trigger. Conversely, it can also be viewed as a manual pointing technique where input is only triggered on targets that the user is simultaneously looking at. GHA can be implemented as an image plane *Gaze&Finger* technique for perspective-based pointing with a finger at a target, or alternatively as *Gaze&Handray* technique with a ray cast from the hand to align a projected cursor on the target. Key to the technique, in either case, is the implicit coordination of eye and hand, as gaze naturally guides manual pointing by looking ahead to objects intended for selection. Even though two modalities are involved, they are integral in a single action of pointing.

Visual alignment of input from gaze and hand is compelling in the way it leverages natural eye-hand coordination, but it is also prone to parallax issues. How finger or cursor lines up with a target will appear different from left versus right eye. Normally, one eye will dominate vision subconsciously, but this is not always the same eye as switches occur dynamically [2]. Gaze input techniques, therefore, rely on an approximation of the line of sight by a vector projected from a central position between the eyes. The discrepancy between the projected gaze vector and the actual line of sight results in an apparent shift of the finger or cursor relative to the target.

The parallax effect becomes more pronounced when one object is close to the eyes, for example, the finger in image plane pointing, while the other is at a greater distance from the eyes.

In prior work, GHA was studied for selection in contextual menus in augmented reality [21], where parallax was mitigated by initially warping the menu closer for finger input, or spawning a cursor at the menu's depth in the scene. In this work, we study GHA for the general case of object selection in 3D without prior adjustment of input or target depth. We evaluate Gaze&Finger where input occurs at a close distance from the eyes, and Gaze&Handray where default cursor feedback is at infinite depth. We compare the techniques with three existing baselines: Handray with pinch gesture to trigger selection[1]; HeadCrusher for image plane pointing and pinch selection of targets in the line of sight [34]; and Gaze&Pinch using gaze to point and pinch to select [32]. All five techniques are compared in a Fitts' Law study design with a circular layout of targets, presented at different levels of depth in the scene to assess the effect of differences in depth between input and target.

Our results contribute the following findings and insights:

- Techniques ordered from high to low throughput are: Gaze&Handray (2.09), Gaze&Pinch (2.06), Gaze&Finger (1.86), Handray (1.39), HeadCrusher (1.32).
- Target depth (and associated parallax) has a detrimental effect on performance of image plane techniques, but less pronounced with Gaze&Finger than with HeadCrusher.
- HeadCrusher was perceived least favourable technique and resulted in the lowest performance.
- Gaze&Handray resulted in highest performance and most preferences, indicating a viable approach next to Gaze&Pinch.
- We find that all techniques except for Handray are affected by target amplitude.
- Overall all three gaze-assisted techniques outperformed manual pointing techniques for our tested target size of 3°.

2 RELATED WORK

2.1 Interaction Techniques in 3D

The two primary interaction metaphors for object selection in a 3D world are the *virtual hand* and the *virtual pointer* [38]. The *virtual hand* affords direct manipulation, i.e., to directly grasp and manipulate objects with the hands – a natural form of communication and easy to learn. *Virtual pointer* adds the capability to interact with objects that are out of reach, through pointing a ray into the distant scene and a confirmatory action. Virtual pointers are typically implemented with controllers (e.g., [1]), but more and more devices support interaction via freehand gestural UIs as controllers may not be available to or preferred by users. This poses a challenge as gestures do not intuitively scale to distant selection.

A currently deployed technique on the HoloLens 2 and Meta Quest is Handray, which integrates two different types of gestures of the nature dimension [35], *metaphorical* pointing with a ray that projects from the user's palm toward the scene and *symbolic* pinch to confirm selection. This addresses gestural pointing at-a-distance, but also comes with constraints. It is susceptible to the *Heisenberg* problem [49]: a pinch gesture will inadvertently affect the pointing ray, making it difficult to accomplish the selection accurately. As well, gestural interfaces are prone to physical exertion (*Gorilla Arm* problem [15]), making it difficult to interact over a longer period of time.

Image Plane techniques introduced by Pierce et al. [34] are a type of hybrid interaction inspired by 2D desktop interaction. They consider the 3D world as a 2D image plane from the user's perspective. Through a virtual hand in the user's line of sight, users can frame far-away objects. One prominent technique is the HeadCrusher, which uses pinch to confirm a selection of the object in the image plane. As the authors discuss, an important challenge of image plane techniques is the parallax effect, which happens when users perceive the close-by hand and distant object at different depths. The technique is used in our study as a baseline to examine the impact of parallax on remote target selection.

2.2 Gaze-based Interaction in 3D

Eye-tracking sensors are becoming a standard feature in head-mounted displays. Gaze input has been shown to be fast and efficient for computer selection tasks [43, 46], and as a suitable modality for a synergistic combination with manual input [51]. There is increasing interest in exploring how gaze can advance the interactive capabilities in XR [36], and in particular to assist manual interaction [17, 21, 32, 50]. However, studies that compare gaze to hands are inconclusive, as some works show superiority of gaze [21, 43, 46], some works find no significant differences in user performance [19, 20, 50], others show the superiority of manual pointing [9, 12, 13, 33]. It comes down to the specific interaction technique and task design for gaze, hand, or combination thereof. E.g., a dwell time can be faster than mouse clicking [43], but dwell time is naturally more susceptible to accidental activations (the *Midas Touch* problem [14]). Yu et al. have compared multimodal gaze and controller to manual-only techniques for object manipulations, finding no performance benefits [50], which can be attributed to the maturity of the controller for remote pointing.

Researchers proposed gaze input to assist hand gestures. As pointed out by Chatterjee et al. [7], gaze and gesture modalities are highly complementary to one another since the eye movements are quick, coarse, and effortless, while the hands are excellent at continuous manipulation and expressive gesturing. A popular variant is to couple two-finger pinch gestures with the gaze-selected target [17, 18, 28, 32, 39] (Gaze&Pinch), as pinch gestures are inherently efficient as input medium [41]. Pfeuffer et al. [32] explored this principle in VR, pointing out that it has distinct advantages to gestures (interaction over distance) and controllers (no need to carry an extra physical device). In a study of gaze pointing in VR, Mutasim et al. found that the use of pinch confirmation is a reasonable alternative to button clicks [28]. As well, Lystbæk et al. [21] showed that it is one of the fastest methods for current AR devices to perform selections. As such, we include it as a multi-modal baseline.

2.3 Spatial Alignment-based Techniques

Early works that could be considered as predecessors are Bier et al.'s Toolglass and Magic Lenses [3] – e.g., to apply a colour theme to a graphical element in the canvas, a two-handed UI is proposed

[1]Point and Commit, Microsoft, https://docs.microsoft.com/en-us/windows/mixed-reality/design/point-and-commit (accessed 12/09/22)

where one hand moves a tool palette over objects, and mouse in the other hand clicks "through" the tool to apply it to the object. The spatial alignment of two layers is a key enabling technique. In 2019, Mardanbegi et al. adopted this concept to 3D and eye-tracking interfaces in the EyeSeeThrough [25] concept. Here, users can spatially align the virtual tool palette in 3D space by hand or head movement, and gazing through a tool toward an object applies the effect. Related is also work that used alignment through eye gaze and head-based raypointing for interaction with 3D UIs, such as Eye&Head Convergence [44] and Look & Cross [45]. A common concept is that gaze provides a *"pre-selection"* – a prerequisite for a selection, that the eyes are on the same target in space as the supplementary ray.

Lystbæk et al. [21] proposed the GHA concept, which specifies interaction enabled through the spatial alignment of both gaze and hand-controlled ray for a command invocation. Direct gestures (Gaze&Finger) and indirect cursor (Gaze&Hand) were evaluated as hand-controlled in the context of remote menu control [21] and close-by text entry [20]. In the menu selection task, users were faster with the techniques than the standard pointing technique in AR (Handray). As such, it raises the question of whether GHA can become useful as a standard interaction technique for remote selection in 3D environments. To assess their suitability, we must first provide a better understanding of three points: (1) The prior study focused on specific menu designs– but what about the speed-accuracy trade-off in a general object selection task? (2) Gaze&Finger introduces potential parallax problems which were circumvented by automatic warping [27], but this is limited to object-based worlds – so, what is the effect of target depth on performance with the general-case Gaze&Finger technique? (3) The Gaze&Hand technique required a two-step approach to place a cursor at the target depth, which however is infeasible in general UIs as this information is not known apriori – How can we adapt the technique that is efficient for general selections?

3 INTERACTION TECHNIQUE DESIGN

This paper investigates the performance of two GHA techniques *Gaze&Finger* and *Gaze&Handray* (Figure 1) in comparison with three baselines for the selection of targets at different distances from the user in the 3D environment. The techniques have differences in their input structure, e.g., gaze-based vs. manual pointing, but partially share properties, such as the use of perspective-based pointing. All techniques involve two steps, namely pointing and selection. The main conceptual components of GHA are 1) the use of the user's gaze direction and an additional hand-controlled manual pointer and 2) the alignment of these two modalities.

3.1 Gaze&Finger

Lystbæk et al. [21]'s Gaze&Finger employs the user's gaze in combination with a manual pointer, starting from between the user's eyes, through the user's index finger, to the scene. The technique appears to the user as if directly touching a target through the image plane. As shown in Figure 1a, the user's gaze is used for target pre-selection while the selection is confirmed by the alignment of the user's index finger with their line of sight. Meaning, the user first looks at the target of interest and then moves their index finger to the same target to confirm selection.

3.2 Gaze&Handray

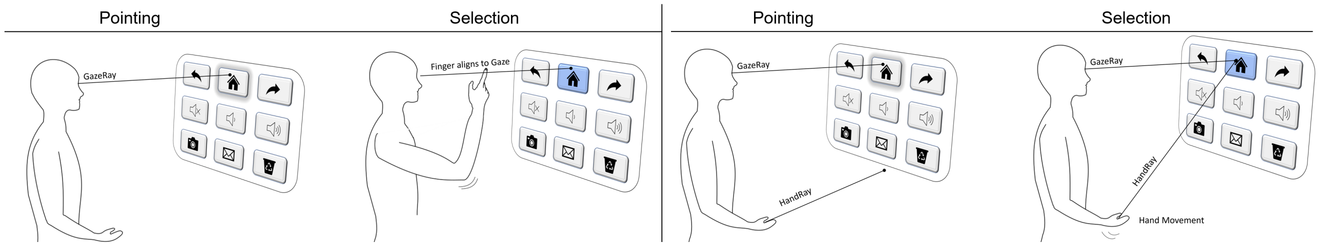
Gaze&Handray is a new interaction technique based on GHA. The concept is similar to Gaze&Hand of [21], but has distinct interaction properties. Gaze&Hand used relative mapping where input and target space are different. This required a method to instantiate a cursor to establish the mapping. Thus, a prior step was introduced where users dwell at an area to instance a cursor. However, their study showed that it did not work well as performance was lower than other multi-modal gaze-based interaction techniques.

Gaze&Handray, in contrast, uses a hand-ray projected from the user's palm for the alignment with gaze. Thus, it is based on the virtual pointer metaphor using absolute raypointing. As shown in 1b, the user's gaze is used for target pre-selection, while the selection is confirmed by the alignment of the user's hand-ray pointer in their line of sight. I.e., the user (1) looks at the target of interest and (2) moves the hand-ray cursor to the same target to confirm selection. The hand-ray used is the same as that of the Handray technique, provided by MRTK, including a cursor at the end of the hand-ray. Users can control the direction of the hand-ray by moving their hand in relation to their body, meaning that the rotation of the user's hand does not influence the direction of the ray.

3.3 Technical details

An important difference in the implementation of the GHA techniques for this paper, as compared to [20, 21], is how alignment of the gaze and manual pointers is determined. In the prior work, the alignment was determined when both pointers hit the same object, which is a simple and robust procedure. This assumes an object-based UI, where it is clear where the targets are. This precludes the possibility to point at any pixel on the screen. For this reason, we enhance their selection algorithm by making it object-independent. We determine alignment to occur when the end-points of the gaze and hand-controlled pointer rays on the target depth's plane are within a threshold. For Gaze&Finger, the vector end-points are where the gaze ray and ray from the head through the index finger intersect on the depth of the target. For Gaze&Handray, the end-points are where the same gaze point and where the hand-ray cursor intersects with the depth of the target. Through pilot tests, we found that $1.5°$ in visual angle can work well. The selection will be triggered at the gaze-selected target.

Furthermore, we apply fixation filtering to filter out quick saccadic movements to limit the likelihood of aligning the two pointers while the eyes are saccadic. We do this by 1) recording the last gaze from $150ms$ ago [4], 2) checking the angle between the raw current gaze and the stored gaze 3) if the angle is less than $4°$, then update the fixation gaze immediately, otherwise disable the fixation gaze pointer as a saccade has occurred. The reason for the $150ms$ is to reflect human visual processing time, while the $4°$ of angle threshold is to allow for smaller eye movements.



(a) Gaze&Finger. **(b) Gaze&Handray.**

Figure 1: Two techniques based on the GHA principle.

	Handray	***HeadCrusher***	***Gaze&Pinch***	***Gaze&Finger***	***Gaze&Handray***
Pointing	Hand-ray casting	Perspective (index finger & thumb)	Eye-gaze ray	Eye-gaze ray	Eye-gaze ray
Selection trigger	Pinch gesture	Pinch gesture	Pinch gesture	Gaze-Hand Alignment (index finger)	Gaze-Hand Alignment (Hand-ray casting)
Modalities	Hand	Hand	Gaze and hand	Gaze and hand	Gaze and hand
Interaction metaphor	Virtual pointer	Virtual hand	Virtual hand	Virtual hand	Virtual pointer
Gesture Type	Motion (rotate hand) + Symbolic (pinch)	Motion (point) + Symbolic (pinch)	Symbolic (pinch)	Motion (point finger)	Motion (rotate hand)
Selection space	3D	Image plane	3D	Image plane	3D
References (same)	HoloLens 2	Pierce et al [34]	Gaze + Pinch [32], Mutasim [28], HoloLens 2	Gaze&Finger [20, 21]	
References (similar)	Meta quest 2, Go-go [37], HOMER [5], Hand Ray-casting [26]		Hales et al. [11], Velloso et al. [47], GG Interaction [40], Gaze+Gesture [7], Gaze-touch [30], Zhang et al. [52]	Air-tap, Sticky Finger [34], EyePointing [42], Look & Cross [45]	Gaze&Hand [20, 21], Look & Cross [45], EyePointing [42]

Table 1: Summary of similarities and differences between the techniques we study.

3.4 Advantages and limitations

A summary of shared and distinct interaction properties of the two techniques is provided in Table 1, where the techniques are set in context to related techniques from the literature. What distinguishes the two GHA techniques is it leverage on eye-hand coordination patterns. Most of the time, the eyes are on the target before users move the manual ray. Thus, this reduces to a technique where users simply point at the target they visually focus on. In contrast to the Handray, the additional step of confirmation with a pinch gesture is eliminated. Thus, there is a potential for a very fast selection. However, with the alignment angle change, there is potential for accidental errors (Midas Touch problem [14]) when the user's gaze and manual pointer happen to be in angle alignment away from the target. Additionally, a limitation with Gaze&Finger is parallax, as the visual occlusion of a target means that the hand and distant object are often at different depths, with further depths leading to more pronounced parallax issues. These lead us to conduct a comparative user study.

4 EVALUATION

We designed a Fitts' Law based user study where participants selected targets at three different depths and amplitudes with five different techniques, to address following research questions (RQ):

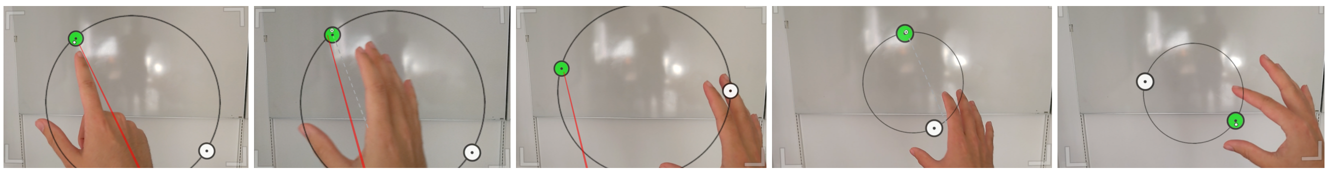
RQ1: How does varying target depth affect the user's performance?. Gaze&Finger has been shown as a promising technique in a prior study, but as an image plane technique, it is principally susceptible to the parallax problem. However, the effect on performance is not clear, e.g. in comparison with image plane techniques such as HeadCrusher that do not rely on gaze alignment.

RQ2: How does the new Gaze&Handray technique fare compare to other gaze and manual techniques? Gaze&Handray provides a new input method that uses gaze and raypointing. Prior research examined a related technique (Gaze&Hand), suggesting a trade-off of reduced physical effort but diminished performance [20, 21].

RQ3: How does varying target amplitude affect the user's performance? For manual approaches, the movement time often rises with amplitude, whereas this is different for gaze pointing. We examine user performance over amplitudes for multimodal strategies that leverage both pointing metaphors, as the outcome is still uncertain.

RQ4: How do multi-modal gaze-assisted techniques compare to manual techniques regarding selection efficiency? As discussed in the Related Work, recent work indicates gaze and gesture-based techniques are superior to gesture-based raypointing [21]. However, as the more general case of manual vs. gaze pointing led to mixed results across studies [17, 31, 43, 51], we provide important, standardised empirical data.

(a) Gaze&Finger. (b) Gaze&Handray. (c) Gaze&Pinch. (d) Handray. (e) HeadCrusher.



(f) Gaze&Finger. (g) Gaze&Handray. (h) Gaze&Pinch. (i) Handray. (j) HeadCrusher.

Figure 2: Examples of the techniques' operation in the user study from outside (a-e) and user perspective (f-j). A red line (f, g, h) indicates the user's gaze.

4.1 Participants

We recruited 20 participants (10 female, 10 male) via email and word of mouth inside and outside of the local university. The users' backgrounds were mixed, with people of various (non-) technical expertise. Participants' age ranged from 20 and 40 ($M = 26,85, SD = 4.95$). All were right-handed, 8 wore glasses, and 5 wore contact lenses. On a scale between 1 and 5, participants rated themselves as having little to moderate experience with AR/VR/XR ($M = 2.80, SD = 1.50$), 3D hand gestures ($M = 2.55, SD = 1.49$) and eye-gaze interaction ($M = 2.20, SD = 1.24$).

4.2 Baselines

We compare against three baselines techniques. Example views of all five techniques in our study are shown in Figure 2.

4.2.1 Gaze&Pinch. Gaze&Pinch is an interaction technique that utilises the user's gaze for pointing, and a pinch close-open gesture as confirmation [7, 32, 47]. The Gaze&Pinch technique uses the MRTK solution with standard parameters. We apply the same fixation filtering as for the GHA techniques to filter out jittery motion.

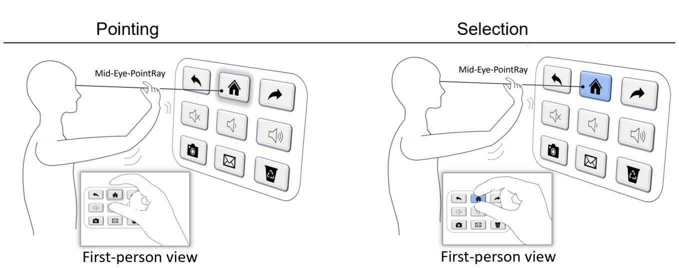


Figure 3: An illustration of the HeadCrusher technique [34] as one out of three baselines of our study.

4.2.2 Handray. Handray is an interaction technique that utilises a hand-ray projected from the user's palm and confirms by a pinch close&open gesture. We use the implementation provided by MRTK, with a cursor at the end of the hand-ray[2].

4.2.3 HeadCrusher. HeadCrusher is an interaction technique that enables pointing by a ray cast from between the user's eyes, through the mid-point of the user's index finger and thumb (Figure 3). Then, the user performs a pinch gesture (close&open) to confirm the selection of the target, as if "crushing" the remote target between the user's fingers. We implemented the technique as in Pierce et al. [34]. We limit the pinch selection point to be at most $2cm$ from the user's thumb as it improved pinch gesture detection. We added a cursor at the pinch point at the depth of the user's fingers, to provide feedback on where the intersection point was. The size of the cursor was 0.3° like MRTK's Handray cursor. The pinch gesture detection is provided by MRTK and uses its standard parameters.

4.3 Apparatus and Experimental Setup

The techniques listed were implemented using the Mixed Reality Toolkit (MRTK) in Unity on the Microsoft HoloLens 2 (43° x 29° FOV), which supports hand and eye tracking (1.5° viewing angle accuracy). During the study, the participants stood 3 meters away from a clear wall in a large, quiet room, cf. Figure 2. The eye tracker was calibrated for each study participant at the start. The study tool showed the four white corners of the FOV to the user. The participants were shown a grey circular area, on which a virtual 2D target appears after each selection. Each target was arranged on a plane. The click-up event applies to the GHA techniques, and the click-down event applies to the pinch techniques. Furthermore, to aid gaze fixation, all targets were designed with a black dot in the centre.

[2]Point-and-commit: https://docs.microsoft.com/en-us/windows/mixed-reality/design/point-and-commit#hand-rays - accessed 13-09-2022

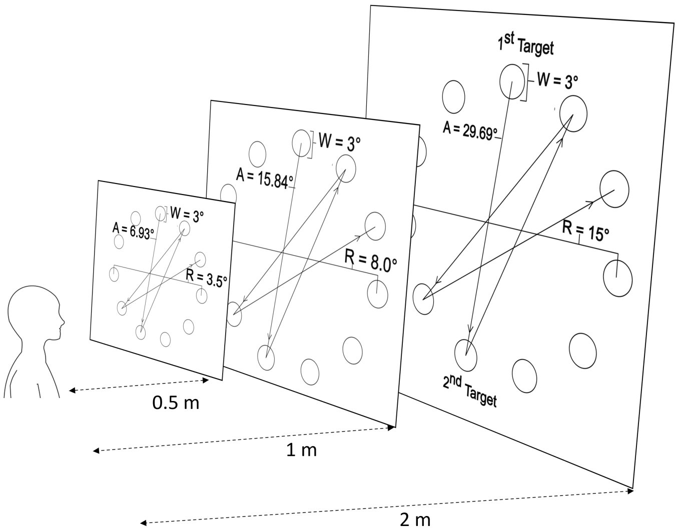


Figure 4: Fitts' Law task design with three target amplitudes (6.93°, 15.84°, 29.69°), and three depth planes (0.5m, 1m, 2m).

4.4 Task

The experimental task is illustrated in Figure 4. The task is a point-and-select task, where the system displays a single target that should be selected. The target was a flat, user-oriented 2D circular shape. We fixed the target size to a value of 3°, as we are primarily interested in depth and amplitude factors. It also minimises the potential effect of eye-tracking inaccuracies in the data collection. After each selection, a new target appeared on the opposite side at the same depth. The target was white by default and received visual feedback when pointing (yellow), successful selection (green), and error selection (red).

4.5 Procedure

Participants were firstly briefed on the study and filled out consent and demographic forms. The system-internal eye tracker calibration with the Microsoft HoloLens was performed at the beginning so that the eye tracking was correctly set for the study participant. The operation of each interaction technique was explained before each run. We used a 1-minute training with six targets so that the users can familiarize themselves with the technique on the one hand, and to avoid learning effects on the other hand. Then, participants performed the tasks with the instruction to be as fast and accurate as possible. After each technique, participants filled out a questionnaire and therefore had a rest of 2-3 minutes between techniques. After all techniques were tested, the study participants finally ranked all five techniques. The study duration was around 50 minutes.

4.6 Study Design

The study used a within-subject design with three independent variables, which order of conditions is counterbalanced. The study was carried out in a calm room while the participant was standing freely. Subjects were free to position their arms close to their body to avoid arm fatigue. We employed a 5×3×3 design of the following variables:

- *Techniques*: We compare the following interaction techniques: Gaze&Finger, Gaze&Handray, Gaze&Pinch, Handray and HeadCrusher.
- *Depth*: How far the target is away from the user (0.5m, 1m, 2m). We chose 0.5m as a typical depth for close-ranged UIs, and 2m as a far depth to induce parallax (farther depths do not visibly increase parallax issues).
- *Amplitude*: We test the following amplitudes: 6.93° (3.5° circle radius), 15.84° (8° circle radius), and 29.69° (15° circle radius). The small distance was chosen as the smallest distance where 3° targets do not overlap, and the largest is chosen as a condition at the boundary of the device's FOV.

Each block involved 11 target selections, where the first target was used for initial target selection and therefore used 10 targets in the data analysis. In sum, this results in:

10 *targets* × 5 *techniques* × 3 *depths* × 3 *amplitudes* = 450 data points per participant.

4.7 Evaluation Metrics

- **Task Completion Time (TCT)**: The time it takes to complete one task successfully. To determine the TCT of each study participant, we captured the time of appearance of a target and the time of successful selection by the user. Note that TCT is also known as Movement Time (MT) in literature and equations.
- **Throughput (TP)**: Throughput is used in the context of Fitts' Law as a dependent variable for point-select techniques to test human performance under the influence of external factors. We used the "Shannon formulation" equation [22] according to the ISO 9241-9 standard:

$$TP = \frac{ID_e}{MT}$$

Where ID_e is the effective index of difficulty, and MT is the movement time determined as follows [23]:

$$MT = a + b \cdot ID_e$$

Where a and b are regression coefficients, and ID_e is computed by:

$$ID_e = \log_2 \left(\frac{A_e}{W_e + 1} \right),$$

using the effective width W_e and the effective amplitude A_e. A_e is the actual physical movement distance from the selection point of the previous target to the selection point of the current target from each user for each trial and is computed as follows [24]:

$$A_e = a + dx$$

Where a is the distance from the center of the prior target to the center of the current target and dx the adjacent of a right-angled triangle (c.f. [24], Fig. 17.8).

- **Effective Width (W_e)**: can be derived from the distribution of "hits". By using the effective target width, a reduced variation in TP can be seen due to the speed-accuracy trade-off. The reduction in the variability of TP is thus influenced by the increase (or decrease) in TCT as well as the effective ID. For the calculation, we use the standard-deviation method [24]. For each trial we

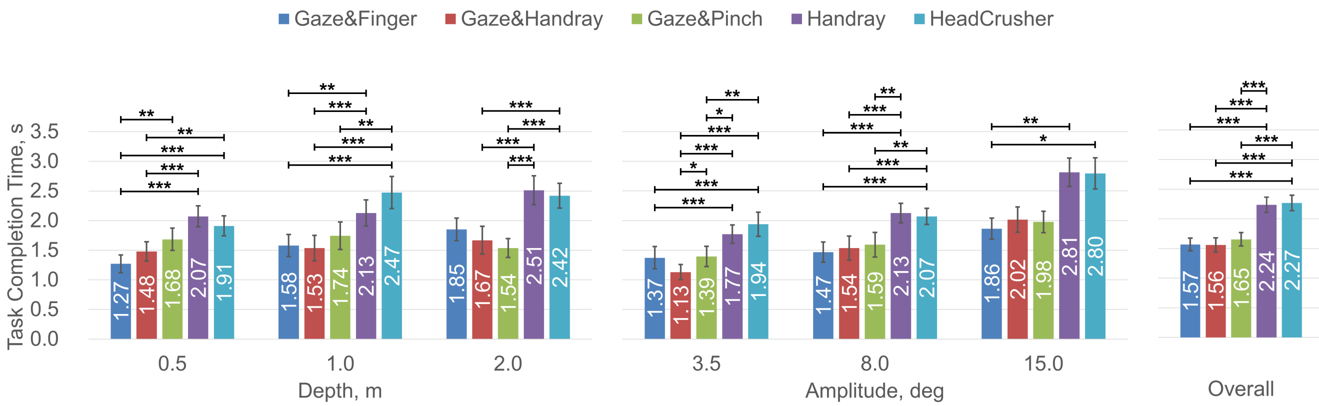


Figure 5: Mean Task Completion Time shown per depth per technique, per amplitude per technique, and per technique overall. Error bars indicate 95% confidence intervals. Statistical significance shown as * for p < .05, ** for p < .01, and * for p < .001. Significant differences across depths are omitted for clarity. The same is true for amplitudes.**

recorded the current selection coordinate point (x, y) and used SD which is a one-dimensional distribution of the selection points.

$$W_e = 4.133 \cdot SD_x$$

Where SD_x is the standard deviation of dx, and represents the measure of accuracy and 4.133 a coefficient.

- **Error Rate (ER)**: The error quotient results from all trials in which there was no successful selection on the displayed target object (i.e., a click outside the target).
- **Hand Movement (HM)**: Hand movement is measured to find out how much the users had to move their hand on average to select the tasks, measured as the accumulated difference in palm position between frames. We used hand movement as a dependent variable because they can be correlated with physical fatigue [50], and it is of interest to understand whether gaze-based techniques can reduce physical effort compared to full manual input.
- **Usability Questionnaire**: We used the NASA-TLX questionnaire (Task Load Index) [8] to measure the subjective task load of the study participants after each interaction technique was performed. By default, the six questions on mental, physical and time demands, perceived success, effort and level of frustration are used, which the user must rate on a 7-point scale Chan et al. [6] from very low to very high. Furthermore, we asked another open-ended question about eye fatigue and the users were able to comment on their user experience with the respective technique. In the end, they ranked the techniques based on their preference.

4.8 Data Analysis

We first have tested the quantitative variables for normality, and applied data transformations (reciprocal transformation or Aligned Rank Transform (ART) [10, 48]) for non-normally distributed factors where appropriate. We conducted a three-way (Technique × Depth × Amplitude) repeated measures ANOVA tests for the statistical assessment of the quantitative data (with Greenhouse-Geisser corrections if sphericity was violated), followed by post-hoc pairwise comparisons with Holm-Bonferroni corrections. In case of the non-parametric data of the questionnaires, we used a Friedman test with post-hoc Conover tests (Holm-Bonferroni corrected).

We applied data filters to reduce tracking-induced errors. First, we found that at times, a double-click was performed instead of a single click. This led to immediately firing a selection after the prior target, without the user having pointed at a target. As the cursor was still on the prior target, it was an obvious error. We, therefore, filtered all trials where the pointer was still on the previous target at the moment of selection. With this, 2.3% of trials were excluded, i.e., 1.6% for HeadCrusher, 0.8% for Gaze&Pinch, 1.7% for Gaze&Handray, 6.4% for Gaze&Finger, 1.2% for Handray. Second, infrequent hand-tracking errors led to outliers in the dependent variable Hand Movement. Here we apply a filter across techniques, where a trial time exceeds the mean+$3 \times SD$. I.e., overall 1.4% (1.3% for HeadCrusher, 0.9% for Gaze&Pinch, 1.4% for Gaze&Handray, 2.9% for Gaze&Finger, 0.4% for Handray).

5 RESULTS

The TCT, TP, W_e, ER, HM, and task load results are shown in Figures 5, 6, 7, 8, 9, and 11 respectively. For brevity, we report only statistically significant effects wrt. factor technique (see supplementary material for full analysis). We analysed block effects on TCT and ER for learning effects. As no statistically significant interaction effects between Technique and Block were found for TCT ($F^{11.57}_{219.97}$=1.53, p=.117) and ER (F^{32}_{608}=0.9, p=.645), we use all blocks for the remaining analysis.

5.1 Task Completion Time (Figure 5)

Regarding the factor technique ($F^{2.83}_{53.68}$ = 16.03, p<.001, η^2=0.22), we find that users were faster with Gaze&Finger, Gaze&Pinch, Gaze&Handray than with Handray and HeadCrusher. The statistical test showed a significant effect for Depth ($F^{1.67}_{31.67}$=25.50, p<.001, η^2=0.024), Amplitude ($F^{1.48}_{28.27}$=154.58, p<.001, η^2=0.200), for Technique× Depth ($F^{4.74}_{90.16}$=9.13, p<.001, η^2=0.033), for Technique×Amplitude ($F^{4.31}_{81.30}$=9.00, p<.001, η^2=0.022) and for Technique×Depth ×Amplitude ($F^{6.77}_{128.69}$=2.24,p<.036, η=0.008). We find that

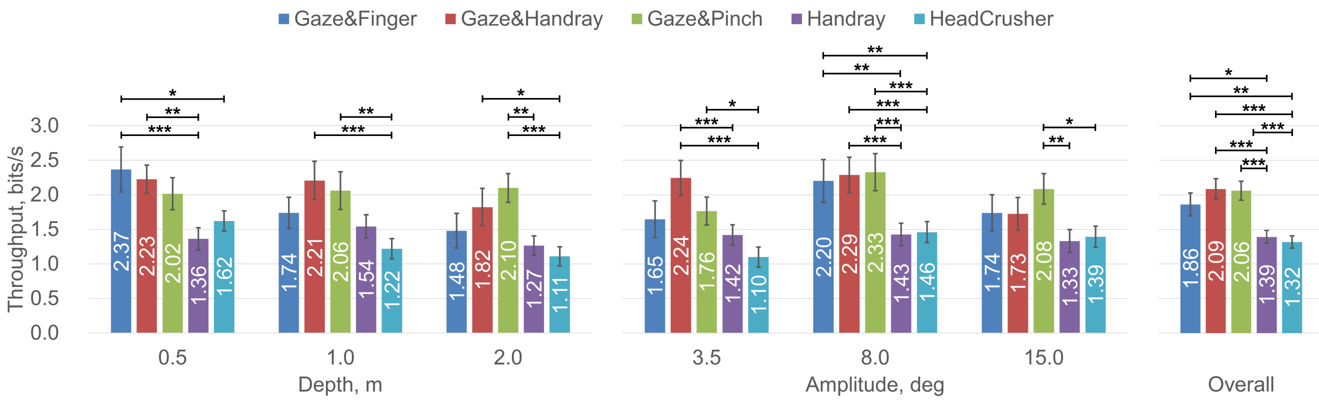


Figure 6: Mean throughput shown per depth per technique, per amplitude per technique, and per technique overall. Statistical significance shown as * for p < .05, ** for p < .01, and * for p < .001.**

there are significant differences between 0.5m and 2m for Gaze&Finger ($Difference(\Delta)$=580ms, t(19)=9.22, p<.001) and HeadCrusher (Δ=508ms, t(19)=3.92, p=.008) as well as between 1m and 2m for Gaze&Finger (Δ=272ms, t(19)=3.81, p=.012) and between 0.5m and 1m for HeadCrusher(Δ=563ms, t(19)=3.59, p=.026) and Gaze&Finger (Δ=308ms, t(19)=5.41, p<.001). HeadCrusher resulted in slower movement time compared to Gaze&Finger(Δ=895ms, t(19)=5.24, p<.001), Gaze&Handray (Δ=939ms, t(19)=6.26, p<.001), and Gaze&Pinch (Δ=731ms, t(19)=3.98, p<.008) in Depth 1m. This also applies to Depth 2m, but only to the techniques Gaze&Handray (Δ=751ms, t(19)=4.96, p<.001) and Gaze&Pinch (Δ=882ms, t(19)=5.10, p<.001) and to Depth 0.5m, but only to Gaze&Finger (Δ=640ms, t(19)=6.26, p<.001) and Gaze&Handray (Δ=432ms, t(19)=3.95, p=.009). The participants were significantly slower with the technique Gaze&Pinch compared to Gaze&Finger (Δ=413ms, t(19)=4.27, p=.003) in Depth 0.5m. Furthermore we found that the participants were significantly slower with the technique Handray compared to Gaze&Handray (Δ=844ms, t(19)=5.01, p<.001) and Gaze&Pinch (Δ=975ms, t(19)=5.15, p<.001) in Depth 2m and also with Gaze&Finger and Gaze&Handray in Depth 0.5m (p<.001) and Depth 1m (p=.009).

Significant differences were found between 3.5° and 15° (p<.001) and as well as between 8° and 15° (p≤.005) for all techniques. For Handray (Δ=354ms, t(19)=4.03, p=.005), Gaze&Handray (Δ=405ms, t(19)=8.91, p<.001) and Gaze&Finger (Δ=94ms, t(19)=4.14, p=.003) there are one between 3.5° and 8°. Gaze&Finger led to less TCT than Handray (Δ=954ms, t(19)=3.93, p=.008) and HeadCrusher (Δ=935ms, t(19)=3.81, p=.012) for Amplitude 15°. Further Gaze&Finger, Gaze&Pinch and Gaze&Handray led to less TCT than Handray and HeadCrusher for Amplitude 8° (p≤.005) and Amplitude 3.5° (p≤.014) but Gaze&Handray led only to less TCT than Gaze&Pinch Δ=264ms, t(19)=3.64, p=.020) in Amplitude 3.5°.

5.2 Throughput (Figure 6)

Throughput provides insight into the user's performance regarding both speed and accuracy. The statistical test showed a significant effect for Technique ($F^{3.17}_{60.38}$=10.05, p<.001, η^2=0.15), for Depth ($F^{1.83}_{34.76}$=19.18,p<.001, η^2=0.031), for Amplitude ($F^{1.68}_{31.87}$=21.66, p<.001, η^2=0.027), for Technique×Depth ($F^{4.48}_{85.05}$=5.10, p<.001, η^2=0.035) and for Technique×Amplitude ($F^{4.28}_{81.30}$=7.70, p<.001, η^2=0.029). Among the techniques, users reached significantly higher throughput (p≤.026) with Gaze&Finger, Gaze&Pinch, Gaze&Handray than with HeadCrusher and Handray. This confirms that gaze-assisted techniques are indeed superior to both manual selection techniques in the tested conditions [21]. For the tested conditions significant differences were found for Gaze&Finger for all three depths (p<.001) and for HeadCrusher between 0.5m and 2m (Δ=0.51bps, t(19)=3.72, p=.017). In comparison to Handray, Gaze&Finger (Δ=1.00bps, t(19)=5.08, p<.001) and Gaze&Handray (Δ=0.86bps, t(19)=4.37, p=.002) resulted in higher throughput for Depth 0.5m, and Gaze&Pinch for Depth 2m (Δ=0.83bps, t(19)=4.22, p=.004). In comparison to HeadCrusher, it was revealed that Gaze&Finger had higher throughput in Depth 0.5m (Δ=0.75bps, t(19)=3.76, p=.018), while for Gaze&Pinch and Gaze&Handray a higher throughput was found for Depth 1m and 2m (p≤.032).

Significant differences were found between 3.5° and 8° for HeadCrusher (Δ=0.36bps, t(19)=3.45, p=.048), Gaze&Pinch (Δ=0.56bps, t(19)=5.38, p<.001) and Gaze&Finger (Δ=0.55bps, t(19)=5.30, p<.001), between 8° and 15° for Gaze&Finger (Δ=0.46bps, t(19)=4.40, p=.001) and Gaze&Handray (Δ=0.56bps), t(19)=5.36, p<.001) and between 3.5° and 15° for Gaze&Handray (Δ=0.52bps, t(19)=4.94, p<.001).

We see a significantly higher throughput with all three gaze-assisted techniques for Amplitude 8° in comparison to both Handray and HeadCrusher (p≤.007). Individual differences were revealed, too. Gaze&Handray led to higher throughput than Handray (Δ=0.83bps, t(19)=4.51, p=.001) and HeadCrusher (Δ=1.15bps, t(19)=6.25, p<.001) for Amplitude 3.5°. While Gaze&Pinch led to higher throughput compared to HeadCrusher for all variations of Amplitude (p≤.030) and compared to Handray only for Amplitude 15° (Δ=0.75bps, t(19)=4.12, p=.006).

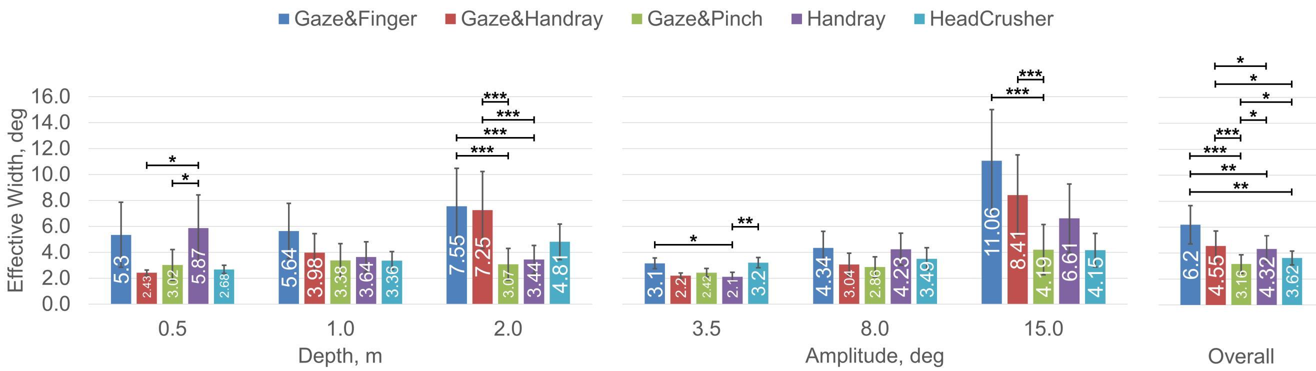


Figure 7: Mean effective width shown per depth per technique, per amplitude per technique, and per technique overall. Statistical significance shown as * for p < .05, ** for p < .01, and * for p < .001.**

5.3 Effective Width (Figure 7)

Effective width provides insight into the pointing accuracy, i.e., how close users selected to the target's center. The statistical test revealed a significant effect for Technique ($F^{3.25}_{61.76}$=13.90, p<.001, η^2=0.070), for Depth ($F^{1.84}_{35.01}$=52.86, p<.001, η^2=0.057), for Amplitude ($F^{1.86}_{35.30}$=62.01, p<.001, η^2=0.081), for Technique×Depth ($F^{5.80}_{110.15}$=19.52, p<.001, η^2=0.084), for Technique×Amplitude ($F^{4.67}_{88.65}$=24.47, p<.001, η^2=0.080), for Depth ×Amplitude ($F^{3.30}_{62.69}$=26.31, p<.001, η^2=0.061) and for Technique×Depth ×Amplitude ($F^{6.33}_{120.24}$=14.16, p<.001, η=0.102).

We found a higher effective width for Gaze&Finger (p≤.005) and Gaze&Handray (p≤.027) compared to Gaze&Pinch, HeadCrusher and Handray. Whereas HeadCrusher (Δ=0.46°, t(19)=2.792, p=0.027) and Handray (Δ=1.16°, t(19)=3.064, p=0.018) led to higher effective width than Gaze&Pinch. We find that there are significant differences between 0.5m and 2m as well as between 1m and 2m for Gaze&Finger (p=.002) and Gaze&Handray (p<.001). There is only a significant differences between 0.5m and 1m for Handray (Δ=2.23°, t(19)=3.64, p=.030). Furthermore, we found that the participants led to higher effective width with the technique Handray compared to Gaze&Handray (Δ=3.44°, t(19)=3.72, p=.022) and Gaze&Pinch (Δ=2.85°, t(19)=3.74, p=.020) in Depth 0.5m. Gaze&Finger and Gaze&Handray led to higher effective width than Handray and Gaze&Pinch in Depth 2m (p<.001). We find that there are significant differences in accuracy between 3.5° and 15° (p≤.001) and between 8° and 15° (p≤.027) for Gaze&Finger, Gaze&Handray and Handray. Furthermore we found that Gaze&Finger (Δ=6.86°, t(19)=5.21, p<.001), Gaze&Handray (Δ=4.22°, t(19)=5.24, p<.001) led to higher effective width than Gaze&Pinch for Amplitude 15° (p<.001). Gaze&Finger (Δ=1.04°, t(19)=3.91, p=.011) and HeadCrusher (Δ=1.09°, t(19)=4.30, p=.002) led to higher effective width than Handray in Amplitude 3.5°.

5.4 Error Rate (Figure 8)

The statistical test revealed a significant effect of Technique ($F^{2.95}_{56.05}$=23.86, p<.001, η^2=0.206), for Depth ($F^{1.60}_{30.44}$=6.79,p=.006, η^2=0.011), for Amplitude ($F^{1.95}_{36.97}$=31.34,p<.001, η^2=0.043), for Technique×Depth($F^{5.23}_{99.36}$=6.01,p<.001, η^2=0.036). We found users exhibited significantly higher error rates with HeadCrusher than all other techniques (p<.001). This shows that the manual image plane techniques leads to a lot more errors. Gaze&Handray, Gaze&Finger, and Gaze&Pinch, led to less errors than HeadCrusher in all depths levels (p≤.008). Handray led only to less errors than HeadCrusher for Depth 1m (Δ=33.70%, t(19)=7.52, p<.001) and 2m (Δ=33.55%, t(19)=7.18, p<.001). However it led to more errors than Gaze&Handray for Depth 0.5m (Δ=7.99%, t(19)=3.797, p<.015). We found that there are only significant differences between all depth levels for HeadCrusher (p<.001) – clearly pointing toward the potential parallax issue.

5.5 Hand Movement (Figure 9)

The statistical test revealed a significant effect of Technique ($F^{2.90}_{55.09}$=26.56, p<.001, η^2=0.178), for Depth ($F^{1.62}_{30.71}$=69.92, p<.001, η^2=0.062), for Amplitude ($F^{1.89}_{35.82}$=385.64, p<.001, η^2=0.291), for Technique×Depth ($F^{4.05}_{77.00}$=11.93, p<.001, η^2=0.034), for Technique×Amplitude ($F^{3.58}_{68.09}$=21.42, p<.001, η^2=0.060), for Depth×Amplitude ($F^{3.30}_{62.78}$=12.02, p<.001, η^2=0.012) and for Technique×Depth×Amplitude ($F^{7.44}_{141.41}$=4.44, p<.001, η^2=0.014). Post-hoc tests showed that users reached significantly better performance in hand movement with Gaze&Pinch compared to Handray (Δ=12.61*cm*, t(19)=7.64, p<.001), HeadCrusher (Δ=9.92*cm*, t(19)=7.62, p<.001), Gaze&Finger (Δ=19.85*cm*, t(19)=9.41, p<.001) and Gaze&Handray (Δ=11.16*cm*, t(19)=6.76, p<.001)which is true for all depths and amplitudes. We find that there are significant differences between 0.5m and 2m (p<.001) and between 0.5m and 1m (p<.001) for Gaze&Finger, Gaze&Handray, HeadCrusher and Handray. We find that there are significant differences in hand movement for all techniques between 3.5° and 15° (p<.001), between 8° and 15° (p≤.021) and between 3.5° and 8° (p<.001) except for Gaze&Pinch.

5.6 Fitts' Law Models and ID (Figure 10)

The techniques are modelled as MT=-0.27 + 0.88ID, R^2=0.97 for Handray, MT=0.96 + 0.49ID, R^2=0.77 for HeadCrusher, MT=0.71 + 0.35ID, R^2=0.65 for Gaze&Finger, MT=0.63 + 0.35ID, R^2=0.90 for

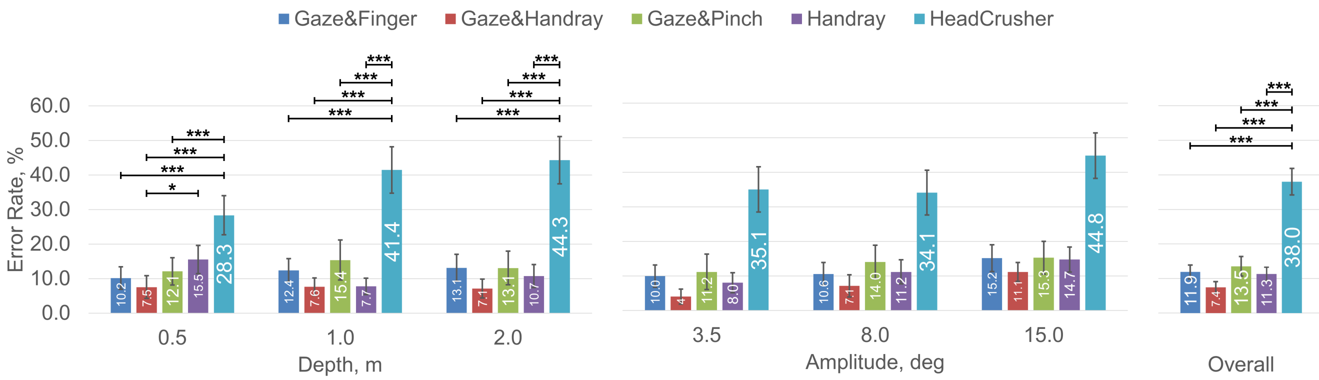


Figure 8: Mean error rate shown per depth per technique, per amplitude per technique, and per technique overall. Statistical significance shown as * for p < .05, ** for p < .01, and * for p < .001.**

Gaze&Pinch and MT=-0.59 + 0.79ID, R^2=0.85 for Gaze&Handray (cf. Figure 10 a).

The model for Handray shows a good fit to Fitts' Law, resulting at a R^2=0.97. As expected, a high b=0.88 value is revealed which captures the dependency of the movement time to the difficulty of the task (i.e., higher times with more difficult tasks). Surprisingly, Gaze&Pinch (R^2=0.90) resulted in high fitness too, even though it only involves little hand-based motor control. Distinct to Handray, the a and b of Gaze&Pinch's model indicate that performance depends less on the difficulty of the task. That the hands merely make a symbolic gesture to affirm the gaze's selection and that the eye movement within the tested amplitudes of view is ballistic might be one explanation. For Gaze&Handray, a high fit is reported (R^2=0.85), with high acceleration (b=0.79), indicating that the hand pointing part of the technique has a high impact on the performance similar to Handray. For HeadCrusher (R^2=0.77) and Gaze&Finger (R^2=0.65), lower R^2 values are reported which could indicate that image plane techniques fit less to the Fitts' Law model.

5.7 NASA TLX (Figure 11)

Participants ranked each technique from most to least preferred. The frequencies for most preferred ranking were Gaze&Handray = 9, Gaze&Pinch = 7, Gaze&Finger = 2, Handray = 1 and HeadCrusher = 1. Post-hoc tests showed that users perceived significantly higher workload with HeadCrusher compared to Handray, Gaze&Pinch, and Gaze&Handray (p<.001). Further, there is a significantly higher workload with Gaze&Finger compared to Gaze&Pinch and Gaze&Handray (p=.002). For Mental Demand (F^4_{76}=3.77,p=.007), Effort (F^4_{76}=7.50,p<.001) and Frustration (F^4_{76}=9.37,p<.001) was HeadCrusher rated significantly as the worst technique by the participants compared to Gaze&Pinch and Gaze&Handray ($p \leq .046$). Furthermore, HeadCrusher was rated significantly as the worst technique for Performance (F^4_{76}=4.53,p=.002) compared to Handray and Gaze&Handray ($p \leq .042$) and also for Physical Demand (F^4_{76}=25.51,p<.001) but compared to Gaze&Handray, Gaze&Pinch and Handray ($p \leq .024$). Regarding Physical Demand, Gaze&Finger was significantly more demanding than Gaze&Pinch and Gaze&Handray ($p \leq .004$).

5.8 User Feedback

5.8.1 Gaze&Finger. Users found the technique was easy and quick to use, especially for short depths (P7: *'I love aiming my gaze because it's reliable.'*. However, when the target was farther away, two cursors appeared and selection became difficult (P2:*'farther ones were tough. The cursor on the finger appears as two'*. As such, we observed users occasionally closing one eye to resolve parallax. For short distances four users were frustrated because of accidental selections: ...*'What I noticed was that when I aim at targets that are very close that I "double tap" the ray'* (P7, P16 similar). We suspect that they moved their finger on the target due to the outstretched arm of 17 users, especially with targets at a short distance. 12 users noted they had less control of the selection because it was happening too fast (P17 *'I did not feel very precise with "crossing the ray", I tried swiping left and right, up and down, slow and fast throughout this condition, and never really felt like I was confidently using it correctly'*). These findings do not confirm the positive reception in prior work [20, 21], likely as our task is not specialized for this technique.

5.8.2 Gaze&Handray. Users found the technique was basically easy to use, e.g., P10 stated: *"It feels natural.* In the beginning, five participants found it a bit difficult to control both rays (gaze and hand). On the one hand, the confirmation with the technology was difficult for long distances. P8 said *'It was super-fast, but for a larger target distance it was more difficult to use the gaze as the field of vision on the HoloLens is so small. I could not always see the targets and had to move my head as well as my gaze'*. On the other hand, it was also prone to errors at short distances because the targets were too close together and accidentally confirmed by hand even before the eyes were on the target. P3 commented *"...I felt a big difference between conditions where the targets were closer or further away. And I also felt that it was easier to accidentally select targets as soon as they appeared before I could move my hand away'*. The participants indicated a speed/error trade-off: they were fast over short Amplitude but made mistakes. The method produced the least amount of errors, those that happened were reported as the result of the handray and eyes being inadvertently aligned before aiming at the target.

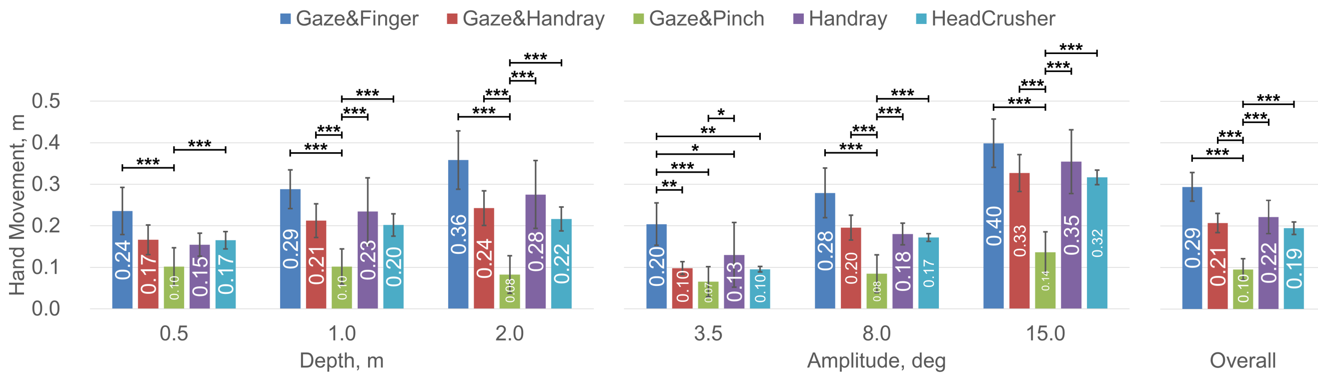
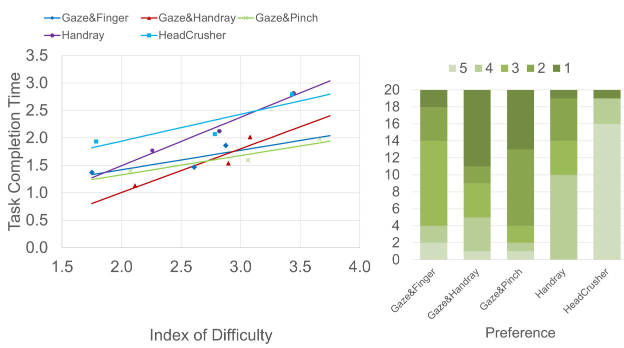


Figure 9: Mean hand movement shown per depth per technique, per amplitude per technique, and per technique overall. Statistical significance shown as * for p < .05, ** for p < .01, and * for p < .001.**



(a) (b)

Figure 10: Fitts' Law models (a) and user preferences (b).

5.8.3 *Gaze&Pinch.* The technique was easy to use with little physical demand because they could position their arm in a comfortable position next to the body. The participants were fast as P8 noted *' It was super fast and easy, my eyes were immediately on the targets without having to think about it too much, and then I just had to hold my arm still and pinch. Minimal effort and minimal fatigue'*. Four participants criticised the gesture, e.g., P9: *'The tapping was harder than I thought, I had to do it more dramatically than I thought'*. Yet, most users were satisfied and quick as of low effort.

5.8.4 *Handray.* Correlating with the quantitative performance results, users noted issues with the technique. Four users found the Handray technique less suitable for longer distances, while five felt less control of the ray, especially for pointing to near and far depths, as that Handray was difficult to move. According to P7: *'When aiming between targets that are far away from each other I felt a disconnect when moving to the next target'*. We observed pronounced head movements and slow arm movements throughout the study, which affected hand tracking and thus usability. As well, users found *'it was hard for me to pinch the fingers'* (P18). 17 out of 20 users slowly opened and flexed their fingers, increasing selection time and cognitive load.

5.8.5 *HeadCrusher.* According to 14 users' feedback, the technique was not easy to use and led to frustration (P17: *'I did not feel confident about this interaction technique'*). Especially if the target had a greater depth, the users noticed two cursors, which made the selection even more difficult (P8: *'I was seeing two cursors about 50% of the time, and I had difficulties seeing which was the right one'*. The other six participants found a way of using the technique correctly (P4: *'Beginning is a bit hard, easier when you get used to it'*). We observed that people were slowly moving their fingers to ensure that the selection will be correct. On the other hand, with closer targets and an outstretched arm that reduces the parallax effect, the technique worked reasonably well (P19: *'I felt like it was easier to hit the targets when my arms were further away from my body or simply closer to the target'*). However, as the hand was positioned at eye level the whole time, it became exhausting over time. 9 users had arm fatigue while using the technique. They always shook their arms before the new block started and were quite frustrated, indicating physical strain. We suspect that this technique would require more practice in general, as the pinch gesture was performed too slowly or the users moved their thumb too much and were unable to keep the cursor on the target consistently.

6 DISCUSSION

In the quest to establish interfaces that allow users to interact efficiently with 3D content in the vast virtual spaces afforded by VR and AR, we investigated interaction techniques for the fundamental point-and-select tasks. The GHA techniques offer a new way to accomplish selections with objects at-a-distance based on coordinated eye and hand inputs. In order to understand their potential merits, we studied Gaze&Finger and Gaze&Handray in contrast to three existing categories of interaction: the current status-quo AR input method (Handray), a representative of image plane techniques (HeadCrusher), and an established multi-modal method (Gaze&Pinch). Our experimental design is based on the ISO 9241-9 standardised Fitts' Law study [22], which allows us to compare findings across studies. In the following, we discuss our main insights of the user study.

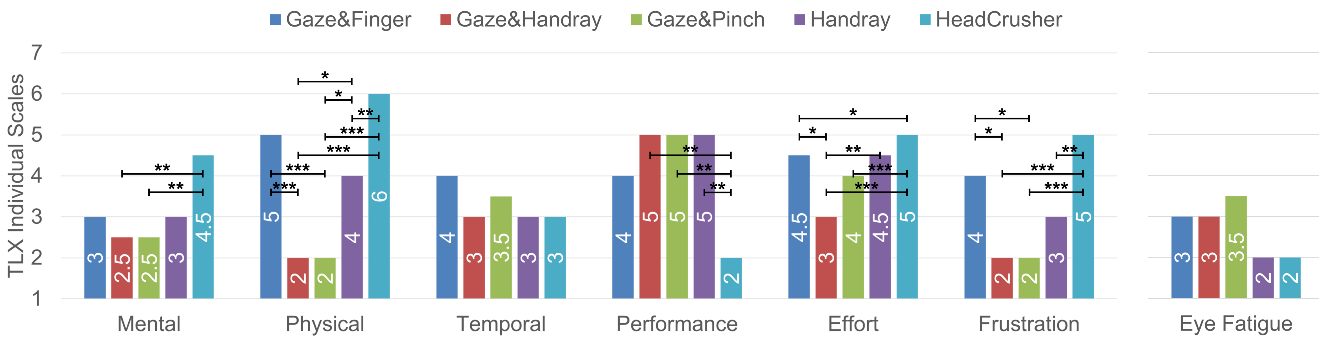


Figure 11: Median ratings of individual NASA TLX scales and for eye fatigue. Statistical significance shown as * for p < .05, ** for p < .01, and * for p < .001.**

6.1 RQ1 How does varying target depth affect the user's performance?

One concern was that the GHA technique Gaze&Finger, as an image plane technique, may be prone to the parallax problem. Our study confirmed this– throughput declined with increasing depth. The response from users validated the parallax effect. Users claimed that Gaze&Finger was rapid and easy to use for short depths, but for more distant targets, users saw double vision of the cursor, which hampered performance. We saw that users occasionally sought to avoid the parallax effect by closing one eye or stretching out their arms, with both strategies significantly reducing the parallax effect. This led to users devoting more time to the selection process.

The parallax problem affected the HeadCrusher to a great extent, with higher performance losses compared to other approaches and throughput declining from close to far Depth conditions. The most noticeable impact is an increase in error rate when compared to the other techniques. HeadCrusher had the highest error rate across all conditions, and users experienced difficulty selecting due to the apparent perceived parallax problem. As a result, we present novel empirical evidence describing how parallax affects the approach, supporting the issues stated by Pierce et al. [34]. Overall, we do not advocate using HeadCrusher's basic implementation for 3D remote pointing.

6.2 RQ2 How does the new Gaze&Handray technique fare compare to other gaze and manual techniques?

As a novel gaze-based approach, we were curious how it compared to Gaze&Finger and Gaze&Pinch, which yielded the best results in Lystbæk et al. [21]'s study. We found no significant changes in throughput across depth between the three techniques, indicating that there is no substantial penalty here. However, we observe an influence on throughput across amplitudes. When compared to the other two, users achieved significantly less throughput at the large amplitude. Thus, as the distance increases, the benefit of Gaze&Handray lowers, but generally, it demonstrates that it is on par with gaze-based approaches.

The effective width analysis reveals discrepancies between different depths and amplitudes, indicating that as distance increases, users become less accurate. A likely reason is the confirmation method affects the stability of the user's gaze on the target. E.g., with Gaze&Pinch, the modalities' inputs are separated, which led to a lower effective width, while the eye-hand coordination needs of Gaze&Finger can distract visual attention on the target.

On the one hand, user feedback showed Gaze&Handray is simple to use and a technique that seems natural and rapid, but on the other hand, it was noted that it was initially difficult to manage both rays (gaze and hand), especially as hand tracking declined and was lost owing to strong head movements. Thus, users stated that Gaze&Handray was problematic for participants over a large amplitude, as distant targets were not visible due to the narrow FOV of the HoloLens, necessitating head movements. However, the technique was ranked as the most liked by users overall, and together with the quantitative findings, we found it to be the best-performing technique. As such, it would be interesting to explore this technique further for raypointing-based techniques. For example, it may be as useful to enhance controller pointing, as an input device that is highly used in current VR devices. As well, to regard real-world devices such as TV remote controls and distant interaction that is, e.g., facilitated through laserpointers [29].

6.3 RQ3 How does varying target amplitude affect the user's performance?

Regarding effects of amplitude, we observe the expected trend that user performance decreases with increasing amplitude. We found that all techniques were affected by amplitude.

One of the conclusions of Lystbæk et al. [21]'s study was that Gaze&Finger was more useful for short distances as one can easily move the finger in close vicinity, whereas Gaze&Pinch led to better performance for long distances to travel as the eyes point independent of distance. In our study, where we systematically varied distance and depth, we could not replicate this result, as no significant differences were found between the three gaze techniques across factors. This can be accounted to the task design, as in their study, the task was designed for the techniques, while we focused on a general study of selection performance. Notably, Gaze&Finger as an image plane technique led to decreasing performance the further away a target is. Thus, it is possible that with more pronounced depth factors, the performance of Gaze&Pinch may become superior.

In this regard, we find clear evidence of physical effort. Users perceived GAZE&FINGER as more physically demanding than the other gaze techniques, and the actual physical movement was higher compared to GAZE&PINCH. We observed that for both techniques, users had consistent hand postures. For GAZE&FINGER, most users outstretched their arms to reduce parallax effects, which led to high physical effort. The technique also revealed an inaccuracy at the largest amplitude (compared to GAZE&PINCH, HANDRAY, and HEADCRUSHER). Thus, given our evaluated context, we attest to a preference for GAZE&PINCH over GAZE&FINGER.

6.4 RQ4: How do multi-modal gaze-assisted techniques compare to manual techniques regarding selection efficiency?

Although multi-modal techniques such as GAZE&PINCH are now available on AR devices by default, there has been no empirical evaluation of their performance compared to manual gesture techniques in 3D environments. As such, it is important to better understand currently used methods, as well as how they compare to potential new techniques. In our study, we find that compared to the two manual techniques a better performance was achieved with all three gaze-assisted techniques for throughput and task completion time.

Users described the gaze-based techniques as faster and easier to use, whereas HANDRAY was reported to need pronounced head and arm movement and HEADCRUSHER was rated as frustrating and difficult to use. Stronger head movements led to a change in hand tracking range, which could have led to the participants experiencing poorer performance and being more frustrated and with it the execution time of the task with the manual-based techniques also increased. On the other hand, we believe that the techniques assisted with gaze are promising for the future and will make it easier for the user to select virtual objects. However, this finding must also be viewed critically. We used a target size of 3° in our study. So the target was still big enough to be selected with gaze, and manual pointing can in principle be more accurate. Looking at the data on effective width, we see that both GHA techniques exhibit larger width than manual inputs at a large amplitude. However, GAZE&PINCH achieved a relatively low effective width, indicating that the method of manual confirmation can affect effective width.

6.5 Limitations

When analysing our results, there are a few caveats to keep in mind. First, we have focused on large target sizes (3°). This allowed us to eliminate potential eye-tracking inaccuracy effects from the results and prioritise the investigation into different depths. Accuracy-enhancing methods can be useful to render gaze more accurate for small target acquisition [16]. Second, we employed Fitts' Law for a standardised comparison, and more studies in real applications are needed to assess real-world applicability. Further, Fitts' Law is a well-established model for 2D UIs, but its applicability to 3D contexts is relatively unclear. In future, it would be interesting to study enhancements, e.g., to vary target depth at each target for a more realistic 3D task, which could potentially amplify parallax effects. Third, we focused on selection tasks only. However, more expressive commands such as drag & drop which are intuitive with the baselines, are currently open and subject to future technique explorations. Fourth, we observed hand tracking and field of view of the hardware affected the user's performance, e.g., when gesture tracking is lost as the hand left the tracking area. With better tracking, user comfort may rise and make the techniques more efficient.

7 CONCLUSION

In this work, we investigated interaction techniques based on GHA for selection of remote targets in AR. We devised a new technique GAZE&HANDRAY that combines gestural raypointing with eye-gaze. We presented a Fitts' Law study that compares five interaction techniques that are based on uni- and multi-modal gaze and gesture inputs. We provide a detailed understanding of performance profiles for each technique regarding speed, error, accuracy and perceived effort. We contribute novel insights into the technique's relative merits, as 1) image plane techniques are particularly susceptible to the Parallax problem, 2) the tested gaze-assisted techniques are highly efficient and fast compared to HEADCRUSHER and HANDRAY, 3) the new technique GAZE&HANDRAY is on-par with regards to performance and was most preferred by users, next to GAZE&PINCH. Our work paves the way to utilise the user's capability to fluidly coordinate eye and hand movement, through techniques that can fundamentally enhance how we conduct the canonical selection task in AR.

ACKNOWLEDGMENTS

This work has received funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grants no. 101021229 GEMINI, and no. 740548 CIO) and by the Innovation Fund Denmark, as part of the Manufacturing Academy of Denmark (MADE) FAST project.

REFERENCES

[1] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique Based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300331

[2] Martin S. Banks, Tandra Ghose, and James M. Hillis. 2004. Relative image size, not eye position, determines eye dominance switches. *Vision Research* 44, 3 (2004), 229–234. https://doi.org/10.1016/j.visres.2003.09.029

[3] Eric A. Bier, Maureen C. Stone, Ken Pier, Ken Fishkin, Thomas Baudel, Matt Conway, William Buxton, and Tony DeRose. 1994. Toolglass and Magic Lenses: The See-through Interface. In *Conference Companion on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 445–446. https://doi.org/10.1145/259963.260447

[4] Pieter Blignaut. 2009. Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, perception & psychophysics* 71 (06 2009), 881–95. https://doi.org/10.3758/APP.71.4.881

[5] Doug A. Bowman and Larry F. Hodges. 1997. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (Providence, Rhode Island, USA) *(I3D '97)*. Association for Computing Machinery, New York, NY, USA, 35–ff. https://doi.org/10.1145/253284.253301

[6] Wesley P. Chan, Geoffrey Hanks, Maram Sakr, Haomiao Zhang, Tiger Zuo, H. F. Machiel van der Loos, and Elizabeth Croft. 2022. Design and Evaluation of an Augmented Reality Head-Mounted Display Interface for Human Robot Teams Collaborating in Physically Shared Manufacturing Tasks. *J. Hum.-Robot Interact.* 11, 3, Article 31 (jul 2022), 19 pages. https://doi.org/10.1145/3524082

[7] Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (Seattle, Washington,

USA) *(ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/2818346.2820752

[8] L. Colligan, H. W. Potts, C. T. Finn, and R. A. Sinkin. 2015. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record, Vol. 84. International journal of medical informatics, 469–476. https://doi.org/10.1016/j.ijmedinf.2015.03.003

[9] Nathan Cournia, John D. Smith, and Andrew T. Duchowski. 2003. Gaze- vs. Hand-Based Pointing in Virtual Environments. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI EA '03)*. Association for Computing Machinery, New York, NY, USA, 772–773. https://doi.org/10.1145/765891.765982

[10] Lisa A Elkin, Matthew Kay, James J Higgins, and Jacob O Wobbrock. 2021. An aligned rank transform procedure for multifactor contrast tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 754–768.

[11] Jeremy Hales, David Rozado, and Diako Mardanbegi. 2013. Interacting with objects in the environment by gaze and hand gestures. In *Proceedings of the 3rd international workshop on pervasive eye tracking and mobile eye-based interaction*. 1–9.

[12] John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. 2018. A Fitts' Law Study of Click and Dwell Interaction by Gaze, Head and Mouse with a Head-Mounted Display. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 7, 5 pages. https://doi.org/10.1145/3206343.3206344

[13] Wen-jun Hou and Xiao-lin Chen. 2021. Comparison of Eye-Based and Controller-Based Selection in Virtual Reality. *International Journal of Human–Computer Interaction* 37, 5 (2021), 484–495.

[14] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

[15] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3328–3339. https://doi.org/10.1145/3025453.3025523

[16] Manu Kumar, Jeff Klingner, Rohan Puranik, Terry Winograd, and Andreas Paepcke. 2008. Improving the Accuracy of Gaze Input for Interaction. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia) *(ETRA '08)*. Association for Computing Machinery, New York, NY, USA, 65–68. https://doi.org/10.1145/1344471.1344488

[17] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173655

[18] Lu Lu, Pengshuai Duan, Xukun Shen, Shijin Zhang, Huiyan Feng, and Yong Flu. 2021. Gaze-Pinch Menu: Performing Multiple Interactions Concurrently in Mixed Reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 536–537. https://doi.org/10.1109/VRW52623.2021.00150

[19] Francisco Lopez Luro and Veronica Sundstedt. 2019. A Comparative Study of Eye Tracking and Hand Controller for Aiming Tasks in Virtual Reality. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 68, 9 pages. https://doi.org/10.1145/3317956.3318153

[20] Mathias N. Lystbæk, Ken Pfeuffer, Jens Emil Sloth Grønbæk, and Hans Gellersen. 2022. Exploring Gaze for Assisting Freehand Selection-Based Text Entry in AR. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 141 (may 2022), 16 pages. https://doi.org/10.1145/3530882

[21] Mathias N. Lystbæk, Peter Rosenberg, Ken Pfeuffer, Jens Emil Grønbæk, and Hans Gellersen. 2022. Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 145 (may 2022), 18 pages. https://doi.org/10.1145/3530886

[22] Ian Scott MacKenzie. 2013. A Note on the Validity of the Shannon Formulation for Fitts' Index of Difficulty. *Open Journal of Applied Sciences* 03 (2013), 360–368. https://doi.org/10.4236/ojapps.2013.36046

[23] Ian Scott MacKenzie. 2015. Fitts' Throughput and the Remarkable Case of Touch-Based Target Selection. In *HCI*. https://doi.org/10.1007/978-3-319-20916-6_23

[24] Ian Scott MacKenzie. 2018. Fitts' Law. (2018), 349–370. https://doi.org/10.1002/9781118976005

[25] Diako Mardanbegi, Benedikt Mayer, Ken Pfeuffer, Shahram Jalaliniya, Hans Gellersen, and Alexander Perzl. 2019. EyeSeeThrough: Unifying Tool Selection and Application in Virtual Environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 474–483. https://doi.org/10.1109/VR.2019.8797988

[26] Mark R Mine. 1995. Virtual environment interaction techniques. *UNC Chapel Hill CS Dept* (1995).

[27] Mark R. Mine, Frederick P. Brooks, and Carlo H. Sequin. 1997. Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*. ACM Press/Addison-Wesley Publishing Co., USA, 19–26. https://doi.org/10.1145/258734.258747

[28] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. *Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality*. Association for Computing Machinery, New York, NY, USA, Chapter 15, 7. https://doi.org/10.1145/3448018.3457998

[29] Dan R. Olsen and Travis Nielsen. 2001. Laser Pointer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. Association for Computing Machinery, New York, NY, USA, 17–22. https://doi.org/10.1145/365024.365030

[30] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-Touch: Combining Gaze with Multi-Touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

[31] Ken Pfeuffer and Hans Gellersen. 2016. Gaze and Touch Interaction on Tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 301–311. https://doi.org/10.1145/2984511.2984514

[32] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[33] Ken Pfeuffer, Lukas Mecke, Sarah Delgado Rodriguez, Mariam Hassib, Hannah Maier, and Florian Alt. 2020. Empirical Evaluation of Gaze-Enhanced Menus in Virtual Reality. In *26th ACM Symposium on Virtual Reality Software and Technology* (Virtual Event, Canada) *(VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 20, 11 pages. https://doi.org/10.1145/3385956.3418962

[34] Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. 1997. Image Plane Interaction Techniques in 3D Immersive Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (Providence, Rhode Island, USA) *(I3D '97)*. Association for Computing Machinery, New York, NY, USA, 39–ff. https://doi.org/10.1145/253284.253303

[35] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-Defined Gestures for Augmented Reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 955–960. https://doi.org/10.1145/2468356.2468527

[36] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-Worn Extended Reality. *ACM Comput. Surv.* 55, 3, Article 53 (mar 2022), 39 pages. https://doi.org/10.1145/3491207

[37] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology* (Seattle, Washington, USA) *(UIST '96)*. Association for Computing Machinery, New York, NY, USA, 79–80. https://doi.org/10.1145/237091.237102

[38] Ivan Poupyrev, Tadao Ichikawa, Suzanne Weghorst, and Mark Billinghurst. 1998. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer graphics forum*, Vol. 17. Wiley Online Library, 41–52.

[39] Katharina Reiter, Ken Pfeuffer, Augusto Esteves, Tim Mittermeier, and Florian Alt. 2022. Look & Turn: One-Handed and Expressive Menu Interaction by Gaze and Arm Turns in VR. In *2022 Symposium on Eye Tracking Research and Applications* (Seattle, WA, USA) *(ETRA '22)*. Association for Computing Machinery, New York, NY, USA, Article 66, 7 pages. https://doi.org/10.1145/3517031.3529233

[40] Kunhee Ryu, Joong-Jae Lee, and Jung-Min Park. 2019. GG Interaction: a gaze–grasp pose interaction for 3D virtual object selection. *Journal on Multimodal User Interfaces* 13, 4 (2019), 383–393.

[41] Martin Schmitz, Sebastian Günther, Dominik Schön, and Florian Müller. 2022. Squeezy-Feely: Investigating Lateral Thumb-Index Pinching as an Input Modality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 61, 15 pages. https://doi.org/10.1145/3491102.3501981

[42] Robin Schweigert, Valentin Schwind, and Sven Mayer. 2019. EyePointing: A Gaze-Based Selection Technique. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 719–723. https://doi.org/10.1145/3340764.3344897

[43] Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of Eye Gaze Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 281–288. https://doi.org/10.1145/332040.332445

[44] Ludwig Sidenmark and Hans Gellersen. 2019. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

[45] Ludwig Sidenmark, Dominic Potts, Bill Bapisch, and Hans Gellersen. 2021. Radi-Eye: Hands-Free Radial Interfaces for 3D Interaction Using Gaze-Activated Head-Crossing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 740, 11 pages. https://doi.org/10.1145/3411764.3445697

[46] Vildan Tanriverdi and Robert J. K. Jacob. 2000. Interacting with Eye Movements in Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 265–272. https://doi.org/10.1145/332040.332443

[47] Eduardo Velloso, Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation. In *Human-Computer Interaction – INTERACT 2015*. Springer-Verlag, Berlin, Heidelberg, 315–330. https://doi.org/10.1007/978-3-319-22668-2_25

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

[49] Dennis Wolf, Jan Gugenheimer, Marco Combosch, and Enrico Rukzio. 2020. Understanding the Heisenberg Effect of Spatial Interaction: A Selection Induced Error for Spatially Tracked Input Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376876

[50] Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2021. Gaze-Supported 3D Object Manipulation in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 734, 13 pages. https://doi.org/10.1145/3411764.3445343

[51] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053

[52] Yanxia Zhang, Sophie Stellmach, Abigail Sellen, and Andrew Blake. 2015. The costs and benefits of combining gaze and hand gestures for remote interaction. In *IFIP Conference on Human-Computer Interaction*. Springer, 570–577.

Received September 2021; revised December 2022; accepted Janaury 2023