Eye-Hand Movement of Objects in Near Space

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

Andreas Asferg Jacobsen
Aarhus University
Denmark
201905139@post.au.dk

Tiare Feuchtner
HCI Group, University of Konstanz
Germany
Aarhus University
Denmark
tiare.feuchtner@uni-konstanz.de

Hans Gellersen
Lancaster University
Lancaster, United Kingdom
Aarhus University
Denmark
h.gellersen@lancaster.ac.uk

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

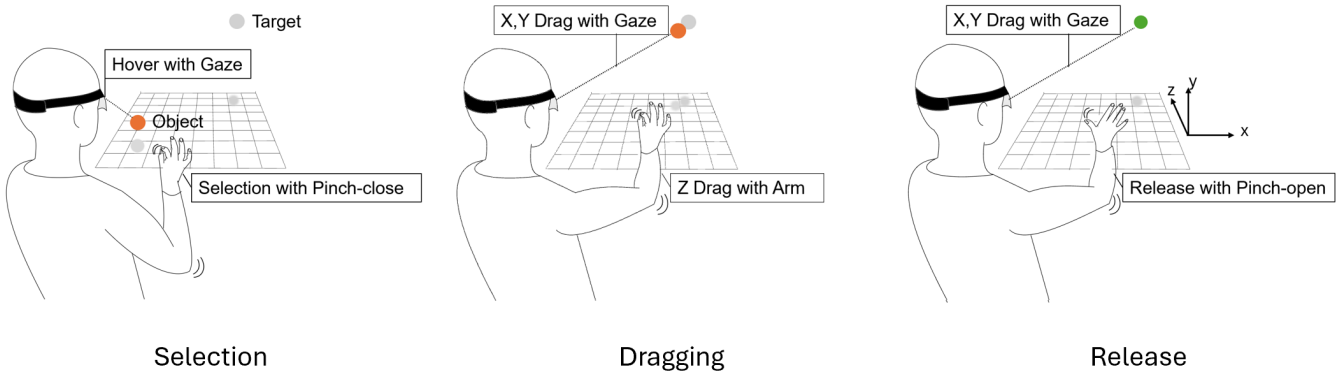


Figure 1: Look&Drop is an interaction technique for eye and hand tracking operated extended reality (XR) interfaces. It provides users with control over a selected object's position at low-effort, which is useful for frequent and prolonged dragging operations. Look&Drop extends Drag&Drop with direct manipulation gestures in near space, as the user can drag objects in X and Y axis by gazing toward the destination, and using the hand to adjust object depth (Z).

ABSTRACT

Hand-tracking in Extended Reality (XR) enables moving objects in near space with direct hand gestures, to pick, drag and drop objects in 3D. In this work, we investigate the use of eye-tracking to reduce the effort involved in this interaction. As the eyes naturally look ahead to the target for a drag operation, the principal idea is to map the translation of the object in the image plane to gaze, such that the hand only needs to control the depth component of the operation. We have implemented four techniques that explore two factors: the use of gaze only to move objects in X-Y vs. extra refinement by hand, and the use of hand input in the Z axis to directly move objects vs. indirectly via a transfer function. We compared all four techniques in a user study (N=24) against baselines of direct and indirect hand input. We detail user performance, effort and experience trade-offs and show that all eye-hand techniques significantly reduce physical effort over direct gestures, pointing toward effortless drag-and-drop for XR environments.

Authors' Contact Information: Uta Wagner, Aarhus University, Denmark, uta.wagner@cs.au.dk; Andreas Asferg Jacobsen, Aarhus University, Denmark, 201905139@post.au.dk; Tiare Feuchtner, HCI Group, University of Konstanz, Germany and Aarhus University, Denmark, tiare.feuchtner@uni-konstanz.de; Hans Gellersen, Lancaster University, Lancaster, United Kingdom and Aarhus University, Denmark, h.gellersen@lancaster.ac.uk; Ken Pfeuffer, Aarhus University, Denmark, ken@cs.au.dk.

This work is licensed under a Creative Commons Attribution International 4.0 License.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0628-8/24/10
https://doi.org/10.1145/3654777.3676446

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Pointing**.

KEYWORDS

eye-tracking, gaze interaction, drag&drop, object manipulation, near space

ACM Reference Format:
Uta Wagner, Andreas Asferg Jacobsen, Tiare Feuchtner, Hans Gellersen, and Ken Pfeuffer. 2024. Eye-Hand Movement of Objects in Near Space. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3654777.3676446

1 INTRODUCTION

With a mouse or touchscreen, small finger and hand movements let us move objects across the entire screen. In contrast, in 3D user interfaces we directly reach out and grab objects, moving them from one location in space to another. This direct physical manipulation creates a strong sense of intuition, mirroring how we interact with the real world. However, large distances and frequent use over time accrue to physical strain and fatigue across multiple parts of the body, from finger to shoulder [15, 24]. Combining the eyes with the hands in the user interface (UI) allows for new forms of gestural interaction that are mapped to gaze-selected objects at a distance, as exemplified in gaze + pinch [9, 23, 28]. This effectively eliminates manual pointing movements by replacing the initial pointing gesture with a simple glance, significantly reducing physical strain [9, 20, 43]. However, current methods strictly separate these actions based on the principle of 'eyes select, hand manipulates' [26, 28]. While this keeps the interaction paradigm simple, it can feel unnatural for complex tasks like moving objects. For example, our eyes naturally dart ahead to where we want to drag or move something. This suggests a powerful opportunity: using eye tracking not only for selection, but also to guide our hand's direct manipulation during tasks like dragging.

While research suggests potential for gaze-based drag and drop interactions, existing studies primarily focus on 2D interfaces and handheld controllers [35, 37, 47]. Controllers can feel cumbersome for direct manipulation of nearby objects, where the natural expectation is to interact directly with our hands. As such, while the effectiveness of eye-hand selection is well-established, the knowledge of eye-hand manipulation remains severely limited for gestural UIs. Addressing this gap is highly informative for UI practitioners, designers, and researchers developing extended reality (XR) interfaces for 3D select-and-manipulate operations.

We investigate 'Look&Drop' interaction techniques that use eye gaze (looking) to specify the direction for the object to travel, instead of moving the hand directly to it, but using the hand only for depth manipulation. In particular, moving an object in 3D would involve (1) selecting the target by looking at it and performing a pinch-in gesture, (2) looking at the drop destination to translate the target in view space, while moving the hand to affect the object's depth, and (3) a pinch-out gesture finishes the task. Our main hypothesis is that this can significantly reduce user fatigue of direct manipulation and make moving objects almost effortless. Several design questions remain to be addressed in realising Look&Drop. First, how should hand movements for depth control be mapped? A one-to-one mapping (direct manipulation) offers intuitive control, while a control-display (CD) gain might improve precision. Second, should object translation in the view space be solely controlled by gaze, or, given that gaze can be wandering, should additional hand input be used to correct and refine gaze position, as suggested in prior research [47, 48]? Finally, how does this gaze-assisted manipulation compare to existing techniques like Gaze + Pinch, where the gaze is used for selection but not the manipulation itself?

To explore these points, we evaluated Look&Drop in a user study with 24 participants, comparing it to baselines of direct manipulation and Gaze + Pinch. We tested four variations of Look&Drop, that form the four unique combinations between 1:1 or CD gain for depth (Z) manipulation, and eyes-only vs. eye-hand dragging (Look&Drop+) for image plane manipulation (XY). The study task involved the movement of objects across two different distances in diagonal 3D directions in front of the user. Our results led to the following key insights:

- Look&Drop resulted in lower perceived physical effort and lower recorded hand motion than direct gestures.
- Look&Drop+ was most preferred overall (66%).
- Look&Drop without manual refinement led to more errors than direct manipulation.
- Gaze-based techniques are faster in selection but slower during the drag-and-drop process compared to direct manipulation.
- Gaze + Pinch is faster for dragging, but demands more physical hand movement compared to Look&Drop techniques.

Our study explored the trade-offs between different interaction techniques, revealing unique strengths and weaknesses for each compared to traditional direct manipulation and Gaze + Pinch. To showcase these techniques in action, we designed several use cases in application probes (Figure 3). For instance, imagine a 3D grid of objects. Look&Drop allows users to quickly select a starting point and destination to create an area selection (Figure 3a). This enables rapid execution of commands like "create," "edit," or "delete" on groups of objects. Alternatively, consider specifying a non-uniform 3D path. Normally, users would specify each point using hand-controlled object movement. Here, users can perform multiple drag-and-drop actions with Look&Drop, seamlessly forming a continuous path (Figure 3b). Finally, we designed a 3D puzzle game inspired by Bejeweled. Look&Drop empowers players to swiftly drag similar objects together to beat the game (Figure 3c).

Our contributions include: (1) The design of Look&Drop as novel gaze-assisted object movement techniques for eye- and hand-tracked XR environments, including the enhanced Look&Drop+ for precise object dropping. (2) An empirical user study that compares Look&Drop to direct and Gaze&Pinch gestures, providing fundamental insight into user performance trade-offs, such as significant reduction of physical effort and how Look&Drop+ addresses late trigger errors. (3) Demonstration of the utility of Look&Drop through application examples for rapid area selection, path generation, and a 3D puzzle game.

2 RELATED WORK

XR researchers have long established the HCI foundations for 3D interaction [6, 22, 31]. Egocentric interaction in XR offers the *virtual hand* metaphor [31] that closely resembles real-world manipulation, allowing users to naturally reach out and grab objects in near space — a key interaction zone for many 3D tasks [1, 12, 46]. Near space is well-suited for interaction with various UI reference frames [21, 25], and in particular for tasks requiring precise gestural object alignment and manipulation [14]. For gaze interaction in virtual

environments, Tanriverdi and Jacob's early work of eye movement based interaction discussed key benefits: lower physical effort, the use of eye movements as a natural pointer, the ability to select remote targets, and the pragmatic benefit of a lightweight sensor integrated in the HMD [36]. They showed that a simple gaze-based dwell-time approach outperforms manual raypointing for remote object selection. However, this method is limited by the Midas touch problem, which can cause ambiguity between looking and intended selection [17]. In our work, we focus on the contrast between hand and eye-hand interaction, a widely available input however underrepresented in HCI research for object manipulation.

A relevant method for our research is Zhai et al.'s Manual And Gaze Input Cascaded (MAGIC) [48], which enhances manual input with gaze. It warps the mouse cursor to the visual attention area on a 2D screen, reducing dragging effort significantly. Extending this, Stellmach and Dachselt introduced Look & Touch, using a handheld touchscreen for cursor refinement on a remote screen [34]. Similarly, Gaze + Gesture [9] employed hand and eye tracking with pinch gestures for cursor refinement on a PC. Both Look & Touch (vs. gaze-only) and Gaze+Gesture (vs. hands-only and gaze-only pointing) demonstrated that eye-hand interaction outperformed single-modality approaches for selection tasks. Inspired by this, we consider ways like MAGIC in our interaction design to potentially enhance dropping precision.

A line of research investigated eye-hand interaction for advanced object manipulation tasks. In Still Looking, Stellmach and Dachselt propose an object translation technique on remote 2D screens where the gaze is coupled with a handheld mobile touchscreen's finger gestures. A single tap selects an object, and a double tap drops an object in place. Their evaluation showed that the gaze variant leads to higher performance than the head-pointing variant. Turner et al. have proposed several techniques for object dragging in a series of user studies focusing on transfer between a local handheld device and a remote display [37–40], where gaze is used to indicate objects of interest and touch gestures such as swipe, tap and hold allow to transfer an object. In their Gaze+RST work, Turner et al., all rotate, scale and translate (RST) operations are investigated through multi-touch gestures on a tablet. Gaze is used to select the object and destination, and several MAGIC-based concepts to integrate eyes and hands are investigated for concurrent RST operations, finding that integration of gaze improves user performance over touch-only manipulation.

Whereas selection is extensively studied, eye-hand object manipulation is underdeveloped in the XR HCI literature. Numerous studies on the selection task showed that gaze + hands and gaze + controller allow for improved user performance with reduced physical effort compared to eyes-only and hands-only input techniques [18–20, 23, 29, 43, 49]. Out of those, we include the Gaze + Pinch technique as an eye-hand baseline in our study. The basic principle is a division of labour between the modalities: the eyes select a target by looking at it, a pinch gesture confirms it, and hand motion offsets the object's position accordingly. Various studies have explored eye-hand manipulation for making menu interfaces easily accessible [10, 29, 32]. For instance, when using a menu to switch colour modes during sketching tasks, both presented eye-controller techniques showed inferior performance to direct manipulation. We add to the empirical knowledge, by assessing eye-hand vs. direct interaction for object movement tasks. Other manipulation tasks outside our scope were as well investigated, such as the specification of a 2D rectangle [33], bi-manual 3D point specification [49], and remote occlusion selection [42].

Closely related are Yu et al.'s gaze-based object manipulation techniques [47], proposing two methods for object translation utilizing gaze. 3D Magic Gaze snaps the object to the gaze during controller movement, followed by manual refinement, while Implicit Gaze adds a dynamically resizing snap area. Their evaluation did not show clear benefits over hands-based dragging. Our work shares the consideration of MAGIC to improve dragging tasks and extends theirs in the following ways. First, we focus on direct manipulation as the natural baseline, rather than controllers, a distinct context as controller pointing is already fast and low-effort. Second, we evaluate 3D object movement using eye-hand input across all dimensions, unlike the prior focus on lateral object positioning at a single depth level.

3 LOOK&DROP INTERACTION DESIGN

We implemented four Look&Drop variations that differ in the mapping of the eye and hand tracking signals to actions for commanding the UI. We compare the interaction techniques with two baseline techniques specifically tailored to 3D Drag&Drop. All techniques consist of four steps based on the input structure of prior work for 3D selection and manipulation of an object [7, 28, 47]: *Indicate, Confirm, Manipulate, and Release.* Users indicate an object by their gaze through raypointing toward the target, confirm a selection by a pinch-in gesture, perform eye-hand object dragging and then release by a pinch-out. The techniques have differences in the mapping for depth control and the addition of a feature for fine positioning, as we detail next.

3.1 Direct vs Indirect Depth Control

Direct manipulation via virtual hand mimics real-world actions and provides physical affordances in manipulations, whereas indirect input affords interaction over distance, variable control-display gains, and occlusion-free interaction [16, 31].

3.1.1 Direct depth control. Employs mapping the depth of the dragged object along the Z-axis and remains consistently at the same level as the hand once the pinch gesture is executed. As a consequence, the object jumps to the current depth level of the hand when pinched. It offers the advantage that the hand can be moved directly to the target depth even before pinching, exploiting the transition of a gaze-selected target to a hand position (similar to GazeGrab that however used full 3D hand manipulation after the transition [47]). As a result, the object immediately jumps to the correct depth and does not need to be moved after selection. Overall, this can feel natural due to the linear (1:1) movement of the object in depth, but the need for a full arm movement in depth could potentially affect physical effort and task time.

3.1.2 Indirect depth control. This establishes an indirect control of the object depth along the Z-axis which is non-linear to the hand. An example of the technique is shown in Figure 1. Once the pinch gesture is executed, the user can indirectly adjust the depth of the

object by moving their hand for- and backward for corresponding object motion.

3.1.3 Visual Feedback. For both techniques, as visual feedback to aid the interaction and depth perception, a vertical plane (for X-Y) contributes by moving continuously with the hand during pinching. This is inspired by Conductor [49] and Gaze&Wall [42], which used a vertical plane to allow users to intersect two-pointer inputs. Here, the plane is a horizontal plane to indicate the current level of depth, and to provide visual cues for the eyes to land on potential dropping destinations. The plane is partially transparent to allow both to visually fixate at the level of depth, as well as to have a see-through effect to perceive distant targets.

3.1.4 Control-Display (CD) Gain. Indirect input typically uses a non-linear input mapping to improve performance [16, 30, 45]. This is important here, as a gaze-selected object is manipulated through indirect hand motion in the depth dimension. Performing a pinching gesture at full range (1:1 mapping) in front of or besides the body can be challenging because the body can hinder the z-movement of the hand, necessitating more clutching. To experiment with this, we employed a simple non-linear CD Gain function using a polynomial based on prior work [50]:

$$\text{movement}_{\text{obj}} = \text{movement}_{\text{hand}} + \text{movement}_{\text{hand}}{}^{2} \cdot \text{multiplier}$$

Where the multiplier is a parameter. This function supports slow movements, allowing the object's motion to closely match the hand's movement with a CD gain approaching 1:1. At faster movements, the object's motion scales proportionally with the hand's movement. Furthermore, the squared part has an upper-limit to avoid unnecessarily fast movements. This also avoids 'losing' the object during frames where hand tracking is lost. We conducted a pilot study (N=3) with 4 different parameter settings (100 to 250) that approximately resemble variations of a linear 2x gain function (Figure 2). We compared speed and accuracy of the techniques. We set for a parameter of 100 as it resulted in better accuracy while maintaining the same speed as larger ones.

3.2 Dropping Refinement: Look&Drop+

Eye-tracking accuracy varies widely across the population, and can affect the selection performance. The main factors that affect accuracy are jitter (imprecision) and slippage (a systematic offset [3]). To enable precise selection, we designed a new variant that allows gaze to adjust the object's lateral position (X,Y), and parallel hand control (X,Y,Z) is employed for refinement when the dragged object approaches the area of visual focus.

This is inspired by the liberal and conservative MAGIC, where eye/hand input is cascaded to coarse/fine-grained input phases [37, 47, 48]. The conservative approach, which uses hand motion to trigger warping, was not suitable for our use case since hand motion is already needed to adjust object depth in our context. The liberal approach warps the cursor to gaze position if the user's gaze deviates a set threshold away from the current cursor. Our tests showed it was ineffective, as users often over- or undershoot their initial gaze saccade, triggering premature warping and requiring additional manual refinement.

In light of this, we developed Look&Drop+, that does not separate eye-hand inputs in two phases but yet addresses the need for additional refinement. At any time, the object sticks to the user's gaze position. This allows to always instantly warp the target to the point of regard. Any jittery object movement caused by eye tracking inaccuracy is eliminated through applying a filter on the gaze signal. We use a 1€ Filter with parameters set at $f_{c_{min}} = 0.9$ and $\beta = 15$. This results in stabilising the gaze position when users fixate on a target within a moment of time.

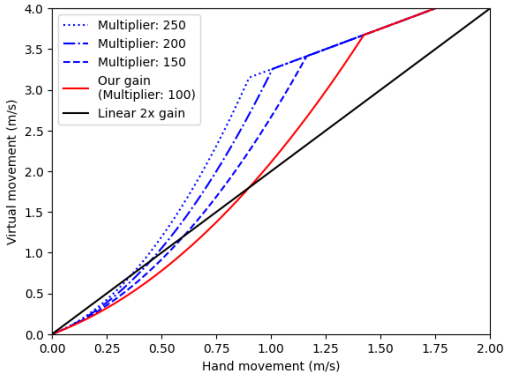


Figure 2: Pilot-tested control-display gain functions, of which one has been used for the study (red).

This leaves mainly eye tracking offset and slippage [3] as error source. This is addressed through lateral hand movement (X,Y), that adds a 1:1 offset to the object's position from the live gaze position. This offset resets when a new fixation falls outside the current fixation area of 4°. This parameter is derived from prior research on MAGIC [5, 48].

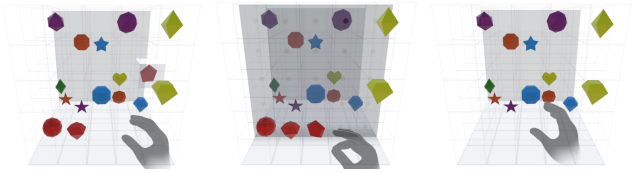
4 3D-GRID APPLICATION EXAMPLES

To demonstrate Look&Drop Techniques in action, we developed an application playground based on a variable 3D grid volume. The application is comprised of cells in which content, e.g. 3D models, can be stored. The cells are discrete object locations, i.e., all targets will snap to one cell after moving. Such a snapping approach has been demonstrated as effective in applications that allow specifically tailored interfaces to the techniques [27, 28]. Overall, this prototype acts as a generic framework that other 3D-Grid applications can be built upon. The size of the volume and cells, are configurable to fit the user's and application's needs. Here we adapt this UI framework to the following examples, as application probes that explore the particular dynamics of eye-hand drag-and-drop. A menu allows switching between the different application examples.

In our prototype, we used a cartesian coordinate system for simplicity and also demonstrate several realistic application scenarios for this where the user is sitting/standing in front of the UI. To support interactions with a user walking around the UI, this could be enhanced by automatic rotation toward the user, and use a spherical polar coordinate system.

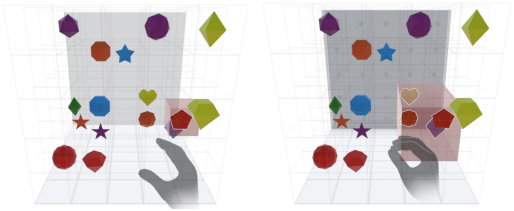
4.1 Gem Game

This application demonstrates the rapid and effortless drag and drop possibility in sequence. The main part of the application is a Drag&Drop based 3D game, similar to the popular browser game



(a) Selection of red gem. (b) Dragging to group of other red gems. (c) Dropping the gem into group.

Figure 3: Drag&Drop operation on a gem into a similarly colored group.



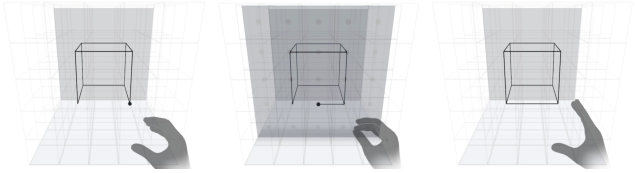
(a) Initial selection of a cell. (b) Dragging increases the selection area.

Figure 4: Selection of four gems using area select.

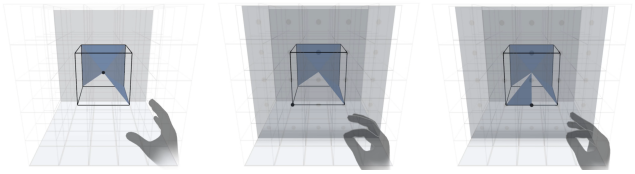
Bejeweled. Gems continuously spawn in random locations in the volume. The objective is to move the gems around, to align 3 of the same color in adjacent cells, which makes them disappear and increases the score. The rate at which new gems spawn increases over time, and the game is over when all cells are filled with gems. Figure 3 shows the Select, Drag and Drop part of moving a gem to a group of similar colored gems. When an object is hovered with gaze, its cell is highlighted, and can be selected with a pinch gesture (Figure 3a). After selection, the transparent wall appears to indicate the current depth of the object, and the selected object can be moved (Figure 3b). Dots are shown on the wall at the cell centers as a visual aid to make it easy to look at cells regardless of the angle to the depth axis. When moving the wall with the hand, it snaps to the center of the cell of the depth layer, and similarly, the selected object, snaps to the centers of the cell where the gaze intersects the wall. When pinch is released, the selected object is dropped, and if it results in a group of 3 or more adjacent gems of the same color, they are removed from the volume (Figure 3c).

4.2 Area Selection

This example illustrates the possibility of using our technique to rapidly select cubic areas within a 3D grid– for follow-up group manipulations such as move, edit, and delete. Multi-object selection is useful in many scenarios, e.g. 3D modelling, interior design, file management, Level design for games, etc. For instance, this method can be considered as the 2D rectangle selection of Shin et al. [33] extended to 3D. It uses two Drag&Drop operations to specify a start and end cell, which makes up the two opposite corners of a rectangular box that is the selection area. Figure 4, shows the steps to form the selection area. First, a pinch gesture is performed anywhere, and the cell at the gaze-wall intersection is highlighted in red. When the desired cell is highlighted, releasing pinch confirms



(a) Specifying first point in line. (b) Dragging second point draws the line. (c) The line is confirmed after realease.



(d) Specifying the first point in triangle. (e) Specifying second point in triangle. (f) Triangle formed with third point.

Figure 5: Lines can be traced with two Drag&Drop operations (a-c). Triangles can be traced with three (d-f).

the first corner (Figure 4a). To increase the selection area, a second pinch gesture is performed, showing the wall again, and the gaze-wall intersection becomes the second corner of the selection area. The selection area is confirmed by releasing pinch, which will select the objects inside (Figure 4b).

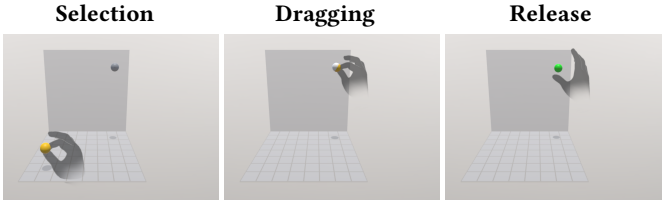
4.3 Point-to-Point Line and Triangle Tracing

This example provides the ability to specify line, triangle, and performed consecutively, it allows the user to create generic 3D shapes. Figure 5 shows both the line and triangle trace modes. For the line, two points are specified with two separate Drag&Drop operations. A visual cue is shown after selection (Figure 5a), from which the line is drawn during until release (Figure 5b). Three points are specified when in triangle tracing mode, where dots are shown after the first (Figure 5d) and second (Figure 5e) Drag&Drop operation. Lastly, the triangle is drawn during the third Drag&Drop operation and confirmed when pinch is released (Figure 5f).

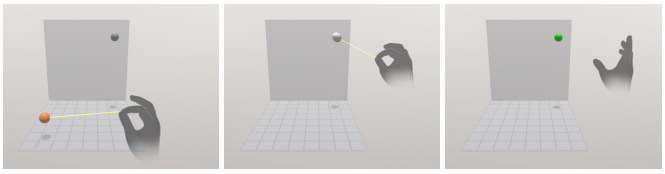
5 USER STUDY

We conduct an empirical user study to formally evaluate the usability of the proposed interaction techniques. Our *baselines* are **direct manipulation** and **Gaze + Pinch** (Figure 6). The *study task* represents an object movement task with diagonal dragging directions similar to prior work [4]. Our *research questions* include:

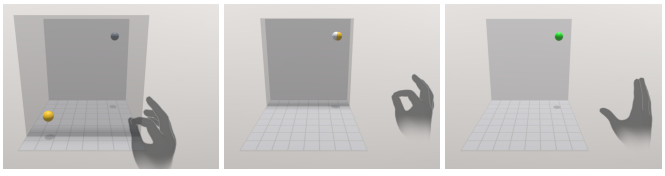
- **Gaze-Assisted vs. Direct:** What are the interaction trade-offs for users when using eye-hand interaction techniques in contrast to direct manipulation? When we inspect prior work, studies of eye-hand interaction techniques (e.g., mouse, controller, and joystick) indicate reduced physical movement and inconclusive findings on task performance [13, 26, 47, 48].
- **Gaze-Assisted vs. Gaze + Pinch:** What are the differences between Gaze + Pinch's indirect hand dragging to the gaze-assisted dragging?
- **Indirect vs. Direct Depth Control:** How does a direct 1:1 mapping for depth control compare to an indirect control-display



(a) DirectPinch Selection. (b) DirectPinch Coarse and Reposition. (c) DirectPinch Release.



(d) Gaze&Pinch Selection. (e) Gaze&Pinch Coarse and Reposition. (f) Gaze&Pinch Release.



(g) Look&Drop Selection. (h) Look&Drop Coarse and Reposition. (i) Look&Drop Release.

(j) Look&Drop+ Techniques Selection. (k) Look&Drop+ Techniques Reposition (l) Look&Drop+ Techniques Release.

Figure 6: Examples of the DirectPinch operation in the user study (a-c), Gaze&Pinch operations (d-f), Look&Drop Techniques (g-i) and Look&Drop+ Techniques (j-l)

gain-based depth control (matching gaze movement to object movement)?

- **Precision Enhancement in Look&Drop+:** How does pointing with Look&Drop+ for specifying the drop position of the object affect the user performance compared to eyes-only specification?

We recruited 24 participants (7 female, 17 male) of mixed backgrounds and technical expertise. The age ranged from 19 and 35 ($M = 25.62, SD = 3.28$), 21 were right-handed, 3 were left-handed, 9 wore glasses. On a scale between 1 and 5, participants rated themselves as moderately experienced with AR/VR/XR ($M = 3.37, SD = 1.01$), 3D hand gestu res ($M = 2.87, SD = 0.99$) and eye-gaze interaction ($M = 2.25, SD = 0.99$). The software were implemented using the OVR toolkit in Unity3D on the Meta Quest Pro (106°x95.57° FOV, 1800x1920 pixels per eye, eye and hand tracking). Eye-tracking accuracy is reported as around 1.5-3° [2, 44]. We use a 1€ Filter to smooth gaze and hand movements, with parameters for gaze set at $fc_{min} = 0.9$ and $\beta = 15$, and for hand movements at $fc_{min} = 0.9$ and $\beta = 90$ based on [49]. The Gaze+Pinch technique uses the user's gaze to point and a pinch gesture (close-open) to confirm, using the MRTK solution with default parameters [43].

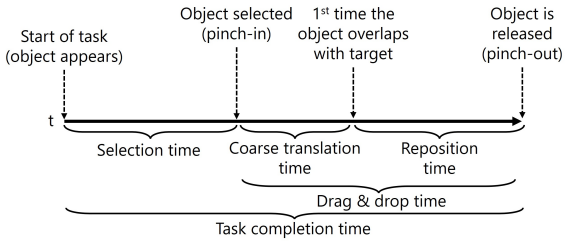


Figure 7: The Drag&Drop task involves several subcomponents that we analyse individually in the experiment.

5.1 Task

Figure 6 illustrates the task where users selected an object and moved it to a corresponding target. Targets appeared in a cubic volume slightly below eye level for comfort and visibility. A new object and target appeared after each trial. A floor grid displayed target shadows for depth perception, and a visible back wall defined the volume's end. Following prior work [4], the object and target were placed randomly in diagonally opposite corners. Users completed 8 repetitions, dragging the object as quickly and accurately as possible to each target. If object selection failed after 30 seconds, a new object replaced it (timed-out attempt). Visual feedback guided users: the target was initially gray, the object orange. Looking at an object turned it yellow (gaze-hover), a successful drop made the target green, and errors turned it red. Objects and targets were tested in two sizes 3° (0.0295m) and 5° (0.0495m), at varying distances (distances: 15cm and 35cm), with conditions randomized within each task block.

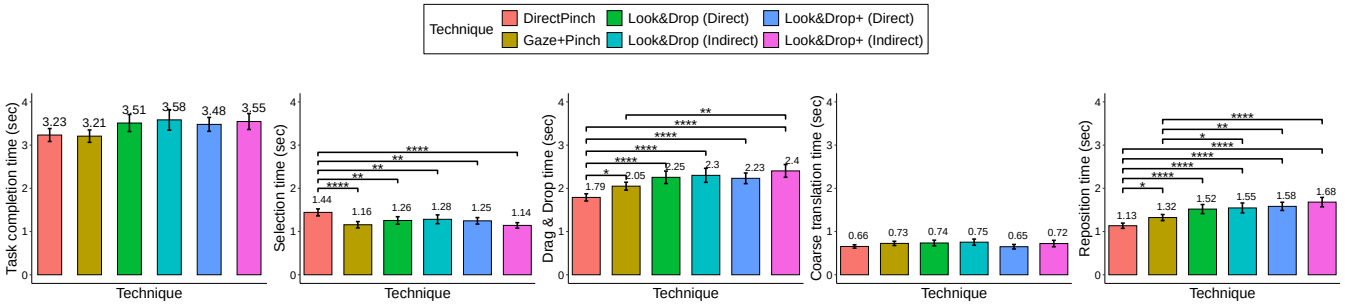
To ensure the balanced distribution we counterbalanced the 6 techniques, while distances and sizes were randomised in each block. Overall, we had 24 *participants* × 6 *techniques* × 2 *distances* × 2 *sizes* × 8 *repetitions* = 4608 data points. Table 1 shows the visual angle for each start position of the dragging object.

Distance (m)	Target Size (m)	Back-top (°)	Back-bottom (°)	Front-top (°)	Front-bottom (°)
0.35	0.05	5.4	5.0	8.6	7.4
0.35	0.03	3.3	3.0	5.2	4.4
0.15	0.05	6.2	5.9	7.5	7.1
0.15	0.03	3.7	3.6	4.5	4.3

Table 1: Study target angles for different conditions

5.2 Procedure

The study participants were sitting in a large, quiet room during the study. At the beginning of the study, participants were briefed about the study and asked to fill out the consent form and demographic questionnaire. The headset was introduced, participants adjusted the headset optimally, and went through an eye tracking calibration. Following this, participants were familiarized with the techniques. Users completed an 8-trial training session with random conditions at the start of each new technique. Learning typically stabilised



(a) TCT. (b) ST. (c) D&DT. (d) CTT. (e) RT.

Figure 8: User study results across time factor for each technique. Error bars denote 95% confidence intervals.

after 2-3 trials. After providing all instructions to the participants, the study began. Participants were instructed to complete the task as fast and accurately as possible. After each run with a technique, participants were allowed to remove the headset and asked to fill out the questionnaire. During this time, participants were given a break of approximately 2-3 minutes. After testing all six techniques, participants ranked them. The entire study lasted approximately 50 minutes.

5.3 Evaluation Metrics

We analyse the following metrics. We decompose the task time into subcomponents, as illustrated in Figure 7.

- **Task Completion Time (TCT)**: Time from object appearance to pinch release.
- **Selection Time (ST)**: Time from object appearance to selection by pinch-in. This is relevant as follow-up manipulations can impact selection times [41].
- **Drag&Drop Time (D&DT)**: The time from selection to release.
- **Coarse Translation Time (CTT)**: The time from selection until the object first approaches within 0.05m of the target.
- **Reposition Time (RT)**: The remaining time after CTT until release.
- **Error Rate (ER)**: The trial is an error if timeout occurs, or the object's center is not within the target's radius.
- **Dropping Accuracy (AC)**: Accuracy is measured as the angle between head-object and head-target rays at the end of the trial.
- **Hand Movement (HM)**: The accumulated difference in palm position between frames serves as the metric for this evaluation, as an indicator for physical effort [43, 47].
- **Preference and Subjective Task Load**: For each condition, we used the NASA-TLX (Task Load Index) questionnaire [11] and questions on about eye and hand fatigue, and rankings at the end of the study.

6 RESULTS

In our statistical analysis, we tested the quantitative variables for normality and applied Box-Cox [8] data transformations for non-normally distributed factors where appropriate. In the following analysis of performance-related measures (sections 6.1 - 6.3), we removed 5 trials (0.11%) from the data due to timeouts (no object was selected and/or dropped at the target within 30s). Further, we excluded 86 trials (1.87%) as outliers (TCT > $Mean + 3 \times SD$). We conducted a three-way repeated measures ANOVA (Technique × Distance × Target Size) for the quantitative data (with Greenhouse-Geisser corrections if sphericity was violated), followed by estimated marginal means post-hoc pairwise comparisons with Bonferroni corrections. For Likert-scale data (section 6.5), we used a Friedman test with Bonferroni corrected post-hoc Conover tests. For brevity, we report statistically significant effects wrt. the factor technique. In all figures, we denote statistical significance with * for $p < .05$, ** for $p < .01$, *** for $p < .001$, and **** for $p < .0001$.

6.1 Task Duration (Figure 8, Figure 10)

No significant effects were reported for ***TCT*** ($F^{3.37}_{74.06} = 1.93$, p=.125, $\eta^2_G = 0.024$), indicating no significant penalty between using the different interaction techniques for the overall task.

Regarding ***ST*** ($F^{3.64}_{80.03} = 6.54$, p<.0001, $\eta^2_G = 0.098$), we find that users were slower in selecting objects with DirectPinch than with Gaze&Pinch, Look&Drop$_D$, Look&Drop$_I$, Look&Drop+$_D$, and Look&Drop+$_I$ (p<.0001). Affirming prior eye-hand selection research, we contribute to the finding that eye-hand input is faster for selection than direct manipulation [20, 43].

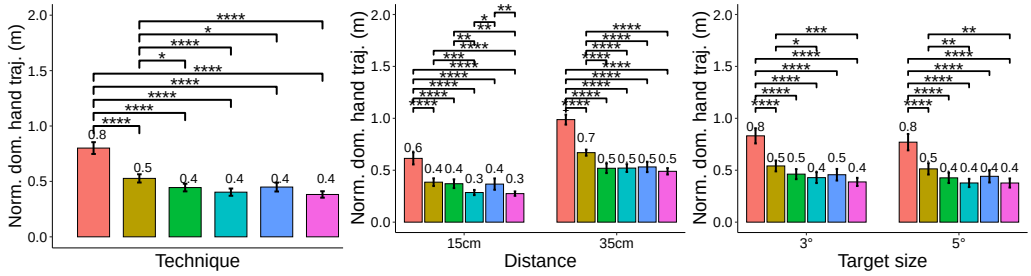
Significant interaction effects were found for Technique × Target Size (F^5_{110}=3.811,p =.003, $\eta^2_G = 0.017$), and Technique×Distance ×Target Size (F^5_{110}=4.48, p<.0001, $\eta^2_G = 0.012$).

With DirectPinch users were significantly slower for selecting objects at a **short Distance of** 15**cm** (p<.001) than with Gaze&Pinch and Look&Drop+$_I$. DirectPinch was also slower for 35**cm** (p<.001) compared to Gaze&Pinch, Look&Drop+$_D$ and Look&Drop+$_I$.

Regarding **Target Size**, all gaze-based techniques were significantly faster in selecting objects with **3°** than DirectPinch (p<.0001). That suggests that smaller objects are more difficult to select directly by hand. For a **Target Size of 5°** Gaze&Pinch was also faster than DirectPinch (p<.01).

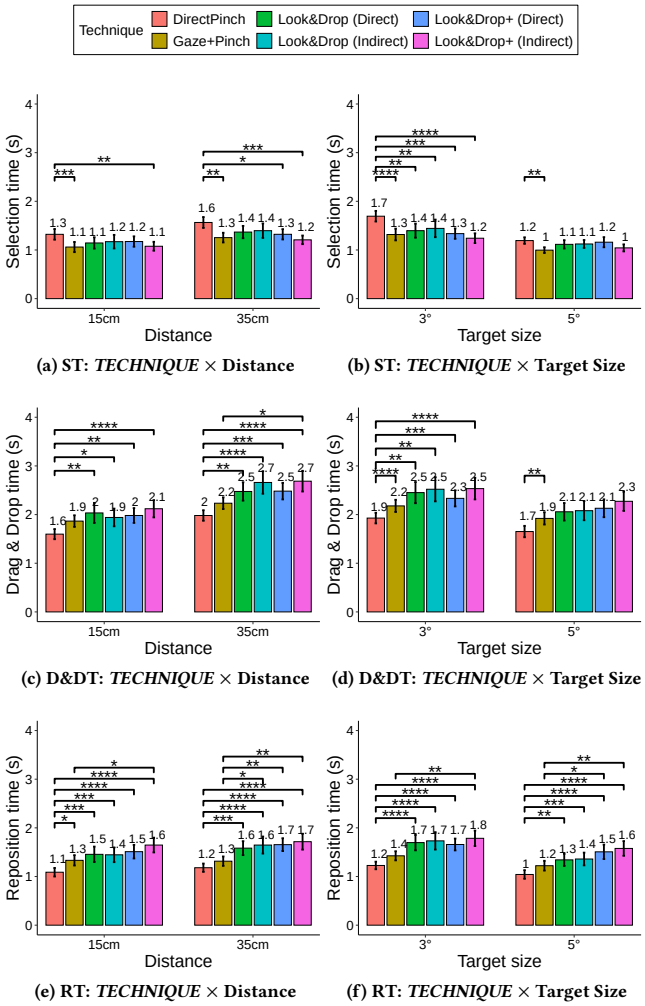
In contrast to ST, our analysis of ***D&DT*** ($F^{3.28}_{72.06} = 9.11$, p<.0001, $\eta^2_G = 0.105$) shows that users were significantly faster in drag and dropping objects with DirectPinch (p<.0001) compared to all the gaze-based techniques. However, Gaze&Pinch was still faster than Look&Drop+$_I$ (p<.01).

Significant interaction effects were found for Technique × Distance ($F^{3.34}_{73.48}$=4.05,p=.008, $\eta^2_G = 0.008$) and Technique × Target Size (F^5_{110}=2.84,p=.019, $\eta^2_G = 0.004$).



(a) Hand Movement *TECHNIQUE*. (b) HM *TECHNIQUE*× Distance. (c) HM *TECHNIQUE*× Target Size.

Figure 9: User study results across techniques for further factors. Error bars denote 95% confidence intervals.



(a) ST: *TECHNIQUE* × Distance (b) ST: *TECHNIQUE* × Target Size

(c) D&DT: *TECHNIQUE* × Distance (d) D&DT: *TECHNIQUE* × Target Size

(e) RT: *TECHNIQUE* × Distance (f) RT: *TECHNIQUE* × Target Size

Figure 10: Mean Selection Time (a, b), Drag&Drop Time (c, d) and Reposition Time (e, f) with significant post-hoc tests. Error bars denote 95% confidence intervals.

With DirectPinch users were significantly faster drag and dropping over a **Distance of** 15**cm** (p<.0001) than with Look&Drop$_D$, Look&Drop$_I$, Look&Drop+$_D$, and Look&Drop+$_I$. DirectPinch was also faster for 35**cm** (p<.0001) compared to Look&Drop$_D$, Look&Drop$_I$, Look&Drop+$_D$ and Look&Drop+$_I$. However, Gaze&Pinch was still faster than Look&Drop+$_I$.

For **Target Size 3°** users were significantly faster with DirectPinch (p<.0001) than all gaze-based techniques. Also for **Target Size 5°** DirectPinch was just faster than Gaze&Pinch (p<.01).

We found no significant effect for ***CTT*** ($F^{3.09}_{67.92}$ = 1.666,p=.181, η^2_G = 0.025). Our findings for ***RT*** (F^5_{110} = 17.15, p<.0001, η^2_G = 0.153) follow the previous pattern: users were faster in repositioning objects at the target with DirectPinch (p<.0001) compared to all gaze-based techniques. However, users spent less time repositioning objects with Gaze&Pinch (p<.0001) than Look&Drop$_I$, Look&Drop+$_D$ and Look&Drop+$_I$.

Significant interaction effects were found for Technique × Distance (F^5_{110}=2.98,p=.015, η^2_G = 0.006) and Technique × Target Size (F^5_{110}=3.771,p =.003, η^2_G = 0.006).

For the **Distance of** 15**cm**, users were significantly faster in repositioning objects at the target with DirectPinch (p<.0001) compared to all gaze-based techniques. However, Gaze&Pinch (p<.05) was still faster than Look&Drop+$_I$. At a **Distance of** 35**cm** DirectPinch (p<.0001) was also significantly faster than Look&Drop$_D$, Look&Drop$_I$, Look&Drop+$_D$ and Look&Drop+$_I$. Furthermore, user reached faster RT with Gaze&Pinch (p<.01) compared to Look&Drop$_I$, Look&Drop+$_D$ and Look&Drop+$_I$.

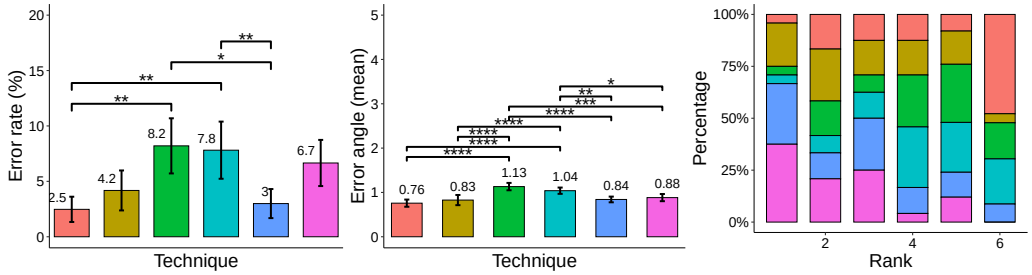
For **Target Size 3°**, DirectPinch (p<.0001) was significantly faster for repositioning objects compared to Look&Drop$_D$, Look&Drop$_I$, Look&Drop+$_D$ and Look&Drop+$_I$. Gaze&Pinch (p<.01) was also faster than Look&Drop+$_I$. Similarly, for **Target Size 5°**, DirectPinch (p<.0001) was faster in the Reposition task compared to Look&Drop$_D$, Look&Drop$_I$, Look&Drop+$_D$ and Look&Drop+$_I$. Furthermore, Gaze&Pinch (p<.01) was faster compared to Look&Drop+$_D$ and Look&Drop+$_I$.

6.2 Error Rate (Figure 11)

ER ($F^{2.66}_{58.62}$ = 3.54, p=.024, η^2_G = 0.044), was significantly lower with DirectPinch compared to Look&Drop$_D$ and Look&Drop$_I$ (p<.01). Also Look&Drop+$_D$ had a significantly lower ER compared to Look&Drop$_D$ and Look&Drop$_I$ (p<.05).

6.3 Hand Movement (Figure 9)

In HM ($F^{3.48}_{76.57}$ = 47.54, p<.0001, η^2_G = 0.461), we find that users moved their dominant hand significantly more with DirectPinch



(a) ER: *TECHNIQUE*. (b) Dropping Accuracy. (c) Preference ranking.

Figure 11: (a) Results on error rate across techniques. Error bars denote 95% confidence intervals. (b) Results on dropping accuracy at trial end. (c) Subjective preference ranking of each technique. A lower rank is better.

compared to all gaze-based techniques (p<.0001). Also Gaze&Pinch required more hand movement than Look&Drop$_D$, Look&Drop$_I$, Look&Drop+$_D$, and Look&Drop+$_I$ (p<.0001).

Significant interaction effects were found for Technique × Distance (F^5_{110}=8.316,p <.0001, η^2_G = 0.028) and Technique×Distance ×Target Size (F^5_{22}=6.48, p=.018, η^2_G = 0.002).

At 15*cm* **Distance**, there was significantly more hand movement with Look&Drop+$_D$, Look&Drop$_D$, Gaze&Pinch, and DirectPinch compared to Look&Drop$_I$ and Look&Drop+$_I$ (p<.0001). For 35*cm* **Distance**, DirectPinch resulted in significantly more hand movement than all the gaze-based techniques (p<.0001). Furthermore, Gaze&Pinch required more movement than Look&Drop$_D$, Look&Drop$_I$, Look&Drop+$_D$ and Look&Drop+$_I$ (p<.001).

For both **Target Sizes** (3°, 5°), DirectPinch resulted in significantly greater hand movement compared to all gaze-based techniques (p<.0001). Further, Gaze&Pinch required more hand movement than Look&Drop$_I$ and Look&Drop+$_I$ (p<.001).

6.4 Dropping Accuracy (Figure 11)

The analysis of accuracy (angle between head-object and head-target rays) at trial end confirms that Look&Drop+ helps with refining the final positioning of the object within the target area, making it more tolerant to positional offsets of the gaze cursor at the end of dropping action induced by late triggers. Statistical analysis of accuracy reveals significant differences (F^5_{110} = 9.58, p<.0001, η^2_G = 0.135). Overall users were more accurate in dropping objects with DirectPinch (0.76°, SD=0.40), Gaze&Pinch (0.83°, SD=0.56), Look&Drop+$_D$ (0.84°, SD=0.31) and Look&Drop+$_I$ (0.88°, SD=0.39) than with Look&Drop$_D$ (1.13°, SD=0.40) and Look&Drop$_I$ (1.04°, SD=0.34)(p<.0001)

6.5 Preference and Task Load (Figure 11-12)

Based on participants' rankings of preference, Look&Drop+$_D$ (37.50%), Look&Drop+$_I$ (29.17%) and Gaze&Pinch (20.83%) were the most often indicated as favorite, with Look&Drop$_I$ (4.17%), Look&Drop$_D$ (4.17%) and DirectPinch (4.17%) bringing up the rear (see Figure 11).

Statistical analysis of fatigue ratings reveal significant differences for eye fatigue ($\chi^2(5)$ = 39.24, p<.0001, W = 0.327) and arm fatigue ($\chi^2(5)$ = 53.49, p<.0001, W = 0.446). DirectPinch was perceived as less fatiguing for the eyes compared to Look&Drop$_D$ (p<.0001) and Look&Drop$_I$ (p<.0001). In contrast however, DirectPinch was perceived as more fatiguing for the hands compared to all gaze-based techniques (p<.0001). This is mirrored by NASA TLX scores on physical demand ($\chi^2(5)$ = 39.06, p<.0001, W = 0.325), according to which, manual DirectPinch was perceived as more demanding compared to all gaze-based techniques (p<.0001). Further, scores on mental demand ($\chi^2(5)$ = 26.13, p<.0001, W = 0.218) show that DirectPinch was perceived as less challenging than Look&Drop$_D$ (p<.0001) and Look&Drop$_I$ (p<.0001).

6.6 User Feedback

DirectPinch was perceived as intuitive and simple but considered the most fatiguing overall. For example, P16 stated *'...pretty tiresome to do for my hand/arm. I could feel I could not do that for many more minutes without pause'*. Additionally, participants found it challenging to select smaller objects using this technique: *'... it was hardest for me to select the orange ball'* (P20).

With **Gaze&Pinch**, users appreciated the combination of hand and gaze interaction (P4: *'I like to use my hands after selecting since it gives me more control over what is happening'*). Despite occasional tracking issues and acceleration-induced errors, Gaze&Pinch was generally perceived as intuitive and familiar. Some users had difficulties moving the object precisely, especially due to hand positioning and acceleration challenges (P21: *'the object was dragged from an awkward angle because of the start position of my hands...'*).

Despite feeling quick, **Look&Drop$_D$** was described as mentally demanding and tiring compared to Look&Drop+ (P3: *'... no fine-tune method making it quite painful to work with, increasing the cognitive load'*). Most users found the release process fatiguing due to difficulties in accurately placing objects using gaze alone.

With **Look&Drop$_I$**, the acceleration feature was seen as helpful in reducing mental workload and enabling natural hand movement. However, its high speed often led to overshooting issues, as described among others by P16: *'it often resulted in overshooting the target, which required some "back and forth" correction before placing...'*.

With **Look&Drop+$_D$** users remarked that it required some learning, but was intuitive and effective afterwards (P3: *'Although less intuitive, it can be learned in minutes and performs better when used to'*). They appreciated the speed of eye-tracking for larger movements and the precision of hand-tracking for depth control. The direct ability to snap to hand level improved control, e.g., P5

stated: *'...it felt more convenient when the wall comes to the position of my hand immediately, instead of staying far'*.

Users found **Look&Drop+$_I$** more efficient than Gaze&Pinch in moving objects, with some appreciating its fine-tuning feature and overall enjoyable experience. E.g. P3 stated, *'... better than just Gaze&Pinch, more efficient on the moving stuff part'*. However, the tendency to overshoot the target was frustrating for some users, affecting their preference for the technique. E.g., P7 emphasised: *'...my least favourite because the object tended to move so much which was hard to focus and irritating.'*

6.7 Main Findings

To provide an overview, we summarise our main findings as follows.

- Overall & Coarse Translation Time: No significant differences.
- Selection Time: All gaze techniques (faster) < than DirectPinch.
- Drag&Drop Time: All gaze techniques (slower) > DirectPinch.
- Reposition Time: All gaze techniques (slower) > DirectPinch; Gaze&Pinch < Look&Drop$_I$, Look&Drop+$_D$ and Look&Drop+$_I$.
- Error Rate: DirectPinch < Look&Drop$_D$ and Look&Drop$_I$; Look&Drop+$_D$ < Look&Drop$_D$.
- Hand Movement: Look&Drop < Gaze&Pinch < DirectPinch.
- Physical effort: All gaze techniques (lower) < DirectPinch.

7 DISCUSSION

This study investigated the impact of integrating eye-tracking technology on direct manipulation interfaces, considered a natural user interface paradigm. We evaluate its effectiveness as an alternative interaction method in near-space environments. Our findings from drag-and-drop tasks reveal no significant difference in overall task completion time between traditional direct manipulation and gaze-assisted techniques.

However, if we decompose to sub-tasks, we find a trade-off: gaze-based techniques are faster in selecting objects, but slower in the object movement. We know that gaze can enhance the selection speed of gestural ray-pointing [20, 43], but we show that eye-hand input can surpass even direct 3D input for selection. This is a significant result, demonstrating that gaze is not only useful for remote targets but also for targets in physical reach. Note that selection times can in principle go further down when the task is not followed by a manipulation [41]. Future research could explore this aspect for a more nuanced understanding.

The compromise is the object movement. This is somewhat surprising, as the act of dragging is in a way a two-step selection. However, adding the third dimension for manipulation renders the task more complex. Direct gestures allow for skilled implicit understanding of 3D interaction through the affordances of the user's hand, whereas with gaze we split up the 3D task structure into separate 2D gaze and 1D hand commands. This may have rendered the task more complex and could have contributed to negating a potential speed benefit of the eyes. Further, while the hand-based operation is reduced to a 1D task, the actual physical travel distance remains almost the same in principle. As we found out, direct manipulation required about 0.8 meters object movement on average on each trial, while all gaze-based techniques were at 0.4-0.5 meters. We showed that the difference can be attributed to the initial hand movement of direct input for the selection. In contrast to Gaze + Pinch, we can compare actual indirect movement time to the gaze-assisted dragging. We see a significant difference of 0.1 meters more for Gaze + Pinch, showing that the dimensional reduction of Look&Drop only saved about 20% of the actual physical hand movement. This is plausible, as the effective physical distance for Gaze&Pinch of the 3D hand path is slightly longer than the 1D depth path. Notably, these findings apply to the conditions tested in our study. When considering manipulations across a larger 3D space, where more movement is required in the XY than in the Z axis, gaze-based dragging will likely become more advantageous.

Regarding the temporal performance for Look&Drop versus Gaze&Pinch, it depends: No differences were found wrt. Look&Drop$_D$, but overall Gaze&Pinch showed a trend of being faster than Look&Drop across specific levels of distance and size. Quantitatively, this indicates a trade-off between physical and temporal movement. However, looking at the user feedback, we can clearly see the preference for Look&Drop techniques over both baselines. At the end of the study 75% of users said they preferred to use one of the Look&Drop techniques, with most of those being the techniques with precision refinement by the hand. Why is that? While the distinction to direct manipulation was noted by all users to revolve around physical effort, Gaze + Pinch was not considered strainful. Yet, both Look&Drop+ techniques were favoured for the perceived intuitiveness of object dragging compared to other methods. This finding expands the applicability of precision-enhancing methods to near-space eye-hand interaction, demonstrating its effectiveness beyond the indirect input use cases investigated so far.

There is, however, one caveat: in our study, the target sizes were relatively small at 3° and 5°. Given that Aziz et al. [3] report eye-tracking errors of up to 3° and that the eye-tracking resolution of our HMD can go up to 3°, a minimum resolution of 6° would have been necessary for our users to experience the techniques with theoretically-ideal eye-tracking input. However, we opted to use 3° as a minimum to be consistent with prior work [43] and facilitate comparison, and as it is possible with our main baseline of direct manipulation. In sum, while we showed clear benefits of Look&Drop for physical effort without time penalties, better sensing the techniques may lead to an even more improved user performance, which demands further study.

Further, our initial hypothesis that CD gain will improve the performance is rejected, as no significant differences were found, moreover, users reported that it led to overshooting. This may have resulted from our specific choice of CD gain and alternatives may improve the performance (like it does for the mouse), however, it is beyond our scope to identify an ideal mapping.

Lastly, we note that both Look&Drop$_D$ and Look&Drop$_I$ were clearly less preferred than the Look&Drop+ variations and also led to a high error rate, which we associate with eye slippage and calibration offset [2]. The object must be moved to the target and held there with the eyes – but the estimated gaze point is not always within the target's boundary. This problem was essentially solved by the Look&Drop+ technique, as users could offset the object's position from the gaze point, leading to the superior results than Look&Drop and preference over all tested techniques.

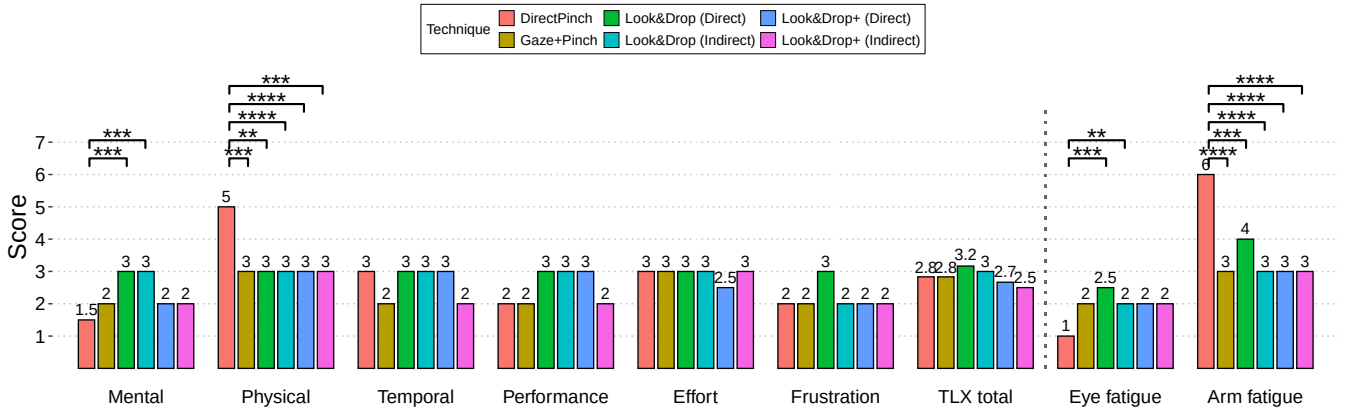


Figure 12: Median results from NASA-TLX and subjective eye and arm fatigue questionnaires.

However, as we noted before, eye-tracking quality should be carefully considered, as more precise sensors for pointing with the eyes may resolve this issue. Moreover, methods that can distinguish slippery from intentional eye movements would allow to further prevent errors, relating to the long-standing Midas Touch problem [17]. Such advances could make it an instant drop, rather than a trial and error to get the estimated gaze indication to remain at the right position. In this context, we refer to our application examples that are specifically designed to account for the system's tracking range. As we demonstrate in our video, an extremely fast and fluid eye-hand manipulation is possible with techniques that do not even support additional precision methods, pointing toward more advanced eye-hand manipulation capabilities in future advanced systems.

8 CONCLUSION

This work contributes a detailed comparison of direct-to-eye-hand manipulation techniques in near space, where direct manipulation is typically seen as the primary input medium. Yet, we find the novel Look&Drop interactions are on par with the common drag-and-drop task. The techniques significantly reduce selection time, but increase dragging time, thereby negating each other's temporal benefits and drawbacks in task completion time. However, given a substantial reduction of physical effort and overall preference by most users, there are clear indicators for the utility of Look&Drop. All in all, our work points to future extensions of eye and hand tracking interfaces, where the status-quo of direct manipulation may well be supported by hybrid eye-hand interactions, even for complex manipulation tasks.

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).

REFERENCES

[1] Ferran Argelaguet and Carlos Andujar. 2013. Special Section on Touching the 3rd Dimension: A survey of 3D object selection techniques for virtual environments. *Comput. Graph.* 37, 3 (may 2013), 121–136. https://doi.org/10.1016/j.cag.2012.12.003

[2] Samantha Aziz and Oleg Komogortsev. 2022. An Assessment of the Eye Tracking Signal Quality Captured in the HoloLens 2. In *2022 Symposium on Eye Tracking Research and Applications* (Seattle, WA, USA) *(ETRA '22)*. Association for Computing Machinery, New York, NY, USA, Article 5, 6 pages. https://doi.org/10.1145/3517031.3529626

[3] Samantha Aziz, Dillon J Lohr, Lee Friedman, and Oleg Komogortsev. 2024. Evaluation of Eye Tracking Signal Quality for Virtual Reality Applications: A Case Study in the Meta Quest Pro. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications* (Glasgow, United Kingdom) *(ETRA '24)*. Association for Computing Machinery, New York, NY, USA, Article 7, 8 pages. https://doi.org/10.1145/3649902.3653347

[4] Ravin Balakrishnan and Gordon Kurtenbach. 1999. Exploring bimanual camera control and object manipulation in 3D graphics interfaces. 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, 56–62. https://doi.org/10.1145/302979.302991

[5] 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

[6] 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

[7] Doug A. Bowman, Donald B. Johnson, and Larry F. Hodges. 1999. Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (London, United Kingdom) *(VRST '99)*. Association for Computing Machinery, New York, NY, USA, 26–33. https://doi.org/10.1145/323663.323667

[8] G. E. P. Box and D. R. Cox. 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26, 2 (1964), 211–252. http://www.jstor.org/stable/2984418

[9] 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

[10] Di Laura Chen, Marcello Giordano, Hrvoje Benko, Tovi Grossman, and Stephanie Santosa. 2023. GazeRayCursor: Facilitating Virtual Reality Target Selection by Blending Gaze and Controller Raycasting. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology* (Christchurch, New Zealand) *(VRST '23)*. Association for Computing Machinery, New York, NY, USA, Article 19, 11 pages. https://doi.org/10.1145/3611659.3615693

[11] 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

[12] Barrett M. Ens, Rory Finnegan, and Pourang P. Irani. 2014. The personal cockpit: a spatial interface for effective task switching on head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3171–3180. https://doi.org/10.1145/2556288.2557058

[13] Ribel Fares, Shaomin Fang, and Oleg Komogortsev. 2013. Can we beat the mouse with MAGIC?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1387–1390. https://doi.org/10.1145/2470654.2466183

[14] Devamardeep Hayatpur, Seongkook Heo, Haijun Xia, Wolfgang Stuerzlinger, and Daniel Wigdor. 2019. Plane, Ray, and Point: Enabling Precise Spatial Manipulations with Shape Constraints. 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, 1185–1195. https://doi.org/10.1145/3332165.3347916

[15] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1063–1072. https://doi.org/10.1145/2556288.2557130

[16] Ken Hinckley. 2007. Input technologies and techniques. In *The human-computer interaction handbook*. CRC Press, 187–202.

[17] 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

[18] Jaejoon Jeong, Soo-Hyung Kim, Hyung-Jeong Yang, Gun A Lee, and Seungwon Kim. 2023. GazeHand: A Gaze-Driven Virtual Hand Interface. *IEEE Access* 11 (2023), 133703–133716.

[19] 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

[20] 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

[21] Pavel Manakhov, Ludwig Sidenmark, Ken Pfeuffer, and Hans Gellersen. 2024. Gaze on the Go: Effect of Spatial Reference Frame on Visual Target Acquisition During Physical Locomotion in Extended Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 373, 16 pages. https://doi.org/10.1145/3613904.3642915

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

[23] 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

[24] Donald A Norman. 2010. Natural user interfaces are not natural. *interactions* 17, 3 (2010), 6–10.

[25] Siyou Pei, David Kim, Alex Olwal, Yang Zhang, and Ruofei Du. 2024. UI Mobility Control in XR: Switching UI Positionings between Static, Dynamic, and Self Entities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 611, 12 pages. https://doi.org/10.1145/3613904.3642220

[26] 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

[27] 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

[28] 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

[29] 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

[30] 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

[31] 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.

[32] 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

[33] Rongkai Shi, Yushi Wei, Xueying Qin, Pan Hui, and Hai-Ning Liang. 2023. Exploring Gaze-Assisted and Hand-Based Region Selection in Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 7, ETRA, Article 160 (may 2023), 19 pages. https://doi.org/10.1145/3591129

[34] Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch: Gaze-Supported Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2981–2990. https://doi.org/10.1145/2207676.2208709

[35] Sophie Stellmach and Raimund Dachselt. 2013. Still looking: investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 285–294. https://doi.org/10.1145/2470654.2470695

[36] 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

[37] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 4179–4188. https://doi.org/10.1145/2702123.2702355

[38] Jayson Turner, Jason Alexander, Andreas Bulling, Dominik Schmidt, and Hans Gellersen. 2013. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *Human-Computer Interaction–INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part II 14*. Springer, 170–186.

[39] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2013. Eye drop: an interaction concept for gaze-supported point-to-point content transfer. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (Luleå, Sweden) *(MUM '13)*. Association for Computing Machinery, New York, NY, USA, Article 37, 4 pages. https://doi.org/10.1145/2541831.2541868

[40] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2014. Cross-device gaze-supported point-to-point content transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, Florida) *(ETRA '14)*. Association for Computing Machinery, New York, NY, USA, 19–26. https://doi.org/10.1145/2578153.2578155

[41] 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

[42] Uta Wagner, Matthias Albrecht, Haopeng Wang, Ken Pfeuffer, and Hans Gellersen. 2024. Gaze, Wall and Racket: Combining Gaze and Hand-controlled Plane for 3D Selection in Virtual Reality. In *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)* (Vancouver, BC) *(ISS '24)*. Association for Computing Machinery, New York, NY, USA.

[43] 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* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3581423

[44] Shu Wei, Desmond Bloemers, and Aitor Rovira. 2023. A Preliminary Study of the Eye Tracker in the Meta Quest Pro. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences* (Nantes, France) *(IMX '23)*. Association for Computing Machinery, New York, NY, USA, 216–221. https://doi.org/10.1145/3573381.3596467

[45] Curtis Wilkes and Doug A. Bowman. 2008. Advantages of velocity-based scaling for distant 3D manipulation. In *Proceedings of the 2008 ACM Symposium on Virtual*

Reality Software and Technology (Bordeaux, France) *(VRST '08)*. Association for Computing Machinery, New York, NY, USA, 23–29. https://doi.org/10.1145/1450579.1450585

[46] Yukang Yan, Chun Yu, Xiaojuan Ma, Shuai Huang, Hasan Iqbal, and Yuanchun Shi. 2018. Eyes-Free Target Acquisition in Interaction Space around the Body for Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Montreal QC</city>, <country>Canada</country>, </conf-loc>) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173616

[47] 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

[48] 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

[49] Futian Zhang, Keiko Katsuragawa, and Edward Lank. 2022. Conductor: Intersection-Based Bimanual Pointing in Augmented and Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 6, ISS, Article 560 (nov 2022), 15 pages. https://doi.org/10.1145/3567713

[50] Qian Zhou, George Fitzmaurice, and Fraser Anderson. 2022. In-Depth Mouse: Integrating Desktop Mouse into Virtual Reality. 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 354, 17 pages. https://doi.org/10.1145/3491102.3501884