Preshaping Hand Behaviour for Direct and Indirect Manipulation of 3D Objects

Thorbjørn Mikkelsen
Aarhus University
Aarhus, Denmark
thormik@cs.au.dk

Qiushi Zhou
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
qiushizhou@hkust-gz.edu.cn

Mar Gonzalez-Franco
Google,
Seattle, U.S.A.
margon@google.com

Hans Gellersen
Lancaster University
Lancaster, U.K.
Aarhus University
Aarhus, Denmark
h.gellersen@lancaster.ac.uk

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

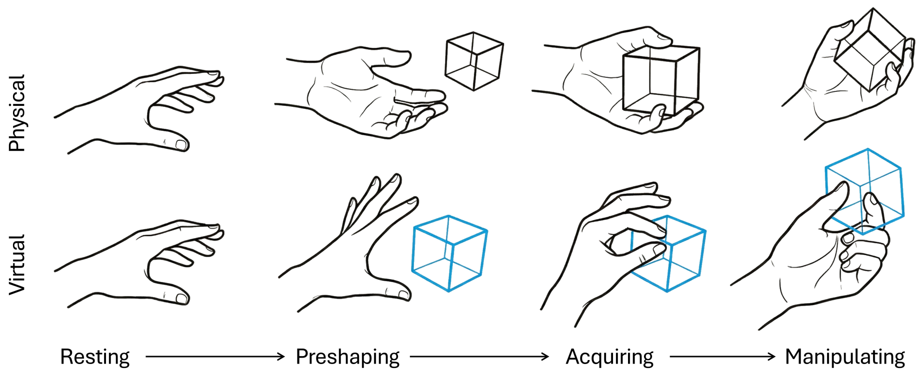

Figure 1: Efficient manipulation relies on effective preshaping: the subconscious adaptation of the hand to an object before contact. Preshaping is informed by the anticipated requirements of the task, and expression thus differs when the task does. Physical hand-object interactions (top) are tuned to accommodate predicted mass and scale. In virtual interactions (bottom), the hand is freed from such concerns, but is governed by the interface. For example, abstractions like the pinch gesture replace natural grasping with object-agnostic poses decoupled from the visual geometry, altering how users acquire and control objects.

Abstract

Effortless manipulation informs and relies on preshaping: the subconscious posing of the hand before grasping. Virtual environments and the design of interaction techniques alters interaction requirements like contact and reach, forcing behavioural adaptation. We present a comparative study investigating preshaping behaviour across direct versus indirect (gaze-assisted) and bare-hand versus controller techniques on a docking task. Results reveal that response patterns scale with anticipated task difficulty, and that direct techniques elicit effective posing of the hand. Indirect techniques shortcut hand transport and in turn lacks the sensory feedback to guide planning, inducing efficient but attenuated responses that necessitate compensatory manipulation and clutching. Notably, controllers that afford in-hand rotation allow users to extend their range of motion. These findings can inform interaction design to better afford preshaping and optimise 3D manipulation tasks.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3793424

CCS Concepts

• **Human-centered computing** → **Gestural input**; **Virtual reality**; **Empirical studies in HCI**.

Keywords

virtual reality, eye-tracking, gaze input, gesture input, 3D manipulation, hand gestures, hand-tracking

ACM Reference Format:
Thorbjørn Mikkelsen, Qiushi Zhou, Mar Gonzalez-Franco, Hans Gellersen, and Ken Pfeuffer. 2026. Preshaping Hand Behaviour for Direct and Indirect Manipulation of 3D Objects. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3772318.3793424

1 Introduction

Manipulation of objects with our hands involves *preshaping* as a preparatory movement during reach. Preshaping is a subconscious motor behaviour that configures hand posture for secure and efficient grasping and the intended manuipulation [40]. For example, when adjusting a knob, we counter-rotate the hand to a larger extent when we intend to turn the knob further. Preshaping is a kinematic response informed by intrinsic (e.g., mass, texture) and extrinsic (e.g., position, orientation) object properties, making physical interactions seem effortless [48, 52]. Virtual 3D manipulation transforms these parameters; the interaction is determined not purely by physics, but also by the interface. Common grasping abstractions such as the pinch gesture modify the hand-object interaction [8], as illustrated in Fig. 1. This enables supernatural manipulations via object-agnostic postures without concerns of stability. Although virtualised, we posit that such manipulations still benefit from preshaping. However, it is not clear how the design of interaction techniques impacts preshaping behaviour and its efficacy in preparation for manipulation.

We propose preshaping to be of relevance to virtual interaction as it can affect the performance of manipulation tasks. The initial position of the hand determines the physical motorspace available for further input during manipulation. Techniques map this input differently to the virtual space, using abstractions, additional modalities, and control-display (CD) gain functions that modify the requirements of the initial position and thus motorspace freedom. During manipulations, the physical hand remains under anatomical constraints, and a poorly prepared starting posture eventually must be compensated for in manipulation [30, 34]. Invariably, when there is no more room for movement, users have to disengage and prepare for additional input, as is typical with the computer mouse. 3D manipulation techniques extensively rely on *clutching* as strategy, to sequentially chunk a manipulation task through repeated acquisition and release [4]. We suggest that clutching closely relates to preshaping, and that a better understanding of preshaping may inform designs that reduce the need for clutching.

The main contribution of this work is a study of preshaping for object manipulation in 3D, guided by three questions: (RQ1) How do users preshape for object manipulation in 3D when they use grasping abstractions? This examines how strongly preshaping is present when input is virtual, and how it relates to clutching. (RQ2) How does preshaping behaviour differ in indirect manipulation compared to direct? We compare direct versus indirect input in how they afford preshaping. (RQ3) How is preshaping behaviour adapted when using a controller rather than the bare hand? This considers the effect that holding a tool may have on preparatory movement. We conducted a virtual reality (VR) user study (N=20) and evaluated 3D hand behaviour on a 6-degree-of-freedom (DOF) docking task. Participants performed the task in four conditions, derived from a combination of direct versus indirect, and bare-hand versus contoller-based input, and exemplifying contemporary 3D input methods. Based on the results, we contribute key novel insights into preshaping in 3D object manipulation:

- Preshaping behaviour is modulated by the interaction technique. Response patterns scale with anticipated task difficulty, and the magnitude and efficacy of the response depend on the technique.
- Results reveal a trade-off between motorspace constraints and sensory feedback. Direct interactions elicit effective responses guided but limited by reach, while indirect techniques enable economical movement but induce less effective hand postures, compromising manipulation.
- A causal link between preshaping behaviour and compensatory strategies. Our work shows that deficits in the preparatory phase, whether biomechanical constraints or lack of planning, directly necessitate increased clutching and corrective adjustments during the manipulation phase.
- Evidence of controllers mitigating preshaping deficits through tangible affordances. We observe that users leverage the physicality of controllers to perform in-hand rotation, extending their range of motion through increased dexterity in a way that bare-hand gestures currently do not support.

These findings advance understanding of motor behaviour in 3D manipulation, and are of practical significance for interaction design, to better afford preshaping, require less clutching and optimise 3D manipulation tasks.

2 Related Work

Preshaping is naturally present in all manual interaction with objects but has not received much attention in HCI. We contextualise our work by drawing on background from other fields, relating preshaping to other work on preparatory movement in HCI, and by motivating our focus on direct versus indirect interactions, with bare hand or controller.

2.1 Hand-Object Interaction

Preshaping has been extensively studied in neuroscience, computer vision, and robotics to understand its components and inform the design of tools and technologies that augment or adopt human manual behaviour [13, 25, 53]. Here, preshaping is defined in relation to the reach, grasp and manipulation phases of physical hand-object interaction. Each of these phases relies on different feedforward and feedback mechanisms for prehension and control of the target [13, 40]. The first two phases are parallel and integrated in the *reach-to-grasp* motor skill [61]. In this context, preshaping refers to the anticipatory posture assumed by the arm, hand, and fingers in response to the object during movement of the upper limb [40]. This behaviour is a response to and prepares the hand for an expected object interaction, thereby setting the immediate context and constraints for the manipulation phase. Preshaping is finely tuned by a lifetime of interactions and is performed subconsciously without mental effort. Literature on the topic is consistent in highlighting the importance of preshaping for effective manipulation, which serves as the principal motivation for our work.

There has been a wide range of work on virtual grasping. Blaga et al. examined how the direct manipulation of objects differs between real and virtual environments for wizard-of-oz translation tasks, finding that virtual grasping posture is modulated by the anticipated intrinsic object properties (e.g., size, shape) [8]. This, and similar work [48], is significant for our work, as it shows transferral of experience from real environments to virtual grasping, but in a critical difference, our work is concerned with grasping abstractions. Within XR, recent works have established taxonomies for grasp-proximate interfaces [3, 8, 50, 51]. These help understand grasping-based input but are focused on the posture assumed for the interaction when an object is acquired, abstracting from the preshaping behaviour that has posture as the outcome. Other works have shown that the performance of virtual grasping depends on visual feedback, to compensate for the lack of physicality [21, 45]. Feedback is also a critical concern for our work, addressed in technique design for our study, as the techniques we consider abstract further from real grasping.

2.2 Preparatory Movement for Interaction

Preshaping is a preparatory movement and, as such, relates to a variety of work in HCI that is concerned with user actions of a preparatory nature. The rise of context-aware interaction, for instance, prompted questions of how users address systems to initiate interaction, as captured in Belotti et al.'s framework [6]. Techniques such as Pre-touch [19], Do That, There [16], GravitySpot [1], and Proxemic Flow [55] all illustrate preparatory action guided by visual or multimodal cues that help users align their bodies or hands with system-defined input regions. Preshaping is different from separate movements guided in that manner, as it describes the finer-grained anticipatory behaviour that is *integral* with hand-object interaction and occurs just before manipulation begins.

While preshaping is implicit with manipulation, it is interesting to consider it as an input signal. Although using different terminology, Mizuno et al. demonstrated the idea of using preshaping gestures as a shortcut to trigger intended manipulations [35]. Other work considering preshaping for feedforward in XR has been conducted by Guiernou et al. [17]. Their work is significant in showing that preshaping can be operationalised in XR, but that it requires adaptation (e.g. grip aperture as feedforward is mismatched when holding a controller).

2.3 Direct and Indirect Interaction, with Hand or Controller

We study preshaping for direct versus indirect manipulation, and with bare hand versus controller, as these represent fundamental differences that could affect preshaping. Direct input offers intuitive, physics-based mappings [23, 46], while indirect input affords occlusion-free input and reduced effort through flexible control-display (CD) mappings [20]. Direct and indirect input methods are frequently compared, with a focus on performance and speed-accuracy trade-offs [12, 33, 49]. In contrast, we aim to understand differences in hand alignment for manipulation tasks, for which we compare direct hand input with gaze-assisted indirect input [42, 58], in both cases with pinch as grasping abstraction. While indirect input can use a hand-ray [38], we use gaze due to its demonstrated performance benefits [34, 57] and its adoption as the default on recent XR devices (e.g., Vision Pro). Hands and controllers are also frequently contrasted (e.g. virtual hand vs. virtual pointer [44]), with hand input characterised as more natural [56] while controllers offer better performance [11, 26, 28, 31]. We focus on the effect of holding a controller on preshaping, which we compare with bare-hand input in both direct and indirect conditions.

Two recent works provide specific motivation. Lystbæk et al. showed how direct and indirect modes of input can be fluidly combined in 3D manipulation of objects, and observed a trade-off between faster acquisition with indirect input and more effective preparation for manipulation with direct input [30]. We hypothesise that the observed difference in performance is rooted in preshaping behaviour. In another study, Mikkelsen et al. observed performance differences in object manipulation with a hand pointer versus gaze+pinch, that appeared to stem from differences in motor space freedom the techniques afforded, thus impacting the extent to which participants were able to prepare hand input for the manipulation task [34]. These observations call for a systematic study to understand how performance differences can be explained by preshaping behaviour.

3 Preshaping in Extended Reality

Preshaping is well defined in other fields, but the concept requires adaptation to XR. In fields such as robotics, preshaping is adopted to describe the kinematics, end-position and possibly splay of the end effector (e.g. hand), and is concerned with mimicking human movement trajectories. Guerniou's work indicates that preshaping is not an a priori good behaviour when it is not adapted to the relationship between the effector (controller) and the object (virtual) [17]. Yet, it has value as an abstraction and motivated the development of domain-specific definitions that adhere to the properties of actors and objects in novel contexts [22, 53]. Accordingly, we abstract from concepts such as hand aperture and finger-object enclosure, as these are not afforded by pinch and controller grasping abstractions.

Figure 2 illustrates our framework for operationalising preshaping in XR. We consider a preparation phase in which we measure translation and rotation of the hand in moving from a neutral position to acquire a 3D object. Acquisition is by pinch gesture or controller trigger, as a grasping abstraction. At the moment of acquisition, we measure preshaping magnitude as how much the hand has moved in the *opposite* direction of the intended manipulation, which can be understood as movement to increase the range for subsequent manipulation. During manipulation, we measure rotation and translation to capture the effect of preshaping on manipulation performance. Finally, at the moment the object is released, we measure how much of the intended manipulation has been achieved, as a measure of preshaping effectiveness. If a manipulation has not been fully accomplished at this point, the object can be re-acquired through clutching, and the amount of clutching required constitutes a further measure for effectiveness.

4 User Study

To investigate our research questions, we designed a controlled 3D docking study. The core objective of our task design was to explore the expression of preshaping in response to a clear rotational challenge, while minimising confounding variables such as complex mental rotation, visual search, and translational difficulty. Participants were tasked with docking an object to a target destination, a task which combined a simple horizontal translation

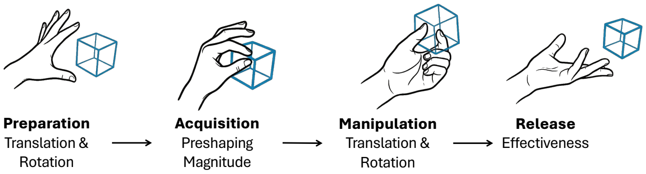

Figure 2: Operationalising preshaping for virtual 3D object manipulation. We measure preparatory movement, preshaping magnitude at the point of object acquisition, manipulation performance, and effectiveness of the combined movements toward completion of a manipulation task. Objects can be reacquired to complete manipulation by clutching, which we also use as a measure of effectiveness.

with a single-axis rotation of varying magnitudes (±45°, 90°, or 135°), informed by previous studies on virtual manipulation and physical preshaping [2, 7, 32]. This controlled design allows us to directly observe how users adapt preparatory movements to the anticipated requirements of the manipulation, providing a basis for analysing the effects of different interaction techniques on preshaping behaviour regarding *directness* and *modality*, counterbalanced as techniques, and *rotations* presented in uniformly random order.

4.1 Technique Implementation

We implemented four techniques to represent a combination of Direct versus Indirect, with Bare-Hand versus Controller. We refer to the two dimensions for short as Directness and Modality:

- VirtualHand emulates physical grasping, as the hand reaches for and acquires objects by pinching in proximity. The object then follows the hand until the pinch is released.
- DirectController functions much the same as VirtualHand, while requiring the user to steer a controller for mediating the grasp. The pinching is consequently substituted with an index-towards-thumb trigger pull.
- Gaze&Pinch delegates object-targeting to eye-gaze, confirming the acquisition with a pinch. Though unconstrained by reach, hand movements are mapped to the object similarly to the Direct conditions.
- Gaze&Controller functions as Gaze&Pinch, but substitutes pinching identical to DirectController.

The four techniques were implemented in Unity using the Meta XR SDK, illustrated in fig. 3. In the following, we describe key properties, by which techniques vary, that may influence preshaping:

Selection and release. For Bare-Hand techniques, pinch-release was adjusted, using the SDK's pinch strength parameter (0.1 confidence) - such that participants encounter fewer frustrating early releases, but may encounter more late releases, based on how the system evaluates the tracked hands. This detail is important as the task promotes hand positions that challenge the tracking system. For Controller techniques, the default halfway press of the index trigger determines selection and release. We chose the index trigger over the inside "grip" buttons of the Quest Touch Pro controllers for consistency with the thumb-index pinch (fig. 4).

Direct Control Mapping. Both Direct techniques, VirtualHand and DirectController, require virtual contact with an object at the time of selection. Both techniques consider contact to be within a ~ 2cm proximity, from the thumb-index pinch point, and from the grab point of the controller, respectively. The point of contact enables the use of grab-points, determining the pivot through which the object rotates 1:1 with the input modality. For VirtualHand, the thumb-index pinch point in space at selection defines this grab-point, and the object within the 2cm proximity. For DirectController, the head of the controller defines the pivot.

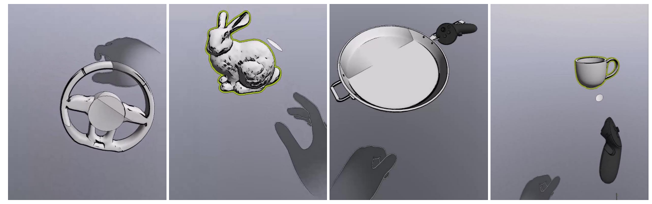

Figure 3: Examples of different hand poses, preshaped into the posture accommodating the object and envisioned interaction of the user via VirtualHand, Gaze&Pinch, DirectController, and Gaze&Controller.

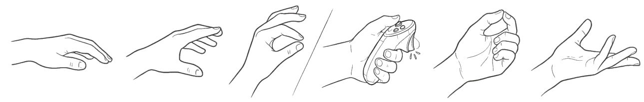

Figure 4: Temporal development of the hand posture. The primary difference between conditions is in *how* the acquisition event is triggered. Bare-hand relies on a pinch gesture, while controller techniques confirm via a trigger press, using a similar index-to-thumb closure.

Gaze-assisted Indirection. For implementing the Indirect Gaze&Pinch and Gaze&Controller techniques, we rely on the current guidelines and prior art [41, 42]. Targeting is enabled by raycasting a forward vector along the averaged eye orientation, and selection is triggered as described above. Consequently, a manipulation can be initiated from any tracked position in the user's motorspace. To explore indirection, we disable selection within 2cm proximity, based on the Direct criteria.

Indirect Control Mapping. The control mapping of the hand determines the spatial relationship between the hand and the object after it is selected. We mimic the 1:1 encoding of physical and virtual manipulation to achieve consistency. For mapping translation, this entails a CD gain that approximates a 1:1 hand-object ratio in visual angles. Informed by prior art, like the HOMER gain, our gain is continuously calculated as the quotient of eye-object and eye-hand vector magnitudes [34, 60]. For mapping rotation, raycasting provides no semantically valid grab-point to inform the pivot. Instead, we follow prior work and map the rotation of the thumb-tip and controller-head to the ego-centric rotation of the object, for Gaze&Pinch and Gaze&Controller, respectively [30, 34].

Addressing Synchronisation/Late-Trigger issues. While the coordination of hand motion with the selection and release of objects is similar for direct techniques, gaze-assisted indirection introduces a fundamental challenge in eye-hand synchronisation:

Figure 5: Example trial sequence. The white object and blue target appear initially overlapping at a reachable chest-level origin, determined by the task variation. The target immediately moves to its final pose via a linear interpolation animation. The destination is offset ±20 cm along the X axis, and ±45, 90, 135° along either X, Y or Z, relative to the object. A trial is completed by manipulating the object to the target through any number of clutches and releasing the docked object indicated by it turning green and freezing.

the late-trigger issue. As gaze determines the target object while outpacing our hands, it becomes necessary to address the timing offset for techniques not to fail during selection. One common compensation strategy is a lookback, where the selection is matched to a historic gaze-point. While task- and user-calibrated offsets have been correctly classified in research, we implement a simpler fixed offset lookback of 85 milliseconds [39].

Visual Feedback. For Controller techniques, we presented the default visualisation: the controller and not the tracked virtual hand holding it — in line with the idea of mediation, rather than occupation of the hand. To address the absence of direct visual anchoring in Indirect techniques, we introduced object-centric cues based on prior work [34]. First, an outline appeared on the object when targeted by the user's gaze. Second, an abstract disc orbited the target to convey the hand's rotational state, mapping the palm's position relative to the thumb. A neutral, palm-inwards hand posture, for example, was represented by the disc appearing on the right-hand side of the target.

4.2 Task Design

To understand users' expression of preshaping and how that impacts the manipulation of virtual objects, we employ a 3D docking task in VR. The object is a white box circumscribing an opaque white Stanford bunny, while the target is a matching translucent blue cube [34, 64]. The cube has a side length of 12.5 cm and appears 50 cm in front of the participant's chest level, approximated based on the HMD position. When the objects appear, the target is animated from the object's starting position to its final destination through linear interpolation of its position and rotation. This presents the user with a solution to the trial, minimising potential difficulty from visual search or complex mental rotation planning [15, 24, 27]. It lasts ≈ 670ms, informed by pilot testing on average acquisition time (N = 12, 987ms), so as not to disrupt final rotation planning.

We designed two variations for the starting positions. In the "centre-out" variation, the user moves the object directly in front of them to a more peripheral target. In the "outside-in" variation, the positions flip, and the user brings an object from the periphery to a central working position. This variation models how manipulations often start with 'object-framing' [30, 37]. The "outside-in" variation is useful for promoting implicit hand control for a less noisy measurement of preparatory movements, as the hand will naturally move towards body-centred positions for all outside-in trials. It consequently acts as a 'control', allowing tests for inter-trial residual movements when compared to other docking variations. We chose not to explicitly control hand position, which may introduce arbitrary movements, whereas we aim to study natural behaviour.

Each trial combines a translation and a rotation component. For translation, the target destination is uniformly randomly offset by ±20 cm along the horizontal axis. This distance was chosen based on pilot tests to represent a simple perceivable translation within reachable space, and the horizontal axis was selected as it covers movements across the body. For rotation, the target orientation is rotated by ±45, 90, 135° along a single axis (X, Y or Z). This allows to investigate clear user responses to the task while avoiding the cognitive load of multi-axis mental rotation, which was studied before [15, 27]. The order of rotations was uniformly random. All trials are possible to finish in a single gesture.

To mitigate the Heisenberg effect [63], completion criteria are checked continuously throughout the manipulation rather than only at the moment of release. A trial is completed when the object is within a Euclidean distance of 2.5 cm and a rotational offset of 10 degrees from the target. Then, the object turns opaque green and locks in place, and the user can then release the object. The next trial begins after a $0.5s$ delay. This affords clutching and fine-grained corrective movements without effects from timing or tracking issues (Figure 5). Overall, we had 4 Techniques × 18 Rotations × 2 Start Positions × 2 Repetitions = 288 trials per participant.

4.3 Apparatus and Implementation

The study was implemented in Unity (`2022.3.27f1`) with Meta XR SDK `v71` for the Meta Quest Pro headset, which features a 90 Hz display and a 30 Hz eye tracker. The SDK offers implementations for the VirtualHand and DirectController techniques, while the Gaze&Pinch and Gaze&Controller techniques used the SDK's tracking data. The reported eye-tracking accuracy for this hardware varies from 1.5° to 3° [5, 59]. To reduce jitter, hand tracking data was smoothed using a 1€ filter [14].

4.4 Procedure

Participants were briefed on the study, and filled out consent and demographics form. A one-minute explanatory video was shown before each new condition. Participants donned the headset and performed eye-tracking calibration. Participants were seated in a fixed chair without armrests. For each technique condition, participants were given a self-paced training period in a task-less playground environment to practice 3D manipulations until they reported feeling comfortable and confident (maximum 2 minutes). They then proceeded to the main task of two blocks. Each block contained 18 centre-out trials, followed by 18 outside-in trials. Between the blocks, a virtual button appeared, allowing for a break. Participants were instructed to complete the tasks as quickly and accurately as possible. After finishing all trials for a technique, participants completed a post-condition questionnaire. After all techniques, they filled out a final post-study questionnaire. The entire session lasted approximately 60 minutes per participant.

4.5 Participants

We recruited 20 participants (12M, 8F) from the local university. They were 22 to 29 years old ($M = 25.85, SD = 1.98$), right-handed and had normal or corrected-to-normal vision. On a scale between 1 (low) and 5 (high), participants rated their experience with related technologies. They reported a medium level with XR ($M = 3.5, SD = 1.15$) and 3D gestures ($M = 3.15, SD = 1.23$), and a medium-low level with gaze interaction ($M = 2.8, SD = 1.47$).

4.6 Evaluation Metrics

We analyse data before, starting, during, and after the first manipulation of the object, based on the measurement framework introduced above (Fig. 2):

- **Preshaping Magnitude**: The hand-target difference at acquisition, reflecting the amount of counter-rotation expressed by the preshaping posture relative to the task. We compute the hand quaternion as the deviation from a consistent neutral hand orientation, with the palm facing inwards.
- **Effectiveness**: The amount of object rotation left toward the target orientation at release, representing the amount of the rotation component of the task solved by the first manipulation, as measure of the effectiveness of preshaping.
- **Preparation Time**: The duration in seconds between when objects appear and the acquisition, reflecting the time taken to perform preshaping before manipulation.
- **Preparation Translation and Rotation**: Hand translation and rotation before the acquisition, reflecting the movement of the hand needed for preshaping.
- **Manipulation Time**: The time elapsed during the manipulation phase after acquisition and before the release, reflecting the effort in the first manipulation as a result of the preceding preshaping.
- **Manipulation Translation and Rotation**: Hand translation and the rotation **during** the manipulation phase, before the release, reflecting the movement of the hand under the constraints of the preshaping.

Data for movements and offsets is accumulated across all axes. While we focus on analysing the first grasp to understand preshaping behaviour and strategy, we use trial-level performance for reference to help interpret preshaping results, including Trial Completion Time and the Clutch Count (number of grasps performed to finish a trial). We also include task load with the NASA-TLX [18], gather qualitative feedback after each technique [18], and collect a Likert-scale (1-7) preference ranking at the end.

4.7 Results

We collected 10,137 manipulations (defined as one grasp-and-release sequence) across 5,760 trials. Results of 5,703 trials (9,959 manipulations) remain after the removal of outliers, due to accidental object selection (24 trials completed faster than 100 milliseconds) and malfunctions (14 accidental triggers of the system interface, 16 for consecutive failures of grasping, and 3 pre-emptive removals of the HMD). Due to violations of normality, we applied the Aligned Rank Transform (ART) prior to analysis [62]. We conducted a 3-way RM-ANOVA with Directness, Modality, and Rotation (Holm-corrected). For rating data, we performed Friedman tests and Wilcoxon signed-rank post-hoc tests (Holm-corrected). A summary of all main and interaction effects is presented in Table 1. We report and expand the significant effects of Directness and Modality in text, while Rotation main and interaction effects are presented in Table 1 and figs. 6 to 14 for brevity. Error bars in figures represent 95% confidence intervals; note the non-zero y-axis, as we focus on comparing interactions.

Table 1: Summary of Aligned Rank Transform RM-ANOVA on the dependent variables. Significant effects are highlighted in grey.

	ANOVA					ANOVA			
Effect	df	F	p	η_p^2	Effect	df	F	p	η_p^2
Preshaping Magnitude					**Manipulation Time**				
D	1, 209	42.79	<.001	**0.170**	D	1, 209	68.07	<.001	**0.246**
M	1, 209	105.00	<.001	**0.334**	M	1, 209	75.04	<.001	**0.264**
R	2, 209	435.75	<.001	**0.807**	R	2, 209	10.98	<.001	0.095
D x M	1, 209	1.05	0.308	0.005	D x M	1, 209	4.85	0.029	0.023
D x R	2, 209	6.05	0.003	0.055	D x R	2, 209	1.76	0.174	0.017
M x R	2, 209	78.13	<.001	**0.428**	M x R	2, 209	5.00	0.008	0.046
D x M x R	2, 209	1.79	0.169	0.017	D x M x R	2, 209	0.02	0.979	0.000
Effectiveness					**Manipulation Translation**				
D	1, 209	6.54	0.011	0.030	D	1, 209	16.18	<.001	0.072
M	1, 209	36.13	<.001	**0.147**	M	1, 209	166.10	<.001	**0.443**
R	2, 209	341.93	<.001	**0.766**	R	2, 209	0.89	0.412	0.008
D x M	1, 209	0.06	0.811	0.000	D x M	1, 209	0.00	0.969	0.000
D x R	2, 209	3.24	0.041	0.030	D x R	2, 209	0.07	0.934	0.001
M x R	2, 209	2.57	0.079	0.024	M x R	2, 209	0.44	0.642	0.004
D x M x R	2, 209	0.03	0.969	0.000	D x M x R	2, 209	0.99	0.374	0.009
Preparation Time					**Manipulation Rotation**				
D	1, 209	28.48	<.001	0.120	D	1, 209	2.53	0.113	0.012
M	1, 209	113.77	<.001	**0.352**	M	1, 209	0.83	0.363	0.004
R	2, 209	1.07	0.344	0.010	R	2, 209	83.25	<.001	**0.443**
D x M	1, 209	30.93	<.001	0.129	D x M	1, 209	1.27	0.261	0.006
D x R	2, 209	0.32	0.727	0.003	D x R	2, 209	2.41	0.092	0.023
M x R	2, 209	0.16	0.850	0.002	M x R	2, 209	1.42	0.245	0.013
D x M x R	2, 209	0.25	0.780	0.002	D x M x R	2, 209	0.17	0.847	0.002
Preparation Translation					**Task Completion Time**				
D	1, 209	339.94	<.001	**0.619**	D	1, 209	188.53	<.001	**0.474**
M	1, 209	25.68	<.001	0.109	M	1, 209	350.62	<.001	**0.627**
R	2, 209	0.58	0.563	0.005	R	2, 209	182.57	<.001	**0.636**
D x M	1, 209	0.04	0.847	0.000	D x M	1, 209	49.68	<.001	**0.192**
D x R	2, 209	0.03	0.968	0.000	D x R	2, 209	3.39	0.035	0.031
M x R	2, 209	0.10	0.906	0.001	M x R	2, 209	15.93	<.001	0.132
D x M x R	2, 209	0.01	0.987	0.000	D x M x R	2, 209	1.75	0.177	0.016
Preparation Rotation					**Clutch Count**				
D	1, 209	48.57	<.001	**0.189**	D	1, 209	15.35	<.001	0.068
M	1, 209	1.27	0.261	0.006	M	1, 209	80.16	<.001	**0.277**
R	2, 209	5.10	0.007	0.047	R	2, 209	279.00	<.001	**0.728**
D x M	1, 209	23.79	<.001	0.102	D x M	1, 209	1.15	0.284	0.005
D x R	2, 209	0.81	0.445	0.008	D x R	2, 209	1.57	0.210	0.015
M x R	2, 209	0.32	0.726	0.003	M x R	2, 209	12.13	<.001	0.104
D x M x R	2, 209	0.43	0.649	0.004	D x M x R	2, 209	0.66	0.520	0.006

D = Directness; M = Modality; R = Rotation. Effect sizes η_p^2 ≥.14 marked **bold**, <.06 grey.

4.7.1 *Results on Preshaping Measures.*

At acquisition, users rotated their hand more with Direct than with Indirect ($p < .001$). At release, however, more rotation remained in the Indirect condition ($p = .011$). Across both metrics, Bare-Hand exhibited larger rotation offsets than Controller (both $p < .001$).

Second, we report the accumulated-data results for the phase before users acquired the object (figs. 8 to 10). Users translated and rotated more, and completed the phase faster, with Direct than with Indirect (all $p < .001$). Comparing modalities, users translated more and required more time with Bare-Hand than with Controller (both $p < .001$).

These effects were supported by interaction effects. The Direct Controller technique produced more rotation than all other techniques ($p < .001$), and both Bare-Hand techniques yielded more

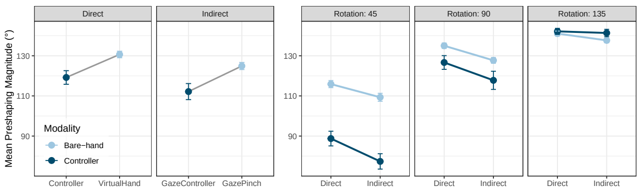

Figure 6: Rotation offset between the target orientation and that of the hand at the moment of grasping, plotted for Direct, Indirect, Bare-Hand, and Controller conditions, and split between three Rotation conditions (45°, 90°, 135°).

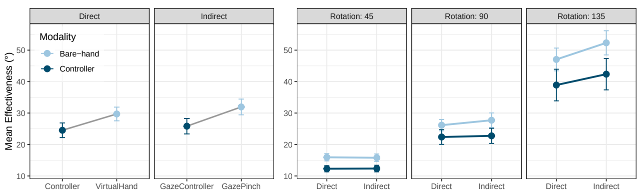

Figure 7: Rotation offset between the target orientation and that of the object upon releasing the first grasp, plotted for Direct, Indirect, Bare-Hand, and Controller conditions, and split between three Rotation conditions (45°, 90°, 135°).

rotation than Indirect Controller ($p < .05$). For time, any Controller version was faster than any Bare-Hand version; within the Bare-Hand techniques, Direct was faster (all $p < .001$).

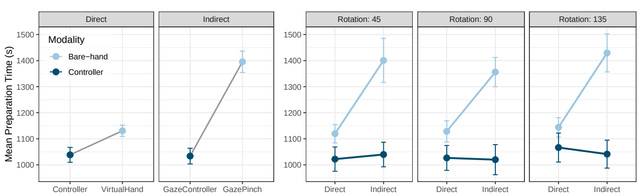

Figure 8: Duration of the preparatory phase from the target appearing to the moment of the first grasp, plotted for Direct, Indirect, Bare-Hand, and Controller conditions, and split between three Rotation conditions (45°, 90°, 135°).

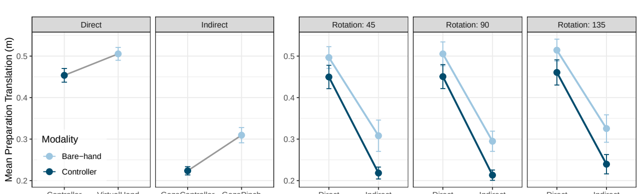

Figure 9: Accumulated hand translation (per frame) from the target appearing to the first grasp, plotted for Direct, Indirect, Bare-Hand, and Controller conditions, and split between three Rotation conditions (45°, 90°, 135°).

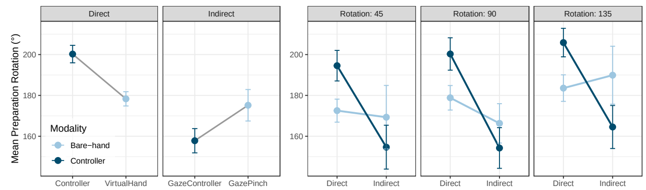

Figure 10: Accumulated hand rotation angle (per frame) from the target appearing to the first grasp, plotted for Direct, Indirect, Bare-Hand, and Controller conditions, and split between three Rotation conditions (45°, 90°, 135°).

Third, we report performance during the manipulation phase after acquisition and before the first release (figs. 11 and 12). Users were faster and translated more with Direct than with Indirect. Across modalities, users were faster and translated less with Controller than with Bare-Hand (all $p < .001$). Interaction effects (all $p < .001$) further showed that Direct Controller was faster than all other techniques, whereas Indirect Bare-Hand was slower than all others.

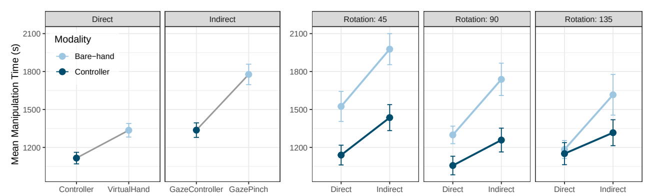

Figure 11: Duration of the manipulative phase from the first grasp to its release, plotted for Direct, Indirect, Bare-Hand, and Controller conditions, and split between three Rotation conditions (45°, 90°, 135°).

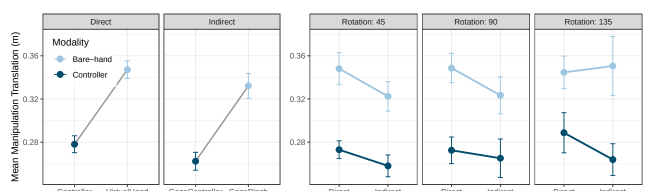

Figure 12: Accumulated hand translation (per frame) from the first grasp to its release, plotted for Direct, Indirect, Bare-Hand, and Controller conditions, and split between three Rotation conditions (45°, 90°, 135°).

4.7.2 Results on Trial-level Metrics (Figure 13 and 14). First, Direct was faster and required fewer clutches than Indirect ($p < .001$). Considering the modalities, Controller was faster and involved fewer clutches than Bare-Hand ($p < .001$).

At the interaction level, users were fastest with Direct Controller and slowest with Indirect Bare-Hand, relative to all other conditions (all $p < .001$). For clutching, users clutched more with Indirect Bare-Hand than with Direct Bare-Hand ($p = .016$)

and both Controller techniques ($p < .001$). They also clutched more with Direct Bare-Hand than with either controller variant ($p < .001$). Between the two controller conditions, Indirect Controller resulted in more clutching ($p = .045$).

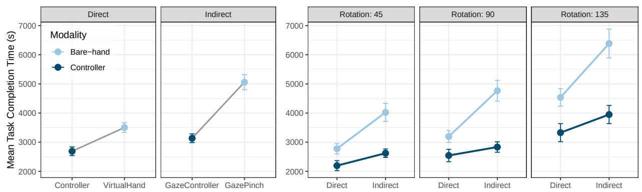

Figure 13: Time taken to complete the trial, plotted for Direct, Indirect, Bare-Hand, and Controller conditions, and split between three Rotation conditions (45°, 90°, 135°).

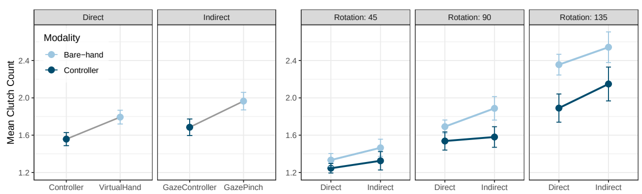

Figure 14: Number of grasps (clutches) taken to complete the manipulation trial, plotted for Direct, Indirect, Bare-Hand, and Controller conditions, and split between three Rotation conditions (45°, 90°, 135°).

4.7.3 NASA-TLX, SEQ, Preferences (figs. 15 to 17). Analysis of subjective feedback revealed significant differences between techniques across most dimensions, including all NASA-TLX workload metrics (all $p \leq .001$) except for Temporal Demand ($p = .076$): Mental Demand ($\chi^2(3) = 20.52$), Physical Demand ($\chi^2(3) = 29.79$), Performance ($\chi^2(3) = 24.96$), Effort ($\chi^2(3) = 28.31$), Frustration ($\chi^2(3) = 24.26$), and overall Total Task Load ($\chi^2(3) = 20.20$). Eye Fatigue was significant, too ($\chi^2(3) = 16.04, p = .001$). A clear pattern emerged from post-hoc comparisons, where the Gaze&Pinch technique consistently induced the highest load. It was rated as inducing significantly higher mental, physical, and total task load, and was also considered more frustrating and effortful than the other techniques. This pattern extended to usability ratings, where Gaze&Pinch was rated as providing a lower sense of control and being more difficult for positioning and rotation. Conversely, the DirectController and Gaze&Controller techniques generally received the most favourable ratings, often being perceived as significantly less demanding and easier to use than Gaze&Pinch and, in several cases, VirtualHand.

For the preference ratings, a significant overall difference was found between the techniques ($\chi^2(3) = 11.92, p = .008$). Post-hoc comparisons (with continuity correction) showed that DirectController was ranked higher than both the Gaze&Pinch ($p = .012$) and the VirtualHand ($p = .036$). Gaze&Controller was also found to be preferred over the Gaze&Pinch ($p = .012$).

4.7.4 User feedback and Observations. User feedback indicates a trade-off between the high physical demand of Direct techniques and the cognitive load or technical fragility of Indirect and Bare-Hand techniques. The most prevalent theme was the significant physical fatigue associated with Direct manipulation, often described as tiring (P5, 8, 15, 19). Despite this physical cost, the haptic feedback and weight of the controller were frequently praised for providing a superior sense of control and tangibility (P2, 5), with one user noting they "have something in my hand to rotate" in contrast to the "grabbing nothing" feeling of Bare-Hand interaction (P2). This sense of control was particularly evident *during* object rotation, where the controller was perceived to offer a greater and more comfortable range of motion compared to the biomechanical limits of the wrist [*during* pinch-based manipulation] (P 1, 4). For some participants, however, the controller provided a less intuitive experience compared with VirtualHand: "It feels most natural, and I don't need to think at all about what I am doing." [VirtualHand] vs. "I have to think what to do when and keep in mind how to use the controller." (P7); "Some of the rotations are awkward to do with the controller, compared to how I would do it with my hands for a real object. With the controller, I can only rotate with my wrist and forearm, not with my fingers."(P10).

While Bare-Hand techniques were praised for reducing physical strain, their drawbacks were twofold. First, users reported frustration with technical reliability, citing inconsistent pinch detection, jitter, and tracking failures that disrupted the workflow (P3, 11, 15, 16). Second, gaze-based techniques introduced a specific cognitive conflict, forcing users to resolve looking at the object for selection versus looking at the target for planning (P1, 8, 10) - "I think a lot of the mental load came from having to look at the object to grab it, while trying to remember the image of the reference object." (P8, Gaze&Pinch). Ultimately, user preference appeared to be governed by a personal balance between the high physical effort but perceived reliability of Direct and Controller interaction, and the reduced fatigue but increased cognitive load and technical challenges of Indirect and Bare-Hand techniques.

Observations confirmed how some users achieved larger rotation during manipulation by employing in-hand rotation of the controller through their fingers, compensating for constraints of preshaping. Other patterns include the supinated (palm up) preshaping posture for underhanded rotation along the vertical axis, uniquely performed with VirtualHand. Finally, while all participants expressed task-informed preshaping behaviour for every condition, it is remarkable how smaller rotations prompted little response for Gaze&Pinch, for which participants defaulted to a pronated, pointing hand posture, echoing previous findings [30, 34].

5 Discussion

In this section, we interpret and discuss the key findings from the study, addressing research questions and implications.

5.1 Preshaping Behaviour and Performance

To address our first research question (RQ1) regarding preparatory hand movements, we examined the extent to which users exhibited preshaping behaviour. Our findings confirm that participants actively preshape their hands in preparation for manipulating virtual

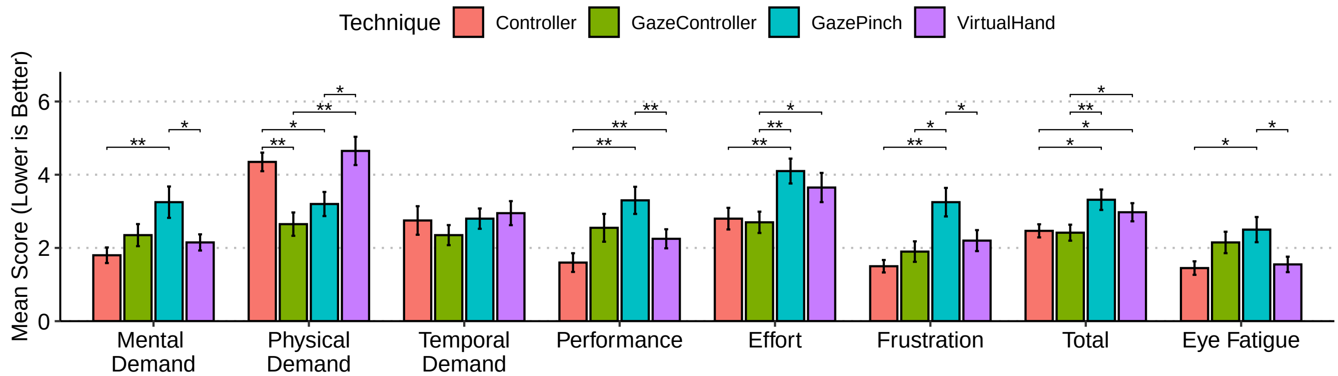

Figure 15: NASA-TLX Task Load ratings on Likert scale 1-7. Direct techniques were perceived as more physically demanding, while Bare-Hand, and especially Gaze&Pinch, induced a higher task load, requiring more effort in use.

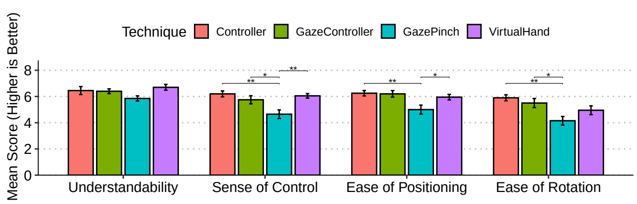

Figure 16: Single Ease Questions rated on Likert scale 1-7. Gaze&Pinch was perceived as harder to use for both translation and rotation.

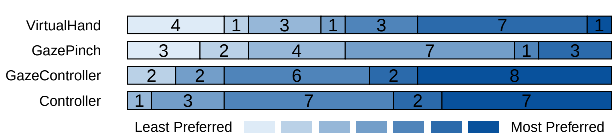

Figure 17: Preference rating for each technique on a scale from 1 (least preferred) to 7 (most preferred).

objects. This is evidenced by a significant main effect of the required object Rotation on our primary preshaping measure, Preshaping Magnitude. Specifically, tasks requiring larger object rotations elicited correspondingly larger preparatory hand rotations from participants, as illustrated in Figure 6. This finding indicates that users adjust the magnitude of their counter-rotational preshaping response to the anticipated difficulty of the manipulation task.

This conclusion is further supported by Preparation Rotation, a measure of the total accumulated hand rotation during the preparatory phase (i.e., between object appearance and the initial grasp). The results for this measure also showed a significant main effect of Rotation, with greater required object rotation leading to more accumulated hand rotation before contact. These results provide evidence that users perform preshaping and that the extent of this behaviour scales with the rotational demands of the task.

In terms of task performance and user strategy, we observed a significant effect of Rotation on the duration of the manipulative phase (Manipulation Time). Remarkably, the manipulation phase was significantly *longer* for tasks with *smaller* Rotation magnitudes. This suggests that when the goal seemed more attainable, users attempted to complete the entire rotation within a single grasp. In contrast, the shorter manipulation phase for larger rotations indicates that users did not plan to complete the task in one go, instead opting for a different strategy. This points to clutching - releasing and re-grasping the object - as the default user strategy for solving complex rotations, a behaviour observed across all conditions. This interpretation is reinforced by the results for Trial Completion Time and Clutch Count, both of which were significantly higher for larger Rotation tasks. This aligns with established findings in HCI on task chunking, where complex actions are broken down into smaller, manageable sub-tasks in both physical and virtual environments [4].

5.2 Preshaping in Direct and Indirect Manipulation

Regarding our second research question (RQ2), our results reveal distinct differences in preshaping behaviour between Direct and Indirect manipulation. The primary finding from Preshaping Magnitude is that Direct techniques induced significantly more rotational preshaping overall than Indirect techniques. Nevertheless, preshaping with Indirect techniques still followed a similar pattern, demonstrating a significant increase in preparatory rotation in response to higher Rotation levels. The difference between the two groups was most pronounced at smaller Rotation levels, where Direct interaction elicited substantially more preparatory movement. As the required rotation increased, the amount of preshaping between the two techniques gradually converged (Figure 6).

We interpret this convergence as a result of biomechanical limitations. At high Rotation levels, the task itself demands that the hand be rotated to an extreme angle for a successful grasp. In Direct techniques, where the hand is physically coupled with the object, it is difficult to rotate much further, thus creating a ceiling effect on preparatory rotation. Conversely, at smaller Rotation levels, participants using Direct techniques exhibited what might be termed 'exaggerated' preshaping - rotating more than minimally necessary. Further analysis of manipulation effectiveness (Effectiveness) shows that for high-rotation tasks, the comparable amount of preshaping in Direct techniques led to more successful rotation completion than in Indirect techniques. This suggests that while the *quantity* of preshaping converges at high difficulties, its *quality* or effectiveness remains superior in Direct manipulation.

In contrast, participants using Indirect techniques consistently performed less preshaping, potentially because the absence of direct visual reference, in combination with the consistent lack of feedback from physical contact, provides a weaker stimulus for anticipatory action [30]. The reduced preparation before the grasp appears to have been compensated for *during* the manipulation. This interpretation is corroborated by the significantly longer Manipulation Time observed with Indirect techniques, where users required more time post-grasp to adjust and complete the rotation. This effect was most apparent in smaller Rotation trials, where participants were more likely to attempt completing the task in a single grasp, as indicated by Clutch Count.

To address RQ2, we conclude that Indirect manipulation induces a preshaping tendency that mirrors Direct manipulation - scaling with task difficulty - but is significantly reduced in magnitude, and perhaps quality, indicated by the defaulting to a pronated posture observed for Gaze&Pinch. This lack of preparatory action is subsequently compensated for with longer-lasting and more demanding post-acquisition manipulation to achieve the same goal.

5.3 Preshaping with Bare-Hand and Controller

In addressing our third research question (RQ3), we found that the input Modality significantly influenced preshaping behaviour. Results from Preshaping Magnitude indicate that Bare-Hand interaction induced more preshaping than Controller-based interaction. This difference was most pronounced at smaller Rotation levels and converged as the required rotation increased. This pattern suggests a tendency for users to "overperform" preshaping with their bare hands when the task is less demanding. However, at higher Rotation levels, the biomechanical difficulty of rotating the hand to extreme angles likely creates the aforementioned ceiling effect, limiting the extent of this preparatory movement. In terms of effectiveness, Effectiveness revealed that as the required Rotation increased, Bare-Hand manipulation became progressively less effective than Controller manipulation. This suggests that while Bare-Hand interaction elicits a greater *quantity* of preshaping, this preparation is less efficient at solving the rotational task compared to the more constrained but precise movements afforded by the Controller.

Further analysis reveals nuances in how this preparation unfolds. The longer Preparation Time with Bare-Hand, particularly in Gaze&Pinch, suggests that participants required more time for cognitive processing - observing the object and planning their grasp - rather than for executing the movement itself. This is supported by the fact that the longer duration did not result in a larger accumulated hand rotation (Preparation Rotation) overall. An interaction effect in Preparation Rotation showed that between Direct techniques, Bare-Hand accumulated less rotation than the Controller yet still produced a greater final preshaping posture (Preshaping Magnitude). This indicates that Bare-Hand movements may be more efficient in reaching the target posture, possibly avoiding the minor, exploratory adjustments sometimes seen with controllers, due to the direct and unambiguous visual reference of the hand's orientation.

Conversely, with Indirect techniques, Bare-Hand interaction resulted in more accumulated rotation (Preparation Rotation), which aligns with its greater preshaping magnitude and longer preparation time in this condition. Despite this increased preshaping, Manipulation Time was significantly longer for Bare-Hand than for Controller, particularly at smaller Rotation levels where more preshaping was observed. This relationship - more preparation leading to longer manipulation - suggests that users have less precise control over their preshaping with bare hands. The larger Manipulation Translation with Bare-Hand further supports this interpretation, indicating less stable or consciously controlled hand movements than with Controller.

To compensate for impaired precision, participants appear to have adopted a strategy of increased clutching. The results from Clutch Count show that participants performed more clutches with Bare-Hand, particularly in tasks with larger rotations. To address RQ3, we conclude that while Controller is associated with more performant preshaping, the magnitude and efficiency of preparatory rotation with Bare-Hand interactions depend on the Directness dimension. Participants appear to compensate for the inherent imprecision of bare-handed manipulation by adopting a strategy of more frequent clutching, while holding a controller uniquely afforded in-hand rotations as a compensatory strategy.

5.4 Implications

Clutching. One item of interaction that is closely related to preshaping is interaction overhead in the form of clutching. In comprehensive surveys on 3D object interactions, such as the one by Argelaguet et al. in 2013 [4], we find that almost 20% of the techniques require some form of clutching as part of their main interaction mechanism. This is in agreement with theory work from Bowman, where it is described how to perform large manipulations on distant objects that users must repeatedly grab, rotate, and release to be able to complete via clutching [9, 10]. And perhaps the only alternative to clutching is a compromise with larger CD-gains. However, that also implies anisomorphic mappings, which are harder to control, like the Go-Go technique [43]. Our results empirically underpin the idea that if we achieve well-informed preshaping, the number of clutches can be minimised, therefore highlighting the implication of preshaping to a full interaction performance. We further show the preshaping magnitude different interaction techniques afford by default, and thus speculate on how to support the preshaping behaviour of users. For example, a system could, based on preparatory motions, infer that a rotation is being prepared and present a momentary visual feedforward of the hand in a task-optimised posture, proactively guiding the user [36, 65]. In extension, it is also worth noting that preshaping can only be facilitated by understandable task requirements, and that any sufficiently complex manipulation suppresses preshaping as mental rotation effort rises, leading to clutching. Consequently, we suggest that designers always consider whether applications could present manipulation tasks in a way that guides the user to better perceive spatial affordances, e.g. by increasing model detail, adding depth cues, and lowering angular disparity whenever possible.

Guiding Hands through Sensory Feedback. It is always the case that hand posture converges to neutral, when spatial reasoning fails, to the detriment of task performance [2, 37]. We can now show how input techniques further complicate preshaping when decoupling targeting (gaze) from the manipulator (hand) as it happens in Gaze&Pinch. Indeed, our results confirmed a hinted "preshaping deficit" for the established gaze-assisted Indirect Bare-Hand [30, 34]. We found that preparatory action had significantly worse carry-over to interaction performance. Meaning, the deficit is paid for through longer, less efficient and effective post-acquisition manipulations. This finding is immediately relevant for the growing number of systems that rely on gaze, and other indirect manual selection methods, like the hand-raypointing, which enforces strict constraints on preshaping [30, 34]. To mitigate potential deficits, indirect manipulation techniques should strive to better benefit from the additional motorspace freedom and could extend on the pre-selection feedback mechanisms included in this and prior work [34, 54]. For example, once a user's gaze dwells on a manipulable object, the system renders a hand-centric visual proxy or an object-local indicator that dynamically responds to the user's hand orientation. This would encourage users to perform the necessary preparatory rotation before locking the connection to the object, reinforcing the preshaping phase for interactions that otherwise rely purely on proprioception. Future work should focus on designing and evaluating such feedback systems to determine how to best guide users toward effective preparation without being intrusive, thereby improving the performance and intuitiveness of a wide range of indirect interaction techniques.

Preshaping as Signal. This study validates that preshaping kinematics encode user intent before a manipulation is executed. Future research could leverage this by developing *anticipatory interfaces* that interpret this continuous stream of kinematic data, not as a precursor to a single grasp type, but as a probabilistic indicator of a desired outcome. Instead of relying on a discrete command-response model (e.g., point, confirm, manipulate), these systems could classify a user's intent based on the dynamics of their preparatory movements. For instance, the velocity and aperture of a hand's approach could dynamically adjust CD-gain for the subsequent manipulation. An expansive, open-handed approach might signal an intent for large-scale translation, causing the system to preemptively raise CD-gain. A tense, precise pincer posture might signal a need for fine-grained rotation, prompting the interface to offer a temporary rotational stabiliser or automatically separate the rotational DOFs [47]. A hesitant or unstable approach might trigger the display of virtual aids, such as rotational guides or snapping points. This reframes XR interaction design from the creation of explicit tools to a symbiotic human-computer interaction, where the system preemptively adapts its state to augment the user's forthcoming actions, reducing cognitive load and improving performance.

5.5 Limitations & Future Work

The different preshaping behaviour between Bare-Hand and Controller could further be affected by the weight of the controller, changing the most efficient movement pattern. Similarly, the tangibility of and thus the ability to rotate the controller further in the hand without rotating the hand itself should be noted as an emergent, unexpected variable. The study and its conditions were designed to ensure controlled variance in the manipulation, to focus on the preceding behaviour, while even minimal differences between techniques, such as the pivot-point, influence coordination throughout the interaction in the abstract docking task. Further study of realistic application use would be interesting to validate the findings. This work provides evidence of behavioural variation between techniques with controlled similarities, e.g. how selection is performed by the closure of the index-to-thumb. Other techniques will provide different constraints in both acquisition and manipulation to which preshaping must adapt. Further studies could examine how grab-points and other grasping gestures or handheld devices may afford different preshaping patterns, e.g., full-hand grasping, gloves or haptic controllers [29]. This work also focused on abstract cubic objects for modelling general manipulation. Our results are likely to generalise to objects with similar properties, and it would be interesting to extend the study to different preshaping cues, such as objects with handles.

This paper did not explore the role of preshaping in far-field indirect interactions. Common techniques used for remote interaction, such as manual ray-pointing, have been shown to severely constrain the freedom of the arm and hand by enforcing a stabilised pointing sub-task for acquiring an object [34]. It would be interesting to investigate the preshaping behaviours and compensations of remote manipulation when the task goes beyond selection and dragging, comparing common techniques with very different motorspace freedom, e.g. hand- or arm-based ray and Gaze&Pinch. Such experiments would contribute fundamental knowledge on how to design interactions for objects outside of reach.

6 Conclusion

This work advances our understanding of interaction in XR and the different behavioural patterns that influence our performance of virtual object manipulation. Highlighting additional trade-offs to the different interaction methods beyond speed and ergonomics, we provide a causal explanation for the fact that some techniques are more prone to clutching than others in 6DOF translation and rotation tasks.

Furthermore, our work emphasises the need to analyse 3D object interactions as a compound of fundamental actions that influence each other; this principle is evident when we consider selection as a separate action from confirmation, which nonetheless has carry-over effects. Our experiments characterise and show the carry-over effects of preparatory and coordination actions in the form of preshaping across Directness and Modality, represented by the manual interaction techniques VirtualHand, DirectController, and the gaze-assisted indirect Gaze&Pinch and Gaze&Controller. We found that preshaping scales with anticipated task difficulty for all techniques, while Directness and Modality further modulate the response patterns. Direct techniques induce intuitive behavioural patterns resulting in effective posing of the hand. Conversely, indirect techniques elicit efficient but attenuated and ineffective responses in the absence of hand transport. Future XR systems should therefore be designed to interpret and support these behavioural patterns for effortless 3D manipulation.

Acknowledgments

This work was supported by funds from a Google Research Gift award ('Multimodal and Gaze + Gesture Interactions in XR'), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant no. 101021229 GEMINI), and the Danish National Research Foundation under the Pioneer Centre for AI in Denmark (DNRF grant P1).

References

[1] Florian Alt, Andreas Bulling, Gino Gravanis, and Daniel Buschek. 2015. Gravityspot: Guiding users in front of public displays using on-screen visual cues. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 47–56.

[2] Caterina Ansuini, Marco Santello, Stefano Massaccesi, and Umberto Castiello. 2006. Effects of End-Goal on Hand Shaping. *Journal of Neurophysiology* 95, 4 (2006), 2456–2465. https://doi.org/10.1152/jn.01107.2005 arXiv:https://doi.org/10.1152/jn.01107.200 PMID: 16381806.

[3] Alejandro Aponte, Arthur Caetano, Yunhao Luo, and Misha Sra. 2024. GraV: Grasp Volume Data for the Design of One-Handed XR Interfaces. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 151–167. https://doi.org/10.1145/3643834.3661567

[4] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136.

[5] 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. arXiv:2403.07210 [cs.HC]

[6] Victoria Bellotti, Maribeth Back, W Keith Edwards, Rebecca E Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 415–422.

[7] Joanna Bergström, Tor-Salve Dalsgaard, Jason Alexander, and Kasper Hornbæk. 2021. How to evaluate object selection and manipulation in vr? guidelines from 20 years of studies. In *proceedings of the 2021 CHI conference on human factors in computing systems*. 1–20.

[8] Andreea Dalia Blaga, Maite Frutos-Pascual, Chris Creed, and Ian Williams. 2025. VR-Grasp: A Human Grasp Taxonomy for Virtual Reality. *International Journal of Human–Computer Interaction* 41, 7 (2025), 4406–4422. https://doi.org/10.1080/10447318.2024.2351719 arXiv:https://doi.org/10.1080/10447318.2024.2351719

[9] 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

[10] Doug A Bowman, Regis Kopper, and Felipe Bacim. 2019. Effortless 3D Selection through Progressive Refinement.

[11] Giuseppe Caggianese, Luigi Gallo, and Pietro Neroni. 2018. The vive controllers vs. leap motion for interactions in virtual environments: a comparative evaluation. In *International Conference on Intelligent Interactive Multimedia Systems and Services*. Springer, 24–33.

[12] Stuart K Card, William K English, and Betty J Burr. 1978. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics* 21, 8 (1978), 601–613.

[13] Alessandro Carfì, Timothy Patten, Yingyi Kuang, Ali Hammoud, Mohamad Alameh, Elisa Maiettini, Abraham Itzhak Weinberg, Diego Faria, Fulvio Mastrogiovanni, Guillem Alenyà, Lorenzo Natale, Véronique Perdereau, Markus Vincze, and Aude Billard. 2021. Hand-Object Interaction: From Human Demonstrations to Robot Manipulation. *Frontiers in Robotics and AI* Volume 8 - 2021 (2021). https://doi.org/10.3389/frobt.2021.714023

[14] Géry Casiez, Nicolas Roussel, and Daniel Vogel. 2012. 1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems. 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, 2527–2530. https://doi.org/10.1145/2207676.2208639

[15] Tobias Drey, Michael Montag, Andrea Vogt, Nico Rixen, Tina Seufert, Steffi Zander, Michael Rietzler, and Enrico Rukzio. 2023. Investigating the Effects of Individual Spatial Abilities on Virtual Reality Object Manipulation. 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, Article 398, 24 pages. https://doi.org/10.1145/3544548.3581004

[16] Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2016. Do that, there: an interaction technique for addressing in-air gesture systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2319–2331.

[17] Tangui Marchand Guerniou, Maxime Jouin, and Jérémy Lacoche. 2023. Smooth Hand Preshaping for Visually Realistic Grasping in Virtual Reality. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 801–802.

[18] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52 (1 1988), 139–183. Issue C. https://doi.org/10.1016/S0166-4115(08)62386-9

[19] Ken Hinckley, Seongkook Heo, Michel Pahud, Christian Holz, Hrvoje Benko, Abigail Sellen, Richard Banks, Kenton O'Hara, Gavin Smyth, and William Buxton. 2016. Pre-Touch Sensing for Mobile Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2869–2881. https://doi.org/10.1145/2858036.2858095

[20] Ken Hinckley and Daniel Wigdor. 2002. Input technologies and techniques. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications* (2002), 151–168.

[21] Shota Imaizumi and Mie Sato. 2024. Effective Visual Feedback for Virtual Grasping. In *ACM SIGGRAPH 2024 Posters* (Denver, CO, USA) *(SIGGRAPH '24)*. Association for Computing Machinery, New York, NY, USA, Article 23, 2 pages. https://doi.org/10.1145/3641234.3671066

[22] Alexandru Itu, Andreea Beraru, and Ionel Staretu. 2008. Developing grasping pre-shaping in virtual environment based on real object shape analysis. *Product Engineering: Tools and Methods Based on Virtual Reality* (2008), 541–554.

[23] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-based interaction: a framework for post-WIMP interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 201–210. https://doi.org/10.1145/1357054.1357089

[24] Robert JK Jacob, Linda E Sibert, Daniel C McFarlane, and M Preston Mullen Jr. 1994. Integrality and separability of input devices. *ACM Transactions on Computer-Human Interaction (TOCHI)* 1, 1 (1994), 3–26.

[25] Marc Jeannerod. 1999. Visuomotor channels: Their integration in goal-directed prehension. *Human Movement Science* 18, 2-3 (1999), 201–218.

[26] Jari Kangas, Sriram Kishore Kumar, Helena Mehtonen, Jorma Järnstedt, and Roope Raisamo. 2022. Trade-off between task accuracy, task completion time and naturalness for direct object manipulation in virtual reality. *Multimodal Technologies and Interaction* 6, 1 (2022), 6.

[27] Alexander Kulik, André Kunert, and Bernd Froehlich. 2020. On motor performance in virtual 3d object manipulation. *IEEE transactions on visualization and computer graphics* 26, 5 (2020), 2041–2050.

[28] 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

[29] Jaeyeon Lee, Mike Sinclair, Mar Gonzalez-Franco, Eyal Ofek, and Christian Holz. 2019. TORC: A virtual reality controller for in-hand high-dexterity finger interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[30] Mathias Nørhede Lystbæk, Thorbjørn Mikkelsen, Roland Krisztandl, Eric J Gonzalez, Mar Gonzales-Franco, Hans Gellersen, and Ken Pfeuffer. 2024. Hands-on, Hands-off: Gaze-Assisted Bimanual 3D Interaction. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST'24)*. Association for Computing Machinery.

[31] Alexander Masurovsky, Paul Chojecki, Detlef Runde, Mustafa Lafci, David Przewozny, and Michael Gaebler. 2020. Controller-free hand tracking for grab-and-place tasks in immersive virtual reality: Design elements and their empirical study. *Multimodal Technologies and Interaction* 4, 4 (2020), 91.

[32] Daniel Mendes, Fabio Marco Caputo, Andrea Giachetti, Alfredo Ferreira, and Joaquim Jorge. 2019. A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, Vol. 38. Wiley Online Library, 21–45.

[33] Shawna Meyer, Oryx Cohen, and Erik Nilsen. 1994. Device comparisons for goal-directed drawing tasks. In *Conference companion on Human factors in computing systems*. 251–252.

[34] Thorbjørn Mikkelsen, Qiushi Zhou, Mathias N. Lystbæk, Yang Liu, Hans Gellersen, and Ken Pfeuffer. 2025. DOF-Separation for 3D Manipulation in XR: Understanding Finger-Wrist Separation to Simultaneously Translate and Rotate Objects. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 1246–1256. https://doi.org/10.1109/ISMAR67309.2025.00130

[35] Shiori Mizuno, Asuka Komeda, Naoko Yoshii, Tomoko Yonezawa, Masami Takata, and Kazuki Joe. 2012. Pre-motion Based VR Object Manipulation: Definitions and Preliminary Experiments. In *AsiaSim 2012: Asia Simulation Conference 2012,*

Shanghai, China, October 27-30, 2012. Proceedings, Part I. Springer, 499–507.
[36] Andreea Muresan, Jess Mcintosh, and Kasper Hornbæk. 2023. Using Feedforward to Reveal Interaction Possibilities in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 30, 6, Article 82 (Sept. 2023), 47 pages. https://doi.org/10.1145/3603623
[37] J. Napier and R.H. Tuttle. 1993. *Hands*. Princeton University Press. https://books.google.dk/books?id=mNQM4jvG7hgC
[38] Tomi Nukarinen, Jari Kangas, Jussi Rantala, Olli Koskinen, and Roope Raisamo. 2018. Evaluating ray casting and two gaze-based pointing techniques for object selection in virtual reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 86, 2 pages. https://doi.org/10.1145/3281505.3283382
[39] Yeji Park, Jiwan Kim, and Ian Oakley. 2025. GazeHandSync: Mitigating Late-Trigger Errors for Seamless Gaze-Hand Interactions. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*. Association for Computing Machinery, New York, NY, USA, Article 12, 8 pages. https://doi.org/10.1145/3715669.3723126
[40] James W Pellegrino, Roberta L Klatzky, and Brian P McCloskey. 1989. Timecourse of preshaping for functional responses to objects. *Journal of Motor Behavior* 21, 3 (1989), 307–316.
[41] Ken Pfeuffer, Hans Gellersen, and Mar Gonzalez-Franco. 2024. Design principles and challenges for gaze+ pinch interaction in xr. *IEEE Computer Graphics and Applications* 44, 3 (2024), 74–81.
[42] 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
[43] 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
[44] I. Poupyrev, S. Weghorst, M. Billinghurst, and T. Ichikawa. 1998. Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. *Computer Graphics Forum* 17 (1998), 12 pages.
[45] Mores Prachyabrued and Christoph W Borst. 2014. Visual feedback for virtual grasping. In *2014 IEEE symposium on 3D user interfaces (3DUI)*. IEEE, 19–26.
[46] Jason L. Reisman, Philip L. Davidson, and Jefferson Y. Han. 2009. A screen-space formulation for 2D and 3D direct manipulation. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology* (Victoria, BC, Canada) *(UIST '09)*. Association for Computing Machinery, New York, NY, USA, 69–78. https://doi.org/10.1145/1622176.1622190
[47] Francielly Rodrigues, Alexander Giovannelli, Leonardo Pavanatto, Haichao Miao, Jauvane C de Oliveira, and Doug A Bowman. 2023. Amp-it and wisdom: Improving 3d manipulation for high-precision tasks in virtual reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 303–311.
[48] Luis F Schettino, Sergei V Adamovich, and Howard Poizner. 2003. Effects of object shape and visual feedback on hand configuration during grasping. *Experimental Brain Research* 151 (2003), 158–166.
[49] Andrew Sears and Ben Shneiderman. 1991. High precision touchscreens: design strategies and comparisons with a mouse. *International Journal of Man-Machine Studies* 34, 4 (1991), 593–613.
[50] Adwait Sharma, Alexander Ivanov, Frances Lai, Tovi Grossman, and Stephanie Santosa. 2024. Graspui: Seamlessly integrating object-centric gestures within the seven phases of grasping. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 1275–1289.
[51] Adwait Sharma, Joan Sol Roo, and Jürgen Steimle. 2019. Grasping microgestures: Eliciting single-hand microgestures for handheld objects. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[52] JOHN F Soechting and Martha Flanders. 1993. Parallel, interdependent channels for location and orientation in sensorimotor transformations for reaching and grasping. *Journal of neurophysiology* 70, 3 (1993), 1137–1150.
[53] Tamara Supuk, Timotej Kodek, and Tadej Bajd. 2005. Estimation of hand preshaping during human grasping. *Medical engineering & physics* 27, 9 (2005), 790–797.
[54] Esen K Tütüncü, Mar Gonzalez-Franco, and Eric J Gonzalez. 2025. HandOver: Enabling Precise Selection & Manipulation of 3D Objects with Mouse and Hand Tracking. *arXiv preprint arXiv:2508.00211* (2025). https://doi.org/10.48550/arXiv.2508.00211
[55] Jo Vermeulen, Kris Luyten, Karin Coninx, Nicolai Marquardt, and Jon Bird. 2015. Proxemic flow: Dynamic peripheral floor visualizations for revealing and mediating large surface interactions. In *IFIP Conference on Human-Computer Interaction*. Springer, 264–281.
[56] 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)*. Association for Computing Machinery.
[57] Uta Wagner, Jinwook Kim, Zhikun Wu, Qiushi Zhou, Mario Romero, Alessandro Iop, Tiare Feuchtner, and Ken Pfeuffer. 2025. A Study of Multimodal Pen+ Gaze Interaction Techniques for Shape Point Translation in Extended Reality. In *A Study of Multimodal Pen+ Gaze Interaction Techniques for Shape Point Translation in Extended Reality*.
[58] Uta Wagner, Mathias N. Lystbæk, Pavel Manakhov, Jens Emil Sloth 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, Article 252, 15 pages. https://doi.org/10.1145/3544548.3581423
[59] 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
[60] 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
[61] Sara A Winges, Douglas J Weber, and Marco Santello. 2003. The role of vision on hand preshaping during reach to grasp. *Experimental Brain Research* 152 (2003), 489–498.
[62] 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* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963
[63] 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*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376876
[64] 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*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445343
[65] Qiushi Zhou, Antony Chacon, Jiahe Pan, and Wafa Johal. 2025. Assisting MoCap-Based Teleoperation of Robot Arm Using Augmented Reality Visualisations. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 1765–1769. https://doi.org/10.1109/HRI61500.2025.10974111