RESEARCH-ARTICLE

Trade-offs in Virtual Grasping: The Interplay of Interaction Fidelity and Object Affordance

JEMIN LEE, Kumoh National Institute of Technology, Gumi, Kyongsangbuk-do, South Korea

JEONGHYEON KIM, Kumoh National Institute of Technology, Gumi, Kyongsangbuk-do, South Korea

HYEONGJUN KANG, Kumoh National Institute of Technology, Gumi, Kyongsangbuk-do, South Korea

HOON RYU, Kumoh National Institute of Technology, Gumi, Kyongsangbuk-do, South Korea

YOUNGWON KIM, Kumoh National Institute of Technology, Gumi, Kyongsangbuk-do, South Korea

Published: 04 December 2025

Citation in BibTeX format

VRST '25: 31st ACM Symposium on Virtual Reality Software and Technology
November 12 - 14, 2025
QC, Montreal, Canada

Conference Sponsors:
SIGCHI
SIGGRAPH

VRST '25: Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology (November 2025)
https://doi.org/10.1145/3756884.3765990
ISBN: 9798400721182

Trade-offs in Virtual Grasping: The Interplay of Interaction Fidelity and Object Affordance

Jemin Lee*
Department of Software, School of Computer Engineering
Kumoh National Institute of Technology
Gumi-si, Republic of Korea
char3941@kumoh.ac.kr

Jeonghyeon Kim*
Department of Software, School of Computer Engineering
Kumoh National Institute of Technology
Gumi-si, Republic of Korea
dnrgusrla1@kumoh.ac.kr

Hyeongjun Kang
Department of Software, School of Computer Engineering
Kumoh National Institute of Technology
Gumi-si, Republic of Korea
johj0816@kumoh.ac.kr

Hoon Ryu†
School of Computer Engineering
Kumoh National Institute of Technology
Gumi, Republic of Korea
elec1020@kumoh.ac.kr

Youngwon Kim†
School of Computer Engineering
Kumoh National Institute of Technology
Gumi, Republic of Korea
kim01@kumoh.ac.kr

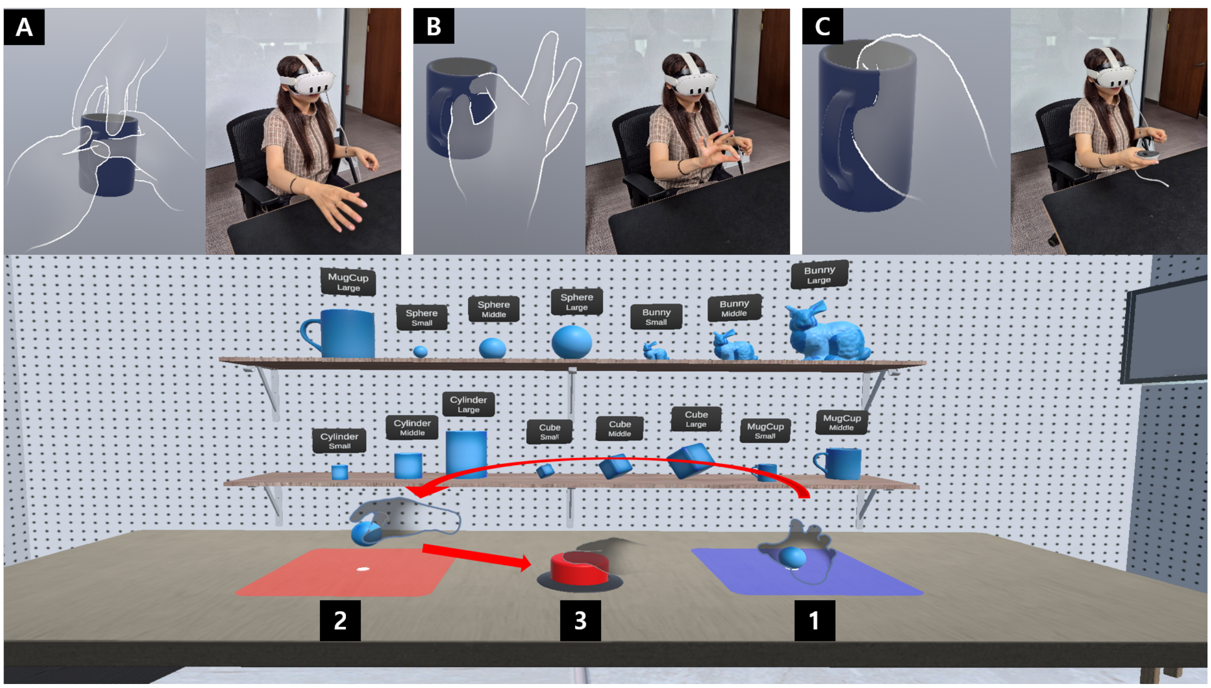

Figure 1: Task flow and grasping methods. A: Plausible Gesture grab, B: Pinch grab, C: Controller grab. Participants 1) picked up the object from the blue area, 2) moved it to the red area, and 3) pressed a button to complete the task.

Abstract

In Virtual Reality (VR), object grasping is a core interaction that critically influences both user immersion and task performance. While contemporary systems offer both high-precision controllers and intuitive hand tracking, they present a trade-off between performance and naturalness. However, empirical guidance for selecting

*These authors contributed equally to this work.
†Corresponding authors.

VRST '25, Montreal, QC, Canada

ACM ISBN 979-8-4007-2118-2/25/11
https://doi.org/10.1145/3756884.3765990

an optimal grasping method for object grasping remains limited. In particular, how object shape and size (as affordance-related factors) modulate this trade-off within a standardized pick-and-place paradigm is underexplored.

We investigate the interplay between interaction fidelity and object shape/size and its impact on user performance and experience in a controlled pick-and-place task. We conducted a within-subjects study with a 3 (grasping modality: controller, pinch, plausible gesture) × 5 (object shape: cube, sphere, cylinder, handled mug, complex model) × 3 (object size) factorial design. We measured objective performance (task completion time, placement accuracy) and subjective experience (NASA-TLX workload, IPQ presence).

Our findings provide evidence-based answers to the scoped question: "Which grasping method is best suited for an object of a given shape and size in a pick-and-place task?" Ultimately, this work offers actionable guidelines to help VR developers design effective and satisfying object-grasping interactions tailored to users' task goals and virtual environments, without claiming a single universally "best" method.

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**; *Haptic devices*; *User interface design.*

Keywords

Virtual Reality, Grasping, Interaction Fidelity, Affordance, Hand Tracking, User Experience, 3D User Interfaces

ACM Reference Format:

Jemin Lee, Jeonghyeon Kim, Hyeongjun Kang, Hoon Ryu, and Youngwon Kim. 2025. Trade-offs in Virtual Grasping: The Interplay of Interaction Fidelity and Object Affordance. In *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3756884.3765990

1 Introduction

Grasping is a fundamental interaction in Virtual Reality (VR), critical to both task success and the quality of the user experience [35, 38]. Modern VR systems provide a range of grasping methods, including the *controller-based grab* using physical buttons, the *plausible gesture grab* that mimics natural hand motions, and the *pinch grab* executed by touching the thumb and index finger [30, 38]. Despite this variety, systematic investigations focusing specifically on object grasping remain limited; many prior studies have examined only subsets of methods or simplified tasks, leaving open the question of when each technique is preferable.

Prior research indicates that controllers, with their high precision and reliability, are advantageous for enhancing *pragmatic quality*, or task performance [25, 35]. In contrast, hand tracking excels at delivering *hedonic quality* by enabling intuitive and natural interactions [4, 29]. VR grasping entails a non-linear performance–immersion trade-off: higher fidelity is not always better; in some contexts lower-fidelity methods are more efficient [13, 37], and misapplied realism can trigger an *Uncanny Valley of Interaction* [1, 32].

This trade-off is further complicated by the concept of *affordance.* Affordance theory posits that an object's physical properties, such as its shape and size, strongly suggest the grasping strategy a user will employ, for instance, a power grip versus a precision grip [14, 34]. Extending this perspective, prior studies have shown that object shape and size can either reinforce performance advantages of controllers or amplify the naturalness benefits of hand tracking. For example, users must adapt grasping strategies depending on the shape (cylinder vs. cube) and size (small, medium, large) of virtual objects when stacking them[27], and geometric properties such as object size and taper angle can alter perceptual interpretation during grasping, thereby influencing how users perceive and execute their grasps.[42]

Consequently, the effectiveness of a grasping method in a pick-and-place context depends not only on the user's action but also on the shape and size of the target object as affordance-related factors. However, empirical research specifically on the interplay between grasping fidelity and object affordance remains scarce.

This study systematically analyzes the interaction between two key dimensions, interaction fidelity and object affordance in grasping. To this end, we present a set of objects with varying affordances, including a sphere, a cylinder, a cube, a handled mug, and a complex, irregular object. By additionally varying the size of these objects, we conduct a fine-grained investigation into the effectiveness and limitations of three grasping techniques: controller-based, pinch-based, and plausible gesture-based. Here, the pinch grab refers to the built-in headset-supported pinch interaction, where virtual penetration of the fingers into objects is permitted, rather than a real-world physical pinch. The ultimate goal of this study is to provide evidence-based guidelines for choosing among grasping methods under specified task and object affordance conditions (including size), evaluated on completion time, errors, perceived naturalness, and workload.

To this end, we pose the following research questions. *First (RQ1)*, how do the three grasping techniques affect task performance, measured by efficiency (completion time) and accuracy, when manipulating objects with diverse affordances? *Second (RQ2)*, how do the three grasping techniques impact the user's subjective experience, measured by cognitive load (NASA-TLX) and the sense of presence and immersion (IPQ)?

2 Related Work

2.1 Controller vs. Hand Tracking

The comparison between controllers and hand tracking is a well-established topic in VR interaction research, often summarized as a trade-off between precision and naturalness [40].

Controllers provide clear, reliable input through physical buttons. This button-based input, with its low latency and haptic feedback (arising from the mechanical haptic sensation of pressing a button), consistently demonstrates superior performance in tasks requiring fine motor skills [24, 25, 30, 31, 35, 38, 40]. These characteristics make controllers highly suitable for precision-critical activities, though the necessity of holding a physical tool can reduce immersion [30].

In contrast, hand tracking mirrors the user's bare hands in the virtual environment, enabling intuitive interactions without a separate controller and offering significant potential for high immersion [26, 35, 40]. Users often prefer the naturalness of hand tracking, particularly in exploratory or slow-paced tasks [29]. This advantage, however, often comes at the cost of performance, as hand tracking can suffer from lower precision, higher latency, and tracking errors during fast or occluded movements [25, 38].

Within this trade-off, the **pinch gesture** is a widely used compromise. Since it primarily measures the distance between the thumb and index finger, it operates reliably even with noisy tracking data [30, 36]. However, using two fingers to grasp most objects is an unnatural method that can conflict with other gestures [30, 36].

2.2 Interaction Fidelity and Sensory Mismatch

Fidelity is a core concept in VR interactions, referring to how accurately a virtual experience mimics its real-world counterpart across multiple dimensions, including visual, kinematic, and haptic [4, 32]. However, merely increasing fidelity does not always lead to a positive user experience, as a mismatch between sensory information can produce negative outcomes [32].

A primary issue is the visual-proprioceptive mismatch, which occurs when there is a discrepancy between the user's actual and visually displayed body position, degrading the sense of control and embodiment [4, 13, 37]. For example, Canales et al. showed that users achieved better task performance with a virtual hand that accurately mirrored their real movements (even allowing object penetration) compared to a hand whose movements were corrected to prevent it [4]. This suggests that sensory consistency may be more critical for performance than strict visual realism. This mismatch can also extend to the Haptic Uncanny Valley, where enhanced haptic feedback, if not congruent with visual cues, paradoxically decreases the subjective sense of realism, highlighting the need for a delicate balance among sensory modalities [1].

2.3 Affordance and Interaction Strategy

Affordance, a concept from psychologist James J. Gibson, refers to the action possibilities an environment or object offers a user [16, 19, 21, 34]. Don Norman extended this into the design field, emphasizing perceived affordances, cues that signal to a user what actions are possible [34].

In grasping, an object's physical properties naturally suggest a certain type of grip. Grasp taxonomies show how object shape and size guide specific grasp types like power and precision grips [9, 14, 22, 33]. The human brain pays more attention to graspable objects and pre-plans the grasp by analyzing these affordances. Neuroscience studies show that even before reaching for an object, humans determine the appropriate hand orientation and finger configuration based on its shape, with gaze guiding the hand's movement [8, 12, 17]. During this process, motor areas of the brain are automatically activated [18, 20].

Particularly interesting is the affordance competition that arises when an object offers multiple affordances simultaneously [5, 6]. The choice of action is ultimately determined by the user's task-oriented goal and is embedded in a continuous loop of **sensorimotor control**, where the brain constantly refines actions based on sensory feedback [10, 11, 41]. Furthermore, modern theories like **predictive coding** propose that the brain actively predicts the sensory consequences of its actions. A mismatch between the predicted sensation (e.g., feeling a solid surface) and the actual sensory input (e.g., no haptics in VR) generates a "prediction error" that requires additional cognitive resources to resolve, potentially impacting performance [7, 15].

Building upon these prior works, our research takes a step further by exploring how the user's interaction modality influences the perception of and reaction to object affordances. We aim to analyze how the intrinsic advantages and disadvantages of each grasping method guide a user's action strategy when interacting with objects that possess complex affordances.

3 System Implementation

We developed a Unity-based VR system on Meta Quest 3 to support three grasping methods: Controller Grab, Pinch Grab, and Plausible Gesture Grab. The design emphasized consistency across conditions so that only the input technique varied.

Controller Grab was implemented through the grip button of the Touch controllers. A virtual hand model was displayed at the controller's position, and grasped objects followed its motion with collisions managed by the physics engine.

Pinch Grab relied on the Quest hand-tracking SDK. A grasp was triggered when the distance between the thumb and index finger fell below roughly 1 cm, automatically attaching the nearest object. Release occurred when the fingers opened. This method enabled reliable precision but felt less natural for larger objects.

Plausible Gesture Grab used full-hand tracking to render a virtual hand that adjusted dynamically to object geometry. Each object defined a broad interaction zone for detecting possible grabs and finer colliders for finger contacts. A grasp was confirmed when sufficient contacts and finger curvature thresholds were met. To enhance realism, the virtual hand conformed to the object surface while preventing penetration, with interpolation applied to maintain smooth motion aligned with the user's hand.

4 Experiment Design

To systematically evaluate grasping interactions in a virtual environment, this study employs a 3 (Interaction Method) $\times$ 5 (Object Shape) $\times$ 3 (Object Size) within-subjects repeated measures design. All participants performed all 45 unique experimental conditions, and each condition was repeated three times per interaction method. Therefore, each participant completed a total of 135 grasping trials (45 conditions $\times$ 3 repetitions), which controls for inter-participant variance and thus increases the statistical power of the study. To counteract learning effects from the order of interaction methods, the presentation sequence was perfectly balanced using a Latin Square design. All procedures were conducted in accordance with ethical guidelines.

The theoretical basis for our experimental design is partially grounded in Biederman's Recognition-by-Components (RBC) Theory [2]. According to this theory, shapes like spheres, cubes, and cylinders are fundamental visual 'alphabets' that humans use to recognize object structures. By using these primitive shapes as experimental stimuli, we test grasping interactions on the very components that the human perceptual system uses to understand objects, thereby ensuring the cognitive validity of our experiment.

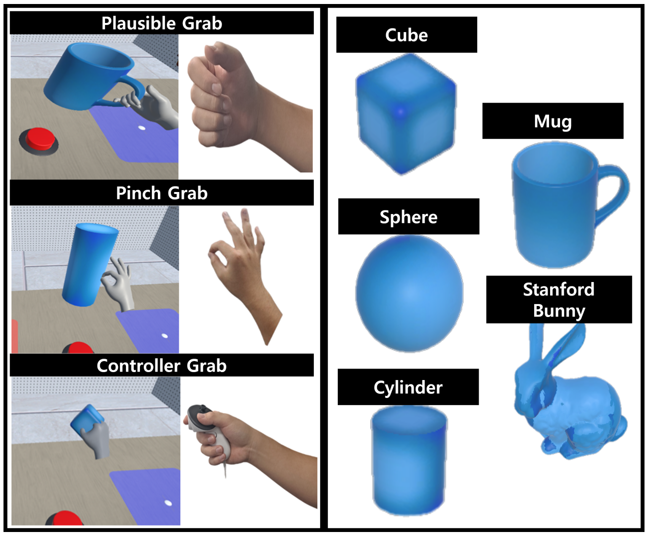

Figure 2: Left: the three interaction methods (Plausible Grab, Pinch Grab, and Controller Grab). Right: the set of stimulus objects (Cube, Sphere, Cylinder, Mug, and Stanford Bunny).

4.1 Independent Variables

The study has three independent variables: interaction method, object shape, and object size.

Interaction Method. Participants performed the task using three different grasping methods (see Figure 2).

Plausible Gesture Grab (High Fidelity). A full-hand hand-tracking modality that simulates the user's finger movements based on physics to grasp an object. This method provides the highest visual and kinematic fidelity and a strong sense of immersion, but it is susceptible to tracking errors and can increase physical fatigue during prolonged use.

Pinch Grab (Medium Fidelity). This symbolic, low-fidelity gesture executes a grasp by pinching the thumb and index finger together. This method is simple and reliable, as shown by Schäfer et al. [38], but can feel unnatural for grasping all types of objects. Although the gesture itself is high-fidelity in the real world, in our implementation penetration into virtual objects is permitted, which reduces its overall fidelity.

Controller Grab (Low Fidelity). An abstract, button-based modality that uses the physical grip button on the VR controller. Although the least realistic in terms of ecological fidelity, it offers distinct haptic feedback, low latency, and high reliability, providing clear advantages in speed and accuracy. For consistency with the other methods, a virtual hand was displayed and animated to form a grasping pose when the grip button was pressed.

Object Shape. To analyze how an object's affordance influences the user's grasping strategy, we selected five object shapes based on an 'affordance spectrum.' These shapes are visually summarized in Figure 2, along with their associated affordance characteristics and expected grasping strategies.

Cube. Note that the affordance descriptions in Figure 2 follow prior work on VR grasping [3], which showed that isotropic objects such as spheres typically elicit power grasps. However, we acknowledge that affordances are modulated by object size; for example, a small sphere may be more naturally grasped with a precision grip than with a full-hand power grip. Likewise, smaller cubes may encourage fingertip precision grasps on edges or vertices, whereas larger cubes are more likely to be enveloped with a power grasp.

Sphere. As an isotropic, non-directional shape, it is intended to purely afford a full-hand power grip [3]. Still, its affordance also depends on scale, with smaller spheres being more suited for precision grasps.

Cylinder. Its elongated, symmetric form partially limits full-hand wrapping and naturally affords a precision grip using the fingertips. Yet, depending on size, a thin cylinder strongly invites precision grasping, while a thicker or larger one can accommodate power-like wrapping with the whole hand.

Mug with Handle. This object provides a clear, learned, and functional semantic affordance: "handles are for grasping." This condition is used to evaluate whether users prioritize an object's meaning over its geometric shape.

Complex Object. We used the Stanford Bunny, an amorphous object with no obvious grasping points. This is to test the hypothesis that general-purpose interaction methods (controller, pinch) would be advantageous when affordances are absent.

Object Size. Object size is a critical factor in determining the grasp type. We used three sizes: small (3–5 cm), medium (7–9 cm), and large (12–15 cm). The small size naturally affords a precision grip, similar to the small objects handled in studies like Hiramatsu et al. [23]. The medium size is designed to fit comfortably in one hand and serves as a baseline for an ideal power grip, facilitating comparisons between interaction methods. The large size is somewhat cumbersome to grasp with one hand and is intended to test the limits of the power grip.

4.2 Dependent Variables

We evaluate the interactions comprehensively by measuring both objective performance and subjective experience metrics.

Task Completion Time (TCT). Measured as the total time in seconds from the moment the experiment participant presses a button to spawn the object until the experiment participant presses a button to confirm successful placement at the target location. The experiment participant confirmed placement once the object was stably positioned within the target area without further adjustment attempts. We deliberately avoided an automatic detection procedure (e.g., collision triggers) because participants were allowed to release and re-adjust the object multiple times before finalizing their placement, which required human judgment to ensure the task was truly complete.

Placement Accuracy. Measured as the Euclidean distance (in cm) on the XZ-plane between the center of the target and the Center of Mass of the placed object. We excluded the Y-axis (vertical) component from this measurement because larger objects tend to produce greater vertical offset errors due to their height, which could distort accuracy comparisons across differently sized stimuli.

Cognitive Load. After the task, participants completed the NASA-TLX questionnaire to measure perceived mental and physical demand, temporal pressure, effort, and frustration.

Presence and Immersion. We used the Igroup Presence Questionnaire (IPQ) to assess how much participants felt present and immersed in the virtual environment.

4.3 Terminology clarification

In this paper, we distinguish between a biomechanical grasp type and a VR interaction implementation. The term *power grip* refers to a grasp taxonomy category from the motor control and ergonomics literature, in which the object is stably enveloped by the whole hand [9, 14, 33]. By contrast, *plausible gesture grab* denotes our hand-tracking based interaction technique implemented in VR. It renders and constrains a virtual hand to follow the user's finger kinematics and object contacts, but it is not bound to a single taxonomic grasp; depending on the object's affordances and size, the resulting pose can resemble a power grip or a more precision-oriented configuration. For completeness, the *pinch grab* in our study is a symbolic two-finger gesture that allows virtual interpenetration.

4.4 Task and Procedure

Task: Participants performed a **'Pick-and-Place'** task, a standard paradigm in HCI research. The goal was to pick up an object from a virtual table and place it accurately at a designated target location.

Procedure: The experiment followed the sequence: [Briefing and Informed Consent] → [Training Session per Interaction Method] → [Main Task Blocks] → [Subjective Questionnaires]. After receiving detailed task instructions, participants provided written informed consent. For each interaction method assigned by the Latin Square, participants first completed a training session consisting of 15 practice trials (5 shapes × 3 sizes) to become familiar with the interaction modality.

The main task was performed as follows: participants pressed a virtual button in front of them to initiate each trial. This action caused the object to appear on the table and simultaneously triggered the start of the TCT measurement. The blue-marked area served only as a consistent starting location for the object to spawn. Participants were instructed to grasp the object that appeared from this area and place it onto the red-marked target area. The distance between the two areas was fixed at approximately 50 cm. Crucially, participants were given no constraints on how or where on the object they should grasp, allowing for natural interaction. To maintain consistency across conditions and reduce task complexity, all grasping actions were performed exclusively with the participant's dominant hand. Upon successful placement, they pressed the same virtual button again to end the trial. At that moment, the next object appeared immediately, and the timer restarted. Each block consisted of 15 such consecutive trials with randomized combinations of object shapes and sizes. An overview of this procedure is illustrated in Figure 1.

Participants repeated this block three times for each interaction method, resulting in a total of 135 trials (15 trials × 3 repetitions × 3 interaction methods). The number of repetitions was intentionally limited to avoid excessive learning or habituation effects that could bias the results. To minimize affordance learning, each condition appeared only once per block.

All subjective questionnaires were administered within the VR environment immediately after each block to minimize breaks in presence and enhance ecological validity. A short debriefing interview was conducted after the completion of all sessions.

4.5 Participants

15 participants (10 male, 5 female, mean age = 24.05, *SD* = 1.39) from the university community participated. All were right-handed with varying VR experience: 10 low, 2 moderate, 3 high. The experiment used Meta Quest 3 in a controlled lab setting with a co-located physical desk providing passive haptic feedback [28]. Participants gave informed consent and received $10 compensation.

5 Results

All data were analyzed using a three-way repeated-measures ANOVA to examine the effects of Grasp Method, Object Shape, and Object Size on the primary dependent variables. We verified the assumption of sphericity using Mauchly's test; where this assumption was violated, we applied Greenhouse–Geisser corrections to the degrees of freedom. Normality was assessed using Shapiro–Wilk tests on model residuals; although some deviations were observed, the ANOVA is robust to moderate non-normality in a balanced within-subjects design [39]. Effect sizes are reported as partial eta squared (η_p^2) and interpreted using Cohen's guidelines where .01, .06, and .14 represent small, medium, and large effects, respectively. For all significant effects, we conducted post-hoc pairwise two-tailed paired t-tests on participant means, with Holm–Bonferroni correction to control the family-wise error rate.

5.1 Results for Task Completion Time (TCT)

The analysis of Task Completion Time (TCT) revealed significant main effects for all three factors, as well as significant two-way and three-way interaction effects.

5.1.1 *Main Effects.*
All three independent variables significantly affected TCT. The effect of grasp method was significant, $F(2, 28) = 68.83, p < .001, \eta_p^2 = .831$. Post-hoc tests confirmed that controller-based grasping yielded the shortest TCTs, followed by pinch and plausible, with all pairwise differences statistically significant ($p_{\mathrm{Holm}} < .001$).

Object shape also had a significant effect, $F(4, 56) = 4.52, p = .0031, \eta_p^2 = .244$, with participants completing tasks faster with simpler shapes (e.g., Cube, Sphere) compared to more complex ones (e.g., Mug, Stanford Bunny) ($p_{\mathrm{Holm}} < .05$).

Finally, object size showed a significant effect, $F(2, 28) = 5.80, p = .0078, \eta_p^2 = .293$, such that medium-sized objects were grasped significantly faster than both small ($p_{\mathrm{Holm}} = .007$) and large ($p_{\mathrm{Holm}} = .005$) ones, while no difference was found between small and large objects ($p_{\mathrm{Holm}} = .89$).

5.1.2 *Interaction Effects.*
Multiple significant interactions were identified, revealing that the effectiveness of grasp methods depended on object properties.

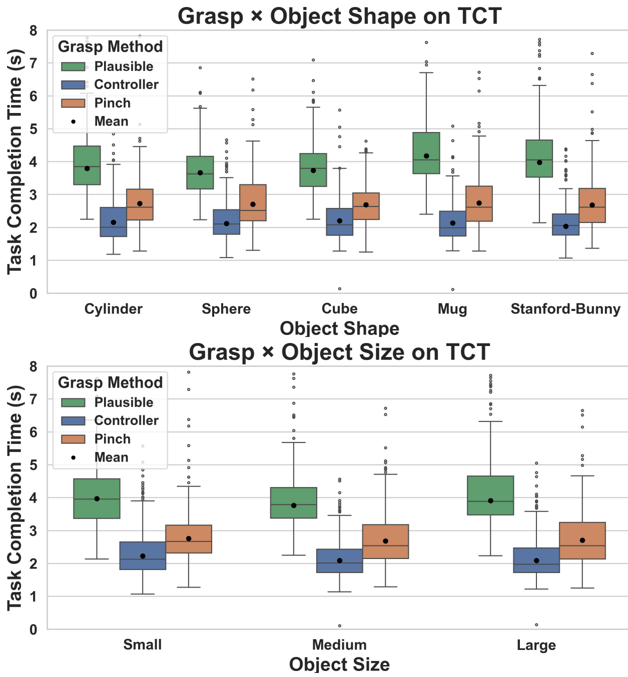

Figure 3: Interaction effects of Grasp Method with Object Shape (top) and Object Size (bottom) on Task Completion Time (TCT).

The interaction between grasp method and object shape was significant, $F(8, 112) = 4.44, p < .001, \eta_p^2 = .24$. While the controller method consistently outperformed the others, the performance gap was more pronounced for complex shapes, especially between controller and plausible grasping of the Mug (see top of Figure 3).

The interaction between grasp method and object size was also significant, $F(4, 56) = 6.41, p < .001, \eta_p^2 = .31$. While the controller method maintained relatively stable TCTs across object sizes, post-hoc analyses revealed that the plausible method required significantly longer completion times for both small and large objects compared to medium ones, whereas the pinch method showed a similar but weaker trend ($p < .05$). As illustrated in the bottom of Figure 3, these differences appear subtle in the boxplot visualization but were statistically reliable.

The interaction between object shape and object size was not significant, $F(8, 112) = 1.99, p = .054, \eta_p^2 = .12$, suggesting that the effect of size on TCT did not vary meaningfully across different object shapes.

Finally, a significant three-way interaction among grasp method, object shape, and object size emerged, $F(16, 224) = 2.82, p = .0004, \eta_p^2 = .17$. This indicates that the combined influence of all three factors determined overall performance. For example, the performance cost of the plausible method was most pronounced for the large Mug, where the combination of a challenging size and a specific affordance (the handle) resulted in the longest completion times, highlighting how the drawbacks of high-fidelity interaction are magnified by complex object properties.

As illustrated in Figure 3, the controller method maintained consistently low TCTs across all shapes and sizes, while plausible grasping showed increased variance and elevated completion times, particularly for complex or larger objects such as the Mug and Stanford Bunny.

Table 1: Holm-corrected p-values for pairwise comparisons across Grasp Methods by Object Shapes and Sizes (Task Completion Time).

	Object Shape					Object Size		
Comparison	Cube	Cylinder	Sphere	Mug	Stanford-Bunny	Small	Medium	Large
Controller vs. Pinch	.005	.007	.009	.003	.002	< .001	< .001	< .001
Controller vs. Plausible	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
Pinch vs. Plausible	.011	.006	.008	< .001	< .001	< .001	< .001	< .001

5.1.3 Post-hoc Analyses. Post-hoc pairwise two-tailed paired t-tests (Holm–Bonferroni-adjusted) were conducted to further investigate the significant interactions, with key results summarized in Table 1.

As shown in Table 1, the controller method was significantly faster than both pinch and plausible methods across every tested object shape and size ($p_{\text{Holm}} \leq .005$). This underscores the robust superiority of the controller method regardless of the object's geometric complexity or scale.

The performance degradation for the plausible method was most severe for complex and large objects. The performance gap between pinch and plausible widened substantially for the Mug and Stanford-Bunny shapes compared to the simple Cube ($p_{\text{Holm}} < .001$). These results highlight the limitations of high-fidelity interaction when dealing with challenging object affordances.

5.2 Results for Placement Accuracy

We conducted a three-way repeated-measures ANOVA to examine the effects of *Grasp Method* (Controller, Pinch, Plausible), *Object Shape* (Cube, Cylinder, Mug, Sphere, Stanford-Bunny), and *Object Size* (Small, Medium, Large) on placement accuracy, defined as the Euclidean distance between the placed object's center and the target location.

5.2.1 Main Effects. The analysis revealed significant main effects for all three independent variables:

The interaction between grasp method and object shape was significant, $F(8, 112) = 5.06, p < .001$. This result indicates that the effect of grasping technique on accuracy was modulated by the shape of the object, with larger differences among grasp methods observed for more complex shapes.

The interaction between grasp method and object size was also significant, $F(4, 56) = 2.61, p = .045$. While the controller method maintained stable performance across all object sizes, accuracy for the pinch and plausible methods declined for small and large objects compared to medium ones.

The interaction between object shape and object size was significant, $F(8, 112) = 35.57, p < .001$. This suggests that the influence of object size on placement accuracy varied across different object shapes, with more pronounced effects observed for irregular geometries.

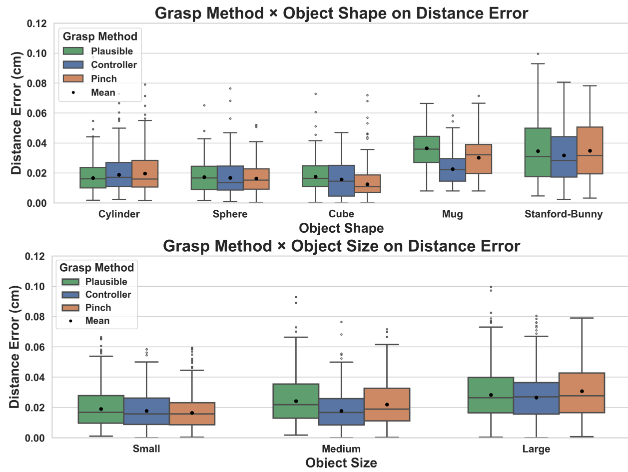

Figure 4: Placement error by grasp method across shape (top) and size (bottom).

Finally, the three-way interaction among grasp method, object shape, and object size was not statistically significant, $F(16, 224) = 1.08, p = .374$. This indicates that their combined effect did not produce additional interactive variance in placement accuracy.

5.2.2 Interaction Effects. In addition to the main effects, several interaction effects were identified.

The interaction between grasp method and object shape was significant, $F(8, 112) = 5.06, p < .001$, indicating that the effect of grasping technique on accuracy was modulated by the shape of the object. Larger differences among grasp methods were observed for more complex shapes.

The interaction between grasp method and object size was also significant, $F(4, 56) = 2.61, p = .045$. While the controller method maintained stable performance across all object sizes, accuracy for the pinch and plausible methods declined for small and large objects compared to medium ones.

The interaction between object shape and object size was significant, $F(8, 112) = 35.57, p < .001$, showing that the influence of size on placement accuracy varied across different object shapes, with more pronounced effects observed for irregular geometries.

Finally, the three-way interaction among grasp method, object shape, and object size was not statistically significant, $F(16, 224) = 1.08, p = .374$, indicating that their combined effect did not produce additional interactive variance in placement accuracy.

Overall, these results indicate that both grasping technique and object characteristics (shape and size) independently and interactively influence placement accuracy. Notably, the *Controller* method demonstrated robust performance across all conditions, while high-fidelity methods like *Plausible* grasping showed greater sensitivity to changes in object geometry and scale.

5.2.3 Post-hoc Analyses. To further examine the significant effects of *Grasp Method* in conjunction with *Object Shape* and *Object Size*, we ran post-hoc pairwise two-tailed paired t-tests with Holm–Bonferroni adjustments.

Table 2: Pairwise p-values for comparisons across Grasp Methods by Object Shape and Size (Placement Accuracy). Values are Holm–Bonferroni corrected from two-tailed paired t-tests.

	Object Shape					Object Size		
Comparison	**Cube**	**Cylinder**	**Mug**	**Sphere**	**Stanford-Bunny**	**Small**	**Medium**	**Large**
Controller vs. Pinch	0.869	0.797	< **.001**	0.357	0.196	0.838	**0.007**	0.142
Plausible vs. Controller	**0.046**	0.211	< **.001**	0.928	0.156	0.071	< **.001**	0.272
Plausible vs. Pinch	**0.014**	0.143	< **.001**	0.558	0.758	0.060	0.054	0.871

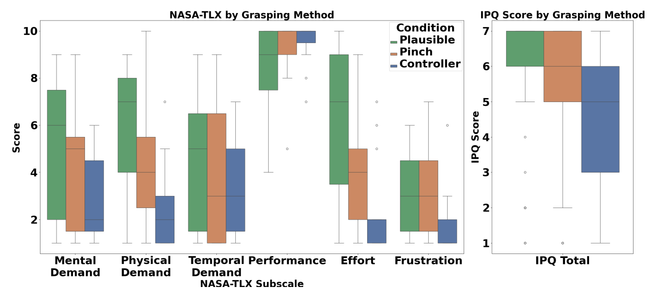

Figure 5: Subjective workload (NASA-TLX) and presence (IPQ) ratings by grasping method. Error bars represent interquartile ranges.

Post-hoc analyses, summarized in Table 2, revealed that the accuracy of grasp methods was highly dependent on the object's shape. The controller method's superiority was only evident for the irregular Mug shape, where it was significantly more accurate than both pinch ($p < .001$) and plausible ($p < .001$). For simpler shapes such as the Cube, Cylinder, and Sphere, there was no significant difference in accuracy between the controller and pinch methods.

The high-fidelity plausible method showed significant limitations only under specific conditions. It was significantly less accurate than the pinch method for the Cube ($p = .014$) and Mug ($p < .001$). However, for other shapes and most size conditions, its accuracy was not significantly different from the pinch method. This suggests that the challenge for high-fidelity interaction is not universal but emerges when handling objects with sharp, regular geometry (Cube) or highly irregular affordances (Mug).

5.3 Subjective Measures

We assessed participants' subjective experience using two established instruments: the NASA Task Load Index (NASA-TLX) for cognitive workload and the Igroup Presence Questionnaire (IPQ) for perceived presence.

NASA-TLX. A repeated-measures ANOVA was conducted on each subscale of the NASA-TLX to evaluate perceived workload across three grasping methods (Controller, Pinch, Plausible). The results revealed significant main effects of Grasp Method on four subscales: Physical Demand, $F(2, 28) = 11.43$, $p = .0002$, $\eta_p^2 = .45$; Effort, $F(2, 28) = 6.81$, $p = .0039$, $\eta_p^2 = .33$; Mental Demand, $F(2, 28) = 5.15$, $p = .0125$, $\eta_p^2 = .27$; and Performance, $F(2, 28) = 3.60$, $p = .0405$, $\eta_p^2 = .21$. In contrast, no significant effects were

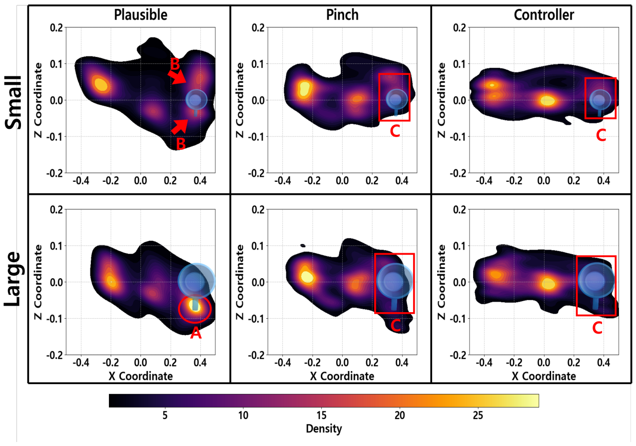

Figure 6: Hand position heatmaps Each heatmap visualizes the spatial density of the average fingertip position for hand-tracking methods and the controller's sensor position.

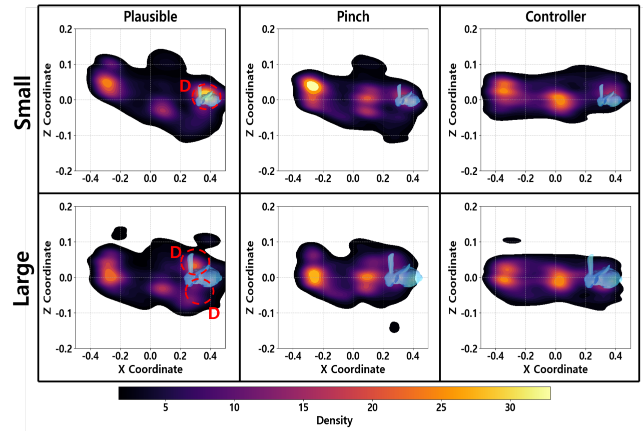

Figure 7: Hand position heatmaps for the Bunny object. The layout and data source definition are consistent with Figure 6. The annotation (D) highlights the key size-dependent grasping strategy for the plausible interaction

found for Temporal Demand, $F(2, 28) = 1.23$, $p = .307$, $\eta_p^2 = .08$, or Frustration, $F(2, 28) = 2.91$, $p = .071$, $\eta_p^2 = .17$.

Post-hoc comparisons (Holm-corrected) showed that the *Plausible* gesture method was rated as significantly more physically demanding than the *Controller* method ($p = .005$, $d_z = 0.93$). All other pairwise comparisons for the remaining subscales did not reach statistical significance after correction. These findings suggest that while the Plausible method may increase perceived physical effort, other aspects of workload did not differ reliably across grasping techniques.

IPQ. To measure the sense of presence, we computed a total IPQ score by summing its 14 items. A repeated-measures ANOVA on this total score showed a significant main effect of *Grasp Method*, $F(2, 28) = 10.57$, $p = .0004$, $\eta_p^2 = .43$. Post-hoc pairwise comparisons (with Holm–Bonferroni correction) revealed that the *Plausible* condition yielded significantly higher presence scores than both the *Pinch* ($p = .0366$, $d_z = 0.60$) and *Controller* ($p = .0077$, $d_z = 0.89$) conditions. Additionally, the *Pinch* method was rated significantly higher than the *Controller* method ($p = .0024$, $d_z = 1.10$), indicating a graded effect of interaction fidelity on the subjective sense of presence. These findings confirm that interaction methods with higher fidelity significantly enhance the user's feeling of being present in the virtual environment. The generally high IPQ scores across all conditions can be explained by our experimental setup: the simple and focused pick-and-place task helped maintain immersion, a consistent virtual hand model was shown in all conditions, and the virtual table and buttons were co-located with a real desk, providing natural passive haptic feedback.

5.4 Plausible Interaction Results

This study aimed to investigate the influence of affordance by visualizing hand dwell points on the XZ plane, derived from 3D hand coordinate data. For analytical consistency, the data was visualized within a unified coordinate range. The DBSCAN algorithm was applied to identify dense clusters, representing areas where the hand remained for extended periods. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was chosen for its key strengths: its proficiency in identifying density-based clusters, its ability to automatically determine the cluster count, and its capacity to detect clusters of arbitrary shapes.

The following analysis is based on the hand position heatmaps presented in Figure 6. To resolve previous ambiguities in object orientation, all visualizations are displayed on the world XZ plane (X-axis right, Z-axis forward), with a semi-transparent 3D model overlaid on each 2D projection. In all conditions, the mug handle and the bunny's head are consistently oriented towards the +X axis.

Dense, concentrated lobes indicate that grasps were drawn to specific "attractors," typically aligned with clear affordances. Conversely, dispersed or bimodal patterns signify user uncertainty or the adoption of alternative, exploratory strategies when a primary affordance is absent or unsuitable.

For the Mugcup, Plausible grasps were intuitively guided by the object's form. In the Large size condition, user interactions were highly concentrated on the handle, the object's primary affordance (Figure 6A). For the Small size, however, a bimodal distribution emerged, indicating that users employed an alternative strategy of grasping both the handle and the cup's body (Figure 6B).

The Bunny's weak affordance led to distinctly size-dependent strategies. For the Small size, users targeted the most prominent feature—the head—as a makeshift handle, resulting in a relatively concentrated pattern; this strategy was not viable for the Large size, leading to scattered and exploratory attempts across the body, which visually confirms user uncertainty when a clear affordance is absent (Figure 7D).

The heatmaps clearly show how the interaction modality can override object affordance. The Controller interaction, in particular, produced a consistent, affordance-agnostic approach pattern that remained nearly identical regardless of the object or its size

(Figure 6C). This visually demonstrates how lower-fidelity, symbolic interactions are governed by their mechanical or gestural constraints, minimizing the influence of object affordance.

6 Discussion

This study examined trade-offs between VR grasping methods based on interaction fidelity and object affordance. Our findings show that method effectiveness varies with object characteristics and user goals. This section interprets these results and provides practical guidelines for VR object grasping.

6.1 The Dominance of Pragmatic Quality

Our first research question (RQ1) concerned the impact of grasping techniques on task performance. The results significantly demonstrate that the **Controller grab was superior in terms of both speed (TCT) and precision (Placement Accuracy)**. This outcome aligns with a significant body of prior work emphasizing the pragmatic benefits of controllers [35, 38]. As illustrated in Figure 3 and Figure 4, the controller's performance remained remarkably stable and efficient regardless of the object's shape or size.

This objective superiority was mirrored in our subjective findings (RQ2). The Controller method consistently scored the lowest on the NASA-TLX for **Physical Demand**, **Mental Demand**, and **Effort** (see Figure 5). This indicates that its abstract, low-fidelity nature minimizes user workload. By bypassing the complexities of hand tracking and physics simulation, a simple controller's button press allows users to achieve their goals with maximum efficiency and minimal cognitive strain. This indicates that in object grasping tasks, low-effort interaction plays a particularly important role for efficient and accurate manipulation.

6.2 Hedonic Quality and the Cost of Fidelity

A core motivation for this study was to understand how object affordance influences the trade-off between performance and naturalness in object grasping tasks. While the Controller excelled in performance, the **Plausible grab provided a significantly higher sense of presence**, as measured by the IPQ (Figure 5). This confirms its value in delivering *hedonic quality* the feeling of being genuinely in the virtual world.

However, this heightened immersion came at a significant cost. The Plausible grab was the slowest and least accurate method and was also rated as the most demanding on all relevant NASA-TLX subscales. Its performance degraded significantly when interacting with objects that had weak or challenging affordances. For instance, TCT and placement error for the Plausible grab increased for complex shapes like the **Mug** and the **Stanford Bunny** compared to simple shapes like the Cube (Figure 3 and 4). Similarly, efficiency decreased for very **small** or **large** objects, which are more difficult to handle with a full-hand grasp.

The performance degradation of the Plausible grab, despite its high immersion, warrants a multi-faceted interpretation. One compelling lens is the **"Uncanny Valley of Interaction"** [32], where the sensory mismatch of seeing the virtual hand make contact without the expected haptic confirmation likely generates a "prediction error" [7]. This conflict can add cognitive load and hesitation, which manifested as increased TCT and reduced placement accuracy.

However, attributing the performance cost solely to this sensory mismatch would be an oversimplification. A more direct factor is the **intrinsic complexity of the control method itself**. Unlike the simple, binary input of a controller button, the Plausible grab requires continuous fine-motor control with many degrees of freedom. This complexity, combined with the lack of physical constraints and potential tracking inaccuracies, directly contributes to the higher workload in the NASA-TLX scores. Therefore, the performance deficit of the Plausible grab likely arises from both a **perceptual-motor conflict** and the **higher intrinsic difficulty** of the control scheme.

6.3 Design Implications and Guidelines

Based on our findings, we offer guidelines to answer: *Which grasping method best fits a given object-grasping context?*

First, for tasks where speed, accuracy, and low cognitive load are paramount (e.g., competitive games, professional training), the **controller-based grasp** should be prioritized. Its robustness across various objects ensures dependable performance in time-sensitive or error-prone environments, making it the safest and most effective choice.

Second, when the goal is to maximize immersion, presence, and naturalness, as in narrative or social VR, the **plausible gesture-based grab** is superior despite its performance trade-offs. It enhances presence but incurs slower completion times, reduced accuracy, and higher physical effort. To mitigate these drawbacks, this method is best suited for medium-sized objects with clear, predictable affordances that support unambiguous grasping actions.

Third, the **pinch grab** offers a practical compromise between fidelity and reliability. More natural than a controller yet less demanding and more stable than a full plausible gesture, it is ideal for tasks requiring a blend of precision and hand interaction, particularly with **small objects** where a precision grip feels natural.

7 Limitations and Future Work

This study has several limitations suggesting directions for future research. First, the sample size of 15 participants may limit statistical power and generalizability. Second, although conducted on Meta Quest 3, the observed trade-offs reflect general interaction categories (abstract, symbolic, and high-fidelity) applicable across platforms, rather than device-specific features. Minor jitter and tracking errors may have influenced the results, warranting replication with more advanced systems. Third, only five basic object shapes were tested, excluding affordances from dynamic, deformable, or tool-like objects. Finally, the task was restricted to pick-and-place actions; future studies should investigate these methods in more complex tasks like assembly or tool use to derive practical design guidelines.

Acknowledgments

This work was supported by the National Information Society Agency (NIA) grant funded by the Korean government (MSIT), and was also supported by Gumi-si and Gyeongsangbuk-do. (Project Title: 2025 Creating an environment for regional, large-scale AI cloud farm verification and expansion)

References

[1] Christopher C Berger, Mar Gonzalez-Franco, Eyal Ofek, and Ken Hinckley. 2018. The uncanny valley of haptics. *Science Robotics* 3, 17 (2018), eaar7010.
[2] Irving Biederman. 1987. Recognition-by-components: a theory of human image understanding. *Psychological review* 94, 2 (1987), 115.
[3] 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.
[4] Ryan Canales, Aline Normoyle, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2019. Virtual grasping feedback and virtual hand ownership. In *ACM Symposium on Applied Perception 2019*. 1–9.
[5] Paul Cisek. 2007. Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362, 1485 (2007), 1585–1599.
[6] Paul Cisek and John F Kalaska. 2010. Neural mechanisms for interacting with a world full of action choices. *Annual review of neuroscience* 33, 1 (2010), 269–298.
[7] Andy Clark. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences* 36, 3 (2013), 181–204.
[8] Raymond H Cuijpers, Jeroen BJ Smeets, and Eli Brenner. 2004. On the relation between object shape and grasping kinematics. *Journal of neurophysiology* 91, 6 (2004), 2598–2606.
[9] Mark R Cutkosky et al. 1989. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on robotics and automation* 5, 3 (1989), 269–279.
[10] Hao Dang and Peter K Allen. 2012. Semantic grasping: Planning robotic grasps functionally suitable for an object manipulation task. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1311–1317.
[11] Charles de Granville, Di Wang, Joshua Southerland, Andrew H Fagg, and Robet Platt. 2009. Grasping affordances: Learning to connet vission to hand action. In *The Path to Autonomous Robots: Essays in Honor of George A. Bekey*. Springer, 1–22.
[12] Loni Desanghere and Jonathan J Marotta. 2015. The influence of object shape and center of mass on grasp and gaze. *Frontiers in psychology* 6 (2015), 1537.
[13] M D'Alonzo, A Mioli, D Formica, L Vollero, and G Di Pino. 2019. Different level of virtualization of sight and touch produces the uncanny valley of avatar's hand embodiment. *Scientific reports* 9, 1 (2019), 19030.
[14] Thomas Feix, Javier Romero, Heinz-Bodo Schmiedmayer, Aaron M Dollar, and Danica Kragic. 2015. The grasp taxonomy of human grasp types. *IEEE Transactions on human-machine systems* 46, 1 (2015), 66–77.
[15] Karl Friston. 2005. A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1456 (2005), 815–836.
[16] James J Gibson. 1978. The ecological approach to the visual perception of pictures. *Leonardo* 11, 3 (1978), 227–235.
[17] Michael A Gomez, Rafal M Skiba, and Jacqueline C Snow. 2018. Graspable objects grab attention more than images do. *Psychological Science* 29, 2 (2018), 206–218.
[18] Scott T Grafton, Andrew H Fagg, Roger P Woods, and Michael A Arbib. 1996. Functional anatomy of pointing and grasping in humans. *Cerebral Cortex* 6, 2 (1996).
[19] James G Greeno. 1994. Gibson's affordances. (1994).
[20] Antonia F de C Hamilton and Scott T Grafton. 2006. Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience* 26, 4 (2006), 1133–1137.
[21] Rex Hartson. 2003. Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & information technology* 22, 5 (2003), 315–338.
[22] Loïc P Heurley, Mohamed Halim Harrak, Ronan Guerineau, Laurent P Ferrier, and Nicolas Morgado. 2023. The visual size is enough to automatically induce the potentiation of grasping behaviours. *Quarterly Journal of Experimental Psychology* 76, 12 (2023), 2749–2759.
[23] Yuichi Hiramatsu, Daisuke Kimura, Koji Kadota, Taro Ito, and Hiroshi Kinoshita. 2015. Control of precision grip force in lifting and holding of low-mass objects. *PloS one* 10, 9 (2015), e0138506.
[24] Chiwoong Hwang, Tiare Feuchtner, and Kaj Grønbæk. 2024. Pseudo-Haptics for Weight Perception in VR: Controller vs. Bare Hand Interactions with Tracking Delay and Vertical Offset. In *2024 10th International Conference on Virtual Reality (ICVR)*. IEEE, 91–99.
[25] Chaowanan Khundam, Naparat Sukkriang, and Frédéric Noël. 2021. No difference in learning outcomes and usability between using controllers and hand tracking during a virtual reality endotracheal intubation training for medical students in Thailand. *Journal of educational evaluation for health professions* 18 (2021).
[26] Chaowanan Khundam, Varunyu Vorachart, Patibut Preeyawongsakul, Witthaya Hosap, and Frédéric Noël. 2021. A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training. In *Informatics*, Vol. 8. MDPI, 60.
[27] ByungMin Kim, DongHeun Han, and HyeongYeop Kang. 2025. Shaping the Future of VR Hand Interactions: Lessons Learned from Modern Methods. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 165–174.
[28] Jeonghyeon Kim, Jemin Lee, Jung-Hoon Ahn, and Youngwon Kim. 2025. ITap: Index Finger Tap Interaction by Gaze and Tabletop Integration. *Sensors* 25, 9 (2025), 2833.
[29] Tanja Kojić, Maurizio Vergari, Simon Knuth, Maximilian Warsinke, Sebastian Möller, and Jan-Niklas Voigt-Antons. 2024. Influence of Gameplay Duration, Hand Tracking, and Controller Based Control Methods on UX in VR. In *Proceedings of the 16th International Workshop on Immersive Mixed and Virtual Environment Systems*. 22–28.
[30] Tiffany Luong, Yi Fei Cheng, Max Möbus, Andreas Fender, and Christian Holz. 2023. Controllers or bare hands? a controlled evaluation of input techniques on interaction performance and exertion in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4633–4643.
[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] Ryan P McMahan, Chengyuan Lai, and Swaroop K Pal. 2016. Interaction fidelity: the uncanny valley of virtual reality interactions. In *Virtual, Augmented and Mixed Reality: 8th International Conference, VAMR 2016, Held as Part of HCI International 2016, Toronto, Canada, July 17-22, 2016. Proceedings 8*. Springer, 59–70.
[33] John R Napier. 1956. The prehensile movements of the human hand. *The Journal of Bone & Joint Surgery British Volume* 38, 4 (1956), 902–913.
[34] A Norman Donald. 2013. *The design of everyday things.* MIT Press.
[35] Ardhika Wida Pangestu, Clara Hetty Primasari, Thomas Adi Purnomo Sidhi, Yohanes Priadi Wibisono, and Djoko Budiyanto Setyohadi. 2022. Comparison Analysis of Usability Using Controllers and Hand Tracking in virtual reality Gamelan (Sharon) based on user experience. *Journal of Intelligent Software Systems* 1, 2 (2022), 89–103.
[36] 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*. 99–108.
[37] Mores Prachyabrued and Christoph W Borst. 2013. Effects and optimization of visual-proprioceptive discrepancy reduction for virtual grasping. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 11–14.
[38] Alexander Schäfer, Gerd Reis, and Didier Stricker. 2022. Comparing controller with the hand gestures pinch and grab for picking up and placing virtual objects. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 738–739.
[39] E Schminder, M Ziegler, E Danay, L Beyer, and M Bühner. 2010. Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution. *European Research Journal of Methods for the Behavioral and Social Sciences* 6, 4 (2010), 147–151.
[40] Francesco Vona, Julia Schorlemmer, Jessica Stemann, Sebastian Fischer, and Jan-Niklas Voigt-Antons. 2025. Hands vs. Controllers: Comparing User Interactions in Virtual Reality Shopping Environments. *arXiv preprint arXiv:2502.05100* (2025).
[41] Daniel M Wolpert, Jörn Diedrichsen, and J Randall Flanagan. 2011. Principles of sensorimotor learning. *Nature reviews neuroscience* 12, 12 (2011), 739–751.
[42] Jian Zhang, Wafa Johal, and Jarrod Knibbe. 2025. Illusion Spaces in VR: The Interplay Between Size and Taper Angle Perception in Grasping. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.