GraspUI: Seamlessly Integrating Object-Centric Gestures within the Seven Phases of Grasping

Adwait Sharma*
as5339@bath.ac.uk
Reality Labs Research, Meta
Toronto, ON, Canada

Alexander Ivanov
alexivan@meta.com
Reality Labs Research, Meta
Toronto, ON, Canada

Frances Lai
flai@meta.com
Reality Labs Research, Meta
Toronto, ON, Canada

Tovi Grossman
tovi@dgp.toronto.edu
University of Toronto
Toronto, ON, Canada

Stephanie Santosa
ssantosa@meta.com
Reality Labs Research, Meta
Toronto, ON, Canada

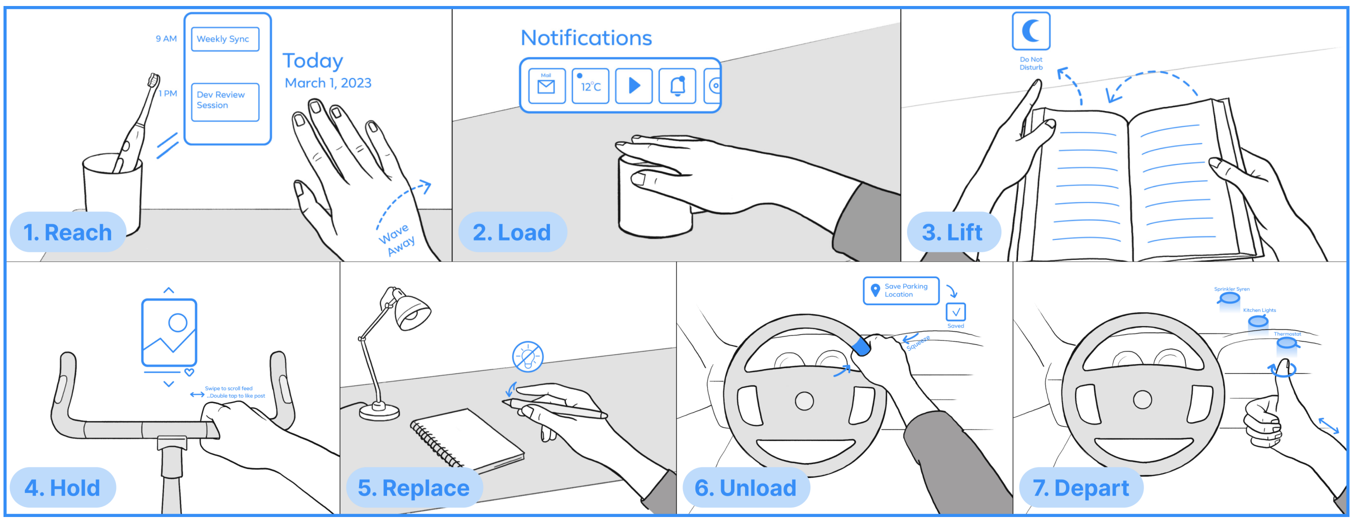

Figure 1: GraspUI is a design space of object-centric gestures spanning the seven distinct phases of grasping that can be seamlessly integrated into existing object interactions. This example demonstrates how a user, Taylor, can employ gestures from the design space throughout her day. (1) In the morning, Taylor checks her schedule for the day ahead as she *reaches* for her toothbrush. (2) Before work, Taylor *loads* her hand atop of her coffee cup, revealing the weather forecast and her unread email. (3) While reading a book during her morning commute, she *lifts* her book to silence her phone to enable distraction-free reading. (4) During her cycling workout, she *holds* her bike's handle and swipes on the handlebar to browse social media and share her progress with her friends. (5) After jotting down meeting notes in her notebook, Taylor taps on her pen, using the *replace* motion to turn off her desk lamp. (6) To ensure she does not forget where she parked, Taylor saves her parking location by squeezing on the steering wheel as she *unloads* her hands from it. (7) After arriving home, Taylor turns on her smart home appliances using mid-air thumb gestures as her hands *depart* the steering wheel.

ABSTRACT

Objects are indispensable tools in our daily lives. Recent research has demonstrated their potential to act as conduits for digital interactions with microgestures, however, the primary focus was on situations where the hand *firmly* grasps an object. We introduce GraspUI, an exploratory design space of object-centric gestures within the seven distinct phases of the grasping process, spanning pre-, during, and post-grasp movements. We conducted ideation sessions with mixed-reality designers from industry and academia to explore gesture integration throughout the *entire* grasping process. The outcome was 38 storyboards envisioning practical applications. To evaluate the design space's utility, we performed a video-based

*Also with University of Bath, United Kingdom.

DIS '24, July 01–05, 2024, IT University of Copenhagen, Denmark

ACM ISBN 979-8-4007-0583-0/24/07
https://doi.org/10.1145/3643834.3661551

assessment with end-users. We then implemented an interactive prototype and quantified the overhead cost of performing proposed gestures through a secondary study. Participants reacted positively to gestures and could integrate them into existing usage of objects. To conclude, we highlight technical and usability guidelines for implementing and extending GraspUI systems.

CCS CONCEPTS

• **Human-centered computing → Gestural input**.

KEYWORDS

design space, input, grasp, grasping process, hand-object manipulation, everyday objects

ACM Reference Format:
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 *Designing Interactive Systems Conference (DIS '24), July 01–05, 2024, IT University of Copenhagen, Denmark.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3643834.3661551

1 INTRODUCTION

People handle an average of 140 objects per day [1]. Prior work has demonstrated that the everyday objects we hold in our hands afford several opportunities for microtasks [4, 33, 42, 63, 64, 76]. For example, a person can activate a dial-like control while holding a cup of coffee by using a thumb rotation gesture to adjust the volume of a music player [59]. Additionally, held objects can serve as a medium to interact with smart home devices [20], head-mounted displays [80], and control systems in healthcare contexts [26].

Using handheld physical objects to perform microtasks via quick and subtle gestures is a promising way to achieve always-available input, with current research focusing on situations where an object is already being *firmly* grasped. However, grasping an object is a multi-step process that involves distinct phases, such as reaching for the object, making contact with it, lifting it, stabilizing it, returning it to its initial position, and removing contact with it [40]. We extended this work by adding a "Depart" phase to signify the hand's moving back to its initial position. The full range of these natural grasping behaviors suggests an unexplored opportunity to utilize hand motion and time for in-situ, context-driven input.

Prior work has explored some of the phases that occur when people use phones [38]. In contrast, we consider the entire process of ballistic hand movements during hand-object manipulation to identify the space of object-centric interactions. By leveraging existing hand movements, the creation of a design space presents a multitude of opportunities for context-sensitive, casual, and fine-grained input. For example, passing one's hand through a calendar app icon while reaching for a toothbrush could trigger a calendar visualization of the upcoming day (Figure 1.1), covering the top of a cup with one's hand could invoke a summary of weather forecast data and unread email (Figure 1.2), or performing mid-air finger gestures while the hand departs from a steering wheel could turn on smart devices to ensure one's home is comfortable upon return (Figure 1.7). One commonality across these gestures is that they seamlessly integrate into existing physical interactions and do not require much additional time to perform.

In this paper, we introduce *GraspUI*, an exploratory design space of object-centric interactions that provides effortless, in-situ input experiences across the seven phases of object grasping (see Figure 2). Employing a Research through Design approach [81], we investigated design opportunities within the GraspUI design space to uncover this new category of object-centric gestures. Specifically, mixed-reality interaction designers participated in iterative design sessions where they were asked to envision real-world applications for microtasks within the phases of grasping. These sessions resulted in 38 storyboards, across 6 different physical objects, that were further refined and are presented herein as a collection of example interactions for each grasping phase.

Our methodology for evaluating the overall performance and technical feasibility of the GraspUI design space employs a threefold process. It focuses on demonstrating user interactions through exploration of the design space, utilizing both novel and replicated examples [47]. Initially, we evaluated the design space in terms of comfort and usefulness using the aforementioned 38 storyboards with 12 end users. Following this, using commodity hardware, we developed an interactive prototype system to demonstrate the technical feasibility of GraspUI with one of the objects (a mug). The system detected the seven phases of grasping while performing microtasks with the physical mug and rendered associated UI widgets using an AR headset. Subsequently, we conducted a second study involving 14 participants to quantify the overhead in terms of time and participants' experience ratings - it took an average of 8.51 seconds to interact with the mug, with a minimal overhead of 0.76 seconds to perform microtasks during the interactions. During the sessions, we also uncovered design strategies that could further accelerate gesture performance. Our evaluations demonstrated that users appreciated the opportunity to engage with a digital interface during these phases. Finally, we present the design implications of integrating microtasks within different gesture phases to support future developers and designers of object-centric gestures.

The main contributions of this research include:

- An exploratory **GraspUI design space** that can enhance everyday objects with additional input opportunities throughout all seven grasping phases, spanning pre-, during, and post-grasp movements.
- The findings from an end user study that demonstrates the **comfort** and **usefulness** of integrating GraspUI gestures across multiple phases and 6 exemplar objects.
- A **working prototype** developed with commodity hardware to enable input capabilities during all phases of grasping with one of the objects.
- The **findings from an empirical study** reveal an average overhead of 0.76 seconds while performing microtasks across all grasping phases, as opposed to when microtasks were not performed.

Noteworthy, our goal in this paper is not to incorporate gestures into every grasping phase and all objects, but instead to introduce a new input dimension that enhances the existing object-centric gestures. The decision about whether to integrate GraspUI within multiple phases will be left to future research endeavors. By exploring the design space and demonstrating a proof-of-concept system, we have laid the foundation for further investigation of future sensing techniques and interaction designs.

2 RELATED WORK

With a research goal of seamlessly integrating hand gesture-based interactions with everyday grasping movements, we review prior work on object-centric input techniques, the process of grasping, dual-purpose interactions, and sensing.

2.1 Object-Centric Input Techniques

Much of the existing literature on designing gestures for objects focuses on the use of gestures once an object is already firmly held. For example, Fitzmaurice et al.'s Graspable User Interfaces paradigm controlled virtual objects using physical artifacts [31]. Tangible Bits, on the other hand, proposed interaction via physical objects and ambient media [39]. Opportunistic Controls leveraged the use of existing objects within an environment for passive haptic feedback and gesture input within Augmented Reality (AR) applications [36]. Clarke et al. utilized a motion-matching and pointing technique as input while holding an object [24, 25]. Graspables explored the holding and manipulation of objects as a user interface [67]. Across these approaches, a variety of possibilities for how firmly held objects could be used for digital input are presented.

Meanwhile, studies by Wolf et al. [76], Corsten [27], Sharma et al. [63, 64], Bardot et al. [10], and Gong et al. [33] explored the usability of grasping microgestures. While the majority of research in this area include gesture designs that involve *holding* an object in the hand, some research has focused on the use of *self-sustained* objects such as bike handles or steering wheels [5, 66]. Additionally, Pohl et al. proposed using the space and objects around a device for casual interaction [59], while Kim et al. conducted a workshop study that explored the use of a everyday objects to control IoT devices – such as tilting a chair to turn on a lamp [44]. GripMarks used handheld objects as mixed reality input surfaces [80]. Recently, Transferable Microgestures were introduced, which utilize the Middle, Ring, and Pinky fingers to enable consistent interactions across various grasps and postures [42]. In social situations, objects have also been found to be useful for interacting with devices [4, 15, 53]. Furthermore, some consumer products have also started to support gestural input. For instance, the Konnect-i Samsonite Jacquard backpack features a touch-sensitive strap [2]).

Our gesture designs build upon this previous research by taking a holistic approach to enable seamless input within the entire seven phases of grasping. We consider the use of held and self-sustained objects to provide users with a wide range of interaction options in our approach, thus expanding the design space of object-centric gestures beyond firmly held object interactions.

2.2 The Process of Grasping

Grasping has been extensively studied in several domains including biomechanics [58], neuroscience [18], psychology [62], and robotics [14]. Previous research on the design of hand gestures when using a firm grasp has relied upon grasp taxonomies that systematically categorize hand postures to enable the grasping of objects with different geometries [52]. In the context of full hand movement, Carfì et al. identified three states of hand-object interaction: Off-hand, In-contact, and Held in-hand [17]. Meanwhile, Rand et al. found a coordinated relationship between hand transport and finger aperture during reach-to-grasp movements [60]. Wimmer proposed a GRASP model, with five factors representing different meanings for the way we grasp an object [74]. In our exploratory GraspUI design space, we build upon these studies and grasp taxonomies to design gestures that are compatible with object geometry and eliminate object-dropping gestures.

Although some research has investigated the potential advantages of sensing pre-contact hand movements, particularly when using mobile devices or tabletop displays, research on this topic is still relatively scarce. For example, explorations into pre-touch sensed finger and hand gestures, before physical contact was made with a mobile phone, have been explored [38]. Zhang et al. explored the use of screen distance for enhanced stylus interaction with tablets [79], and Medusa offered visual guidance whenever a hand hovered over a multi-touch tabletop [6]. While these approaches illustrate the benefits of the pre-contact phase, we contribute to this literature by demonstrating how other under-utilized phases within the grasping process are also feasible for performing microtasks. Our work makes use of visual and tactile sensory signals as control points during object manipulation tasks, which are identified by mechanical contact events representing sub-goals, as described by Johansson and Flanagan [40].

2.3 Dual-Purpose Interactions and Microtasks

The human brain controls the grasping process through separate visuomotor pathways (reach and grasp), as described by the Dual Visuomotor Channel Theory [7]. Leveraging these two pathways, we explored encouraging input into the entire grasping process to serve a dual purpose with minimal additional overhead. Previous research has used a similar idea to achieve dual-purpose interaction with a device, such as using speech for communication and input [51] or completing microtasks while unlocking a phone [68]. Furthermore, micro-tasking has been shown to lead to higher quality outcomes and increased productivity [8, 9, 21], and enabled users to accomplish crowdsourcing tasks without burdening them [69]. Our empirical evaluation echos prior work, showing a minor increase in the time needed to perform microtask-based interactions while grasping. This reveals users can complete additional tasks throughout all phases of grasping.

2.4 Capturing Hand-Object Interaction

GraspUI interactions can be implemented using various sensing technologies and are not limited to the technique in our interactive prototype. However, it is worth noting that simultaneously tracking un-instrumented objects and hands presents an ongoing research challenge due to occlusions caused by holding and manipulating an object [57]. A large body of prior work involves using camera-based approaches [82] or attaching sensors to objects, such as capacitive sensors [34]. Researchers have also explored various methods for free-hand tracking [48, 65]. While much prior work has focused on adding sensing capabilities to objects or tracking hands individually, it's important to note that current solutions still struggle to achieve robust tracking for both un-instrumented objects and hands. Some recent devices can improve freehand tracking precision [56], although robustly tracking objects and hands often requires the use of markers [70, 80]. Drawing inspiration from earlier works that highlighted the advantages of body-worn projectors

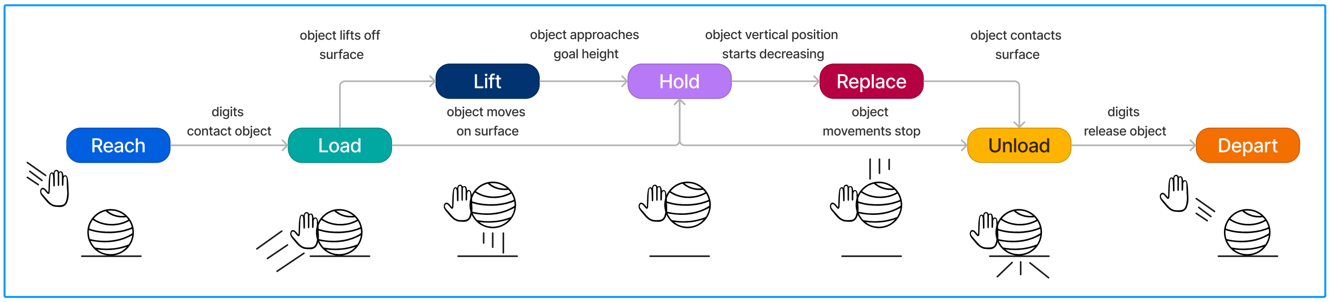

Figure 2: GraspUI enables users to perform microtasks during each phase of grasping an object (i.e., Reach, Load, Lift, Hold, Replace, Unload, and Depart). As some objects are too large or heavy to be picked up, the Lift and Replace phases would be skipped while interacting with such objects.

[75] and unlocked numerous novel applications, we incorporated visual feedback into the GraspUI's interactive prototype. This provides users an opportunity to interact with interfaces within all phases of grasping.

To overcome the challenges of tracking and providing visual feedback, we adopted a hybrid tracking approach in our interactive GraspUI prototype. This approach pairs stereo webcams with a Microsoft HoloLens 2 AR headset, allowing us to simultaneously track hands and objects, while providing visual feedback without the need to instrument the hand or the object.

3 GraspUI: DESIGN SPACE

While gesture design encompasses a range of methodologies, from expert-led to user-led [77], our work explores the feasibility of incorporating gestures into the seven different grasping phases and then evaluates user receptiveness to them. To systematically understand the practicality and challenges associated with integrating interaction within all seven grasping phases, we created a design space of object-centric interactions.

3.1 The Seven Phases of Grasping

Although existing object-centric gestures tended to focus on the user *holding* an object, there is an opportunity to explore the full design space across different phases of grasping. By breaking down hand movements into individual phases of grasping, we can help designers and developers design more effective and intuitive interactions with physical objects. In Johansson and Flanagan's hand-object manipulation phases [40], six distinct phases of grasping are presented to describe the way we interact with objects. Building upon this work, GraspUI proposes gestures for each phase, and adds a "Depart" phase to signify an additional opportunity for interactions - when the hand returns to its initial position. The seven phases thus include (Figure 2):

(1) **Reach**: The initial phase of the grasp where the hand begins moving towards the object before any contact occurs with the object.
(2) **Load**: Occurs when the digits of the hand contact the object and ends when the object begins to lift off the surface it is resting on.
(3) **Lift**: Occurs as the object is being raised and ends when the object reaches an intended goal height.
(4) **Hold**: Occurs while the object is firmly grasped.
(5) **Replace**: Occurs when the object is being lowered and ends when the object is almost touching the surface it will rest on. This is the opposite of the Lift phase.
(6) **Unload**: Occurs when the object is supported by the surface and the task involving the object concludes. The digits of the hand begin to break contact with the object. This is the opposite of the Load phase.
(7) **Depart**: The termination phase of the grasp when the hand moves away from the object and ceases to have any contact with the object. This is the opposite of the Reach phase.

In the remaining sections of this paper, we demonstrate the benefits of incorporating gestural interactions into various phases of grasping, going beyond the traditionally focused "hold" phase. By leveraging these different phases, designers and developers can create seamless object-centric gestures that enable users to perform input during primary activities with minimal overhead (see Figure 1 for an illustration of GraspUI in action).

3.2 Ideation Sessions

To gain insight into daily interactions that might be possible within the seven phases of grasping, we conducted eight ideation sessions over a two month period with two groups of expert mixed reality researchers and practitioners. All participants possessed several years of academic and/or industry experience in the area of mixed reality user interfaces.

We collaborated with the first group of four experts during seven 1-hour semi-structured design sessions. To guide their design process, we initially reviewed existing literature on grasp taxonomy and hand-object interaction [17, 52, 64, 74, 76]. We then discussed common activities that users could perform while interacting with digital devices and handling everyday objects, both indoors and outdoors. To better understand the contextual factors that might impact the design process, we considered various scenarios, such as in cafes or while on the move. We referred to the top 20 most frequently performed hand activities based on previous research [45] and selected activities that involved the use of a single object.

This resulted in six objects: a cup, pen, book, bicycle handle, toothbrush, and steering wheel. After identifying common applications and objects, these activities were used as exemplars to generate ideas for interaction. Our aim while designing these interactions was to seamlessly blend a user's physical actions with an object and provide opportunities for digital input with minimum effort. As a result of this process, we generated 49 potential use cases by the end of our sessions.

We then conducted a single 2-hour session with the second group of four experts to narrow down the ideas to those that would be feasible to implement using existing technologies.

During our workshops, we followed a research-through-design approach [81] and adapted the scenarios to include a wide range of control commands. Our experts intentionally incorporated slightly less ergonomic gestures – such as forming a circle with the thumb for critical commands like muting and unmuting the microphone – to minimize the likelihood of unintended activations [72]. Conversely, they also suggested using gestures with more direct mappings. For example, placing an object on top of app icons displayed on a table to switch to the respective app. The sketches and notes from the sessions were then probed by two co-authors, resulting in a set of 38 storyboards.

3.3 Sample GraspUI Applications

The 38 storyboards from our design sessions were categorized based on the grasp phase during which they occurred while using the six objects. In the following section we discuss these storyboards, highlighting how GraspUI could provide unique benefits in everyday life (see supplementary material for high-resolution version).

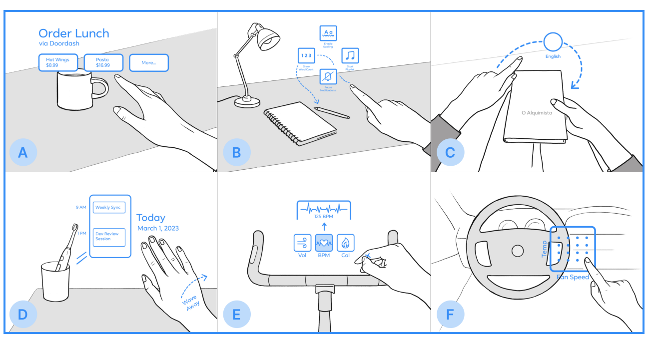

Figure 3: Reach: (A) selecting a lunch option, (B) muting a phone and starting a word counter, (C) translating text, (D) hiding a calendar window, (E) monitoring health, and (F) controlling the temperature and fan speed. Interactions include passing the hand through a command or mid-air hand gestures.

Reach: As the hand nears an object, the user's intent is mirrored in its posture [52], with the trajectory revealing potential targets [32]. In our storyboards, the proposed interactions during the Reach phase often involved adjusting the hand trajectory to pass through a widget to trigger a desired behavior while reaching for an object to grasp (e.g., Figure 3A, B, C, E). The car storyboard (Figure 3F) also showcases how cross-through interactions can extend to multi-dimensional controllers. As the viewer passes their hand through a position on a 2D grid interface, they simultaneously select both the AC fan speed and temperature before grabbing the steering wheel of the car. The Reach phase also provides an opportunity for mid-air hand gestures to occur, such as waving a notification away (Figure 3D), giving a thumbs up, or performing a finger tap.

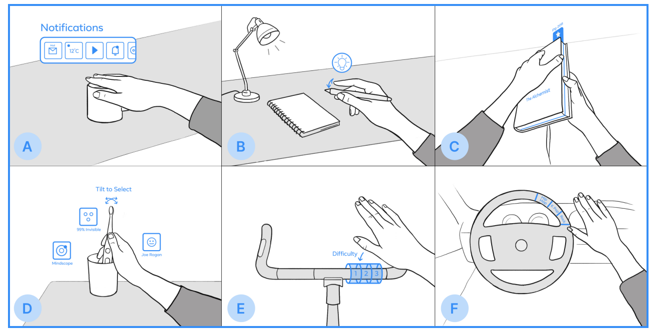

Figure 4: Load: (A) glancing notifications, (B) turning on a lamp, (C) displaying bookmarks, (D) selecting podcasts, (E) setting workout difficulty, and (F) playing a specific music genre. Interactions include unique grasps, finger gestures, or object re-orientation.

Load: With 27 bones, the human hand provides over 25 degrees of freedom (DOFs), facilitating dexterous finger movements and creating diverse hand poses [41]. As depicted in our storyboards (Figure 4), this phase provides interaction opportunities just before an intended activity with an object occurs. In this moment, the object remains supported by the surface, overcoming the constraints associated with holding an object in hand. Making contact with an object at a particular location, via a unique grasp, can act as an input mechanism for subsequent actions to occur (Figure 4A, C, D, F). Load interactions can also support the performance of gestures with the object using individual fingers (Figure 4B) or adjusting the orientation of an object for selection tasks (Figure 4D).

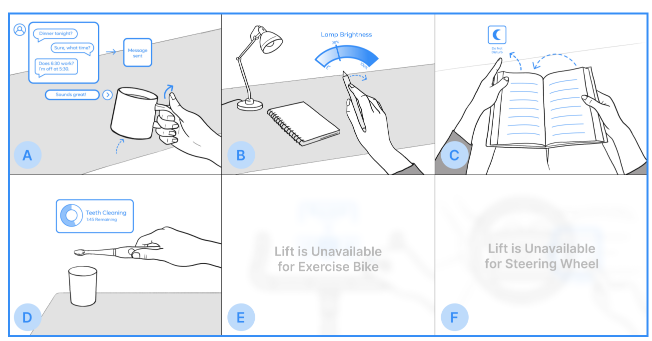

Figure 5: Lift: (A) responding to a message, (B) adjusting a lamp's brightness, (C) enabling a Do Not Disturb mode, and (D) activating a teeth-cleaning timer. Interactions include finger gestures and object re-orientation.

Lift: Lifting occurs in the time between picking up an object and the object reaching its needed position (Figure 5). In this phase, the constraints of finger movements caused by grasping the object are introduced. Our storyboards depict opportunities for various single-finger gestures (Figure 5A, C, D), and the ability for the orientation of a pen to adjust a continuous input slider (Figure 5B) [44]. Lifting the object enables users to employ motion-coupling techniques for fine-grained continuous input [25]. Notably, not all objects have the ability to be lifted in this way as they could be too heavy or large to raise.

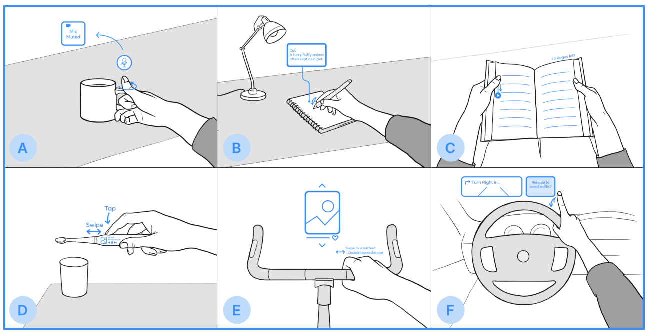

Figure 6: Hold: (A) muting a mic, (B) opening a dictionary, (C) activating audio narration, (D) controlling audio playback, (E) scrolling a social media feed, and (F) accepting a re-routing suggestion. Interactions include finger gestures.

Hold: Interactions when an object is already being held have been well explored in the literature [10, 33, 42, 63, 76], often through the use of subtle finger gestures. For completeness, we still include the storyboards that were generated for this phase. As with prior work, our storyboards also include finger tapping and swipe gestures for controlling virtual content (Figure 6B, C, D, E, F). Of particular interest, hold phase interactions have the potential to change state in relation to how an object is currently being used. For example, whenever a user is about to take a sip of coffee, an option to mute a microphone during a video call could be presented then invoked using a rapid thumb gesture (Figure 6A).

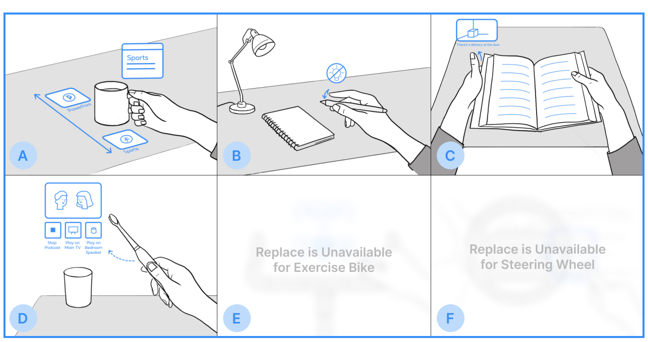

Figure 7: Replace: (A) selecting an app icon, (B) turning off a desk lamp, (C) dismissing a notification, and (D) selecting an output device. Interactions include mapping placement locations, finger-based gestures, or passing an object through a command.

Replace: While opposite to the Lift phase, one possible interaction during the Replace phase is to map different placement locations to specific digital actions, which can then be used to activate various applications through the use of visual cues (Figure 7A). Additionally, finger-based gestures could be employed to execute discrete input commands (Figure 7B, C). Similar to the Reach phase, where a hand can pass through a widget, the object can pass through a widget to select an appropriate command, such as choosing which smart speaker to use to play a song when returning a toothbrush to its holder (Figure 7D).

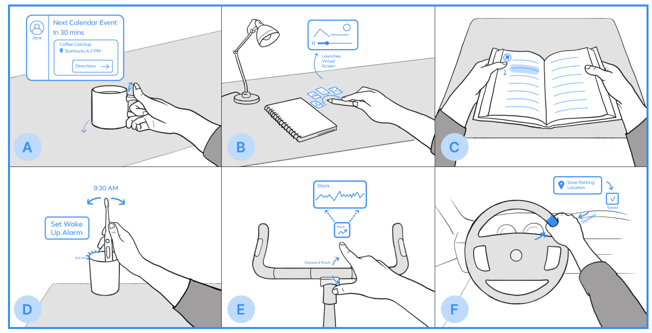

Figure 8: Unload: (A) dismissing a calendar notification, (B) selecting a TV show, (C) Bookmarking a page, (D) setting a timer, (E) expanding a stock app, and (F) saving a parked car location. Interactions include finger gestures, object re-orientation, or pressure.

Unload: Given the stationary state of the object and its support on a surface, various modalities can be utilized during the unload phase, such as finger gestures (Figure 8A, E), object re-orientation (Figure 8B, D), and pressure (Figure 8C, F). It is worth noting that previous research has also referred to applying pressure as a form of input during the Hold phase [64]. As can be seen in our storyboards, these actions can be used to seamlessly transition to the next task. For instance, a double tap on an object, like a cup, can trigger the display of the next event in a user's calendar.

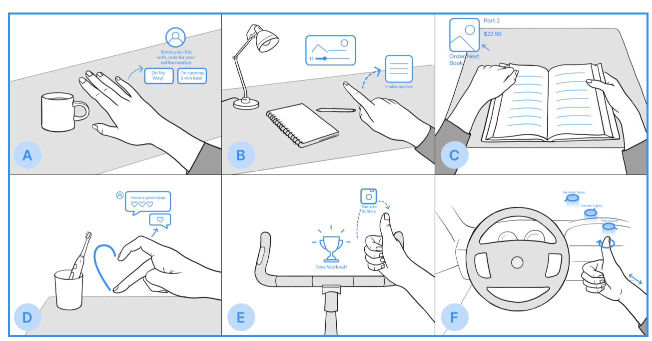

Figure 9: Depart: (A) sending an ETA update, (B) enabling captions, (C) ordering the next book, (D) replying with a heart emoji, (E) sharing workout progress, and (F) setting a home's temperature. Interactions include passing the hand through a command or mid-air hand gestures.

Depart: Similar to the Reach phase, the Depart phase is useful for displaying menus with discrete options (Figure 9). The hand may pass through a button from its back side (Figure 9A, C), or the viewer may use different hand poses to switch modes (Figure 9B, D, F). Our storyboards depict several applications when Depart may take place, such as prompting the user to share a message with friends after completing a cycling workout (Figure 9E).

3.4 Summary and Takeaways

This design space is intended to be exploratory in nature, and is by no means exhaustive or definitive. On the contrary, our hope is that the storyboards resulting from our research through design process will serve as *"design exemplars, providing an appropriate conduit for research findings to easily transfer to the HCI research and practice communities"* [81].

From the presented storyboards, we observed a set of recurring themes worth noting. While some interactions were *phase-independent*, being common across several phases, others were *phase-specific* and only applicable in certain phases. For example, placing an object at a specific location on a surface is only applicable in the Replace phase, while finger gestures on an object can be performed in the Load, Lift, Hold, Replace, and Unload phases. Additionally, some interactions were *semantically related* to the physical objects, such as bookmarking a page by interacting with the page, while others were *temporally related*, such as muting a microphone before sipping coffee. Finally, most interactions were *object-independent*, being applicable on any type of object (e.g. passing through a widget during Reach), while others were *object-specific*, such as covering a coffee cup during Load.

4 EVALUATING THE DESIGN SPACE

To investigate *gesture integration* throughout different grasping phases for various objects, we performed an initial evaluation of the GraspUI design space from the end-users perspective. We employed a video-based evaluation method for demonstration [47], following an approach similar to prior research [4, 28, 37, 42, 61].

4.1 Participants

Twelve participants were recruited from outside our institution (i.e., 7 males, 5 females; μ = 30 years; $\tilde{x}$ = 28 years, range = 21-40 years; all right handed). Participants had diverse backgrounds, including software and mechanical engineering, homemaking, advertising, customer service, etc., thus bringing a wide range of perspectives to the evaluation. Participants were compensated with a $75 USD gift card, and the survey took approximately 1 hour to complete.

4.2 Study Setup and Procedure

Participants were initially informed about the seven phases of grasping using a textual description and a brief introductory video. They were then prompted to assess each phase and object individually via a web-based survey interface that we developed. The survey interface illustrated the primary tasks (e.g., sipping coffee from a cup) and microtasks (e.g., dismissing a calendar notification) described in our storyboards. During each survey response round, descriptions of these tasks were described using text alongside accompanying storyboard illustrations of possible interactions and looped video clips of possible gestures. The storyboard images showed the virtual overlays, while the videos showcased the hand movements only, with no virtual overlays.

For each round of the storyboards, we asked participants to rate how well the presented gesture could be seamlessly integrated into

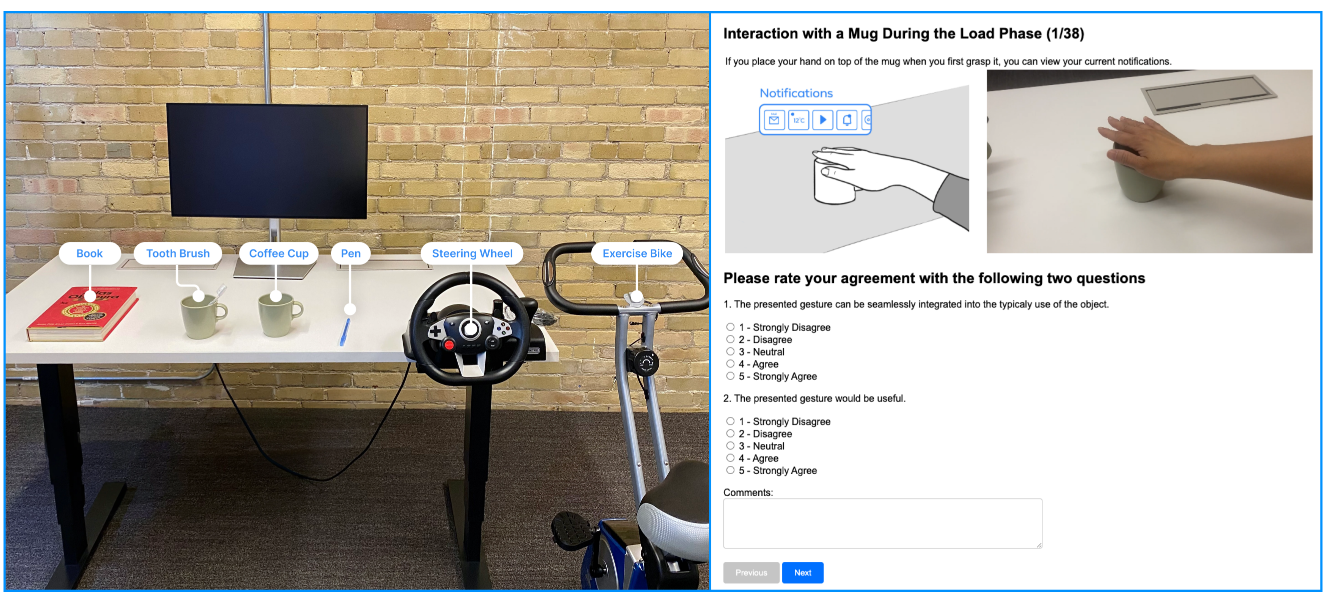

Figure 10: Our study setup invites participants to interact with a book, toothbrush, coffee cup, pen, steering wheel, and exercise bike (left). On the computer monitor, a web application displays a survey prompting participants to rate presented GraspUI storyboards' interactions.

Question 1

The presented gesture can be seamlessly integrated into the typical use of the object

	Reach	Load	Lift	Hold	Replace	Unload	Depart
Cup	4 (4 - 5)	4 (2.75 - 5)	4 (4 - 4.25)	3.5 (3 - 5)	5 (4 - 5)	4 (3 - 4.25)	4 (3 - 5)
Pen	4 (3.75 - 5)	4.5 (4 - 5)	5 (4 - 5)	5 (4 - 5)	5 (4 - 5)	5 (3.75 - 5)	4 (3.5 - 5)
Book	3 (3 - 4.25)	5 (4 - 5)	4 (3 - 4.25)	5 (4 - 5)	3.5 (2 - 4.25)	5 (4.75 - 5)	3 (2 - 4)
Toothbrush	5 (5 - 5)	4 (3 - 5)	4 (3 - 5)	4 (4 - 4.25)	4 (4 - 5)	4 (3.75 - 4.25)	2 (1 - 3.25)
Bicycle Handle	5 (4 - 5)	4 (3.75 - 5)		4.5 (4 - 5)		4 (3.75 - 5)	4.5 (4 - 5)
Steering Wheel	4 (3.75 - 5)	4.5 (4 - 5)		5 (4.75 - 5)		5 (4 - 5)	4 (2.75 - 4)

Question 2

The presented gesture would be useful

	Reach	Load	Lift	Hold	Replace	Unload	Depart
Cup	4.5 (4 - 5)	4 (4 - 5)	4 (3 - 5)	3.5 (3 - 5)	5 (4 - 5)	4 (3 - 5)	3.5 (3 - 4.25)
Pen	4.5 (4 - 5)	5 (4 - 5)	4.5 (4 - 5)	5 (4 - 5)	5 (4 - 5)	4.5 (2.75 - 5)	4 (2.75 - 5)
Book	3.5 (3 - 4.25)	5 (5 - 5)	3.5 (3 - 4.25)	4.5 (4 - 5)	4 (3 - 5)	5 (5 - 5)	2.5 (2 - 4)
Toothbrush	5 (4.75 - 5)	4 (3 - 4.25)	4.5 (3 - 5)	4.5 (4 - 5)	4 (4 - 5)	3 (2.75 - 4.25)	2 (2 - 3.25)
Bicycle Handle	5 (4 - 5)	4 (3.75 - 5)		5 (4 - 5)		4 (3.75 - 5)	4.5 (4 - 5)
Steering Wheel	4.5 (3.75 - 5)	4.5 (4 - 5)		5 (4.75 - 5)		5 (4 - 5)	4 (3 - 4)

Strongly disagree 1 2 3 4 5 *Strongly agree*

Figure 11: The median and inter-quartile ranges represent (left) the integration of GraspUI into the typical use of the object and (right) the usefulness of gestures. Empty cells indicate that the phase was not possible using those objects. (Higher is better.)

the typical use of the object and if the presented gesture would be useful. These questions align with factors that prior research has also utilized for subjective evaluations when developing new interaction techniques with objects [37, 42, 63]. Participants provided their responses on a 5-point Likert scale, where 1 denoted "strongly disagree," and 5 denoted "strongly agree". A free-form text field was included below the scale to encourage participants to provide additional comments and suggestions to improve the gesture designs. During the study, participants also had access to each of the six physical objects (Figure 10) to (optionally) perform the gestures for themselves if desired. The order of objects was counterbalanced using a balanced Latin square and the order of phases for each object was randomized.

4.3 Results

Overall, 456 responses were collected during the study. From these responses, we computed the median and interquartile range for the Likert scale questions (Figure 11).

Generally speaking, participants expressed a positive view towards the GraspUI interfaces. They found the gestures and micro-task applications contextually well-aligned, with comments such as *"very easy and straightforward"* (P1) and *"natural and useful!"* (P4). In at least 4 out of 7 phases, all objects received a median rating of 4 or higher for both Likert scale questions. These results highlight the potential of various grasp phases beyond the commonplace Hold phase, which have been underexplored in prior literature. Across all seven grasp phases, the Pen object consistently received higher ratings (ranging from 4 to 5) for both integration and usefulness. Conversely, the Toothbrush object received a notably lower score of 2 for the Depart phase on both questions, while receiving higher ratings (between 4 to 5) for the first five phases. We also observed a similar pattern with the Book object. This variation across phases may be attributed to gesture design, as pointed out by participants: *"(the heart) is an awkward gesture that I've never used or done otherwise"* (P10). Additionally, when compared to all other phases, the Depart phase received the lowest rating for liftable objects but was found relatively more useful when self-sustained objects (Bicycle Handle and Steering Wheel) were used.

Participants also appreciated the Load phase gestures during their initial contact with the six objects. The Load phase consistently achieved higher integration and usefulness scores, with participants describing them as *"easy to use"* (P3). These scores may be attributed to the object's presence on a sturdy surface, which facilitates improvements to gesture performance. A similar pattern was observed for the Unload phase, which occurred as the fingers were about to release the object. Although the Hold phase has been emphasized in previous gesture research, it did not receive high ratings, particularly with the Cup object. In future work, we recommend conducting tailored studies that consider both the grasp phase and the object for the design of suitable gestures.

5 INTERACTIVE PROTOTYPE

In our storyboarding sessions, we generated a *breadth of Novel Examples* to showcase the numerous ways our toolkit might be used, and our initial study shows promise for the design space through a video-based evaluation. Here, our goal is to *demonstrate* a practical implementation of GraspUI via an end-to-end interactive prototype and gather feedback from end-users [47]. The prototype utilizes commodity hardware and enables input opportunities across all grasping phases. While hand tracking in AR is improving [56], it is still non-trivial to detect specific grasp gestures, especially when an occluding physical object is involved. As such, we elected to implement our prototype with a single object (coffee cup) to ensure each of the phases and gestures would be accurately detected, as shown in the previous storyboards (Section 3.3).

5.1 Hardware Setup

The prototype uses a stereo-configuration consisting of two Logitech C922 Pro Stream Webcams to individually track the hand and object, along with a mixed-reality headset (Microsoft HoloLens 2)

Figure 12: Interactive GraspUI prototype consisting of a Microsoft HoloLens 2 and two Logitech webcams to simultaneously track the hand and object.

to display the interface components (Figure 12). All computations were performed on a Lenovo ThinkStation P620 desktop PC with 16 GB of RAM and NVIDIA RTXTM graphics cards. The webcams were connected to the PC via USB cables and positioned at opposite corners of a table to provide full coverage of the cup from one side and the hand from the other.

5.2 Implementation

To demonstrate the cup-based use cases that were designed for all 7 grasping phases, the implementation needed to know one's hand state, identify the cup's state, recognize hand gestures, and provide timely interfaces for interaction. We thus developed a hybrid system that combines off-the-shelf hardware, classification models, and heuristics. We included 3 classification models: an *Object tracker*, a *Static hand pose tracker*, and a *Dynamic hand pose tracker*. Additionally, we utilized the Hololens' built-in hand tracker to precisely track hand positions in Unity.

5.2.1 Object Tracker. MediaPipe Objectron [50] was used to track the cup's 3D coordinates within the webcam images and classify its position relative to the table using a threshold-based classifier. The classifier had three classes: in the air, towards the user, and away from the user. During testing, this approach effectively tracked the cup's position without the need for additional training data.

5.2.2 Static Hand Pose Tracker. MediaPipe's hand tracking with 21 3D landmarks and a TensorFlow [3] neural network were used to classify four hand poses. The neural network had four fully connected layers with a softmax activation function. The four hand pose classes were: grasping the object, covering the object, thumbs up while holding the object, and freehand.

5.2.3 Dynamic Hand Pose Tracker. We reused the webcam feed from the static hand-pose tracker to classify six hand motions. We input MediaPipe's 21 3D landmarks into a Sequential Neural Network model with an LSTM network with 16 hidden units, two fully connected layers, and a dense layer with ReLU activation function. The final layer had a softmax activation function. The six hand motion classes were: reaching for an object, departing from an object, holding an object, returning an object, circling the thumb while holding an object, and tapping an object with the thumb.

Our real-time implementation incorporated heuristics defined using the empirically selected 3D position of the cup and hand. The heuristics were then combined with the trained classifiers' output to prevent misclassification and automatically fade-in/out interfaces.

5.3 Supported Interactions

Our interactive prototype supported passing the hand through menu items, using finger gestures, and placing the cup on a table in a specific location. During the Reach and Depart phases (Figure 13), users could pass their hand over the buttons from the front or back respectively. During the Load phase, the user could place their hand onto the cup to view notifications, while the Lift, Hold, and Unload phases required that a thumb gesture be performed. During the Replace phase, the cup could be placed in different locations on the table to trigger the commands of the mapped microtasks. We intentionally chose to implement a diverse range of gestures as representative examples for each individual grasping phase. While many interactions are possible, these gestures were notably designed to take place during individual instances of grasping, rather than numerous gestures taking place sequentially in a single grasping motion. We do anticipate however, that future work may reveal that combinations of gestures may also present unique opportunities for GraspUI interfaces.

6 USER EVALUATION

An empirical evaluation was conducted with end users to assess the user experience of the GraspUI gestures and to measure the overhead cost of performing the associated microtasks, in comparison to a baseline when using objects without microtasks. Participants' willingness to use the proposed interactions was also assessed. The methodology employed is akin to Slide to X [68], where a regular activity was substituted with a simple voluntary task.

6.1 Participants

Fourteen participants were recruited from within our institution and outside to participate in the study (i.e., 6 males, 8 females; μ = 27 years; $\tilde{x}$ = 26 years, range = 22-41 years; 11 right handed, two ambidextrous and one left-handed). The participants represented a diverse range of professions, including university professors, computer science graduate students, software engineers, human resources professionals, and graphic designers. They were from different cultural backgrounds across Europe, North America, and Asia. To ensure that there was variation in participants' arm lengths, we measured the average arm length from the shoulder to the middle finger's tip, with the average measurement being 74.02 centimeters (σ = 3.9 centimeters).

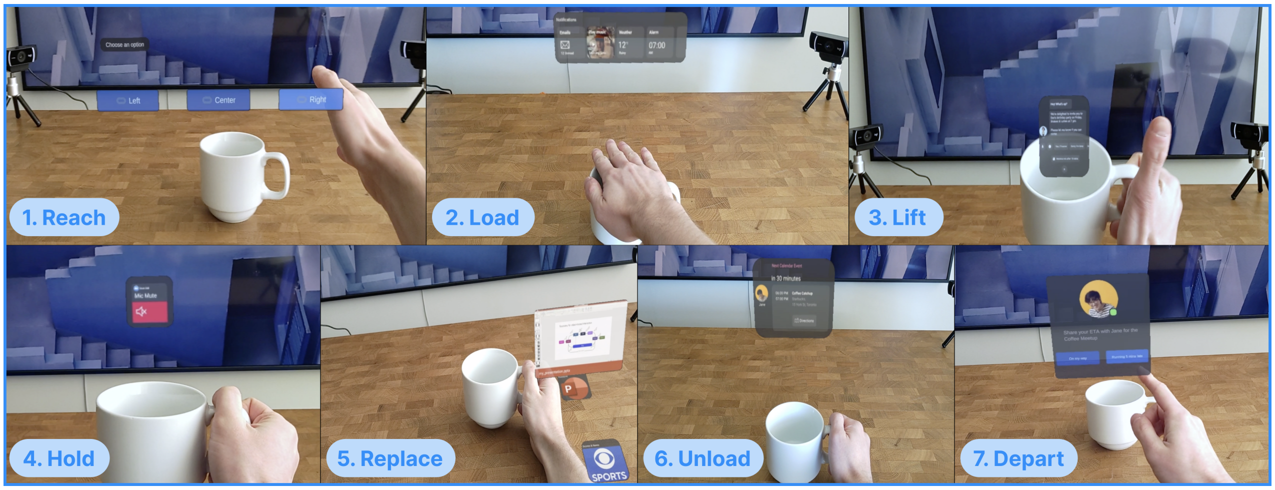

Figure 13: Demonstrations of the interactions possible with the GraspUI prototype during a given phase of grasping. Derived from our storyboard sessions, each example is designed to take place during a single instance of grasping, rather than sequentially across all phases. (1) Reach phase: The viewer's hand passes through a menu option to select left, center, or right buttons while reaching for the cup. (2) Load phase: Covering the top of the cup displays a notification center overview. (3) Lift phase: A thumbs up gesture responds to a message while lifting a cup. (4) Hold phase: A thumb circle gesture while holding the cup mutes a video conference call's microphone. (5) Reach phase: Placing the cup on the PowerPoint or Scoreboard app icons triggers the respective application. (6) Unload phase: A thumbs up gesture shows the calendar entry for the next event as the user releases the cup. (7) Depart phase: A user may pass through a button from the back side to send a pre-text with Left button for "On my way" and Right "Running 5 mins late".

6.2 Apparatus

The study was conducted in a quiet room and participants were instructed to sit comfortably in a chair with their hands on their laps following an introduction from the experimenter. The interactive prototype from Section 5 and a regular coffee mug were used during the study. The cup was filled with water and a passive sticker was attached to the table for consistent placement across trials.

6.3 Task and Procedure

In each trial, participants were asked to complete a *primary task*, which consisted of their arm being placed on their lap, reaching towards the cup on the table, taking a sip, placing it back on the table, and bringing their arm back to their lap. During some trials, participants completed *one* of the seven grasp phase interaction conditions (i.e., Reach, Load, Lift, Hold, Replace, Unload, and Depart), whereas during others they only completed the task activity without any additional microtask (i.e., the Baseline condition). Participants signaled the start and end of their hand movements by saying 'start' and 'stop', which were used to measure the duration of the trial. Before data collection began, participants were allowed to try each of the interactions. All trials were video recorded for later analysis and trials with classification accuracy issues, as identified by the experimenter, were noted and repeated.

To compare GraspUI with the Baseline condition, participants were split into two groups. One group performed all GraspUI interactions first, followed by the baseline condition, while the other group performed the baseline condition and then the GraspUI interactions. Participants performed five trials for each grasp phase interaction condition. The order of the grasp phase interaction conditions was counterbalanced across all seven phases. To better understand the interactions, we investigated seven input variations as prescribed in the original storyboards. These variations were implemented for the Reach phase with the Left, Middle, and Center buttons (Figure 5.1), the Replace phase with a vertically placed Scoreboard and Powerpoint applications (Figure 5.5), and the Depart phase with Left ("On my way") and Right ("Running 5 mins late") buttons (Figure 5.7).

We employed a think-aloud protocol to collect participant comments during the study. After completing trials for each of the 7 grasp phases, participants were asked to complete a NASA-TLX questionnaire and rate the disruption of the interaction on a scale from 1-5, where 1 denoted "very low," and 5 denoted "very high". The study took 90 minutes to complete and 14 participants $\times$ 5 trials $\times$ (1 Baseline + 7 Grasp phases) = 560 total trials were performed.

Although our interactive prototype enabled multi-class interaction classification across all phases, we evaluated each grasping phase separately by instructing participants to interact during only one phase at a time. We disabled the always-available Reach condition buttons and the early display of icons during the Return condition during trials that were evaluating the other five grasp phases.

6.4 Data Analysis

Our prototype employed training-based models for hand tracking. Thus, to train these models, we gathered data from a single user for ten trials across each phase. It is worth mentioning that this

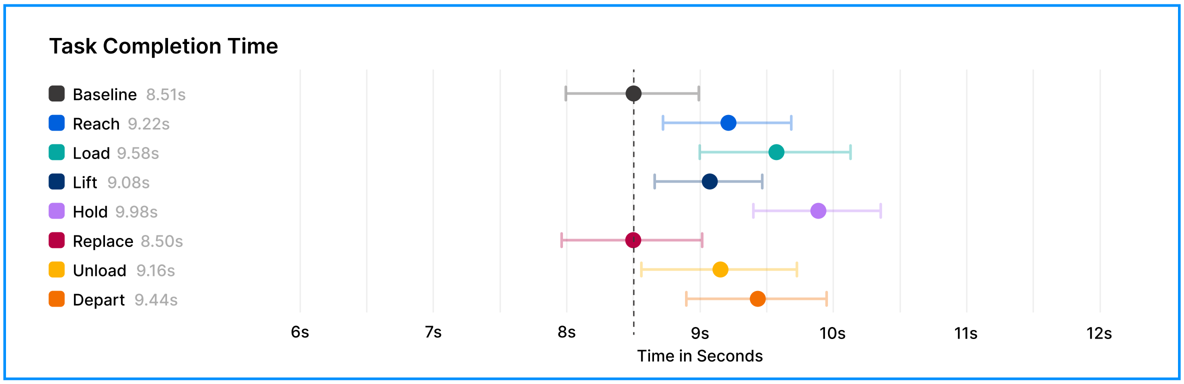

Figure 14: A comparison of average microtask completion times within all seven phases of grasping, performed in the study, with a Baseline condition (performing the primary activity without microtask). Error bars depict 95% confidence intervals.

user served as an additional participant specifically for training the classifiers and was not included in the formal evaluation. Using this participant's data, we obtained F1 scores of 0.99 and 0.85 for the Static and Dynamic hand-pose trackers, respectively, through an 80-20 evaluation. During the study, we observed that the recognition system failed to accurately predict the correct gesture or phase in 41 out of 560 trials (7.32%). Note that we did not gather any training data from the fourteen participants.

7 RESULTS

After implementing the interactive prototype and collecting data from the second user study, we now present our results to quantify the overhead in terms of time and the participants' experience ratings collected in different phases of grasping.

7.1 Overhead Costs

Our findings indicate that the average baseline task completion time was 8.51 seconds (.95 CI= [8.00, 9.00]), compared to an average task completion time for all grasp phase interaction conditions of 9.27 seconds (.95 CI= [9.07, 9.46]). This led to a difference of 0.76 seconds (Figure 14) which is lower than prior microtask integration work (i.e., 0.76 seconds versus 1.07 seconds, respectively) [68]. Notably, our approach focuses on objects, whereas previous work primarily centered on phones. Nonetheless, we share the common goal of integrating microtasks into everyday interactions.

ANOVA did not yield significant results but we outline notable trends observed during the study. When considering each grasp phase interaction condition individually, microtasks performed during the Hold phase led to the slowest task completion time (μ = 9.89 seconds, .95 CI= [9.41, 10.37]). Meanwhile the fastest task completion time was observed during the Replace condition (μ = 8.50 seconds, .95 CI= [7.98, 9.03]). Although most previous research on hand-object interaction has focused on the Hold phase condition, our results demonstrated that the other grasp phase interaction conditions resulted in lesser microtask completion times. In particular, the average microtask completion time during the Replace condition was similar to the Baseline condition (i.e., 8.50 and 8.51 seconds, respectively). These findings suggest that GraspUI can be optimized to minimize the time required to complete microtasks while grasping. By reducing or eliminating the need for interaction during certain grasping phases, users may be able to complete microtasks more efficiently. For example, users can quickly launch an application on their PC while simultaneously placing their coffee mug on a table.

Moreover, we also observed that the interface's position affected the task completion time within the same grasp phase interaction condition. For instance, during the Replace condition, the Score-board app icon that appeared towards the user on the table resulted in task completion time of 7.99 seconds (.95 CI= [7.20, 8.78]), and was on average 0.5 seconds faster than the Baseline condition. The Reach condition – prompting the user to de-route their hand to pass through a button and then reach toward the object – invoked a similar variation: left μ = 9.69 seconds, (.95 CI= [8.84, 10.54]); center μ = 9.13 seconds, (.95 CI= [8.2, 10.05]); right μ = 8.84 seconds, (.95 CI= [7.99, 9.7]). As the cup's handle was on the right side, we observed that the right button in the Reach phase was the most optimal for input as well as more ergonomic for our right-handed participants when grabbing the object; even moreso than the center button. We observed similar patterns during the Depart phase, where the "On my way" button was on the left side and the "Running 5 min. late" button was on the right side. Overall, these results demonstrate that incorporating an interface into different phases of hand-object interaction resulted in a minimal increases in task completion times, and thus minimal overhead.

7.2 Users' Subjective Ratings and Feedback

We analyzed the NASA-TLX ratings across the 7 grasping phases using a Friedman Test (Figure 15). The results revealed a significant variance among the grasp phases (χ^2 = 39.50, p < 0.001). Post-hoc pairwise comparisons using a Bonferroni correction showed that different grasp phases resulted in significantly different workloads. Specifically, performing microtasks in the Reach and Unload phases differed significantly in the disruptive dimension (p < 0.01), while the Hold and Load phases differed significantly in the physical demand dimension (p < 0.01). Qualitative feedback provided by participants also sheds light on potential sources of difficulty during the Reach and Load phases, e.g., *"don't see the cup when it is on the display (icons)"* (P4) and *"if it is a hot drink, covering on it will be* *uncomfortable"* (P1). Some participants reported difficulties with the left button during the Reach condition, saying "*it feels unnatural*" (P7). Similarly, participants expressed concerns about covering the cup with comments such as "*if it was a hot coffee, it would be difficult*" (P14), but showed acceptance if it was another object.

In general, however, the median workload for each phase was 2 or lower, which suggests that performing microtasks wasn't very demanding and participants liked the idea of performing microtasks throughout different phases of grasping. For example, they stated "*It's so cool, I'd love to text my friends or click likes on posts with this*" (P8) and "*gestures are quite incorporated*" (P2). Participants also exhibited a strong preference for the interactions during the Load and Unload conditions, e.g., "*the simplicity and the immediacy was nice*" (P1). Although we devoted considerable effort to designing interfaces to match the interactions envisioned in the storyboards, some participants commented on the affordances of the buttons. In particular, during the Depart phase, they displayed an urge to press the button and cross-through from the front rather than the intended rear-side, e.g., "*We have a tendency to press from the front and not from behind*" (P3).

Overall, the evaluation highlights the benefits of GraspUI and suggests that there are opportunities for future research to explore additional interface designs within each grasping phase.

	Reach	Load	Lift	Hold	Replace	Unload	Depart
Mental Demand	1.5 (1-2)	1 (1-1)	1 (1-2)	1.5 (1-2)	1.5 (1-2)	1 (1-1)	2 (1-2)
Physical Demand	2 (1-2)	1 (1-1)	2 (1.25-2.75)	2 (1-2.75)	1.5 (1-3)	1 (1-1.75)	1 (1-2)
Temporal Demand	1 (1-1)	1 (1-1)	1 (1-2)	1 (1-2)	1 (1-1)	1 (1-1)	1 (1-1.75)
Performance	1 (1-2)	1 (1-1)	1 (1-2)	1 (1-2)	1 (1-2)	1 (1-1.75)	1 (1-2)
Effort	1.5 (1-2)	1 (1-1)	1.5 (1-2)	1.5 (1-3.75)	1 (1-2.75)	1 (1-1)	1.5 (1-2)
Frustration	1.5 (1-2)	1 (1-1.75)	1.5 (1-2)	1 (1-2)	1 (1-2)	1 (1-1)	1 (1-2)
Disruptive	2 (1.25-2)	2 (1-2)	1.5 (1-2)	2 (1.25-3)	2 (1-2)	1 (1-1)	1.5 (1-2)

Very high
5
4
3
2
1
Very low

Figure 15: The median and inter-quartile ranges of NASA-TLX ratings for the 7 grasping phases (lower is better).

8 DISCUSSION

Our results illustrate that our design space supports the integration of microtasks within various phases of grasping. It is worth noting that the duration-based overhead costs associated with GraspUI was lower than previous work, which incorporated an additional input during the phone unlocking gesture [68]. The average time overhead noted in previous work was 1.07 seconds, while our approach achieved a lower overhead of 0.76 seconds. Another interesting point to consider is that, while the majority of prior work on hand-object interactions has focused primarily on the holding phase, the findings from both of our user studies show that the other grasp phases may offer more opportunities for better gesture integration, increased utility, and reduced overhead. This would enable GraspUI to be used in a variety of applications, including as a means to provide explanations for AI outputs [78], enabling quick input actions for computing devices [9], facilitating civic data collection [55], simplifying data labeling for algorithms [23], streamlining reviewing tasks [35], and realizing the interfaces depicted in Sci-Fi movies [22]. In this section, we provide recommendations based on our design sessions, prototype development, and two user evaluations to help researchers, designers, and engineers develop future object-centric interfaces. We also outline the limitations of our work.

8.1 Design Implications

We first discuss how researchers and designers can apply GraspUI to seamlessly integrate object-centric gestures.

8.1.1 Hand, Object, and Environment State Knowledge. This research investigated the utility and challenges associated with incorporating interaction into various phases of the grasping process. Our studies showed that information about the user's hand, the object being grasped, and environmental settings were crucial to improving user experiences. For instance, during the initial stages of developing our prototype, we observed that delayed user interface visualizations hindered users' ability to respond effectively. To address this, we incorporated hand and object information into our implementation, allowing us to trigger UI elements at more opportune moments. To further enhance user experience, we recommend that designers thoroughly investigate the five factors outlined in the GRASP model [74]. As AR technology continues to advance, future devices may incorporate full-body tracking capabilities. In a future version of GraspUI, this full-body tracking could be leveraged to offer users tailored recommendations as they approach objects or even direct their gaze toward them. Expanding GraspUI interfaces may include adding undo options and ensuring unobtrusive interactions. Furthermore, a promising next step could be to adopt optimization approaches that consider users' cognitive load while placing objects in the environment [49].

8.1.2 Overcoming Legacy Biases. During our second study, we observed that users initially hesitated to interact with interactive components during the Reach phase. Similarly, during the Depart phase, where gestures were intended to guide the hand across an interface, we noticed instances where users momentarily reverted to selecting an item instead of passing their hand through the reverse side of a button for activation. This behavior can be attributed to participants being accustomed to using their index finger to point at, and select, options. To assist users, we dynamically adjusted the interface brightness and modified the button visuals to indicate that the selection was recognized. We also iteratively made adjustments to the target sizes and positioning of the visual cues in the system to enhance the user experience. Although designing optimal interfaces falls outside the scope of our work [73], given that this is an underrepresented but crucial aspect in AR research, our findings could serve as a starting point for designing prompts with everyday objects.

8.1.3 Utilizing Multiple Phases for Complex Interactions. During our user evaluation, we systematically examined interactions within all seven phases of grasping. We believe that focusing on each individual phase is crucial for this initial study. Nevertheless, our findings indicate that there could be more than one phase that might be suitable for integrating microtasks depending on the object. This offers designers the opportunity to deconstruct large tasks into manageable segments and leverage GraspUI in diverse application domains. For example, as the user reaches for the dumbbells, they display today's strength training workout on a pair of smart glass. Once the user lifts the dumbbells, they activate a timer and begin counting the user's reps. While holding the dumbbells, the user can swipe on the handles to change the music or adjust the volume of their headphones. This allows the user to stay focused on their workout without having to fumble with their phone or other devices. After finishing their set, the user can replace the dumbbells on the rack to save their workout data. GraspUI can be used to track the user's progress over time and provide personalized recommendations for the next workouts. Here, it would be essential to consider the unique characteristics of each phase and how they can be used in harmony to create a seamless and effective user experience during the workout.

8.2 Limitations and Opportunities

The objective of this project is to lay the foundation for advancing object-centric gestures within the seven phases of grasping. We anticipate that researchers and designers can leverage the design space and our findings to develop more innovative interactions. Next, we discuss some limitations and suggest future directions.

8.2.1 Refining Gesture Designs. To the best of our knowledge, no previous research has investigated interactions in different grasping phases. In our design space, we incorporate prior research and designers' feedback, prioritizing affordances based on object geometry and grasp type to eliminate physically impossible or object-dropping gestures. We also prioritize natural hand-object interactions and augment them with additional functionalities. Rapid movements during existing hand-object interactions may cause ambiguity with input gestures however, making it challenging to understand user intent within grasp phases. To further understand user intent within grasp phases, future work should integrate other evaluation methods, such as expert interviews and elicitation studies. These methods have been extensively employed in designing gestures during the Hold phase [10, 33, 64, 76]. We suggest refining interaction techniques tailored to specific phases and objects as the immediate next step. By utilizing μGlyph's comprehensive database of finger movements and contact surfaces [19], designers can create a wide variety of microgesture interactions. Although our research did not specifically focus on avoiding false activations, certain phases, such as Load, present opportunities due to dedicated input locations on less frequently used surfaces [16]. Moreover, decomposing complex tasks into sub-tasks across multiple objects and grasping phases presents an exciting avenue for future research.

8.2.2 Extending GraspUI Implementation and Evaluation. Our interactive prototype lays the technical groundwork for the GraspUI gesture recognition architecture, overcoming challenges with simultaneous un-instrumented object and hand tracking [29, 82]. This proof-of-concept system enabled us to evaluate GraspUI's feasibility across all phases with a coffee mug. Meanwhile, our end-user video study demonstrated the generalizability of the design space across other objects and grasp phases. However, we acknowledge that this methodology may not perfectly capture user experience within a real system where gestures are sensed in real-time.

To further assess GraspUI's practicality, we propose several future research directions. First, expand the system implementation to incorporate more objects, including larger and fixed ones. Second, tailor sensing technologies and feedback to specific grasping phases and object characteristics. For instance, previous studies have explored the use of electromyography (EMG) [11–13, 43, 54] and non-invasive electroencephalograms (EEG) [46, 71] to identify hand movements during particular phases of the grasping process. Third, explore alternative feedback methods such as audio or haptic cues, and continuous feedforward gesture guides [30]. Lastly, quantify the time overhead associated with GraspUI interactions compared to traditional methods, such as retrieving the phone from pocket, navigating to the relevant application, and executing the required task.

9 CONCLUSION

In this research, we presented GraspUI, an exploratory design space for object-centric gestures that enabled seamless input opportunities throughout the entire hand-object manipulation process. In collaboration with mixed reality designers from academia and industry, we developed 38 storyboards. These storyboards cover all seven distinct phases within our design space and illustrate the diverse application possibilities for gesture integration with six everyday objects. To evaluate the utility of the design space, 12 end users participated in a video-based assessment. Additionally, we demonstrated the technical feasibility of GraspUI through an interactive prototype. In our second study with 14 participants, we measured the overhead cost of performing microtask interactions within all grasp phases. Our findings show that performing microtasks during the grasp phases resulted in a minimal duration overhead of 0.76 seconds, when compared to the baseline condition of performing these activities without any microtask. Overall, the participants reacted positively to the proposed gestures and findings indicated that enabling input opportunities within existing hand movements can result in interfaces that are more efficient and seamlessly integrated with everyday activities.

ACKNOWLEDGMENTS

We express our gratitude to Daniel Clarke and Raj Sodhi for providing valuable feedback. Special thanks to Michelle Annett for proofreading. We also thank all the participants who generously gave their time to participate in our study. Lastly, we would like to thank the fellow research interns from the 2022 batch for their insightful discussions and constructive feedback.

REFERENCES

[1] 2015. Paula Zuccotti's 'every thing we touch' photo project catalogs daily life in objects. https://www.bloomberg.com/news/articles/2015-11-27/paula-zuccotti-s-every-thing-we-touch-photo-project-catalogs-daily-life-in-objects
[2] 2023. Jacquard by Google - Saint Laurent. https://atap.google.com/jacquard/products/ysl/
[3] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/

[4] Fraser Anderson, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2015. Supporting Subtlety with Deceptive Devices and Illusory Interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. https://doi.org/10.1145/2702123.2702336

[5] Leonardo Angelini, Francesco Carrino, Stefano Carrino, Maurizio Caon, Omar Abou Khaled, Jürgen Baumgartner, Andreas Sonderegger, Denis Lalanne, and Elena Mugellini. 2014. Gesturing on the Steering Wheel: a User-elicited taxonomy. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '14)*. https://doi.org/10.1145/2667317.2667414

[6] Michelle Annett, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2011. Medusa: a proximity-aware multi-touch tabletop. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. https://doi.org/10.1145/2047196.2047240

[7] Michael A Arbib. 2011. *Perceptual Structures and Distributed Motor Control*. Wiley Online Library. https://doi.org/10.1002/cphy.cp010233

[8] Tal August, Shamsi Iqbal, Michael Gamon, and Mark Encarnación. 2020. Characterizing the Mobile Microtask Writing Process. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20)*. https://doi.org/10.1145/3379503.3403541

[9] Nikola Banovic, Christina Brant, Jennifer Mankoff, and Anind Dey. 2014. ProactiveTasks: the short of mobile device use sessions. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14)*. https://doi.org/10.1145/2628363.2628380

[10] Sandra Bardot, Surya Rawat, Duy Thai Nguyen, Sawyer Rempel, Huizhe Zheng, Bradley Rey, Jun Li, Kevin Fan, Da-Yuan Huang, Wei Li, and Pourang Irani. 2021. ARO: Exploring the Design of Smart-Ring Interactions for Encumbered Hands. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21)*. https://doi.org/10.1145/3447526.3472037

[11] Iason Batzianoulis, Sahar El-Khoury, Elvira Pirondini, Martina Coscia, Silvestro Micera, and Aude Billard. 2017. EMG-based decoding of grasp gestures in reaching-to-grasping motions. *Robotics and Autonomous Systems* 91 (2017). https://doi.org/10.1016/j.robot.2016.12.014

[12] Iason Batzianoulis, Nili E Krausz, Ann M Simon, Levi Hargrove, and Aude Billard. 2018. Decoding the grasping intention from electromyography during reaching motions. *Journal of Neuroengineering and Rehabilitation* 15, 1 (2018). https://doi.org/10.1186/s12984-018-0396-5

[13] Iason Batzianoulis, Ann M Simon, Levi Hargrove, and Aude Billard. 2019. Reach-to-grasp motions: Towards a dynamic classification approach for upper-limp prosthesis. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE. https://doi.org/10.1109/NER.2019.8717110

[14] Jeannette Bohg, Antonio Morales, Tamim Asfour, and Danica Kragic. 2013. Data-Driven Grasp Synthesis—A Survey. *IEEE Transactions on Robotics* 30, 2 (2013), 289–309. https://doi.org/10.1109/TRO.2013.2289018

[15] Ahmet Börütecene, Idil Bostan, Ekin Akyürek, Alpay Sabuncuoglu, Ilker Temuzkusu, Çaglar Genç, Tilbe Göksun, and Oguzhan Özcan. 2018. Through the Glance Mug: A Familiar Artefact to Support Opportunistic Search in Meetings. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '18)*. https://doi.org/10.1145/3173225.3173236

[16] Samarth Brahmbhatt, Cusuh Ham, Charles C Kemp, and James Hays. 2019. Contactdb: Analyzing and predicting grasp contact via thermal imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '19)*. https://doi.org/10.1109/CVPR.2019.00891

[17] 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* (2021). https://doi.org/10.3389/frobt.2021.714023

[18] Umberto Castiello. 2005. The neuroscience of grasping. *Nature Reviews Neuroscience* 6, 9 (2005). https://doi.org/10.1038/nrn1744

[19] Adrien Chaffangeon Caillet, Alix Goguey, and Laurence Nigay. 2023. μGlyph: a Microgesture Notation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. https://doi.org/10.1145/3544548.3580693

[20] Michael Chamunorwa, Mikołaj P Wozniak, Sarah Vöge, Heiko Müller, and Susanne CJ Boll. 2022. Interacting with Rigid and Soft Surfaces for Smart-Home Control. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (2022). https://doi.org/10.1145/3546746

[21] Justin Cheng, Jaime Teevan, Shamsi T Iqbal, and Michael S Bernstein. 2015. Break It Down: A Comparison of Macro- and Microtasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. https://doi.org/10.1145/2702123.2702146

[22] Victor Cheung and Alissa N Antle. 2020. Tangible Interfaces and Interactions in Sci-Fi Movies: A Glimpse at the Possible Future of TUIs through Fictional Tangible Systems. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '20)*. https://doi.org/10.1145/3374920.3374942

[23] Kabdo Choi, Hyungyu Shin, Meng Xia, and Juho Kim. 2022. AlgoSolve: Supporting Subgoal Learning in Algorithmic Problem-Solving with Learnersourced Microtasks. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. https://doi.org/10.1145/3491102.3501917

[24] Christopher Clarke, Alessio Bellino, Augusto Esteves, Eduardo Velloso, and Hans Gellersen. 2016. TraceMatch: a computer vision technique for user input by tracing of animated controls. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. https://doi.org/10.1145/2971648.2971714

[25] Christopher Clarke and Hans Gellersen. 2017. MatchPoint: Spontaneous Spatial Coupling of Body Movement for Touchless Pointing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. https://doi.org/10.1145/3126594.3126626

[26] Vanessa Cobus and Wilko Heuten. 2019. To Beep or Not to Beep? Evaluating Modalities for Multimodal ICU Alarms. *Multimodal Technologies and Interaction* (2019). https://doi.org/10.3390/mti3010015

[27] Christian Corsten, Ignacio Avellino, Max Möllers, and Jan Borchers. 2013. Instant User Interfaces: Repurposing Everyday Objects as Input Devices. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces (ITS '13)*. https://doi.org/10.1145/2512349.2512799

[28] Enrico Costanza, Samuel A Inverso, Rebecca Allen, and Pattie Maes. 2007. Intimate interfaces in action: Assessing the usability and subtlety of EMG-based motionless gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. https://doi.org/10.1145/1240624.1240747

[29] Zicong Fan, Omid Taheri, Dimitrios Tzionas, Muhammed Kocabas, Manuel Kaufmann, Michael J Black, and Otmar Hilliges. 2023. ARCTIC: A Dataset for Dexterous Bimanual Hand-Object Manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '23)*. https://doi.org/10.1109/CVPR52729.2023.01244

[30] Katherine Fennedy, Jeremy Hartmann, Quentin Roy, Simon Tangi Perrault, and Daniel Vogel. 2021. OctoPocus in VR: Using a dynamic guide for 3D mid-air gestures in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* (2021). https://doi.org/10.1109/TVCG.2021.3101854

[31] George W Fitzmaurice, Hiroshi Ishii, and William AS Buxton. 1995. Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '95)*. https://doi.org/10.1145/223904.223964

[32] Tamar Flash and Neville Hogan. 1985. The Coordination of Arm Movements: An Experimentally Confirmed Mathematical Model. *Journal of Neuroscience* (1985). https://doi.org/10.1523/jneurosci.05-07-01688.1985

[33] Weilun Gong, Stephanie Santosa, Tovi Grossman, Michael Glueck, Daniel Clarke, and Frances Lai. 2023. Affordance-Based and User-Defined Gestures for Spatial Tangible Interaction. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. https://doi.org/10.1145/3563657.3596032

[34] Tobias Grosse-Puppendahl, Christian Holz, Gabe Cohn, Raphael Wimmer, Oskar Bechtold, Steve Hodges, Matthew S Reynolds, and Joshua R Smith. 2017. Finding Common Ground: A Survey of Capacitive Sensing in Human-Computer Interaction. In *Proceedings of the 2017 CHI conference on human factors in computing systems (CHI '17)*. https://doi.org/10.1145/3025453.3025808

[35] Kurtis Heimerl, Brian Gawalt, Kuang Chen, Tapan Parikh, and Björn Hartmann. 2012. CommunitySourcing: engaging local crowds to perform expert work via physical kiosks. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '12)*. https://doi.org/10.1145/2207676.2208619

[36] Steven J Henderson and Steven Feiner. 2008. Opportunistic controls: leveraging natural affordances as tangible user interfaces for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology (VRST '08)*. https://doi.org/10.1145/1450579.1450625

[37] Seongkook Heo, Michelle Annett, Benjamin J Lafreniere, Tovi Grossman, and George W Fitzmaurice. 2017. No Need to Stop What You're Doing: Exploring No-Handed Smartwatch Interaction.. In *Proceedings of the 43rd Graphics Interface Conference (GI '17)*. https://doi.org/10.20380/GI2017.14

[38] 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 (CHI '16)*. https://doi.org/10.1145/2858036.2858095

[39] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. https://doi.org/10.1145/258549.258715

[40] Roland S Johansson and J Randall Flanagan. 2009. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience* 10 (2009). https://doi.org/10.1038/nrn2621

[41] Lynette A Jones and Susan J Lederman. 2006. *Human hand function*. Oxford University Press.

[42] Nikhita Joshi, Parastoo Abtahi, Raj Sodhi, Nitzan Bartov, Jackson Rushing, Christopher Collins, Daniel Vogel, and Michael Glueck. 2023. Transferable Microgestures Across Hand Posture and Location Constraints: Leveraging the

Middle, Ring, and Pinky Fingers. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23).* https://doi.org/10.1145/3586183.3606713

[43] Salman Mohd Khan, Abid Ali Khan, and Omar Farooq. 2021. EMG based classification for pick and place task. *Biomedical Physics & Engineering Express* 7 (2021). https://doi.org/10.1088/2057-1976/abfa81

[44] Chang-Min Kim and Tek-Jin Nam. 2022. Exploration on Everyday Objects as an IoT Control Interface. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference (DIS '22).* https://doi.org/10.1145/3532106.3533472

[45] Gierad Laput and Chris Harrison. 2019. Sensing Fine-Grained Hand Activity with Smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* https://doi.org/10.1145/3290605.3300568

[46] Richard T Lauer, P Hunter Peckham, and Kevin L Kilgore. 1999. EEG-based control of a hand grasp neuroprosthesis. *Neuroreport* 10, 8 (1999). https://doi.org/10.1097/00001756-199906030-00026

[47] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation strategies for HCI toolkit research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* https://doi.org/10.1145/3173574.3173610

[48] Lik-Hang Lee and Pan Hui. 2018. Interaction Methods for Smart Glasses: A Survey. *IEEE Access* 6 (2018). https://doi.org/10.1109/ACCESS.2018.2831081

[49] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19).* https://doi.org/10.1145/3332165.3347945

[50] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. In *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019.* http://arxiv.org/abs/1906.08172

[51] Kent Lyons, Christopher Skeels, Thad Starner, Cornelis M. Snoeck, Benjamin A. Wong, and Daniel Ashbrook. 2004. Augmenting Conversations Using Dual-Purpose Speech. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04).* https://doi.org/10.1145/1029632.1029674

[52] Christine L MacKenzie and Thea Iberall. 1994. *The grasping hand.* Elsevier.

[53] Ville Mäkelä, Johannes Kleine, Maxine Hood, Florian Alt, and Albrecht Schmidt. 2021. Hidden Interaction Techniques: Concealed Information Acquisition and Texting on Smartphones and Wearables. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21).* https://doi.org/10.1145/3411764.3445504

[54] Chiara Martelloni, Jacopo Carpaneto, and Silvestro Micera. 2009. Characterization of EMG patterns from proximal arm muscles during object-and orientation-specific grasps. *IEEE Transactions on Biomedical Engineering* (2009). https://doi.org/10.1109/TBME.2009.2026470

[55] Thomas Maskell, Clara Crivellaro, Robert Anderson, Tom Nappey, Vera Araújo-Soares, and Kyle Montague. 2018. Spokespeople: Exploring Routes to Action through Citizen-Generated Data. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* https://doi.org/10.1145/3173574.3173979

[56] Microsoft. 2023. *Hand Interaction Examples.* https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/features/example-scenes/hand-interaction-examples?view=mrtkunity-2022-05

[57] Franziska Mueller, Florian Bernard, Oleksandr Sotnychenko, Dushyant Mehta, Srinath Sridhar, Dan Casas, and Christian Theobalt. 2018. GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '18).* https://handtracker.mpi-inf.mpg.de/projects/GANeratedHands/

[58] John R Napier. 1956. The Prehensile Movements of the Human Hand. *The Journal of Bone and Joint Surgery British Volume* 38, 4 (1956). https://doi.org/10.1302/0301-620X.38B4.902

[59] Henning Pohl and Michael Rohs. 2014. Around-Device Devices: My Coffee Mug is a Volume Dial. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices &; Services (MobileHCI '14).* 81–90. https://doi.org/10.1145/2628363.2628401

[60] Miya K Rand, YP Shimansky, Abul BMI Hossain, and George E Stelmach. 2008. Quantitative model of transport-aperture coordination during reach-to-grasp movements. *Experimental Brain Research* 188 (2008). https://doi.org/10.1007/s00221-008-1361-5

[61] Julie Rico and Stephen Brewster. 2010. Gesture and voice prototyping for early evaluations of social acceptability in multimodal interfaces. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI '10).* https://doi.org/10.1145/1891903.1891925

[62] David A Rosenbaum, Ruud J Meulenbroek, Jonathan Vaughan, and Chris Jansen. 2001. Posture-based motion planning: applications to grasping. *Psychological review* 108 (2001). https://doi.org/10.1037/0033-295x.108.4.709

[63] Adwait Sharma, Michael A Hedderich, Divyanshu Bhardwaj, Bruno Fruchard, Jess McIntosh, Aditya Shekhar Nittala, Dietrich Klakow, Daniel Ashbrook, and Jürgen Steimle. 2021. SoloFinger: Robust Microgestures while Grasping Everyday Objects. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21).* https://doi.org/10.1145/3411764.3445197

[64] 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 (CHI '19).* https://doi.org/10.1145/3290605.3300632

[65] Roy Shilkrot, Jochen Huber, Jürgen Steimle, Suranga Nanayakkara, and Pattie Maes. 2015. Digital Digits: A Comprehensive Survey of Finger Augmentation Devices. *Comput. Surveys* (2015). https://doi.org/10.1145/2828993

[66] Yanke Tan, Sang Ho Yoon, and Karthik Ramani. 2017. BikeGesture: User Elicitation and Performance of Micro Hand Gesture as Input for Cycling. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17).* https://doi.org/10.1145/3027063.3053075

[67] Brandon T Taylor and V Michael Bove Jr. 2009. Graspables: grasp-recognition as a user interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09).* https://doi.org/10.1145/1518701.1518842

[68] Khai N. Truong, Thariq Shihipar, and Daniel J. Wigdor. 2014. Slide to X: Unlocking the Potential of Smartphone Unlocking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14).* https://doi.org/10.1145/2556288.2557044

[69] Rajan Vaish, Keith Wyngarden, Jingshu Chen, Brandon Cheung, and Michael S. Bernstein. 2014. Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14).* https://doi.org/10.1145/2556288.2556996

[70] Vuforia. 2023. VuMarks. https://library.vuforia.com/objects/vumarks

[71] Stephan Waldert, Hubert Preissl, Evariste Demandt, Christoph Braun, Niels Birbaumer, Ad Aertsen, and Carsten Mehring. 2008. Hand Movement Direction Decoded from MEG and EEG. *Journal of Neuroscience* 28 (2008). https://doi.org/10.1523/JNEUROSCI.5171-07.2008

[72] Martin Weigel, Vikram Mehta, and Jürgen Steimle. 2014. More than touch: understanding how people use skin as an input surface for mobile computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14).* https://doi.org/10.1145/2556288.2557239

[73] Thomas J Williams, Simon L Jones, Christof Lutteroth, Elies Dekoninck, and Hazel C Boyd. 2021. Augmented Reality and Older Adults: A Comparison of Prompting Types. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21).* https://doi.org/10.1145/3411764.3445476

[74] Raphael Wimmer. 2010. Grasp sensing for human-computer interaction. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '11).* https://doi.org/10.1145/1935701.1935745

[75] Katrin Wolf, Markus Funk, Pascal Knierim, and Markus Löchtefeld. 2016. Survey of interactive displays through mobile projections. *International Journal of Mobile Human Computer Interaction* 8 (2016). http://doi.org/10.4018/IJMHCI.2016100102

[76] Katrin Wolf, Anja Naumann, Michael Rohs, and Jörg Müller. 2011. A taxonomy of microinteractions: Defining microgestures based on ergonomic and scenario-Dependent requirements. In *13th International Conference on Human-Computer Interaction (INTERACT '11).* https://doi.org/10.1007/978-3-642-23774-4_45

[77] Haijun Xia, Michael Glueck, Michelle Annett, Michael Wang, and Daniel Wigdor. 2022. Iteratively designing gesture vocabularies: A survey and analysis of best practices in the HCI literature. In *ACM Transactions on Computer-Human Interaction (TOCHI '22).* https://doi.org/10.1145/3503537

[78] Xuhai Xu, Anna Yu, Tanya R. Jonker, Kashyap Todi, Feiyu Lu, Xun Qian, João Marcelo Evangelista Belo, Tianyi Wang, Michelle Li, Aran Mun, Te-Yen Wu, Junxiao Shen, Ting Zhang, Narine Kokhlikyan, Fulton Wang, Paul Sorenson, Sophie Kim, and Hrvoje Benko. 2023. XAIR: A Framework of Explainable AI in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23).* https://doi.org/10.1145/3544548.3581500

[79] Yang Zhang, Michel Pahud, Christian Holz, Haijun Xia, Gierad Laput, Michael McGuffin, Xiao Tu, Andrew Mittereder, Fei Su, William Buxton, and Ken Hinckley. 2019. Sensing Posture-Aware Pen+Touch Interaction on Tablets. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* https://doi.org/10.1145/3290605.3300285

[80] Qian Zhou, Sarah Sykes, Sidney Fels, and Kenrick Kin. 2020. Gripmarks: Using Hand Grips to Transform In-Hand Objects into Mixed Reality Input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20).* https://doi.org/10.1145/3313831.3376313

[81] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07).* https://doi.org/10.1145/1240624.1240704

[82] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. 2023. Object Detection in 20 Years: A Survey. *Proc. IEEE* (2023). https://doi.org/10.1109/JPROC.2023.3238524