Affordance-Based and User-Defined Gestures for Spatial Tangible Interaction

Weilun Gong
Reality Labs Research at Meta
weilungong@meta.com

Stephanie Santosa
Reality Labs Research at Meta
ssantosa@meta.com

Tovi Grossman
University of Toronto
tovi@dgp.toronto.edu

Michael Glueck
Reality Labs Research at Meta
mglueck@meta.com

Daniel Clarke
Reality Labs Research at Meta
danclarke@meta.com

Frances Lai
Reality Labs Research at Meta
flai@meta.com

ABSTRACT

Although mid-air hand gestures have been widely adopted by VR/AR products (e.g., Quest 2 and HoloLens), some drawbacks remain due to their lack of tangibility and tactile feedback. Opportunistic Tangible User Interfaces could address these shortcomings by repurposing existing objects in one's physical environment. However, there has yet to be a systematic investigation of the gestures that would be desirable when using opportunistic objects or how such gestures would be impacted by such objects. In this work, we conducted an elicitation study to investigate the desirability of object and gesture combinations across a variety of interactions. The results contribute (1) an opportunistic tangible UI gesture set for spatial interfaces, and (2) an Affordance-Based Object Selector Scheme that identifies ideal objects for tangible input given a desired input gesture, based on that object's physical affordances. Arising from these findings is the vision of the Adaptive Tangible User Interface, which supports the on-the-fly composition of tangible interfaces based on the affordances found in the physical environment and a user's input task.

CCS CONCEPTS

• **Human-centered computing**; • **Gestural input; Mixed/augmented reality**;

KEYWORDS

Embodied Interaction, Tangible User Interface, Input Technique

ACM Reference Format:
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 *Designing Interactive Systems Conference (DIS '23), July 10–14, 2023, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3563657.3596032

DIS '23, July 10–14, 2023, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-9893-0/23/07…$15.00
https://doi.org/10.1145/3563657.3596032

1 INTRODUCTION

Researchers and designers in human-computer interaction have conducted extensive research into spatial interfaces, which enable interactions within our real 3D environment, often with the support of AR and VR headsets. In such interfaces, mid-air hand gestures are often used to interact with virtual content. However, mid-air gestures lack tangibility and haptic feedback, resulting in usability challenges that have been identified in past research [33, 39]. Despite subsequent studies that have shown that the use of physical objects can improve spatial input [3, 41], mid-air input remains a common interaction technique within mixed reality environments.

These challenges have motivated researchers to explore ways in which physical objects can be used to provide tangibility for spatial interfaces. However, existing approaches, such as shape-changing interfaces [16, 40, 58] or custom haptic devices [73], often require complex mechanical structures and are impractical for portable AR and VR form factors. Alternatively, researchers have proposed Opportunistic Tangible Interfaces [19, 21, 30–32, 38, 59, 66], which use objects within an existing physical environment for tangible interaction. However, previous research has only focused on targeted designs of such interfaces for specific use cases and objects, which are not scalable or adaptable to the diverse scenarios users encounter in everyday life. To overcome the scalability and adaptability challenges that Opportunistic Tangible Interfaces have, we focus on two core research questions. First, we investigate the relevant characteristics of tangible objects that impact their appropriateness for different types of input. We define such characteristics as affordance factors. Second, we seek to identify desirable Tangible User Interface (TUI) gestures for different input commands and understand how they can be determined by their associated affordance factors and adapted across different objects.

To investigate these fundamental questions, we designed an elicitation study [26, 74] for tangible input across a range of physical objects and tasks. In our study, participants performed gestures for a set of referent tasks with ten different objects representing a diverse range of affordance factors. The elicitation study approach allowed us to identify a range of tangible input gestures for various task-object combinations. Based on the quantitative and qualitative results from the study, we generated a gesture set that covered different dimensions of input for spatial computing and identified the relevant affordance factors that objects needed to have for each gesture. We then proposed an affordance-based object selector scheme, which identified ideal objects for tangible input given a desired input gesture. Finally, we outlined and illustrated our vision for an Adaptive Tangible User Interface system that repurposes the physical environment around the user on-the-fly to compose TUIs for spatial interaction. In summary, our research makes the following contribution to the field of Opportunistic Tangible User Interfaces:

- An elicitation study that understands appropriate gestures across a range of physical objects and tasks.
- A user-defined object-based gesture set that can be used for a range of 2D and 3D spatial interface tasks.
- An Affordance-Based Object Selector Scheme to determine appropriate objects to use with specific gestures.
- The vision of an Adaptive Tangible User Interface system illustrated with a mockup use case scenario.

2 BACKGROUND AND RELATED WORK

Relevant prior work includes TUIs for spatial interfaces, opportunistic tangible user interfaces, taxonomies of tangible objects, interfaces, and gestures, elicitation studies for user-generated input, and the concept of affordance in HCI.

2.1 Tangible User Interfaces for Spatial Interfaces

The lack of tangibility inherent in spatial interfaces has been one of the main challenges hindering the development of such interfaces. Ishii et al. defined tangible user interfaces [34], which transform digital information and interfaces into physical forms and enable users to leverage their existing abilities to manipulate physical things when using digital UIs. TUIs have been used for a range of spatial interfaces. A full review is beyond the scope of this research; however, readers are directed to Bouzbib's survey on the topic [9]. Examples of TUIs include adding tangibility to AR Interaction with 3D printed tangible Interfaces in an urban planning context [78], using TUIs to increase engagement for smartphone-based virtual reality [14], using fan-inspired shape-changing TUIs to add tactile feedback to VR [73], and adding directional force feedback to virtual racket sports using compressed air [68]. However, these projects were based on customized hardware designs and were developed for specific tasks and contexts. In this work, we wish to enable the development of scalable and adaptable TUI interfaces that can determine appropriate tangible gestures based on a user's intended tasks and utilize objects in the user's environment as a platform for tangible interaction.

2.2 Opportunistic Tangible User Interfaces

Ubiquitous computing research has previously explored methodologies to employ the affordances [24] of physical objects in the process of TUI design [66]. Within the context of mixed reality, this approached was explored by using otherwise unused everyday objects in one's physical environment as *Opportunistic TUIs* to provide haptic feedback to users [30]. Hettiarachchi and Wigdor's Annexing Reality work [31], for example, explored opportunistic interfaces for VR, using computer vision techniques to scan one's physical environment to find tangible objects that could be used as proxies for virtual objects. Different sensing technologies and interface authoring methods have also been explored to enable opportunistic TUIs, including using computer vision [19, 32], voice commands [22], AR markers [21, 38], or by tracking hand skin deformations [59]. For example, Du et al. [22] introduced *Ad hoc UI*, a prototyping toolkit that enables users to convert surrounding objects into opportunistic interfaces in real time, however, the toolkit focuses on 2D flat, rigid objects. More generally, most of the work in this field only investigated opportunistic TUIs for a limited range of contexts, consisting of case-by-case and object-specific interface designs. In our work, we seek to a provide a generalizable framework towards the vision of an Adaptive Tangible User Interface that can compose opportunistic tangible user interfaces based on one's current task and the existing affordances in a user's physical environment. The findings from our study enabled us to take two major steps towards this vision: (1) An opportunistic tangible UI gesture set for spatial interfaces, and (2) An affordance-based object selector scheme which maps input gestures to ideal objects.

2.3 Taxonomies of Tangible Objects, Interfaces, and Gestures

Previous work has explored ways to categorize tangible objects and interfaces. Researcher have proposed TUI taxonomies based on factors such as their affordances [23], mechanisms [16], and shape [56]. Roudaut et al. increased the fidelity of these taxonomies by proposing the Morphee [58] and Morphees+ [40] frameworks of shape resolution features of deformable objects. However, these prior taxonomies have not been designed to specifically inform which affordances and input possibilities are offered by different objects. The present research forms a taxonomy of everyday objects based on their *affordance factors*, i.e., the characteristics that impact their appropriateness for different types of input.

Gesture classifications and taxonomies in general have also been widely explored under different contexts, such as Mid-Air Hand Gestures [1], surface-based touch gestures [75] and gestures in Human Computer Interaction in general [37]. Others have looked at categorizing the types of gestures that may be possible when the hands are holding an object, such as a steering wheel [2, 20]. Sharma et al. [65] created a taxonomy of micro gestures that could be made when different hand grasps were used. Zhou et al. explored TUI gestures for AR while users were holding an object [77]. The present research builds upon this body of work to form a gesture taxonomy specifically for opportunistic object-based interaction for spatial interfaces, where little work has been done.

2.4 Elicitation and Wizard of Oz Studies for User-Generated Input

Asking users to create input systems and gesture sets has become a popular methodology in HCI. In an elicitation study, users are prompted with referents, which are the result of an action, and are asked to perform possible corresponding signs, which are the cause of the effect. This process (also called User-Derived Interface Design) was validated by Good et al. to refine a standard command-line mail interface [26]. It also was used by other researchers including Wobbrock et al. [27] and Nielsen et al. [51] in research exploring hand gestures for surface computing. Villarreal-Narvaez provided a survey of 216 gesture elicitation studies, demonstrating the prominence of this relatively new design methodology [70]. However, we are unaware of prior work using elicitation studies for Opportunistic Tangible User Interfaces.

The Wizard of Oz technique is another commonly used method due to its ability to simulate new system capabilities [67]. For example, Maulsby et al. explored the prototyping of an intelligent agent using a Wizard of Oz approach [46]. Robbe [57] studied gesture and speech input for PC applications. Voida et al. also used a Wizard of Oz approach to explore hand gestures for AR interfaces in the workplace [71]. The Wizard of Oz approach can be particularly valuable for elicitation studies, as it can allow users to feel like the system is actually recognizing their newly defined gestures. For instance, Connell et al. applied Wizard of Oz in an elicitation study to define body gestures for a Kinect-based interface [17]. In the present research, the Wizard of Oz method was used alongside the elicitation study methodology to enable participants to see the real-time effects of their gestural input.

2.5 Affordance

Affordance has been a popular concept in HCI and large amount of work has been done to explore frameworks to understand affordances. The term "affordance" was first introduced by Gibson [25] and then introduced to the HCI community by Norman [52] to describe an object's perceived and actual properties that determine how it can be used. Bærentsen et al. expanded this paradigm by suggesting that activity theory can be used as a frame of reference for the concept of affordance [4]. Kaptelinin et al. calls for a mediated action perspective on affordances, where a socio-cultural framework is employed to understand technology affordances as possibilities for human actions mediated by cultural means [36]. In our work, we adopt the framing proposed by Kaptelinin et al. and apply the concept of affordance to refer to the *instrumental affordances* of objects, which are the possibilities of action on objects to conduct input. We explicitly focus on how different physical characteristics on objects influence their appropriateness for different types of inputs. It is important to distinguish the concept *affordance* from the concept of *metaphor,* which we also refer to in this work. We adopt Fishkin's definition of metaphor within the context of tangible user interfaces, which refers to the association of a system effect of a user action as analogous to the real-world effect of similar actions[23]. While metaphors are distinct from affordances, we explore their potential influence on a user's perception of affordance.

3 USER STUDY

To develop a gesture set for opportunistic tangible interfaces, arrive at a taxonomy of object affordances, and derive an Affordance-Based Object Selector Scheme so that the gesture set scales to other contexts, we conducted a user study using an adaption of the elicitation study methodology [27, 51, 75]. In our study, each participant was shown the effect of a gesture input (e.g., a button is clicked) and was asked to perform the gesture that they believe could cause that effect (e.g., tapping on the surface of an object). We refer to the effect of the gesture as the referent [47]. Similar to prior elicitation studies [17], Wizard of Oz control was used to enable participants to perceive the effects of their gesture once it was performed. The study utilized 10 physical objects that were chosen to represent a range of possible affordance factors. Each object was used to perform 12 referents, which were chosen to represent a range of input requirements.

3.1 Participants

Twenty-two paid participants between 23-67 years old (i.e., 12 male, 10 female; mean 34 years; median 32 years; 2 left-handed) were recruited to participate in the study. Participants were from diverse professional backgrounds (e.g., engineering, art, law, finance, business). Participants with a background in human-computer interaction, AR/VR, or user interface design were excluded, to avoid any biased data due to previous UI design experience. All participants were compensated at a rate of $75USD/h and recruited by a 3rd party study recruitment company operating in major North American cities. We did not have any selection criteria related to cultural background or socioeconomic status, and did not collect such information. We discuss these factors and potential impact further in our discussion section.

3.2 Selection of Referents

To select a representative range of referent tasks, we relied on prior work that provided characterizations of 2D and 3D input tasks. The first property considered was the *dimensionality* of the task content. While our interest was in spatial input, it is important to consider how 2D interface widgets, such as menus and control panels, are often used within spatial user interfaces [10] and can be found in all AR/VR headsets on the market. The next property, *Input Degrees-of-Freedom* (DoF) was derived from Buxton's input taxonomy, which categorized tasks based on the number of continuous dimensions to be controlled [12]. In addition to one, two, and three DoF tasks [12], we also included a 6 DoF task, commonly performed when using 3D user interfaces [76]. Finally, we considered the flow of the task, i.e., whether or not the referent tasks were *discrete or continuous*, another characteristic of input from Buxton's taxonomy of input [13] and surface-based input [75].

Given these 3 key task characteristics, we arrived at the 12 referents, presented across 6 interface scenes (Table 1, Figure 2b). The first four scenes and seven referents consisted of 2D content, and the remaining two scenes and five referents consisted of 3D content. Arguably the most critical component of any 2D or 3D interface is object selection [42, 44], so, the first chosen referent was a simple on/off control of a button. The next referents were the 1D discrete previous/next control of a carousel UI [60, 75] and the 1D continuous control of a slider. The next four referents were conducted within a web page scene, which represented the types of floating 2D UI panels that may be present in a mixed reality environment [10]. The scene included two 1-DoF referents (scrolling, zooming), a 2-DoF task (pointer control) and a 6-DoF[1] task (ray casting). The three referents for 3D object manipulation (i.e., scale, rotate, and move) were selected as they are canonical tasks for 3D manipulation [42]. Finally, a 3D scene (flight simulator) was included to investigate tasks which require the use of several input commands simultaneously, a key element of many 3D user interfaces [35, 42, 50]. In this scene, users needed to control acceleration

[1]Technically ray casting requires only 5 DoF, however the literature often classifies it as 6 DoF, as the input controls both 3D position and orientation [6].

Table 1: The list of referents presented to participants grouped by scenes and labelled with dimensionality.

Scene	Referent	Scene Dimensionality	Input Degrees-of-Freedom	Flow
Press a button	on/off	2D	0	Discrete
Control a carousel UI	previous/next	2D	1	Discrete
Control a slider	increase/decrease	2D	1	Continuous
Navigate a web page	scroll: up/down	2D	1	Discrete
	zoom: in/out	2D	1	Continuous
	move pointer: up/down/left/right	2D	2	Continuous
	ray casting	2D	6	Continuous
3D object manipulation	scale: up/down	3D	1	Continuous
	rotate: roll/yaw/pitch	3D	3	Continuous
	move: x/y/z	3D	3	Continuous
Flight simulator	acceleration (1D): increase/decrease direction(3D): roll/yaw/pitch	3D	4 (compound)	Continuous

(1D input) and direction (3D input) at the same time. While the referent set does not exhaust all combinations of the three referent properties, it was carefully chosen to include a range of the most commonly used input commands and represented a range of key task characteristics, with the goal of enabling generalizations to other tasks with similar characteristics.

3.3 Selection of Objects

A key differentiating factor between this work and other elicitation studies is that instead of one single interaction medium [11, 23, 28], we wanted to elicit gestures across a range of objects with varying characteristics. As described by prior research, the properties of a tangible object can invoke any number of metaphorical links that may guide a user to use the object in different ways [23]. These properties may relate to its geometry (e.g., a mug is cylindrical and can be rotated like a knob) or semantic meaning (e.g., the tip of a pen is like a pointer). We define an *affordance factor* as a relevant characteristic of a tangible object that impacts its appropriateness for different types of input. Thus, our goal in selecting an object set was to choose objects with a range of possible affordance factors so that the object set could be as scalable as possible. To determine the relevant affordance factors, we reviewed characteristics used within prior taxonomies of shape-changing and tangible user interfaces [8, 11, 16, 28, 40, 53, 56, 58, 61, 72]. Based on this review, we arrived at the following set of potential key affordance factors. For each listed affordance factor, we cite prior work where it has been utilized for input.

Potential Geometric Affordance Factors: Surface Size and Curvature [40, 53, 58]; Edge Length and Curvature [40, 55]; Body Shape [8, 28, 56, 61, 72]; Body Grabability [11, 28, 72].

Potential Kinetic Affordance Factors: Movable Structures [8, 16]; Elasticity [40, 56].

We selected ten objects (Figure 1) that possessed a diverse range of properties across the affordance factors listed above. This list was derived through several internal workshops and iteration sessions by the research team. The research team first ran a workshop to narrow in on 30 common daily life objects. Afterwards, 10 out of the 30 objects were selected the in a subsequent iteration sessions

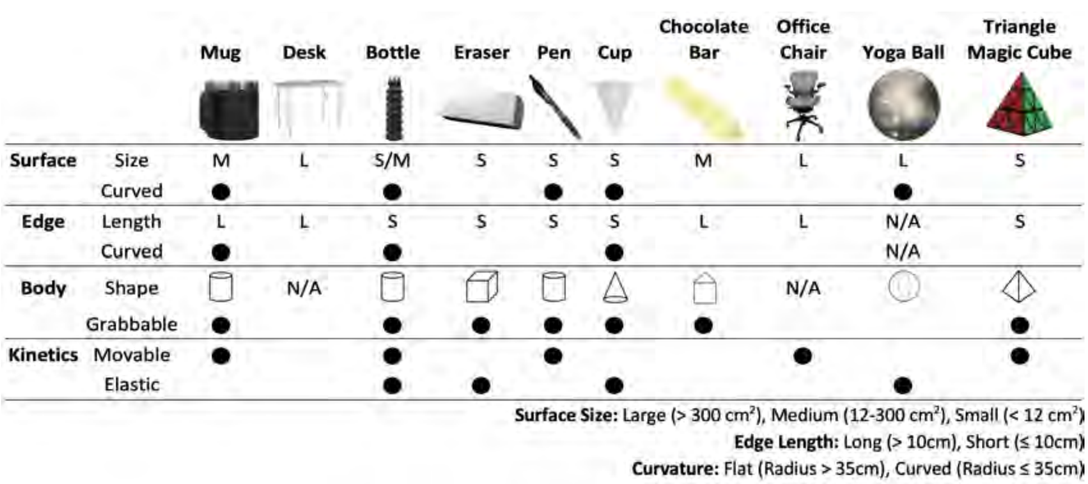

Figure 1: List of objects used by participants, labelled with the potential affordance factors they have.

to form an object set in which each affordance factor was roughly balanced. To delineate geometric sizes, we used the size of a regular human hand (250 cm2) [54] as standard: an object with a much smaller, similar and much bigger volume were considered as small, medium and large object respectively. For length, an edge longer than the average breadth of hand (10 cm) [54] was considered as long while others are labelled to be short. For curvature, the team decided to use Radius ≤ 35cm as the threshold to consider a surface or edge to be curved. For the purpose of external validity, and because of the potential importance of metaphorical links [23], we chose objects that were typically accessible and common in daily life. The only exception was the Triangle Magic Cube, which is less common in daily life, but specifically selected to cover a wider range of shapes. As with our selection of referents, we do not consider this to be an exhaustive list of all object types or potential affordance factors. However, by choosing objects based on these key affordance factors, our hope that the object set would be scalable and applicable to other objects with similar characteristics. We discuss this issue of scalability, for both objects and referents, in greater detail in Section 6.

3.4 Procedure

The study was conducted at a table in a lab with one monitor and ten everyday objects (Figure 2). At the start of the study, participants were shown the ten objects. Participants then viewed all referents for each scene. Participants were then asked to perform one gesture for each of the referents, repeated with every object. Participants completed all 12 referents in the same fixed order for one object before moving onto the next object. The ten objects were presented in a randomized order. A think aloud protocol was applied, which required participants to explain why they performed their gestures and what characteristics they saw within the object that made them think they could perform such a gesture. For each object-referent combination, the participant demonstrated their gesture twice. Once to explain the gesture and have it captured by the facilitator and once to see its resulting user interface effects using the Wizard-of-Oz setup. At the end of each scene, participants answered a survey question: "*If you needed to pick one gesture to use for this kind of interaction for the rest of your life, which gesture would you choose and why?*" This question allowed participants to report their most desirable gesture for each referent across all objects. With 22 participants, 11 referents (the two *flight simulator* referents were merged for the purpose of data collection as they were performed simultaneously), and 10 objects, $22 \times 11 \times 10 = 2420$ gestures were performed. A total of 213 gesture data points were discarded when it was clear to the facilitator that the participant didn't understand the referent. These data points mostly came from two participants that struggled with the study in general, especially with referents that had a 3D scene dimensionality (e.g., understanding the distinction between 3D movement and zoom). With the discarded data removed, there were a total of 2207 gestures analyzed in this study. For each participant, the study took between 1.5 – 2 hours.

A facilitator controlled the monitor to display the effect of gestures (i.e., referents) using a Wizard of Oz approach. We elected to use a monitor instead of a head-mounted display to improve the output quality and to ensure participants' gestures would not be impeded by head-mounted display limitations (e.g., reduced touch accuracy in VR [62] and restricted field-of-view in AR devices [7]). The referents were illustrated with three interactive interfaces that were created using Figma (referent 1-7), SketchUp (referent 8-10) and flight simulator game (referent 11-12). These interactive interfaces showed recorded animations of the referents or were controlled by a facilitator in real-time. For example, for referent 1, the session started with a standard introduction from the facilitator: "Task 1, Imagine you are turning this button on or off. Here is the result of the input." Then, the facilitator played the animation, which illustrated a button turning on and off on the monitor. After the animation was played, the interface on the monitor was reset to the original status, the participant could then start to perform their gesture with an object from the object set, and then their gesture proposals were recorded. After the participant conveyed his or her idea, the facilitator said: "Now try to use this gesture to control the interface." When participants started performing the gesture, their hand movements were closely monitored. The facilitator controlled the interface to offer real-time visual feedback to the participant. Participants' hand movements and think-aloud audio were recorded throughout the study. Two camera angles were used (i.e., one third person view through a desk-mounted camera and one first person view through a head-mounted camera). A third

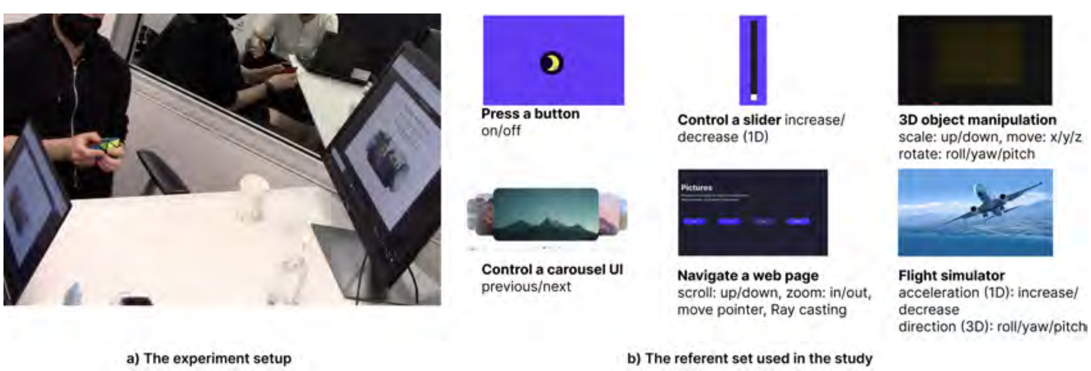

Figure 2: a) The experiment setup. A user works on the carousel control task using the Triangle Magic cube. The effect of the gesture is displayed on the monitor. b) The 12 referents used in the task were presented across six different UI scenes. For flight simulator, image is not from the actual game, but is representative of the referent shown to the user.

hand-held camera operated by the facilitator was used for close-up shots when the participants were manipulating small objects. In addition, the facilitator observed each session and took notes on the user's behavior.

A *priming* technique was applied to improve the data quality and reduce legacy bias [49]. At the beginning of the study, the facilitator introduced possible interaction paradigms, including touch, slide, press, squeeze, move, rotate and using movable structures. These gestures were demonstrated on a few other objects that were not from the study object set. Moreover, the participants were told not to consider the technical feasibility of any potential interface.

3.5 Data Coding and Analysis

Approximately 43 hours of recorded video data were analyzed and labelled with the description of the gesture and related object affordance factors. The analysis was conducted by three researchers who worked independently but coordinated frequently to ensure a consistent process was being followed. Gesture labels included the name of the gesture (e.g., squeezing, tapping, rotating), the hand usage (e.g., right index finger, whole right hand), the location on the object (e.g., on the cap of the bottle, the tip of the pyramid), and the pressure that was applied (i.e., no pressure, discrete pressure, continuous pressure). Labels were inferred from both the video data and the think-aloud audio (e.g., pressure/squeeze came from the think-aloud data). Each gesture was also labelled with any relevant affordance factor, based on the participant's comments on what led them to perform the observed gesture on the object (e.g., "This desk has a big flat surface, I want to touch it like using a trackpad"). Think-aloud data was labelled directly from the audio recordings of the video, not from a text transcription.

An open coding [18] reprocess was performed to categorize and group gestures together. Three researchers independently labelled the first 10% of gestures for every referent and compared their results. The three researchers had a high consensus (Cohen's kappa $\kappa = 0.82$), which suggested the labelling was reliable. The researchers then continued to finish coding the remaining data. A similar open coding process was used to label and categorize the affordance factors related to each gesture (with a Cohen's kappa $\kappa = 0.83$). Once the gesture and affordance factors were labelled, we calculated the agreement level of the gestures, which measured the level of consensus among the participants for each referent [74], the distribution of gestures for each referent, and the distribution of affordance factors for each gesture.

4 RESULTS

We now describe the results from our study, which include the taxonomy of the observed gestures (Section 4.1), the agreement rates of the observed gestures (Section 4.2), the associated user-defined gesture set (Section 4.3), an analysis of the gesture distributions across referents (Section 4.4), an analysis of the affordance factors (Section 4.5), and the derivation of the Affordance-Based Object Selector Scheme (Section 4.6).

4.1 Gesture Taxonomy

While previous work has proposed taxonomies of input [12, 23], little work has been done to create a gesture taxonomy specific for object-based spatial tangible interactions. As such, we first present a taxonomy to summarize and categorize the observed gestures to gain a more systematic understanding of object-based input. Based on the 2207 gestures that participants performed, we identified four classification categories that applied to these gestures (Table 2).

We observed gestures ranging in their degrees-of-freedom (i.e., 0D, 1D, 2D, 3D, 6D) which is a well-known property of input [12]. Here, we define 0D gestures as those that either switch the status of a value between on and off or just express the confirmation of an action, such as tapping on a desk's surface or squeezing a water bottle to press a virtual button. We also observed both discrete and continuous gestures, also a known property of input [12], sometimes referred as flow [75].

More specific to tangible interaction, we observed a range of expression strategies. With *hand-centric* gestures, participants used their hands as the primary medium to communicate, while only using objects as a medium to afford their hand movements. Examples included tapping on the surface of any object for *on*/off and sliding along any object with a flat surface to *increase/decrease*. With *object-centric* gestures, participants expressed their input commands via a status change on an object. Examples included squeezing a water bottle for *on/off* or rotating an object to control the orientation of a virtual object. With *object and hand-centric* gestures, participants simultaneously expressed their input command with a combination of the two strategies above. One example was tapping on a pen while rotating it for *on/off* and *ray casting* in a 3D pointing scene.

Finally, the gestures performed by participants arose from various metaphorical links [23]. Gestures based on *physical world metaphors* transferred previous experiences interacting with physical objects. One example was to rotate a chocolate bar like a steering wheel to control the direction of an airplane in a video game. Gestures based on *digital world metaphors* were derived from the participant's previous experience interacting with digital content. One example was to use the surface of a desk like a touchpad for a *move pointer. Abstract* gestures were not based on previous experiences or metaphors. Instead, they were subjectively defined by the user. An example was to use a double tap for *next* or a single tap for *previous.* This taxonomy will be used to ground the discussion of the resulting gesture set and how the gestures relate to an object's affordance factors.

4.2 Agreement Rates

To identify a set of widely applicable and generalizable object-based gestures for different types of input commands, we first calculated the agreement rate for each type of object to determine if there was a high level of consensus between participants' proposed gestures. Participants were considered in agreement if they proposed the same gesture types for a referent and object combination. We adopted the definition of agreement rate from the Agreement Analysis Toolkit [69], where *P* was the set of all proposals for referent-object combination *r*, *|P|* the size of the set, and *Pi* the subsets of identical proposals from *P*:

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \subseteq P} \left(\frac{|P_i|}{P}\right)^2 - \frac{1}{|P|-1} \tag{1}$$

The outcome of the agreement analysis indicated that a gesture set could be generated with a high level of consensus among the participants (Table 3). Agreement rates ranged from 0.1 to 1.000, with a mean across all objects

Table 2: Taxonomy of object-based gestures for spatial interaction based on collected gestures.

Category	Sub-category	Example Object-Based Gesture
Degrees-of-Freedom	0 DoF	Tap on the table to press a bottom
	1 DoF	Slide on table to control a slider
	2 DoF	Moving the fingers along the surface of a desk like using a trackpad to control a webpage
	3 DoF	Push a mug forward to move a virtual cube forward in a 3D modelling environment
	6 DoF	Control the position and orientation of a pen to point as virtual ray
	Compound	Rotate and squeeze a bottle to control the direction and acceleration in a flight simulator
Flow	Discrete	Use the index finger to swipe on an armchair surface to go to the next item in a carousel UI
	Continuous	Slide along a chocolate bar to control a continuous volume slider
Expression Strategy	Hand-centric	Use the index finger to swipe on a surface to control a carousel UI
	Object-centric	Squeeze a water bottle to "confirm" an action
	Object and hand centric	Use a pen to point at a slider and use the finger to slide along the surface of the pen to control the slider
Metaphorical Link	Physical	Hold a chocolate bar like a steering wheel and rotate it to control direction in a flight simulator
	Digital	Moving fingers on the surface of a desk like using a trackpad to control a webpage
	Abstract	To control a carousel, single tap for next, double tap for previous

Table 3: Agreement rates for each referent, displayed for each object type.

Referent	Chocolate Bar	Office Chair	Cup	Pyramid Cube	Desk	Pen	Eraser	Water Bottle	Mug	Yoga Ball	MEAN	STD
On/Off	0.71	0.51	0.34	0.39	1.00	0.69	0.60	0.28	0.49	0.33	0.53	0.22
Next/Previous	0.28	0.28	0.20	0.32	0.75	0.40	0.30	0.15	0.29	0.23	0.32	0.16
Increase/Decrease	0.45	0.58	0.42	0.30	1.00	0.53	0.46	0.30	0.46	0.37	0.49	0.20
Scrolling	0.27	0.58	0.17	0.15	0.68	0.53	0.29	0.10	0.28	0.20	0.32	0.20
Move Cursor	0.42	0.74	0.39	0.29	1.00	0.63	0.42	0.41	0.28	0.23	0.48	0.24
Zoom In/Out	0.23	0.38	0.13	0.13	0.68	0.15	0.22	0.13	0.18	0.27	0.25	0.17
Move a 3D Object	0.46	0.29	0.56	0.49	0.63	0.65	0.65	0.52	0.83	0.49	0.56	0.15
Rotate a 3D Object	0.62	0.60	0.91	0.81	0.91	0.66	0.91	0.75	0.82	0.81	0.78	0.12
Scale a 3D Object	0.32	0.37	0.17	0.14	0.90	0.21	0.22	0.17	0.24	0.23	0.29	0.22
Flight Direction	0.62	0.18	0.46	0.53	0.82	0.44	0.52	0.52	0.60	0.28	0.50	0.18
Flight Acceleration	0.22	0.21	0.19	0.19	0.43	0.17	0.20	0.16	0.19	0.18	0.21	0.08
MEAN	0.42	0.43	0.36	0.34	0.80	0.46	0.43	0.32	0.42	0.33		
STD	0.17	0.18	0.23	0.21	0.19	0.20	0.22	0.21	0.24	0.18		

and referents of 0.430 (SD = 0.19), which the Agreement Analysis Toolkit classifies as high agreement (between 0.3 and 0.5) [69].

4.3 Set of Gesture Types

Observed gestures were categorized according to the characteristics described in the object-based gesture taxonomy (Table 2). Figure 3 illustrates the main set of 14 gestures that occurred more than 5% of the time for at least one referent. Only 1.2% of observed gestures did not fall within this set, such as Bending (0.3%) and Rotate Chair (0.1%). Gestures were named with terms that described the action while also adopting existing names when possible. The total number of times a gesture was observed, and the total number of participants that produced each gesture, are also provided in Figure 3. It should be noted that certain gestures were observed for multiple referents, while others were not. This distribution is described in more detail in Section 4.4.

4.4 Gesture Distributions and Preferences

While the analysis of agreement rates demonstrated a high level of consistency across participants, an analysis of the actual gestures performed is still needed. To determine the most appropriate gesture for each referent, we grouped all the identical gestures for each referent. The resulting distribution of gestures across referents is illustrated in Figure 4. The flight simulator referents are omitted from Figure 4, as participants were asked to perform a combination of gestures. They will be discussed separately below (Figure 5).

In Figure 4, we highlight the most popular gesture for each referent, which we classify as the primary gesture for this referent. The identification of these primary gestures was validated by the qualitative data i.e., for each referent, all gestures with the highest representation were also selected as "the preferred gesture" by all participants.

In contrast to other user-defined gesture sets, we cannot rely on a single gesture type (with the highest occurrence level) as the only representation for each referent [60, 75]. This is because opportunistic objects would not always afford the most popular gesture for a specific input command. For example, a user may not be able to find a grabbable object needed to rotate a virtual object for Object Rotate (i.e., the most ideal gesture for this task). As such, we also highlight the second most popular, or secondary, gesture for each referent. This secondary gesture, a backup per se, could be used in

Figure 3: The user-defined gesture set for spatial tangible interactions. In parenthesis is the number of times the gestures were observed and the number of participants that produced the gesture at least once, across all objects and referents.

cases where the primary gesture is not possible with the objects currently in the user's environment. Below we highlight some key insights gained from these gesture preferences.

Single-handed gestures were dominant, but more bimanual gestures were proposed for 3D and compound commands. Overall, 87% of the observed gestures used one hand, while 13% of proposed gestures were performed bimanually. Bimanual gestures were observed more often for three referents: *Move 3D Object (21% bimanual), Rotate 3D Object (22.1%) and Drive Airplane (45%)* – all of which had 3D scene dimensionality (Table 1). Object Move and Object Rotate were the most common gesture types to be performed bimanually (22% and 52%, respectively). This pattern relates to literature that has shown the appropriateness of bimanual input for 3D tasks [5].

Participants' gestures were influenced by their previous experiences. Our work confirms a key advantage of TUIs – the ability of users to draw on previous experiences with metaphorical links [23]. For referents with a 2D scene dimensionality, the most common gestures, such as Finger Tap, Finger Swipe, and Finger Slide, were similar to gestures commonly used with digital devices such as smartphones and laptops (*Digital Metaphorical Link*). For referents with a 3D scene dimensionality, participants applied their experiences manipulating physical objects, such as Object Rotate and Object Move (*Physical Metaphorical Link*). Examples were imagining an object to be a physical representation of a digital object (e.g., mapping the movement of the physical object to the digital one), or imagining the object was a known physical controller (e.g., holding a chocolate bar like a steering wheel to control the direction of an airplane).

Participants attempted to provide congruent spatial relationship mappings. In scene 3, participants were asked to create gestures to *increase/decrease* the value on a slider. There were two types of sliders, one vertical and one horizontal. Overall, 87% of participants proposed the *Finger Slide* gesture that aligned with the direction of the sliders on the screen. This demonstrates the importance of supporting congruent spatial relationship mappings [45] in opportunistic TUIs.

Users applied different strategies when performing compound gestures. For the flight simulate scene, controlling *acceleration* and *direction* were a combination of the input commands *increase/decrease* and *rotate 3D object,* except they had to be performed simultaneously. While users' proposed gestures for *control direction* and *rotate 3D object* were similar, there was a noticeable difference between the results of *acceleration* and *increase/decrease.* For *increase/decrease,* Finger Slide (52.6%) and Object Slide (35.3%) were used over 85% of the time. However, for *acceleration*, Squeeze (24.8%) and Object Move (24.2%) were also used as they were easier to perform while rotating an object, whereas Object Slide was not used at all as it conflicted with Object Rotate (Figure 5).

4.5 Observed Affordance Factors

Once we understood which gestures were used and their distribution across referents, we examined the affordance factors that influenced participants' choice of gestures.

4.5.1 Taxonomy of Object Affordance Factors. During the user study, participants were asked to explain why they proposed each gesture and what properties of the object influenced their choice of gesture (e.g., a flat surface or its grabbable size). These object characteristics were identified as affordance factors. We analyzed the recorded video data and compiled a list of affordance factors using the factors mentioned in the think-aloud data. The affordance factors were labelled and grouped into nine factors (Figure 6).

These nine affordance factors were then divided into four categories, including geometric, kinetic, semantic and ergonomic. The semantic factors were further sub-divided into three categories: Sharp Tip and Button Like, which were frequently observed across several gestures, and Metaphorical, which represented other semantic factors that did not exceed 5% representation for any individual referent (e.g., Steering Wheel-Like, Handle-Like,

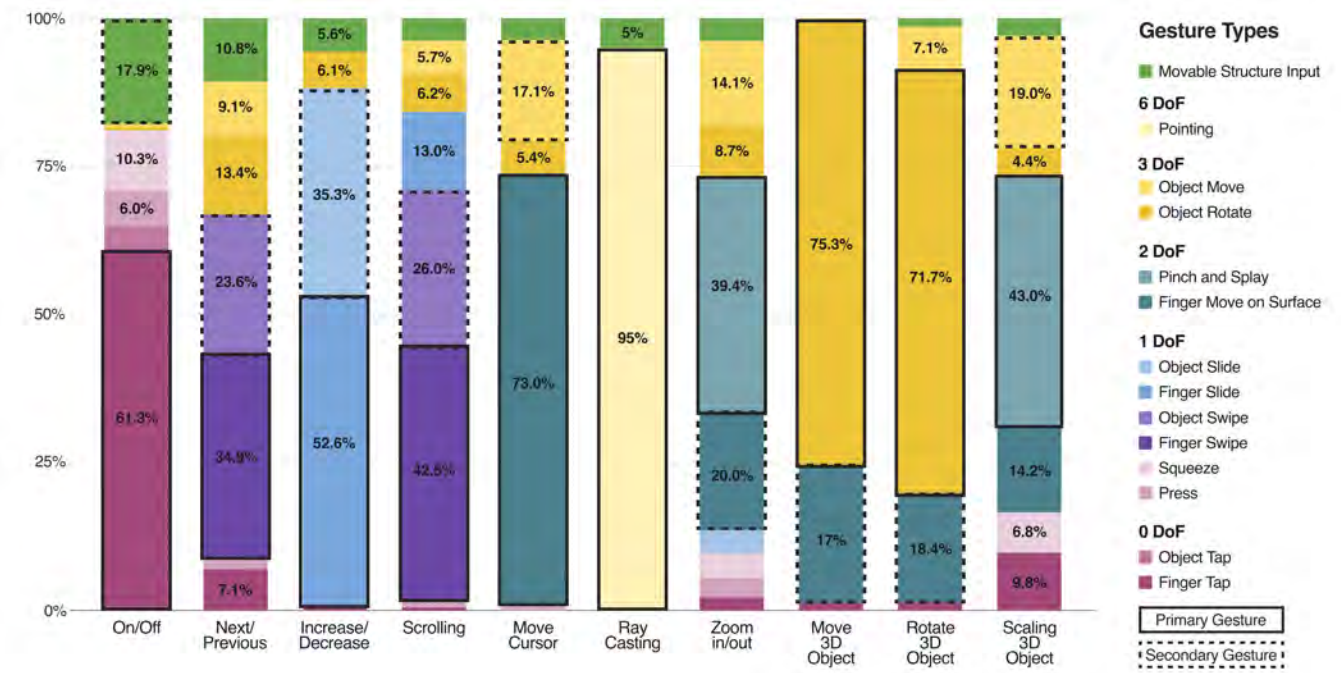

Figure 4: The distribution of gestures with the primary and secondary gestures for each referent highlighted. The flight simulator referents are omitted from this figure, as participants were asked to perform a combination of gestures. They are discussed in Figure 5.

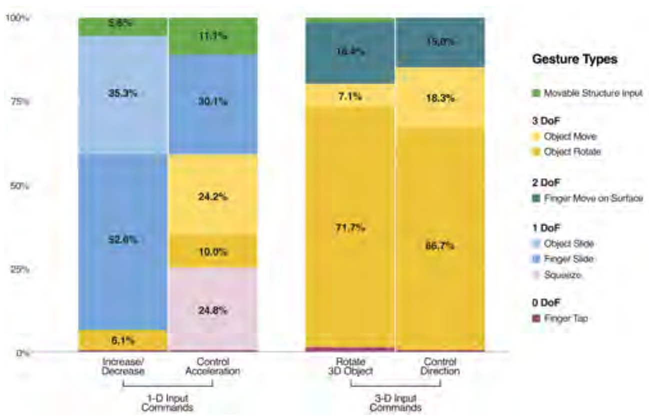

Figure 5: A comparison of gesture selections for equivalent referents when elicited individually vs in combination. *Increase/decrease* and *Rotate 3D Object* were performed individually, while *Control Acceleration* and *Direction* were performed simultaneously.

Knob-Like). Some of these affordance factors (e.g., size and curvature for surface) closely match the initial potential affordance factors, which are criteria used to select the objects for the study (Figure 1), while some other factors, such as geometric shape of the entire object body (e.g., cylinder or pyramid), were found not influence participants' chosen gestures. To some extent this validates the decision to use real work objects for the study and not abstract, geometric shapes.

4.5.2 Affordance Factors Across Gestures. The distribution of affordance factors showed a high alignment among participants, as the top affordance factor for each gesture had more than 50% representation, which we consider to be the primary affordance factor (Figure 7). We also identify secondary affordance factors for instances where the primary affordance factor may not be possessed by nearby objects. Objects that possess both the primary and secondary affordance factors could potentially be even more suitable for the associated gesture. Pointing was the only gesture that had no secondary affordance factor. Below we discuss key insights on how the affordance factors influenced gesture preferences.

Hand-centric gestures were surface-based. All hand-centric gestures, including *Finger Tap*, *Finger Swipe*, *Finger Slide, Pinch & Splay,* and *Finger Move on Surface*, had *Surface* as their primary affordance factor. Some participants reported that they utilized their existing experience interacting with smartphones to create these hand-centric gestures.

Semantic characteristics played an important role for hand-centric gestures. As reported in prior work, metaphorical links played an important role in influencing users' gestures [23]. For example, participants were drawn to *Button-like* areas of objects when performing *Finger Tap* (12.4%). Similarly, *Long Edge* was a secondary affordance factor for *Finger Slide* (83%), as participants reported that "*the shape of a long edge looks like a slider*".

Influence of visual guidance. While the size of a surface was not a strong influence, 8.3% of participants wanted the surface to be "small or constrained" to perform a *Finger Tap* (such as the small surfaces on the sides of the chocolate bar or the center of rings on a yoga ball surface). This may indicate that visual guidance or a signifier can be useful to inform the user where to perform a gesture.

Influence of metaphors. Semantic affordance factors were commonly identified, appearing as a primary or secondary gesture for 8 of the 14 gestures, often due to metaphorical links [23]. For example, *Object Swipe*, *Object*

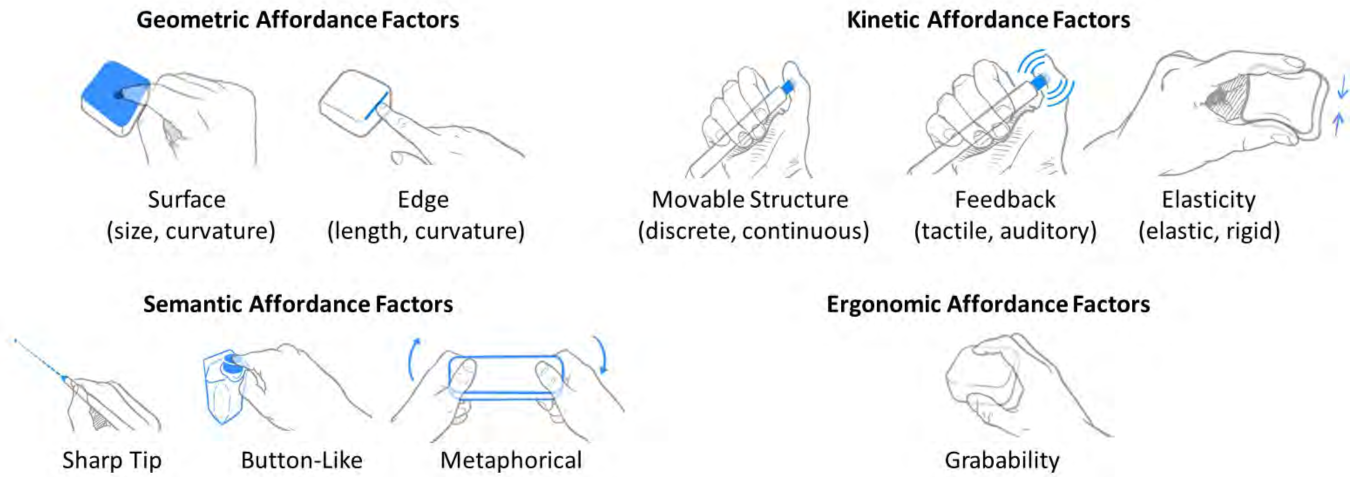

Figure 6: The taxonomy of object affordance factors and their related variables, grouped by category.

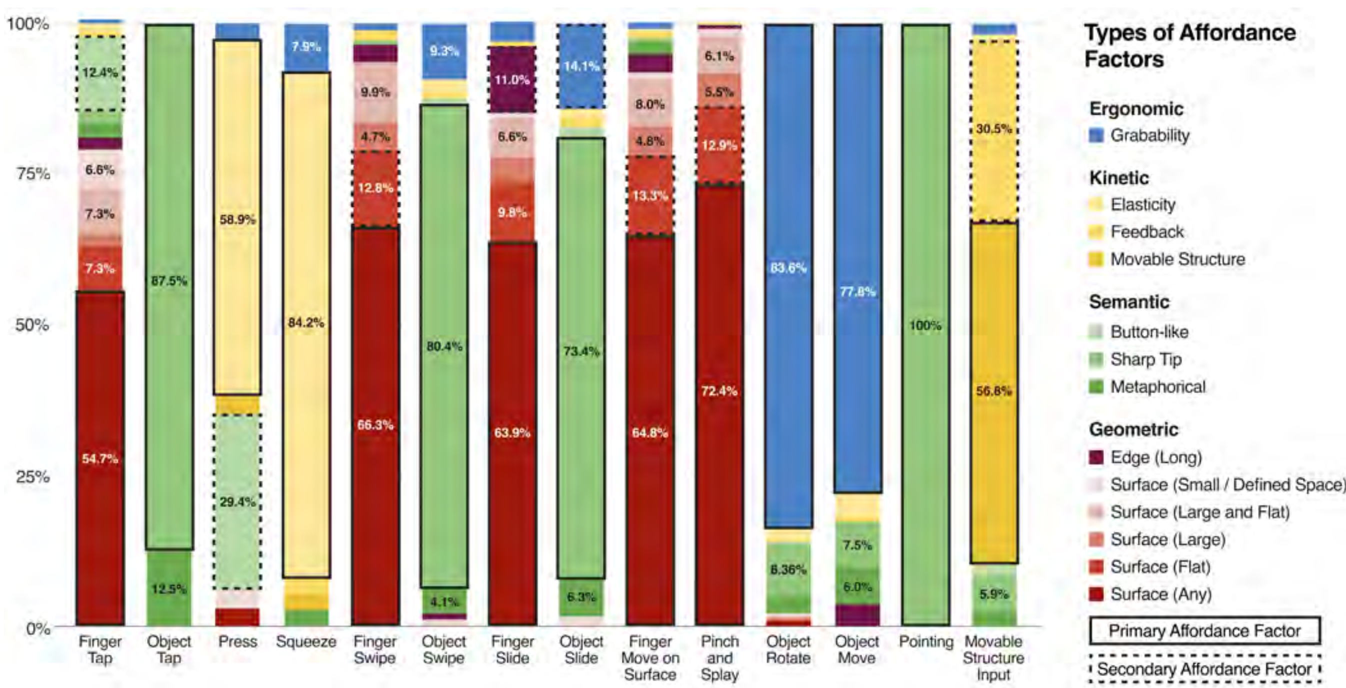

Figure 7: Percentage of affordance factors in each gesture category. Primary and secondary affordance are highlighted for each gesture.

Slide, and *Pointing* all had *Semantics-Sharp Tip* as the primary affordance factor. Participants reported that the sharp tip of an object was often used as a "pointer" in daily life. One example is that teachers often use the tip of a marker to point at a whiteboard. Some objects with a sharp tip, such as a stylus, can also be used as an input device for sliding and swiping on touchscreens.

Deformation and tactile feedback led to gesture preferences. Both *Squeeze* and *Press* had *Elasticity* as their primary affordance factor. Although participants could perform these gestures using rigid objects, most participants (84.2% and 58.9%, respectively for *Squeeze* and *Press*) preferred to have an elastic object due to its deformation and tactile feedback.

Table 4: The Affordance-Based Object Selector Scheme identifies object affordance factors for the 14 gesture types. Secondary affordance factors are shown in square brackets. Thresholds for geometric measurements are provided at the bottom of the table.

Gesture	Primary and [Secondary] Affordance Factors	Example Object
Finger Tap	Objects with a surface [and button-like part]	Top surface of a bottle's cap
Object Tap	Objects with a sharp tip [and metaphorical link]	A pen
Press	Objects with elastic surface [button-like part]	A water bottle's side surface or its cap
Squeeze	[Grabbable] Objects with an elastic surface	A water bottle's side surface
Finger Swipe	Objects with a [flat] surface	A desk
Object Swipe	[Grabbable] Objects with a sharp tip	A pen
Finger Slide	Objects with a surface [and a long edge]	A desk
Object Slide	[Grabbable] Objects with a sharp tip	A pen
Pinch and Splay	Objects with a [flat] surface	A desk
Finger Move on Surface	Objects with a [flat] surface	A desk
Object Rotate	Grabbable objects [with a sharp tip]	An eraser
Object Move	Grabbable objects [with a sharp tip]	An eraser
Pointing	Objects with a pointer-like tip	A pen
Movable Structure Input	Objects with a movable structure [providing feedback]	The button on the back of a pen

Surface Size: Large (> 300 cm^2), Medium (12-300 cm^2), Small (< 12 cm^2)
Edge Length: Long (> 10cm), Short ($\leq$ 10cm)
Curvature: Flat (Radius > 35cm), Curved (Radius $\leq$ 35cm)

4.6 Affordance-Based Object Selector Scheme

Based on the above analysis of affordance factors, we derived a lookup table that can determine suitable objects for each gesture type (Table 4). We define this table as an *Affordance-Based Object Selector Scheme*, as it selects objects based on their affordance factors. For each gesture, we list the associated primary object affordance factor (Figure 7) and an example object. We also include the secondary affordance factors, which indicate additional preferences. For example, for *Finger Tap*, while some participants did not discern the shape of the object, some preferred to tap on a button-like region. These secondary affordance factors are marked with square brackets in the table. By developing this affordance-based object selector scheme, our results can be generalized to arbitrary objects, by considering their associated affordance factors.

5 DESIGN IMPLICATIONS: ADAPTIVE TANGIBLE USER INTERFACES

From our analysis, we developed a user-defined gesture set for different types of spatial interface input that demonstrated a high level of agreement. Also, we created the Affordance-Based Object Selector Scheme to identify and prioritize opportunistic objects for each gesture in the gesture set. We now have the foundation to design an adaptive system that could be applicable in different physical environments for different spatial interaction tasks.

We now propose our vision of Adaptive Tangible User Interfaces (ATUI) which could repurpose the physical environment around a user to support opportunistic tangible input. This vision is similar in spirit to the concept of *Ad hoc UI [22]*but would intelligently map UIs to objects based on their affordance factors. ATUIs would need to proactively scan, interpret, and identify objects within one's physical environment that have the affordance factors needed for users to accomplish their goals. Imagine if one wants to turn up the volume of music they are listening to while using AR glasses – in our vision, they could simply rotate a nearby cup using their hand. In another example, if one finds that they are running low on milk, they could squeeze a milk jug to add milk to their shopping list. Several challenges must be overcome to realize this vision, which we now discuss.

5.1 Object and Affordance Factor Detection

The ability to process users' surroundings and recognize objects and their affordances is fundamental to any ATUI system. This would involve proactively scanning the environment when needed and applying the Affordance-Based Object Selector Scheme (Table 4), to form a dynamic list of objects in the environment that can afford the required gestures. Such a system would likely apply computer vision techniques and require the system to recognize and label both the objects (e.g., an elastic yoga ball) and the geometry of the objects (e.g., a large and flat surface on a desk) to form a list of all affordance factors in the environment. Past work on obtaining 3D meshes representing the geometry of the physical environment, could be of use [15].

5.2 TUI Composition

Such a system also needs to determine desirable actions based on the user's current interaction goals and compose the associated TUI in real-time. With this TUI composition capability, a system could recognize potential input tasks based on the existing spatial UI (e.g., one opens a 3D model in AR with the potential input of rotating, moving and scaling) and produce a list of suitable gestures for the input tasks based on our gesture distribution results (Figure 4). The system would then locate suitable nearby objects that afford these gestures, using our Affordance-Based Object Selector Scheme (Table 4). Once the TUI is composed, visual prompts would be rendered on top of the object to provide feedback to the user. The user could then activate the object as a controller and begin the interaction.

5.3 Hand and Object Tracking and TUI Gesture Recognition

The ATUI system we envision would require both object recognition and hand tracking to be interactive. Existing research has explored using computer vision to recognize and track the status of an object and hand gestures [48]. Wrist devices can also be used to track hand gestures that are outside the camera's field of view [59]. The system may also need a way to sense object deformations to support gestures like squeeze. Existing research on object-constrained gestures could be utilized to help address these challenges [64].

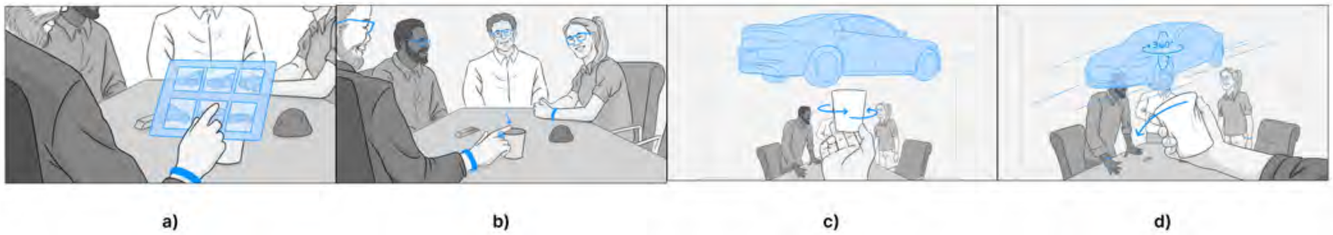

Figure 8: Mockup of the ATUI vision. a) Tom loads a 3D hologram of a car. b) An indicator is displayed on the cup to activate the ATUI system. c) Tom rotates the cup to steer the car. d) Tom squeezes the cup to drive the car.

5.4 Example Use Case - Virtual Object Manipulation with A Paper Cup

We now describe an example use case of an adaptable tangible user interface. We first discuss the user experience, followed by the underlying system behavior.

5.4.1 User Flow. A sample user flow is illustrated in Figure 8. Tom receives a car design proposal sent from his colleague. He wants to check the 1:1 3D hologram in his AR glasses, but it is hard to control such a large hologram with in-air hand gestures.

The system displays an indicator on a paper cup so Tom realizes the cup can be used as a controller. He double-taps the top of the cup to activate it in the ATUI system. Some additional visual signifiers then appear, indicating how he can interact with the cup. Tom holds the cup and rotates it, and the 1:1 hologram of the car rotates with it. When Tom squeezes the cup, the hologram of the car starts to drive, and Tom can rotate the cup to control its direction.

5.4.2 Behind the Scenes. When Tom opens the car design hologram, the system captures the input tasks required by the application. In this case, the potential input tasks are: 1) Rotate the hologram, 2) Move the hologram, 3) Scale the hologram, 4) Control the acceleration of the car, and 5) Exit. Based on the understanding of the preferred gesture distributions (Figure 4), the system identifies that the desirable object-based gestures for these tasks are: 1) Object Rotate, 2) Object Move, 3) Finger Slide, 4) Squeeze, and 5) Finger Tap. Using the Affordance-Based Object Selector Scheme (Table 4), the system determines that the preferred object to accommodate those gestures should: 1) be grabbable, 2) have a flat surface, and 3) have an elastic body. With its affordance detection capability, the system scans the environment and identifies a paper cup on the desk that has all the affordance factors needed. The system then renders the visual signifiers on the cup and prepares the cup to be activated. Object and hand tracking technologies are then used to track the user's gestures.

6 DISCUSSION, LIMITATIONS, AND FUTURE WORK

A key motivation for this research was that prior Opportunistic TUIs for spatial interfaces tend to focus on designing interfaces for specific use cases and objects, so the outcomes are hard to scale to other objects and environments. As such, the generalizability of this research is important to discuss, and, in particular, our choice of referents and objects used in the study.

By choosing an object set for our study that represented a spectrum of characteristics, all informed by prior research, our hope was that the results would be applicable to not just the objects themselves, but to other objects with similar affordance factors. While future studies could test this hypothesis, the think-aloud feedback did indicate user's gestures were guided by the affordance factors of the objects rather than the specific objects themselves. This provides some level of confidence that the results would scale to other objects with similar characteristics. We also chose a wide range of referents for our study representing both 2D and 3D input tasks that are commonly used in spatial interfaces. While certainly not exhaustive, we believe the insights gained from our study of this set of 12 referents should be sufficient to infer appropriate gestures for many spatial UI actions. Finally, the development of the Affordance-Based Object Selector Scheme (Table 4) should, itself, allow for the scalability of our results. This scheme should help designers identify and prioritize suitable objects for input tasks based on the objects' affordance factors. Furthermore, it should enable the concept of an ATUI, where the results could be utilized within a scalable system that composes an adaptive, on-the-fly TUI based on the task and objects in an environment. While we hypothesize these contributions may help with the scalability of our research, future work is needed to further validate the generalizability of the results, and a true implementation of the ATUI vision may be the best test of scalability. In order to demonstrate the scalability of our results, we considered several examples to reflect how the affordance-based object selector can be applied to objects that were not included in the user study. For example, rather than utilizing the paper cup, a soda can or an apple can be employed for virtual object manipulation mentioned in section 5.4. The *Finger Tap* gesture can be performed on both the soda can and the apple using a surface, which is the primary affordance factor for this gesture. Similarly, both objects can be rotated as they are both capable of being grasped, making them suitable for the *Object Rotate* gesture.

The results of our user study have yielded some interesting and innovative findings with regards to the gesture set. We introduced a total of 14 gesture types that can be applied across tasks involving 0-6 Degrees of Freedom (DoF). Although certain gesture types for 2D tasks, such as Finger Slide and Finger Tap, exhibit similarities with results existing elicitation studies for traditional 2D touch interfaces [75], a substantial number of scalable novel gestures were discovered for 3D input or using object properties such as elasticity or movable structures, which have not been extensively explored in the literature.

There are several other topics that warrant further research. One limitation of this research is that we didn't compare the usability of the object-based gestures with mid-air hand gestures. It is possible that in some cases object-based gestures have advantages over mid-air hand gestures, while in other cases they may not. For example, users could feel that it is easier to pinch their fingers to confirm an action instead of tapping on an object when they are not already holding the object. A better understanding of these trade-offs could be used within a dynamic UI that combines mid-air and object-centric gestures. In addition, we did not consider the impact of objects' spatial locations. The distance an object is to a user could be used to influence the composition of a TUI. Prior research on reachability could also provide a guiding model [29, 63].

Additionally, our study used a traditional desktop display to provide feedback to the user to ensure hardware limitations would not constrain the gestures that users performed. We adopted this approach of using desktop display for 3D gesture user study because its ability to simulate 3D tasks successfully and produce highly transferable results to 3D environments [43]. Moreover, all referents within 0 – 2 dimensions normally appear as 2D floating windows in AR/VR, which is analogous to a desktop display. Our referents under higher dimensions were also less dependent on stereoscopic vision as all referents doesn't require world-lock 3D content and participants were in a stationary position. That said, future work should determine if there are any substantial differences in preferences when utilizing a head-worn display.

Moreover, participants of our user study were recruited in major north American cities. Future work should also consider a more diverse culture and geographical background to explore how culture or socioeconomic status may impact gesture preferences, especially given the influence which culture may have on the concept of affordance [36] and metaphor [23].

Finally, to further increase the scalability of our research, quantitative data about how the parameters of an affordance factor (e.g., length, curvature) impact performance data (e.g., performance time, comfort, accuracy), could be collected and utilized by the ATUI.

7 CONCLUSION

This paper presented a study of opportunistic tangible user interface gestures and object affordance factors that led to a user-defined object-based spatial input gesture set and an Affordance-Based Object Selector Scheme. The taxonomy and gesture set were based on the 2207 gestures and affordance factors proposed by 22 participants in an elicitation study. This work led to a systematic understanding of the desirable object-based gestures for different input tasks in spatial computing, the characteristics an object needed to afford those gestures, and insights into how metaphorical links influenced users' gesture preferences. As a direct implication of the findings, we discussed requirements for an Adaptive Tangible User Interface, which aims to build a system that composes a desirable tangible UI on-the-fly based on the affordances in the physical environment and a user's input task.

REFERENCES

[1] Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindlbauer, Alexandra Ion, and Shengdong Zhao. Understanding Mid-Air Hand Gestures: A Study of Human Preferences in Usage of Gesture Types for HCI. 10.

[2] 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), 1–8. https://doi.org/10.1145/2667317.2667414

[3] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), 5643–5654. https://doi.org/10.1145/3025453.3025474

[4] Klaus B. Bærentsen and Johan Trettvik. 2002. An activity theory approach to affordance. In Proceedings of the second Nordic conference on Human-computer interaction (NordiCHI '02), 51–60. https://doi.org/10.1145/572020.572028

[5] Ravin Balakrishnan. 2001. Issues in bimanual interaction for computer graphics. University of Toronto, CAN.

[6] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique based on Raycasting. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), 1–12. https://doi.org/10.1145/3290605.3300331

[7] Hrvoje Benko, Eyal Ofek, Feng Zheng, and Andrew D. Wilson. 2015. FoveAR: Combining an Optically See-Through Near-Eye Display with Projector-Based Spatial Augmented Reality. In Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15), 129–135. https://doi.org/10.1145/2807442.2807493

[8] Lonni Besançon, Paul Issartel, Mehdi Ammi, and Tobias Isenberg. 2017. Mouse, Tactile, and Tangible Input for 3D Manipulation. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), 4727–4740. https://doi.org/10.1145/3025453.3025863

[9] Elodie Bouzbib, Gilles Bailly, Sinan Haliyo, and Pascal Frey. 2022. "Can I Touch This?": Survey of Virtual Reality Interactions via Haptic Solutions: Revue de Littérature des Interactions en Réalité Virtuelle par le biais de Solutions Haptiques. In Proceedings of the 32nd Conference on l'Interaction Homme-Machine (IHM '21), 1–16. https://doi.org/10.1145/3450522.3451323

[10] D.A. Bowman and C.A. Wingrave. 2001. Design and evaluation of menu systems for immersive virtual environments. In Proceedings IEEE Virtual Reality 2001, 149–156. https://doi.org/10.1109/VR.2001.913781

[11] Evren Bozgeyikli and Lal Lila Bozgeyikli. 2021. Evaluating Object Manipulation Interaction Techniques in Mixed Reality: Tangible User Interfaces and Gesture. In 2021 IEEE Virtual Reality and 3D User Interfaces (VR), 778–787. https://doi.org/10.1109/VR50410.2021.00105

[12] William Buxton. 1983. Lexical and pragmatic considerations of input structures. ACM SIGGRAPH Computer Graphics 17, 1: 31–37. https://doi.org/10.1145/988584.988586

[13] William Buxton. 2002. Human input to computer systems: theories. Tech. Technol. 6: 1–13.

[14] Jorge Cardoso and Jorge Ribeiro. 2020. VR Book: A Tangible Interface for Smartphone-based Virtual Reality. In MobiQuitous 2020 - 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous '20), 48–58. https://doi.org/10.1145/3448891.3448901

[15] Hyunwoo Cho, Sangheon Park, Chanho Park, and Sung-Uk Jung. 2021. Efficient Mapping Technique under Various Spatial Changes for SLAM-based AR Services. In Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (VRST '21), 1–3. https://doi.org/10.1145/3489849.3489916

[16] Marcelo Coelho and Jamie Zigelbaum. 2011. Shape-changing interfaces. Personal and Ubiquitous Computing 15, 2: 161–173. https://doi.org/10.1007/s00779-010-0311-y

[17] Sabrina Connell, Pei-Yi Kuo, Liu Liu, and Anne Marie Piper. 2013. A Wizard-of-Oz elicitation study examining child-defined gestures with a whole-body interface. In Proceedings of the 12th International Conference on Interaction Design and Children (IDC '13), 277–280. https://doi.org/10.1145/2485760.2485823

[18] Juliet M. Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. Qualitative Sociology 13, 1: 3–21. https://doi.org/10.1007/BF00988593

[19] 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), 71–80. https://doi.org/10.1145/2512349.2512799

[20] Tanja Döring, Dagmar Kern, Paul Marshall, Max Pfeiffer, Johannes Schöning, Volker Gruhn, and Albrecht Schmidt. 2011. Gestural interaction on the steering wheel: reducing the visual demand. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11), 483–492. https://doi.org/10.1145/1978942.1979010

[21] Adam Drogemuller, James Walsh, Ross T. Smith, Matt Adcock, and Bruce H Thomas. 2021. Turning everyday objects into passive tangible controllers. In Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21), 1–4. https://doi.org/10.1145/3430524.3442460

[22] Ruofei Du, Alex Olwal, Mathieu Le Goc, Shengzhi Wu, Danhang Tang, Yinda Zhang, Jun Zhang, David Joseph Tan, Federico Tombari, and David Kim. 2022. Opportunistic Interfaces for Augmented Reality: Transforming Everyday Objects into Tangible 6DoF Interfaces Using Ad hoc UI. In Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22), 1–4. https://doi.org/10.1145/3491101.3519911

[23] Kenneth P. Fishkin. 2004. A taxonomy for and analysis of tangible interfaces. Personal and Ubiquitous Computing 8, 5: 347–358.

[24] William W. Gaver. 1991. Technology affordances. In Proceedings of the SIGCHI conference on Human factors in computing systems Reaching through technology - CHI '91, 79–84. https://doi.org/10.1145/108844.108856

[25] James J. Gibson. 2014. The Ecological Approach to Visual Perception: Classic Edition. Psychology Press.

[26] Michael D. Good, John A. Whiteside, Dennis R. Wixon, and Sandra J. Jones. 1984. Building a user-derived interface. Communications of the ACM 27, 10: 1032–1043. https://doi.org/10.1145/358274.358284

[27] Michael D. Good, John A. Whiteside, Dennis R. Wixon, and Sandra J. Jones. 1984. Building a user-derived interface. Communications of the ACM 27, 10: 1032–1043. https://doi.org/10.1145/358274.358284

[28] Daniel Harley, Aneesh P. Tarun, Daniel Germinario, and Ali Mazalek. 2017. Tangible VR: Diegetic Tangible Objects for Virtual Reality Narratives. In Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17), 1253–1263. https://doi.org/10.1145/3064663.3064680

[29] Alan Hedge. 2013. Anthropometry and Workspace Design. 8.

[30] 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), 211–218. https://doi.org/10.1145/1450579.1450625

[31] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), 1957–1967. https://doi.org/10.1145/2858036.2858134

[32] Valentin Heun, Shunichi Kasahara, and Pattie Maes. 2013. Smarter objects: using AR technology to program physical objects and their interactions. In CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13), 961–966. https://doi.org/10.1145/2468356.2468528

[33] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. A survey of design issues in spatial input. In Proceedings of the 7th annual ACM symposium on User interface software and technology (UIST '94), 213–222. https://doi.org/10.1145/192426.192501

[34] Hiroshi Ishii. 2008. Tangible bits: beyond pixels. In Proceedings of the 2nd international conference on Tangible and embedded interaction (TEI '08), xv–xxv. https://doi.org/10.1145/1347390.1347392

[35] Robert J. K. Jacob, Linda E. Sibert, Daniel C. McFarlane, and M. Preston Mullen. 1994. Integrality and separability of input devices. ACM Transactions on Computer-Human Interaction 1, 1: 3–26. https://doi.org/10.1145/174630.174631

[36] Victor Kaptelinin and Bonnie Nardi. 2012. Affordances in HCI: toward a mediated action perspective. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), 967–976. https://doi.org/10.1145/2207676.2208541

[37] Maria Karam and m.c Schraefel. 2005. A Taxonomy of Gestures in Human Computer Interactions. Electronics and Computer Science.

[38] Annie Kelly, R. Benjamin Shapiro, Jonathan de Halleux, and Thomas Ball. 2018. ARcadia: A Rapid Prototyping Platform for Real-time Tangible Interfaces. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), 1–8. https://doi.org/10.1145/3173574.3173983

[39] Mohamed Khamis, Ludwig Trotter, Ville Mäkelä, Emanuel von Zezschwitz, Jens Le, Andreas Bulling, and Florian Alt. 2018. CueAuth: Comparing Touch, Mid-Air Gestures, and Gaze for Cue-based Authentication on Situated Displays. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, 4: 1–22. https://doi.org/10.1145/3287052

[40] Hyunyoung Kim, Celine Coutrix, and Anne Roudaut. 2018. Morphees+: Studying Everyday Reconfigurable Objects for the Design and Taxonomy of Reconfigurable UIs. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), 1–14. https://doi.org/10.1145/3173574.3174193

[41] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), 1–9. https://doi.org/10.1145/3173574.3173919

[42] Joseph J. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan Poupyrev. 2017. 3D User Interfaces: Theory and Practice. Addison-Wesley.

[43] Minkyung Lee and Mark Billinghurst. 2008. A Wizard of Oz study for an AR multimodal interface. In Proceedings of the 10th international conference on Multimodal interfaces (ICMI '08), 249–256. https://doi.org/10.1145/1452392.1452444

[44] I. Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. Human-Computer Interaction 7, 1: 91–139. https://doi.org/10.1207/s15327051hci0701_3

[45] I. Scott MacKenzie. 2013. Human-Computer Interaction: An Empirical Research Perspective. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[46] David Maulsby, Saul Greenberg, and Richard Mander. 1993. Prototyping an intelligent agent through Wizard of Oz. In Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93, 277–284. https://doi.org/10.1145/169059.169215

[47] David Mcneill. 1994. Hand and Mind: What Gestures Reveal About Thought. Bibliovault OAI Repository, the University of Chicago Press 27. https://doi.org/10.2307/1576015

[48] Stan Melax, Leonid Keselman, and Sterling Orsten. 2013. Dynamics based 3D skeletal hand tracking. In Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D '13), 184. https://doi.org/10.1145/2448196.2448232

[49] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. Interactions 21, 3: 40–45. https://doi.org/10.1145/2591689

[50] Marvin Everett Mundel. 1978. Motion and Time Study: Improving Productivity. Prentice-Hall.

[51] Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. 2004. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In Gesture-Based Communication in Human-Computer Interaction (Lecture Notes in Computer Science), 409–420. https://doi.org/10.1007/978-3-540-24598-8_38

[52] Donald A. Norman. 1988. The Psychology of Everyday Things. Basic Books.

[53] Payod Panda, Charles Ho, and Derek Ham. 2021. Morphaces: Exploring Morphable Surfaces for Tangible Sketching in VR. In Creativity and Cognition (C&C '21), 1–11. https://doi.org/10.1145/3450741.3465387

[54] Stephen Pheasant and Christine M. Haslegrave. 2005. Bodyspace: Anthropometry, Ergonomics and the Design of Work, Third Edition. CRC Press.

[55] Philip Quinn, Seungyon Claire Lee, Melissa Barnhart, and Shumin Zhai. 2019. Active Edge: Designing Squeeze Gestures for the Google Pixel 2. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), 1–13. https://doi.org/10.1145/3290605.3300504

[56] Majken K. Rasmussen, Esben W. Pedersen, Marianne G. Petersen, and Kasper Hornbæk. 2012. Shape-changing interfaces: a review of the design space and open research questions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 735–744. https://doi.org/10.1145/2207676.2207781

[57] Sandrine Robbe. 1998. An empirical study of speech and gesture interaction: toward the definition of ergonomic design guidelines. In CHI 98 Conference Summary on Human Factors in Computing Systems (CHI '98), 349–350. https://doi.org/10.1145/286498.286815

[58] Anne Roudaut, Abhijit Karnik, Markus Löchtefeld, and Sriram Subramanian. 2013. Morphees: toward high "shape resolution" in self-actuated flexible mobile devices. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 593–602. https://doi.org/10.1145/2470654.2470738

[59] Julius Cosmo Romeo Rudolph, David Holman, Bruno De Araujo, Ricardo Jota, Daniel Wigdor, and Valkyrie Savage. 2022. Sensing Hand Interactions with Everyday Objects by Profiling Wrist Topography. In Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '22), 1–14. https://doi.org/10.1145/3490149.3501320

[60] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11), 197–206. https://doi.org/10.1145/1978942.1978971

[61] Kadek Ananta Satriadi, Jim Smiley, Barrett Ens, Maxime Cordeil, Tobias Czauderna, Benjamin Lee, Ying Yang, Tim Dwyer, and Bernhard Jenny. 2022. Tangible Globes for Data Visualisation in Augmented Reality. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), 1–16. https://doi.org/10.1145/3491102.3517715

[62] Daniel Schneider, Verena Biener, Alexander Otte, Travis Gesslein, Philipp Gagel, Cuauhtli Campos, Klen Čopič Pucihar, Matjazz Kljun, Eyal Ofek, Michel Pahud, Per Ola Kristensson, and Jens Grubert. 2021. Accuracy Evaluation of Touch Tasks in Commodity Virtual and Augmented Reality Head-Mounted Displays. In Symposium on Spatial User Interaction (SUI '21), 1–11. https://doi.org/10.1145/3485279.3485283

[63] Stacey Scott. 2003. Territory-based interaction techniques for tabletop collaboration.

[64] 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), 1–15. https://doi.org/10.1145/3411764.3445197

[65] Adwait Sharma, Joan Sol Roo, and Jürgen Steimle. 2019. Grasping Microgestures: Eliciting Single-hand Microgestures for Handheld Objects. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3290605.3300632

[66] Jennifer G. Sheridan and Gerd Kortuem. 2006. Affordance-Based design of physical interfaces for ubiquitous environments. In Proceedings of the Third international conference on Ubiquitous Computing Systems (UCS'06), 183–199. https://doi.org/10.1007/11890348_15

[67] Aaron Steinfeld, Odest Chadwicke Jenkins, and Brian Scassellati. 2009. The oz of wizard: simulating the human for interaction research. In Proceedings of the 4th ACM/IEEE international conference on Human robot interaction - HRI '09, 101. https://doi.org/10.1145/1514095.1514115

[68] Ching-Yi Tsai, I-Lun Tsai, Chao-Jung Lai, Derrek Chow, Lauren Wei, Lung-Pan Cheng, and Mike Y. Chen. 2022. AirRacket: Perceptual Design of Ungrounded, Directional Force Feedback to Improve Virtual Racket Sports Experiences. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), 1–15. https://doi.org/10.1145/3491102.3502034

[69] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), 1325–1334. https://doi.org/10.1145/2702123.2702223

[70] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies? In Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20), 855–872. https://doi.org/10.1145/3357236.3395511

[71] Stephen Voida, Mark Podlaseck, Rick Kjeldsen, and Claudio Pinhanez. 2005. A study on the manipulation of 2D objects in a projector/camera-based augmented reality environment. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05), 611–620. https://doi.org/10.1145/1054972.1055056

[72] Philipp Wacker, Oliver Nowak, Simon Voelker, and Jan Borchers. 2019. ARPen: Mid-Air Object Manipulation Techniques for a Bimanual AR System with Pen & Smartphone. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), 1–12. https://doi.org/10.1145/3290605.3300849

[73] Koki Watanabe, Fumihiko Nakamura, Kuniharu Sakurada, Theophilus Teo, and Maki Sugimoto. 2021. X-Wing: Propeller-Based Force Feedback to Head in a Virtual Environment. In SIGGRAPH Asia 2021 XR (SA '21 XR), 1–2. https://doi.org/10.1145/3478514.3487622

[74] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the guessability of symbolic input. In CHI '05 Extended Abstracts

on Human Factors in Computing Systems (CHI EA '05), 1869–1872. https://doi.org/10.1145/1056808.1057043

[75] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1083–1092. https://doi.org/10.1145/1518701.1518866

[76] Shumin Zhai. 1995. Human performance in six degree of freedom input control [microform].

[77] 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), 1–11. https://doi.org/10.1145/3313831.3376313

[78] Extending AR Interaction through 3D Printed Tangible Interfaces in an Urban Planning Context | The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology. Retrieved September 16, 2022 from https://dl.acm.org/doi/10.1145/3332167.3356891