Scale Impacts Elicited Gestures for Manipulating Holograms: Implications for AR Gesture Design

Tran Pham[1], Jo Vermeulen[2], Anthony Tang[1], Lindsay MacDonald Vermeulen[3]
[1] University of Calgary, Canada
[2] Department of Computer Science, Aarhus University, Denmark
[3] Alexandra Institute, Denmark
phamt@ucalgary.ca · jo.vermeulen@cs.au.dk · tonyt@ucalgary.ca · lindsay.vermeulen@alexandra.dk

ABSTRACT
Because gesture design for augmented reality (AR) remains idiosyncratic, people cannot necessarily use gestures learned in one AR application in another. To design discoverable gestures, we need to understand what gestures people expect to use. We explore how the scale of AR affects the gestures people expect to use to interact with 3D holograms. Using an elicitation study, we asked participants to generate gestures in response to holographic task referents, where we varied the scale of holograms from desktop-scale to room-scale objects. We found that the scale of objects and scenes in the AR experience moderates the generated gestures. Most gestures were informed by physical interaction, and when people interacted from a distance, they sought a good perspective on the target object before and during the interaction. These results suggest that gesture designers need to account for scale, and should not simply reuse gestures across different hologram sizes.

Author Keywords
Gestures; augmented reality; gesture elicitation; HoloLens.

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces — interaction styles, user-centered design.

INTRODUCTION
Gestural interaction for augmented reality (AR) remains idiosyncratic, where users may need to re-learn gestures and skills for each system they use. The commoditization of AR technology is making AR experiences, where people interact with virtual objects and content overlaid atop a tracked model of the world, a reality today. While many manufacturers deliver specialized handheld controllers to manipulate the virtual objects and scenes, these remain dedicated pieces of hardware with manufacturer-specific affordances. To overcome this challenge, many researchers still aim to design freehand gestures to allow users to create, modify, and interact with virtual 3D objects in physical spaces. The problem is that we do not yet have a common framework for designing gestures for such AR experiences, much less a common gesture set for this type of 3D object interaction.

To help develop a common framework for AR gestures, we are specifically interested in this paper with how different scales of AR affect people's expectations for gestures. The AR experiences we have seen over the years have varied widely in terms of the scale of the virtual objects and scenes. For instance, researchers have explored AR experiences on tabletops (e.g. [3, 63]), as mid-air holograms (e.g. [43, 60]), at room-scale (e.g. [34, 59]), and even at city- or world-scale (e.g. [29, 45, 49]). Generally, we are interested in designing easily-guessable and memorable gestures for interacting with 3D AR content. This discoverability is important for near-future AR scenarios, for example: in "walk-up and use" AR in a museum, or in casual usage scenarios such as a new home owner using a simple AR tool to explore furniture arrangement for a room. We build on prior work that has articulated several new dimensions for AR gesture design [40], where we explore specifically whether the scale of an AR experience affects gestures people expect to use. Specifically, should designers use the same gesture for (say) moving a small AR object to also move a large AR object? Or alternatively, does a designer need to account for different object sizes when designing such movement gestures? Furthermore, does this change if the object is attached to a surface such as a wall or table, and do people use different kinds of gestures if the content is room-scale?

We conducted an elicitation study to address our research question about the impact of scale on the discoverability of gestures in AR. In this study, participants were presented referents of different AR scales and asked to propose a gesture for the referent. A referent is the desired action that the proposed gesture invokes, e.g. "zoom in." We recruited 16 participants to create these gestures in response to "before" and "after" visuals of operations on 3D objects that they could view through a head-mounted AR display (each were 3D AR scenes). We used this elicitation study to understand what kind of gestures people would expect to use to enact the operations. By analyzing this data, we gain

DIS '18, June 9–13, 2018, Hong Kong
 ACM ISBN 978-1-4503-5198-0/18/06…$15.00
https://doi.org/10.1145/3196709.3196719

insight into which gestures people expect to use across different scales, and insight into designing a discoverable gesture set that matches people's expectations. To explore the impact of scale, we visualized operations at three different AR scales: a small hologram floating in mid-air, a city model attached to a physical table, and another where virtual furniture was laid out in a small physical room.

To foreshadow our results: rather than relying on the same gesture at different scales, participants generated new gestures for each hologram size. We found that because the gestures people generated had a largely physical nature (i.e. rather than symbolic gestures), participants used gestures to manipulate the virtual objects' affordances. Consequently, the size of the object or scene moderated the size of the gesture they performed. Finally, we observed that for smaller objects, people moved towards the object to operate on it directly—as if to touch it (*proximally*), while for larger objects and scenes, people relied more on perspective-based gestures to operate on the object from a distance (*distally*). Our findings extend those from Piumsomboon et al. [40] by characterizing the nature of gestures in relation to scale of virtual objects, and by providing a further characterization of their "*locale"* dimension (i.e. where the gestures are performed [40]) that designers can use.

Our work contributes an additional set of considerations for designers who are building AR experiences where people interact with 3D objects and scenes. Specifically, they suggest that to design discoverable gestures for AR experiences where virtual objects are manipulated, designers need to account for scale and apparent physical affordances of the virtual objects. Additionally, where multiple scales are involved, multiple gestures may be required for the same operation at each scale of interaction.

RELATED WORK

We begin by considering how researchers have envisaged different scales of AR experiences, demonstrating that these scales are markedly different in terms of size, the envisioned use case, and how people interact with the AR content. We then discuss how freehand gestural interaction with AR remains of interest for both practical and accessibility reasons, but that gestures have still been frequently designed around recognition technologies rather than being focused on discoverability. Finally, we describe how elicitation studies have been used to addressed this "guessability" problem for gestures. We synthesize these learnings from the community, and show how our work addresses the gap between AR gestures and the wide variety of scales that we expect AR to be used.

Scale of AR experiences. Researchers and visionaries have demonstrated that AR experiences will vary widely in terms of the scale of the virtual objects and the experience. Some experiences are small, table-based experiences, where the entirety of the objects under control are the size of one's palm (e.g. [63]). Some are tabletop experiences, where objects interact with surfaces (e.g. [3, 7]). Other AR experiences are room-scale, with virtual objects attached to multiple walls and large surfaces (e.g. [19, 20, 34, 37, 59]). Finally, some AR experiences are world-scale, where the digital content is anchored to large objects in the world (e.g. landmarks [29, 45], or infrastructure [49]).

Each of these AR scales presents different interaction challenges for acting on or affecting the digital content. To some extent, this interaction is also mediated by how we view AR content. Some approaches use head-mounted glasses with digital overlay (e.g. [34, 36, 43]) or projection-based (spatial) AR [6] (e.g. [20, 59]). With handheld, see-through displays such as tablets or smartphones (e.g. [1, 9, 49]), interaction techniques rely on the stability of a handheld see-through display to facilitate mediated "touching" (e.g. as in Apple's ARKit showcase game "The Machines" [9]). For larger-scale content, we have seen ray-based interaction techniques, where people use either their fingers, arms or another device to point at the distant objects to be interacted with (e.g. [11, 35, 38, 44]).

Given the disparate range of AR scales and experiences, it seems unlikely that we can adopt a "one-size-fits-all" approach to interacting with virtual content in AR. This presents the challenge that people cannot easily reapply learnings from one experience to the next. Our work aims to understand whether this idiosyncrasy due to size is unnecessary, that is, whether we can reasonably expect that people will expect to interact with content at different scales in the same way.

Freehand Gestural Interaction in AR. Many researchers are actively seeking to develop freehand gestures for interacting with virtual content in AR, which are considered natural user interfaces [4]. Early work considered the use of tangible proxies such as cards or wands for interacting with virtual content, where manipulating physical proxies would result in similar actions in the virtual world (e.g. [5, 22, 48]). These were later augmented via multimodal approaches integrating speech and gestures (e.g. [15, 17, 23, 25, 35, 42]). Recent work on freehand gestures has been accelerated by the commoditization of depth cameras and sensing technologies optimized for hand posture and gesture recognition (e.g. [33, 24]). These freehand gestures leverage reality-based interaction mental models for manipulating virtual content [18], for instance by pushing virtual objects with one's hands (e.g. [3, 16, 41]).

The challenge is that many of these gestures are designed for optimal recognition by the sensing technologies, rather than by being motivated by human characteristics (e.g. discoverability). For instance, [10, 21, 26] rely on overhead cameras to track the position and posture of hands. This means that because of the positioning of the cameras, certain types of gestures are not accurately recognized, while some gestures are effectively impossible to detect properly (e.g. if they are obscured by other parts of one's hands) [10, 26, 27, 28]. While advances in machine learning may help to ameliorate the sensing problems, the

issue remains that the underlying design of these gesture sets is mainly motivated by identifying easily distinguishable and recognizable gestures.

Elicitation Approaches for Discoverable Interaction. To design discoverable, memorable gestures for new interaction contexts, many researchers have now turned to "elicitation studies." Rather than relying on designers (i.e., experts) create gestures for actions, participants propose gestures that they would expect to use in response to system actions. Gesture elicitation has been used for many domains including surface computing [31, 62], AR [25, 40], deformable interfaces [55], TV controls [57], omni-directional video [46], multi-display environments [50], mobile motion gestures [47], back-of-device gestures [52] and above-device gestures for smartwatches [51]. Follow-up work has demonstrated that this approach generates a better gesture set than an expert-generated one [30].

Elicitation studies for AR have identified a working set of gestures for AR (e.g. [25, 40]). Piumsomboon et al. [40] focus on interactions with "table-sized" AR content, where gestures included both manipulation on the objects (e.g. resizing, moving, etc.) along with editing, simulation, browsing, selection and menu interaction. The study helped identify two new parametric categories to classify gestures unique to AR: whereas Wobbrock et al. [62] characterized gestures in terms of Form, Nature, Binding and Flow, Piumsomboon et al. [40] added two new parameters: Symmetry (i.e. how the hands are used) and Locale (where the gestures are performed). Beyond this, the authors also propose a working set of freehand gestures for AR.

Gesture elicitation studies for AR have not adequately accounted for scale of the AR experience. Critically, while the working set of gestures of Piumsomboon et al. [40] is likely to be discoverable, the gestures were elicited for "baseball-sized" referents that appeared on or just above a large table. Notably, the authors observed that the size of the AR object affected the number of hands used to manipulate the objects (e.g. for objects palm-sized or smaller, only one hand; for objects larger, two hands). Unfortunately, the study only investigated relatively small-scale AR (in comparison to bigger experiences such as [20, 34, 59]). Thus, while large portions of the proposed gesture set are likely to be effective for tabletop AR (i.e. actions that interact with AR "above" the objects themselves such as menu selection, and abstract actions such as editing or simulation), it is unclear whether they adequately account for the range of scales we have already seen AR experiences take, which we investigate in this paper.

Synthesis. AR experiences already exist that address a wide range of scales; however, we do not understand how differences in scale impact people's expectations of how to interact with holograms. Freehand gestural interaction with 3D content in AR is likely to be a common approach, as it does not require specialized external hardware. Our current understanding of how to design these gestures, based on elicitation studies leaves unclear whether the gestures should be consistent across scales, or if they should vary across scale. We address this gap in the present study, where we consider three distinct settings and scales for AR (in mid-air, anchored to a surface and at room-scale). As we will see, these settings impact the character of the gestures that participants create, and this is even more pronounced with room-scale AR, extending earlier findings in [40].

STUDY DESIGN

To understand the impact of scale on people's expectations of how to interact with gestures, we used a reformulated form of an elicitation study. As with other domains of inquiry, an elicitation method presents participants with a "before" and "after" visualization, whereupon participants are asked to design a gesture that would move the system from one state to another. Our use of this method helped us to uncover underlying mental models people have of how 3D objects and scenes would/should be manipulated using freehand gestures, thereby giving us knowledge to design learnable, memorable gestures for these situations. Whereas prior work with elicitation studies provided participants with each referent in turn, we reformulated this approach using what we called "scenarios." These scenarios asked participants to imagine themselves completing an entire series of tasks (i.e. multiple referents), so that the referents were elicited in context with one another.

Scale. Each scenario grouped together a series of referents at a specific "scale" of interaction, each with its own constraints: *Mid-Air*, *Surface*, and *Room*. The apparent visual size of the virtual objects differed in each scenario as well as the size of the scene (the volume containing all objects expressed in length × width × height). Figure 1 illustrates these scenarios within the context that the participants experienced them in the study. The "mid-air" scale of interaction, for instance, used a 14 cm × 31 cm × 25 cm model of a home. The "surface" scale of interaction used a small city model with buildings and streets on a tabletop prop (26.5 cm × 97 cm × 56 cm). This scale also afforded interaction with the tabletop, since the hologram appeared to sit on the tabletop itself. Finally, the largest scale of interaction, the "room", used a physical three walled-room (with an open side) with dimensions of (549 cm × 303 cm × 255 cm) that contained virtual furniture such as a couch, shelves, and a coffee table. Each of these scales are illustrated in Figure 1. We chose each of these scales to represent major classes of interactive AR experiences: the mid-air scale abstracts tasks that involve a mid-air hologram (e.g. a car in automotive design), whereas the surface scale maps to augmented document tasks (e.g. [56]), while the room-scale scenario abstracts furniture re-organization in a physical space.

Scenarios. To give participants a motivating purpose in the study, participants were provided with a series of tasks to accomplish within each scenario, where each referent made sense in the context of the greater scenario. Each of these

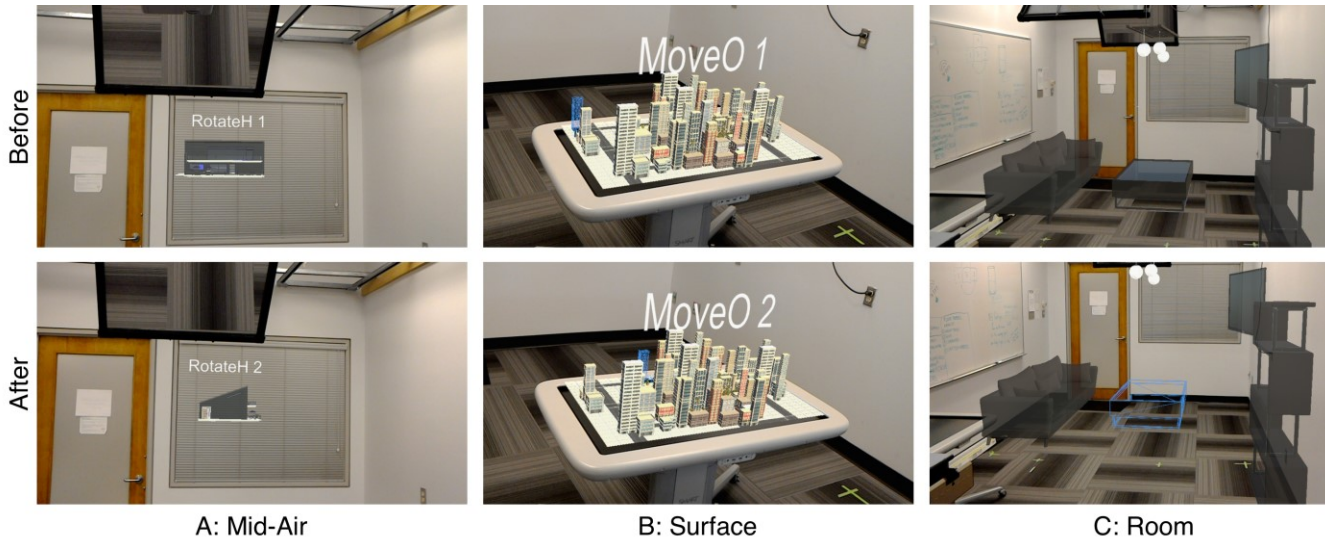

Figure 1. We used three different scenarios in our study. Illustrated here are sample before and after referents for each scenario: (A) *Mid-Air* – a virtual house model in mid-air (14 cm × 31 cm × 25 cm), "rotate scene" referent; (B) *Surface* – a small city model with buildings and streets on a tabletop prop (26.5 cm × 97 cm × 56 cm), "move building" referent; (C) *Room* – a small physical room with virtual furniture such as a couch, shelves, and a coffee table (549 cm × 303 cm × 255 cm), "select coffee table" referent.

"tasks" was represented as a referent pair of "before" and "after" scenes. Thus, the same base set of referents was presented in each scenario, and each scenario added its own set of referents that fit the scenario.

To get a sense of how the scenarios worked, we will provide an overview of the initial set of referents for each scenario (note that each scenario was considerably longer). In the *Mid-Air* scenario, the participant remodeled a house in several steps. First, they were to rotate the scene so that the home faced the right direction, then they would scale the depth of the home to fit the lot. Then, they were to move some objects out of the way before duplicating the front door, and then moving the duplicated door onto the back wall (so the home would have a back entrance). For the *Surface* scenario, participants were to change the plan for a few city blocks. Here, a participant would select a set of buildings, then, duplicate one of them. The duplicate building needed to be moved to a new block, where it would need to be rotated to fit into the block. Then, a particularly ugly building was to be deleted. Finally, in the *Room* scenario, participants were to redecorate the room they stood in. First, they were to select the bookshelf, and rotate it to create a more intimate space. Then, they were to shrink the TV on the wall so it did not occupy as much wall space, and then move the coffee table against the wall.

Referents. Each scenario enacted a different number of referents that made sense for the scenario, and all scenarios shared a common "base set" of referents. A *referent* is the desired effect of an action (e.g. zoom in) for which participants are asked to propose a gesture (e.g. pinch) [30]. We extracted 17 basic operations from 3D modelling tutorials (e.g. for SketchUp [54]) that reflected a common set of operations on 3D objects or scenes (select object, duplicate object, move object, rotate object, delete object, undo, scale x-dimension, scale y-dimension, scale z-dimension, move scene, scale scene uniformly, separate layers, select group of nearby objects, scale object uniform, select multiple objects, select layer, rotate scene). From these referents, we selected referents that were suitable for the scenario, and adapted them as appropriate; for instance, "Select object" became "Select a plant" in the *Mid-Air* scenario, "Select a building" in the *Surface* scenario, and "Select a piece of furniture" in the *Room* scenario. Not all referents were represented in every scenario (e.g. "Rotate scene" did not make sense in the room scenario). To ensure that we had a common set of referents where we could compare gestures across each scale and scenario, we identified seven referents to act as the "base set": select, duplicate, move, rotate, delete, undo, and scale (x-dimension). Each of these referents was represented in each of the scenarios, and formed the basis of our analysis.

Although some elicitation studies show a transitional animation between referents (e.g. [40]), we chose instead to "slideshow" the referents as in [50]. By presenting using a strict "before" and "after" approach, we free participants of this potential bias, allowing them to consider discrete variations if this suited their mental model. We believe the animation pre-supposes how the gesture ought to operate on the object/scene (i.e. as smooth continuous operations).

Apparatus. Participants viewed the referents using a Microsoft HoloLens head-mounted display, where the objects in the scene were anchored to fit the scenario (e.g. anchored to a physical table, the wall, or the ground). Participants could move around the scene, and the holograms would update accordingly, with the virtual

objects remaining anchored to the real world. Participants were presented the start and end state in sequence.

Participants. We recruited sixteen participants (10 male, 6 female) with an average age of 23 years. Five participants had prior experience with the HoloLens (four of these had only played with demos), none of the others had prior experience. Of these 16 participants, four reported limited experience with AR/VR systems (i.e. demos at stores), while 11 reported having none. All participants were students, and their backgrounds were broad: computer science, statistics, occupational therapy, neuroscience, chemistry, engineering and kinesiology.

Method. Scenarios were presented to participants in counter-balanced order. For each referent, the participant saw the "before" and "after" scene, and was asked to generate a gesture that they expected would move the system from one state to the next. Participants could ask to toggle between the "before" and "after" scenes (controlled by the experimenter) as many times as they liked. They were then asked to rate the suitability of the gesture for the referent, and the ease of performing the gesture. Participants were given a break between each scenario.

Data Collection. We collected four sources of data: first, demographic information from a pre-study questionnaire; second, a video capturing participants' gestures created during the study; third, participants' confidence scores in the suitability and ease of use of the gestures they created (as they went), and fourth, responses to a post-study semi-structured interview. Video footage was recorded from in front of the participants.

Analysis. As we collected data, we used iterative video coding of the gestures that participants generated, identifying recurring gestures and themes, and properties that described the gestures (e.g. number of fingers used, shape of the gesture). This coding process was iterative, where we generated provisional codes, revising them as we identified better phrases and ideas to capture what we observed. We developed a tool to simultaneously view gestures across participants and scenarios and facilitate comparison. Using this tool, all co-authors viewed videos of participants' gestures to develop a coding scheme to capture and describe the gestures. We then used axial coding [8] to explain how lower level codes were related to one another. Critically, while our analytic frame was influenced by all the gestures from participants, our analysis examines *only* the base set of seven referents common to all scenarios, as these are the only ones that could be compared.

FINDINGS

We focus our analysis in this paper on the subset of seven referents that appeared across all three scenarios: select object, move object, rotate object, duplicate object, undo, delete object, scale x-dimension (which we from now on refer to as "scale" for simplicity). After describing the gestures our participants generated, we discuss the physical nature of many of these gestures and how the apparent visual affordances of the objects influenced the character of the gestures. We then describe how the size and scale of scene influenced the gestures, and how participants used perspective-based gestures for larger scenes.

Collected Gesture Data and Agreement Rate. We analyzed 261 gestures generated by the 16 participants, where we identified 64 unique gestures (some gestures were lost because they were obscured by a participant's body, or were not captured due to the camera's field of view). Participants' gestures primarily made use of their fingers, hands, arms, as well as their body position (and their body orientation). Only two participants made use of gestures where the posture of the hand was intentionally part of the gesture (e.g. using two fingers rather than one, or using their pinky and thumb). In general, participants' gestures relied on the visual position of fingers/hands in relation to one another, or on the viewing angle to the object. For instance, for the *Select* referent, some participants walked up to the object and tapped on the object (e.g. with one finger, or with an open palm). Others used what we call a perspective gesture, where they used the same tapping gesture, but rather than walk to the virtual object, changed their position or viewing angle so that their hand would appear between their eye and the object.

Table 2 shows the agreement rates (*AR*) for each of the seven common referents across the three scenarios. Based on the qualitative classification scheme from [58], most of the gestures for these referents would be considered as having medium agreement.

Gesture Themes. Rather than simply code the gestures that we observed, we saw that a coding of the *themes* underlying participants' gestures was more insightful. First, the gestures sometimes varied, but only in very minor ways, and the strict coding of the gestures distinguished between instances in ways that were not useful. Second, we found that the themes seemed to express a mental model of how the participant thought about the objects and their own role in the scene. Consequently, we found the themes to be a much more useful analytic unit in our analysis.

Gestures varied in terms of how participants executed an idea, even if thematically, the idea was the same. For instance, for the *delete* referent, one theme we observed was to squish the object, and this was performed in three

Referent	Gestures			Themes		
	Air	Surface	Room	Air	Surface	Room
Select	0.383	0.183	0.108	0.542	0.650	0.475
Duplicate	0.092	0.100	0.100	0.158	0.167	0.142
Move	0.225	0.200	0.258	0.300	0.308	0.317
Rotate	0.192	0.175	0.125	0.242	0.275	0.433
Delete	0.217	0.167	0.125	0.225	0.208	0.192
Undo	0.092	0.075	0.108	0.208	0.217	0.192
Scale	0.183	0.167	0.183	0.425	0.483	0.508

Table 2: Agreement Rate (*AR*) for referents and themes across scenarios. Shading shows low, medium, high agreement [58].

slightly different ways: one participant clapped, another pinched her fingers, while another closed his fist. Here, each of these gestures is distinct in how it is executed (and would be coded independently in our original coding process); however, given that they are conceptually similar, it makes sense to group them together in the same theme.

A gestural theme seemed to indicate a particular mental model of how the participant could operate and affect the world. For instance, with the *delete* referent, we classified three other themes: pick and throw (fingers grabbing an object and throwing; fist closing and throwing; grabbing the object with two hands and throwing), wiping the scene away (full arm wipe; one hand wipe), and symbols (drawing an X; slashing across the object). Each of these is suggestive of a different kind of mental model for the scene and the objects in the scene: whereas the "squish" implies a certain omnipotent strength, pick-and-throw implies a certain strength to the viewing perspective of the user (i.e. that throwing the object out of the view deletes it rather than throwing it behind oneself), the wiping theme implies a "scene"-based approach (e.g. as an image in a slideshow viewer), whereas the symbolic theme takes on a more conventional "operation on an object" model.

Table 3 summarizes the 33 themes that we identified. As illustrated in Table 2, our coding for the themes that underlie the gestures shows a considerably higher rate of agreement than the gestures themselves, where we counted each gesture uniquely when different body parts were used (e.g. fingers vs. hands vs. fists vs. arms).

Nature of Gestural Themes: Physical Interaction

The nature of most gestural themes was physical: participants visibly seemed to act on the objects themselves—virtually grabbing, touching, and pushing the objects in various ways. These are reminiscent of ideas from Reality-Based Interaction [18], where the themes of body awareness and environmental skills seemed to underlie many gestures (e.g. tossing an object to delete it, pushing an object to move it, stretching sides of an object to scale it, squishing an object to delete it, and so on).

Theme	Variants	Referents
Tap	With fingers, step forward and use finger, with palm	Select
Grab	With finger and thumb, as a fist	Select, Delete
Circle	Finger (pointing), using finger and thumb (to draw a selection area)	Select
Pulling it apart	Two finger tips pulled apart, hand to hand pulling it apart, two handed grab and pull, upside-down U-shape, hold original and pull a second hand out	Duplicate, Scale
Double pinch and pull	Two finger pinch, then pull apart	Duplicate
One handed tap and move over	Using fingers to tap, using palm to tap	Duplicate
Point at object, then point at second location	Using fingers	Duplicate, Move
Pick up and place in new location	With fingers, with arms, with whole body (i.e. walk)	Duplicate, Move
Grab, pull over and release	Twist hand sideways	Move
Pinch and move	Using fingertips, using fist, using finger and pinching with thumb	Move
Push with open hand	One hand, two hands	Move
Pick up and toss	Physically move, using only arms	Move, Delete
Twisting like a lightbulb	Twisting like a handle	Rotate
Hands on front and back, twist both (like a box)	With thumb and pinky, with thumb and index	Rotate
Point and twist hand		Rotate
Twist body (as if body is object)		Rotate
One point to anchor, other finger to twist around	With finger, with hands/fists	Rotate
Grab corner or edge, and move	Arm or body	Rotate
Push aside	Toss away, toss behind, upside down U	Delete
Squish	Clap with hand, with fist, double fist, twist and fist, push into ground	Delete
X symbol	Make a crossing gesture	Delete
Wipe	Using forearm	Delete, Undo
Cross off (like a straight stroke)		Delete
Backward arrow	Drawn with finger, arm, or point with thumb	Undo
Tap in top left		Undo
Reverse action		Undo
Three finger swipe		Undo
Double tap on object		Undo
Hands at sides of object and symmetric pull/squish (squeeze)	Pointer fingers, hands, fingers, palms, fists	Scale

Table 3. Each of the 33 gesture themes we observed from our participants, their variants, and the referents for which they were generated.

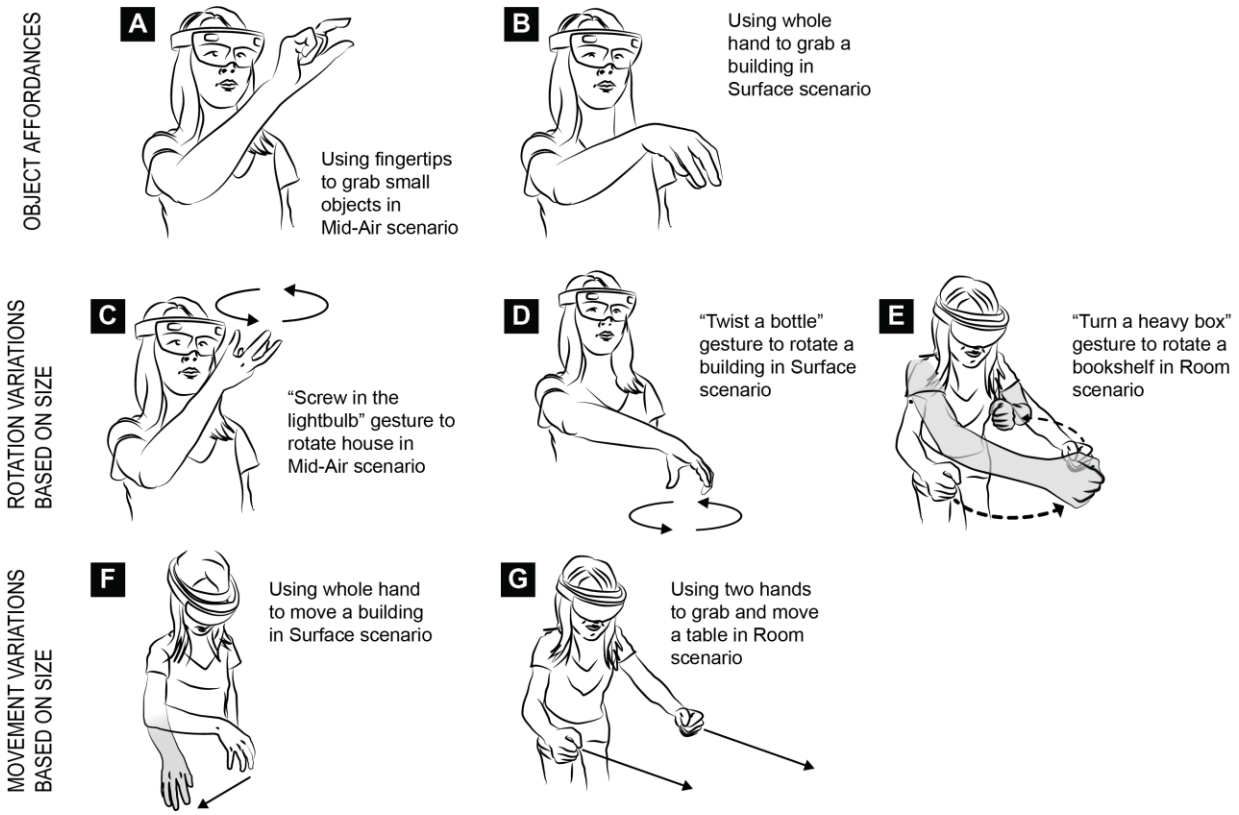

Figure 2. Examples of gestures that we observed, illustrating: (A-B) Variations for grabbing objects based on physical features (affordances); (C-E) Variations based on size for the *Rotate* referent; and (F-G) Variations based on size for the *Move* referent.

Our results indicate that 79% of the gestures were physical in nature, and this is in variance to Piumsomboon et al. [40], who report a more balanced percentage of physical gestures (only 39% of the gestures). We suspect that this is due to the nature of the referents that we selected in our study, where all referents strictly operated on the virtual objects in the holographic scene, whereas Piumsomboon and colleagues [40] also explored menu-based interaction. We offer here a more detailed exploration of the nature of these gestures, and how the participants executed these gestures within the context of the virtual objects.

Participants' gesture choices were influenced by the physical features of the virtual objects they were manipulating. As illustrated in Figure 2, peculiarities of specific models themselves would change the gestures, where participants would grab protruding parts of objects (e.g. six participants grabbed tall buildings as handles, as shown in Figure 2-B), or used finger tips to pick up small objects (rather than trying to grab it with a whole-hand grip), shown in Figure 2-A. For instance, participants generated gestures for the *Rotate* referent with small conceptual variants on the theme of turning the virtual object. Three participants in the *Mid-Air* scenario rotated the house with their wrist, as if twisting a lightbulb from a ceiling (Figure 2-C). In the *Surface* scenario, two participants rotated buildings by placing their hand on the top of a building and twisting it, like with a bottle placed on a table (Figure 2-D), while in the *Room* scenario three used a two-handed twisting gesture to rotate the bookshelf (as if to grab and rotate a heavy box), illustrated in Figure 2-E.

Participants maneuvered themselves into positions where they could easily see large parts of objects they were trying to manipulate. As with physical objects, being able to grab an object is easier when large parts of the object are visible, rather than when visibility is limited. When target objects were obscured by other virtual objects (as in the *Surface* scenario, where some buildings obscured others), people walked around the table, craned their heads or physically moved towards objects to get a better view of the target object before enacting a gesture. One reason this might have happened was to give themselves a better "target area" to gesture at. This relates to another peculiar result, where most participants (13/16) *never* touched the table surface for any referent in the *Surface* scenario (the remaining three only touched it once for the *Move* referent). This is in line with our observation of gestures being dependent on the virtual objects' affordances, where people's behaviour seems to reflect real life. For instance, it does not make sense to reach 'through' or 'under' a physical object on a table, so people did not do this with virtual objects either. The easiest way to grab a skyscraper in a physical city model is to grab it by the top or middle, which is what we observed people doing (Figure 2-B). P6 articulates the influence of this physicality: "*I think [for] stuff that is more*

relatable to the physical world, … like moving stuff around and rotating it, you probably don't need an extra menu."

Participants varied their gestures within an interaction theme as a response to the size of the object/scene, and their distance from an object. Notably, participants only sometimes used the exact same gesture across all scenarios for the same referent: 27 times out of 96 opportunities (16 participants × 7 referents). Most of these happened with more abstract, non-spatial referents (eight times with undo, nine times with delete, nine times with duplicate). With a given participant, we usually observed gestures that varied on the same theme for the referent across different scales (e.g. using fingers to "grab and move" a small object in the *Mid-Air* scenario, then using one hand to move an object in the *Surface* scenario, and finally using two arms to grab and move the object in the *Room* scenario (Figure 2 C, D, E).

We mainly observed non-physical gestures only when the referents did not have a clear physical analogue (e.g. abstract referents such as delete, duplicate, undo). For these, participants found it difficult to come up with gestures, and that resulted in mainly metaphorical or symbolic gestures (21% in our dataset). P6 again suggested that more sophisticated gestures might be necessary for such referents: "*If you want to duplicate [an object] or something, the gesture should be a bit more [complicated]. For example, I crossed my arms to delete stuff. I just haven't thought about anything for duplicating stuff though."* A common suggestion was to include a menu for those more abstract referents. P7: *"I like the idea of having a menu on the side, so that you can more easily [perform] undo, or delete. […] Maybe a menu on the side to give you more flexibility with what you're doing."*

Size Matters for Spatial Operations

The size of objects and the virtual scene influenced participants' overall behaviour, including how they moved around and which gestures they performed. As shown in Figure 1, the *Room* scenario had the largest objects (furniture) and virtual scene size, followed by the *Surface* (city model) and *Mid-Air* scenario (house model). Whereas participants would use less bodily movement for the *Mid-Air* scenario, they tended to use more gross bodily movement (i.e. either physically moving, or using their hands or arms) for gestures in the *Room* and *Surface* scenarios. For instance, four participants changed from using a single-handed gesture for the *Rotate* referent in both the *Mid-Air* and *Surface* scenario to a two-handed gesture in the *Room* scenario. Similarly, for the *Move* referent, P4, P7 and P11 grabbed the virtual object of interest with both hands (Figure 2-E), and then proceeded to physically "carry" it to its new location in the *Room* scenario, but only used hand gestures for the other two scenarios. These changes were most dramatic when participants switched to the *Room* scenario compared to *Mid-Air* and *Surface*.

Most striking was that these differences persisted even in the number of fingers or hands being used: the smaller the virtual objects, the fewer fingers and hands that were involved. For the *Scale* referent, we observed 29 instances of a "squeeze" theme gesture. In the *Mid-Air* scenario, six participants created gestures that used their fingers, one used a hand, and four used two hands; in the *Room* scenario, only three used two fingers, three used one hand, seven used two hands, and one used both arms for the gesture. These differences are striking, since virtual objects do not weigh anything—they just appear larger.

Most participants (12/16) confirmed in the post-study interview that the experience of size of these objects impacted their gesture choices. For instance, P8 describes the difference between the scales as working with "*Lego versus Ikea.*" Similarly, P11 reported, "*Yes! [nodding] Yeah, I think the scale definitely matters. When it's really small, it gets awkward to handle around with [large gestures].*" We asked participants whether this was because it was hard to be precise with small objects. P4 mentioned: "*Yeah, something like a couch, I can move it like I move a couch […]. With the model couch in that tiny house, I'd do something else.*" Participants generally talked about the room-scale scenario as having large objects and the other two scenarios as having smaller objects.

With abstract, non-spatial referents, size did not generally influence the gesture choice. For instance, for *Undo* and *Delete*, the common theme was to create a symbolic gesture (e.g. gesturing to the left to "undo" an action—much like undo icons in document editors) or a metaphorical gesture (e.g. tossing the object out of view, or "wiping" the view).

"Physically" Touching the Objects from a Distance

Although not strictly due to the size of the virtual objects, when using physical gestures, participants used both proximal gestures and distal gestures. Sometimes, participants would gesture extremely close to the target object—so close that the participant appeared to be touching the object: a *proximal* gesture (Figure 3-A). Other times, participants generated the gesture from a distant location, where the gesture was far from the virtual object (e.g. pointing at or grabbing something that was clearly out of reach): a *distal* gesture (Figure 3-B). Some gestures included a distal gesture component in addition to a proximal one (e.g. "picking up" an object using a proximal gesture and throwing it to a location using a distal gesture): a *mix* gesture. Figure 3-C shows a mix gesture where a distal and proximal gesture are combined to grab a coffee table from a distance and pull it closer. Finally, we noted that symbolic and metaphorical gestures (see [62]) did not occur meaningfully as a function of the proximal/distal dichotomy, thus we coded them in a separate category.

Distal gestures where still physical in nature, and operated on objects that were out of reach, and relied heavily on the participant's view and perspective on the object. As illustrated in Figure 3-B, participants lined up their perspective so that their hands or fingers appeared in line with their view of the object. Participants seemed to interact

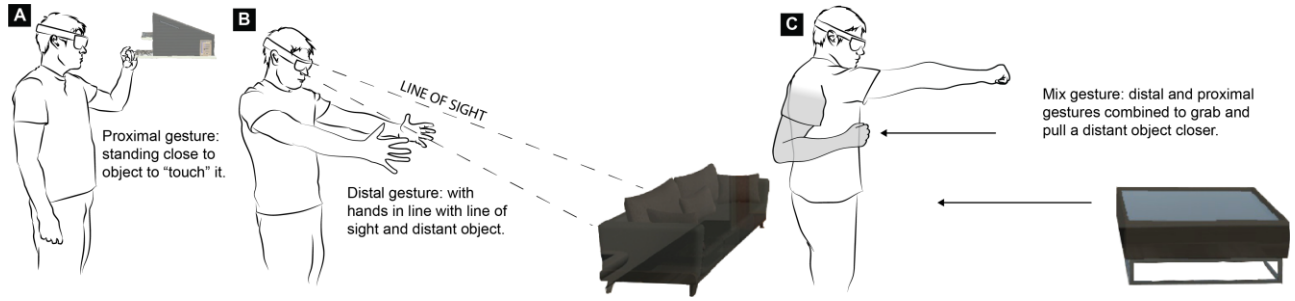

Figure 3. We saw three kinds of object interactions: (A) *proximal*, where people seemed to touch the object directly, (B) *distal*, where they interacted from a distance with a good perspective on the object, and (C) *mixed* approaches combining both aspects.

with the 2D projection that the 3D object made on the headset's image plane (e.g. as in [39]). This perspective issue is significant, as it meant that participants would change either their head position or their physical position to give them a good perspective on the object before enacting the gesture. We observed three kinds of reasons for head and physical movements: first, to reduce occlusion by other virtual objects; second, so that the virtual object was visually "large" in the headset (as opposed to being too small to interact with), and third, the opposite—to reduce the visual size of the object by moving one's head away from it (get too close, and it is hard to see anything else).

Participants moved away from the objects only rarely, but when they did, it was due to a poor field of view. In the *Room* scenario, the scene extended beyond people's viewing angle (i.e. when standing in the middle of the room, one needs to explicitly turn one's head to see all the furniture). Within this scenario, five participants stood near the back of the room—just so they could see the entire scene without needing to move around (or move their heads). We also observed that participants did this to allow them to see the object's entirety (e.g. for a *Rotation* referent), or the origin and the destination at the same time (in a *Move* referent). What is striking here is that this type of movement occurred far less for smaller scenes (i.e. the *Mid-Air* and *Surface* scenarios); rather, for these scenarios, we suspect that the movement was mainly to reduce occlusion (e.g. of the target virtual object by another).

This issue of field of view was also identified by participants in the post-study interview. P3: "*I think [with larger scenes make it] harder to see everything at once. […] Instead of seeing everything on [the display], you have to move your body […] to look at certain places to actually see what you're talking about.*" Similarly, P7 mentioned: "*I think if it's confined to an area, then you're more easily able to manipulate and see the changes. If you're needing to back up and look around and figure out what's going on, it's not as straightforward as if it was a confined [space].*"

Besides physically moving in the space (i.e. walking around) to get a better perspective on the objects before enacting the gestures, we also observed many instances where participants used movement as a part of the gesture itself. As discussed earlier, three participants walked from one location to another as part of a *Move* referent in the *Room* scenario. This suggests that the physical location is an enactment of the overall gesture (i.e. rather than the gesture simply being a function of hand movements).

The distribution of proximal, distal and mixed gesture locales varies and is related to the size of the virtual objects and scenes. Figure 4 illustrates the relative distribution of each gesture type across each referent and scenario. For *Mid-Air* scenarios, participants were generally proximal to the target, many appearing to carefully touch the objects with their fingers, whereas as the scale of the scene grows larger (i.e. into *Surface* and *Room* scenarios), the relative frequency of distal gestures increases fairly dramatically.

Our characterization of the "locale" of the gesture extends the articulation in [40], as it explains why some of these gestures happen *distally* (i.e. "in-the-air" in [40]) and also shows that this changes as a function of scale. Whereas with prior work [40], the percentage of "in-the-air" gestures might be explained as a function of menu interaction, we observe that these distal gestures in our study also occur for practical reasons such as ease of interaction (i.e. being able to see origin and location at the same time, or being able to see the objects of interest, both of which become more important at larger scales). We also see that many of these gestures incorporate movement within them.

DISCUSSION

Our interest is in designing gestures for 3D objects/scenes in augmented reality, and our findings show that the scale

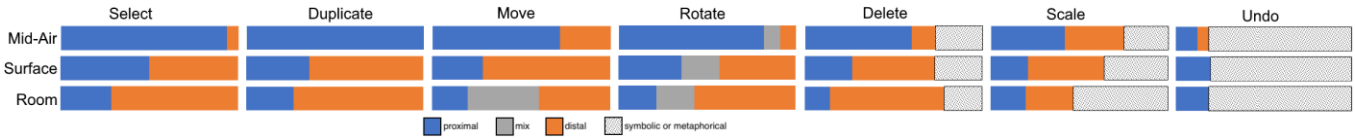

Figure 4. The relative distribution of proximal, mixed, distal and symbolic gestures in our dataset.

of the holographic AR objects and scene clearly impact how people expect to interact with them. The findings suggest that deriving a unified gesture set in this context is probably inappropriate: participants acted on the virtual objects in the scene as if they were physical, thus the gestures need to account for the apparent affordances of the objects, their size, and the person's perspective on the objects and the scene. This has implications for how we design systems that interpret gestures—particularly, that they need to account for the way that the scene is visualized, and the nature of the scene itself.

Physicality. Participants interacted with the virtual objects as they would with physical objects in the real world. They used the object's available affordances, and when objects became larger, increased the size of their gestures accordingly. The effect we observed appears to be in line with Aslan et al.'s observation that object semantics such as perceived physical weight impact the size of people's touch gestures on a tablet [2], which replicated earlier findings from neuroscience on how object semantics activated related motor tendencies [13, 14]. These findings suggest that scale matters for interaction and extends prior work around gestures for AR in more confined spaces. People's tendency to act with virtual objects in a physical way suggests that gestures that account for people's environmental skills and appreciation for naïve physics [18] will be easiest to adopt. Thus, rather than designing a unified AR gesture set, a possible approach may be to use aliasing [12, 61] (i.e., having multiple 'gesture synonyms' that people can use), to address the common gestural themes that we observed. This would address discoverability, since these themes were primarily comprised of different gestures for different scales of objects and scenes in a logical progression.

Proximal and Distal Interactions. Participants interacted with objects in two distinct ways: proximally, moving towards the target object to operate on it directly, and distally, positioning themselves with a good perspective on the object to interact from a distance. This type of interaction has already been foreshadowed to some extent by techniques that make use of perspective (e.g. [53]); however, we believe there is far more at work here. While our results suggest that designers should support both ways of interaction, we observed that a few participants always interacted from a distance, or always up close, hinting at the fact that personal preferences may play a role here.

Legacy Bias. A common issue with gesture elicitation studies is legacy bias [30], where participants propose legacy-inspired interactions. We observed this phenomenon too. Many participants used a "squishing" theme for scaling that looked like pinch-to-zoom in mobile devices. We also saw whole-arm "swipe" gestures to move objects, like gestures in contemporary sci-fi movies. Finally, many of the gestures had a corresponding action in the real world. We were somewhat less concerned about legacy bias in our study: we were interested in discoverability. It should not be surprising that legacy plays into people's first "guesses" into how they can interact with the scene.

Applicability. We are not implying that every AR application should support the gestures we extracted from our data, nor that these gestures are the most effective or efficient. Our interest was in investigating gestures people would propose in AR experiences for casual use (e.g. discoverable gestures that people could remember even when they used the AR experience once or twice a year) at different interaction scales. We envision this work being applicable to situations as using an AR application for redecorating one's home, or to repair a leaky faucet. Thus, we were focused on understanding the straightforward ways people would propose to interact with 3D AR content.

Freehand vs. Instrumented. Because we do not know whether freehand gestures will be the most dominant form of interaction for AR, it is difficult to predict the applicability of our findings beyond freehand gestures. Many consumer-grade VR and AR systems now make use of instrumented interaction, where users hold a specialized controller with buttons and manipulators (i.e. traditional 3DUI in Billinghurst's classification [4]). While we see these as a crutch (i.e. owing to poor sensing for freehand gestures), we cannot be sure whether our findings will ultimately apply to these kinds of instrumented interactions.

Limitations and Future Work. Some of our findings may be tied to the hardware limitations of the HoloLens, e.g., its limited field-of-view (FOV) [32]. People may have tried to keep larger holograms in the FOV by interacting distally. An interesting direction for future work is replicating this study for future and improved headsets and for different AR experiences such as handheld AR (e.g. with ARKit [1]) or spatial AR (e.g. RoomAlive [20]). Follow-up studies can explore if people's use of proximal vs. distal gestures would change, and if the physical nature of interaction still holds.

CONCLUSIONS

This paper explored whether people expect to use the same gesture across AR experiences at different scales. In an elicitation study with an AR headset, we asked participants to generate gestures in response to referents in three different scenarios, each at a different scale. The three main findings from our study are: (i) for most gestures, participants seemed to use physical interaction to operate on the virtual objects; (ii) the size of the virtual objects and the scene influences the gestures people perform; and (iii) people tend to touch smaller objects but increasingly interact from a distance as the size of the objects and scene grows larger. Our results suggest that AR designers ought to consider the delivery context and affordances of their AR content when designing gestures for easy discoverability.

ACKNOWLEDGEMENTS

We thank our anonymous participants and NSERC for their support.

REFERENCES

1. Apple. ARKit. Accessed September 9th, 2017. https://developer.apple.com/arkit/
2. Ilhan Aslan, Martin Murer, Verena Fuchsberger, Andrew Fugard, and Manfred Tscheligi. 2013. Drag and drop the apple: the semantic weight of words and images in touch-based interaction. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (TEI '13). ACM, New York, NY, USA, 159-166. DOI: https://doi.org/10.1145/2460625.2460650
3. Hrvoje Benko, Ricardo Jota, and Andrew Wilson. 2012. MirageTable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 199-208. DOI: http://dx.doi.org/10.1145/2207676.2207704
4. Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Found. Trends Hum.-Comput. Interact.* 8, 2-3 (March 2015), 73-272. DOI: http://dx.doi.org/10.1561/1100000049
5. Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The MagicBook: a transitional AR interface. *Computers & Graphics*, *25*(5), 745-753. DOI: https://doi.org/10.1016/S0097-8493(01)00117-0
6. Oliver Bimber and Ramesh Raskar. 2005. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A. K. Peters, Ltd., Natick, MA, USA.
7. Volkert Buchmann, Stephen Violich, Mark Billinghurst, and Andy Cockburn. 2004. FingARtips: gesture based direct manipulation in Augmented Reality. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia* (GRAPHITE '04), Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 212-221. DOI: http://dx.doi.org/10.1145/988834.988871
8. Juliet Corbin and Anselm Strauss. 2015. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Sage. ISBN: 9781412997461.
9. Directive Games. The Machines. Accessed January 5th, 2018. http://themachinesgame.com/
10. Bruno Fernandes and Joaquin Fernández. 2009. Bare hand interaction in tabletop augmented reality. In *SIGGRAPH '09: Posters* (SIGGRAPH '09). ACM, New York, NY, USA, , Article 98 , 1 pages. DOI: http://doi.acm.org/10.1145/1599301.1599399
11. Tiare Feuchtner and Jörg Müller. 2017. Extending the Body for Interaction with Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17). ACM, New York, NY, USA, 5145-5157. DOI: https://doi.org/10.1145/3025453.3025689
12. G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. 1987. The vocabulary problem in human-system communication. *Commun. ACM* 30, 11 (November 1987), 964-971. DOI: http://dx.doi.org/10.1145/32206.32212
13. Maurizio Gentilucci, Francesca Benuzzi, Luca Bertolani, Elena Daprati, and Massimo Gangitano. 2000. Language and motor control. *Experimental Brain Research*, 133(4), 468-490.
14. Scott Glover, David A. Rosenbaum, Jeremy Graham, and Peter Dixon. 2004. Grasping the meaning of words. *Experimental Brain Research*, 154(1), 103-108.
15. Gunther Heidemann, Ingo Bax, and Holger Bekel. 2004. Multimodal interaction in an augmented reality scenario. In *Proceedings of the 6th international conference on Multimodal interfaces* (ICMI '04). ACM, New York, NY, USA, 53-60. DOI: http://dx.doi.org/10.1145/1027933.1027944
16. Otmar Hilliges, David Kim, Shahram Izadi, Malte Weiss, and Andrew Wilson. 2012. HoloDesk: direct 3d interactions with a situated see-through display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 2421-2430. DOI: http://dx.doi.org/10.1145/2207676.2208405
17. Sylvia Irawati, Scott Green, Mark Billinghurst, Andreas Duenser, and Heedong Ko. 2006. "Move the couch where?": developing an augmented reality multimodal interface. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality* (ISMAR '06). IEEE Computer Society, Washington, DC, USA, 183-186. DOI: http://dx.doi.org/10.1109/ISMAR.2006.297812
18. Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-based interaction: a framework for post-WIMP interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08). ACM, New York, NY, USA, 201-210. DOI: https://doi.org/10.1145/1357054.135708
19. Brett R. Jones, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2013. IllumiRoom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 869-878. DOI: https://doi.org/10.1145/2470654.2466112
20. Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. 2014. RoomAlive: magical experiences

enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (UIST '14). ACM, New York, NY, USA, 637-644. DOI: https://doi.org/10.1145/2642918.2647383

21. Ed Kaiser, Alex Olwal, David McGee, Hrvoje Benko, Andrea Corradini, Xiaoguang Li, Phil Cohen, and Steven Feiner. 2003. Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th international conference on Multimodal interfaces* (ICMI '03). ACM, New York, NY, USA, 12-19. DOI: http://dx.doi.org/10.1145/958432.958438
22. Hirokazu Kato, Mark Billinghurst, Ivan Poupyrev, Kenji Imamoto, and Keihachiro Tachibana. Virtual object manipulation on a table-top AR environment. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality* (ISAR 2000). IEEE, 111-119. DOI: https://doi.org/10.1109/ISAR.2000.880934
23. Mathias Koelsch, Ryan Bane, Tobias Hoellerer, and Matthew Turk. 2006. Multimodal Interaction with a Wearable Augmented Reality System. *IEEE Comput. Graph. Appl.* 26, 3 (May 2006), 62-71. DOI: http://dx.doi.org/10.1109/MCG.2006.66
24. Leap Motion. Accessed January 3rd, 2018. https://www.leapmotion.com/
25. 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). ACM, New York, NY, USA, 249-256. DOI: http://dx.doi.org/10.1145/1452392.1452444
26. Jae Yeol Lee, Gue Won Rhee, and Dong Woo Seo. 2010. Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment. *The International Journal of Advanced Manufacturing Technology*, *51*(9), 1069-1082. DOI: https://doi.org/10.1007/s00170-010-2671-x
27. Taehee Lee, and Tobias Hollerer. 2007. Handy AR: Markerless inspection of augmented reality objects using fingertip tracking. In Proceedings of 11th IEEE International Symposium on Wearable Computers (ISWC '07). IEEE, 83-90. DOI: https://doi.org/10.1109/ISWC.2007.4373785
28. Taehee Lee and Tobias Höllerer. 2009. Multithreaded Hybrid Feature Tracking for Markerless Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 15, 3 (May 2009), 355-368. DOI: http://dx.doi.org/10.1109/TVCG.2008.190
29. Blair MacIntyre, Alex Hill, Hafez Rouzati, Maribeth Gandy, and Brian Davidson. 2011. The Argon AR Web Browser and standards-based AR application environment. In *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality* (ISMAR '11). IEEE Computer Society, Washington, DC, USA, 65-74. DOI: http://dx.doi.org/10.1109/ISMAR.2011.6092371
30. 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 (May 2014), 40-45. DOI: https://doi.org/10.1145/2591689
31. Meredith Ringel Morris, Jacob O. Wobbrock, and Andrew D. Wilson. 2010. Understanding users' preferences for surface gestures. In *Proceedings of Graphics Interface 2010* (GI '10). Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 261-268.
32. Microsoft. Hologram Stability. Accessed March 23rd, 2018. https://docs.microsoft.com/en-us/windows/mixed-reality/hologram-stability
33. Microsoft. Kinect for Windows. Accessed January 3rd, 2018. https://developer.microsoft.com/en-us/windows/kinect
34. Microsoft. Skype for HoloLens. Accessed January 3rd, 2018. https://www.microsoft.com/en-us/hololens/apps/skype
35. Alex Olwal, Hrvoje Benko, and Steven Feiner. 2003. Senseshapes: Using statistical geometry for object selection in a multimodal augmented reality. In *Proceedings of The Second IEEE and ACM International Symposium on Mixed and Augmented Reality* (ISMAR '03). IEEE. 300-301. DOI: https://doi.org/10.1109/ISMAR.2003.1240730
36. Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (UIST '16). ACM, New York, NY, USA, 741-754. DOI: https://doi.org/10.1145/2984511.2984517
37. Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 1716-1725. DOI: https://doi.org/10.1145/2818048.2819965
38. Wayne Piekarski and Bruce H. Thomas. 2002. Using ARToolKit for 3D hand position tracking in mobile

outdoor environments. In *Augmented Reality Toolkit, The first IEEE International Workshop.* IEEE. DOI: https://doi.org/10.1109/ART.2002.1107003

39. Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. 1997. Image plane interaction techniques in 3D immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics* (I3D '97). ACM, New York, NY, USA, 39-ff.. DOI: http://dx.doi.org/10.1145/253284.253303
40. Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-Defined Gestures for Augmented Reality. In *IFIP Conference on Human-Computer Interaction* (INTERACT '13). Springer, Berlin, Heidelberg. 282-299. DOI: https://doi.org/10.1007/978-3-642-40480-1_18
41. Thammathip Piumsomboon, Adrian Clark, Atsushi Umakatsu, & Mark Billinghurst. 2012. Poster: Physically-based natural hand and tangible AR interaction for face-to-face collaboration on a tabletop. In *Proceedings of IEEE Symposium on 3D User Interfaces* (3DUI '02). 155-156. IEEE. https://doi.org/10.1109/3DUI.2012.6184208
42. Thammathip Piumsomboon, David Altimira, Hyungon Kim, Adrian Clark, Gun Lee, and Mark Billinghurst. 2014. Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In Proceedings of the *IEEE International Symposium on Mixed and Augmented Reality* (ISMAR '14). IEEE. 73-82. DOI: https://doi.org/10.1109/ISMAR.2014.6948411
43. Christina Pollalis, Whitney Fahnbulleh, Jordan Tynes, and Orit Shaer. 2017. HoloMuse: Enhancing Engagement with Archaeological Artifacts through Gesture-Based Interaction with Holograms. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction* (TEI '17). ACM, New York, NY, USA, 565-570. DOI: https://doi.org/10.1145/3024969.3025094
44. Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th annual ACM symposium on User interface software and technology* (UIST '96). ACM, New York, NY, USA, 79-80. DOI: http://dx.doi.org/10.1145/237091.237102
45. Gerhard Reitmayr, and Dieter Schmalstieg. 2004. Scalable techniques for collaborative outdoor augmented reality. In *Proceedings of the 3rd IEEE and ACM international symposium on mixed and augmented reality* (ISMAR'04). IEEE.
46. Gustavo Alberto Rovelo Ruiz, Davy Vanacken, Kris Luyten, Francisco Abad, and Emilio Camahort. 2014. Multi-viewer gesture-based interaction for omni-directional video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 4077-4086. DOI: https://doi.org/10.1145/2556288.2557113
47. 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). ACM, New York, NY, USA, 197-206. DOI: https://doi.org/10.1145/1978942.1978971
48. Christian Sandor, Alex Olwal, Blaine Bell, and Steven Feiner. 2005. Immersive Mixed-Reality Configuration of Hybrid User Interfaces. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality* (ISMAR '05). IEEE Computer Society, Washington, DC, USA, 110-113. DOI: http://dx.doi.org/10.1109/ISMAR.2005.37
49. Gerhard Schall, Erick Mendez, Ernst Kruijff, Eduardo Veas, Sebastian Junghanns, Bernhard Reitinger, and Dieter Schmalstieg. 2009. Handheld Augmented Reality for underground infrastructure visualization. *Personal Ubiquitous Comput.* 13, 4 (May 2009), 281-291. DOI: http://dx.doi.org/10.1007/s00779-008-0204-5
50. Teddy Seyed, Chris Burns, Mario Costa Sousa, Frank Maurer, and Anthony Tang. 2012. Eliciting usable gestures for multi-display environments. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces* (ITS '12). ACM, New York, NY, USA, 41-50. DOI: http://dx.doi.org/10.1145/2396636.2396643
51. Shaikh Shawon Arefin Shimon, Courtney Lutton, Zichun Xu, Sarah Morrison-Smith, Christina Boucher, and Jaime Ruiz. 2016. Exploring Non-touchscreen Gestures for Smartwatches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 3822-3833. DOI: https://doi.org/10.1145/2858036.2858385
52. Shaikh Shawon Arefin Shimon, Sarah Morrison-Smith, Noah John, Ghazal Fahimi, and Jaime Ruiz. 2015. Exploring User-Defined Back-Of-Device Gestures for Mobile Devices. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '15). ACM, New York, NY, USA, 227-232. DOI: http://dx.doi.org/10.1145/2785830.2785890
53. Adalberto L. Simeone, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2016. Three-Point Interaction: Combining Bi-manual Direct Touch with Gaze. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (AVI '16), Paolo Buono, Rosa Lanzilotti, and Maristella Matera (Eds.). ACM, New York, NY, USA, 168-175. DOI: https://doi.org/10.1145/2909132.2909251

54. Trimble, Inc.. SketchUp: 3D modeling for everyone. Accessed September 9th, 2017. https://www.sketchup.com/

55. Giovanni Maria Troiano, Esben Warming Pedersen, and Kasper Hornbæk. 2014. User-defined gestures for elastic, deformable displays. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (AVI '14). ACM, New York, NY, USA, 1-8. DOI: https://doi.org/10.1145/2598153.2598184

56. John Underkoffler and Hiroshi Ishii. 1999. Urp: a luminous-tangible workbench for urban planning and design. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (CHI '99). ACM, New York, NY, USA, 386-393. DOI: http://dx.doi.org/10.1145/302979.303114

57. Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. 2014. Leap gestures for TV: insights from an elicitation study. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (TVX '14). ACM, New York, NY, USA, 131-138. DOI: https://doi.org/10.1145/2602299.2602316

58. 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). ACM, New York, NY, USA, 1325-1334. DOI: https://doi.org/10.1145/2702123.2702223

59. Andrew D. Wilson and Hrvoje Benko. 2010. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology* (UIST '10). ACM, New York, NY, USA, 273-282. DOI: https://doi.org/10.1145/1866029.1866073

60. Andrew D. Wilson and Hrvoje Benko. 2017. Holograms without Headsets: Projected Augmented Reality with the RoomAlive Toolkit. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '17). ACM, New York, NY, USA, 425-428. DOI: https://doi.org/10.1145/3027063.3050433

61. 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). ACM, New York, NY, USA, 1869-1872. DOI: http://dx.doi.org/10.1145/1056808.1057043

62. 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* (CHI '09). ACM, New York, NY, USA, 1083-1092. DOI: https://doi.org/10.1145/1518701.1518866

63. Yan Xu, Sam Mendenhall, Vu Ha, Paul Tillery, and Joshua Cohen. 2012. Herding nerds on your table: NerdHerder, a mobile augmented reality game. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '12). ACM, New York, NY, USA, 1351-1356. DOI: http://dx.doi.org/10.1145/2212776.2212453