Did I Do It Differently? Comparing Gesture Elicitation in Virtual and Augmented Reality

Jieqiong Li*
Colorado State University

Zahra Borhani†
Colorado State University

Matthew Buckman‡
Colorado State University

Adam Sinclair Williams§
Colorado State University

Yanlin Guo¶
Colorado State University

Mayra Donaji Barrera Machuca‖
University of Calgary

Andreea Danielescu**
Wordly, Inc, Seattle

Francisco Raul Ortega††
Colorado State University

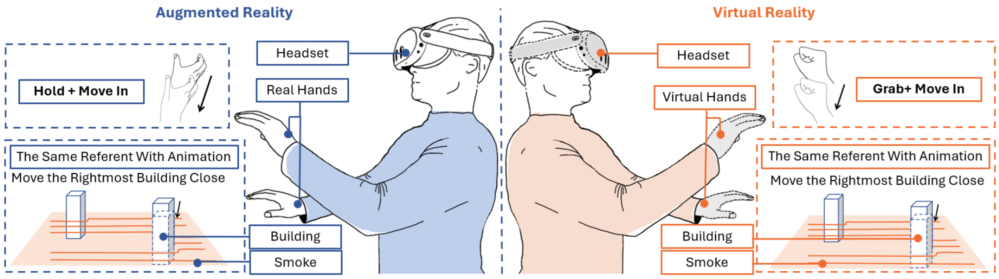

Figure 1: In both studies, the participant wears the headset and designs gestures after being presented with the referents with animation (e.g., "Move the Rightmost Building Close"). Both studies used the same participants, referents and environment (two buildings and smoke). Left: Augmented Reality (AR) study environment, where participants view their own hand. Right: Virtual Reality (VR) study environment, where participants view a virtual hand.

Abstract

Gesture elicitation is a common way to map commands, e.g., actions or referents, to gestures by reaching an agreement between participants. However, it is uncertain how gestures elicited in Virtual Reality (VR) transfer to Augmented Reality (AR) and vice versa. Using a wind-simulation virtual environment that supports gesture interaction with concepts, we aim to explore and compare gesture-elicitation results from participants in AR and VR for the 33 referents presented. By analyzing gesture elicitation for different types of referents across the two technologies within five referent categories, this study shows that, except for the translation category, participants' gestures in AR and VR were highly similar. The specific result for the translation category is likely due to its reliance on the physical surface. This also provides a foundational context for deeper understanding in future research.

Index Terms: Virtual Reality, Augmented Reality, Gestures, Elicitation Study, Human Computer Interaction

1 Introduction

Elicitation studies are widely used in Human-Computer Interaction (HCI) [7, 50, 61], serving as a participatory, human-centered design method to identify the action most commonly agreed upon by participants [53, 58]. For gestures, past work has used elicitation studies in Augmented Reality (AR) [25, 45] and Virtual Reality (VR) [18, 38] to identify gestures for three-dimensional (3D) object manipulation, such as translation, rotation, and scaling, as well as system-level gestures like menu switch in TV. However, **an open question is whether gesture sets are consistent and available across both AR and VR**. Previous AR/VR comparison studies [21, 22, 23] examined technical transferability; they overlooked whether gesture design would differ across levels of immersion, which distinguishes our work. Li et al. also examined differences in immersion between AR and VR, which may lead to distinct preferences among user groups [26]. **Our research fills this gap by investigating the consistency and availability of gestures in AR and VR**, as the similarity of gestures across technologies not only affects the continuity of the user's experience across modalities [22] but also influences the learnability of interaction methods [61]. Therefore, understanding whether users' gesture preferences across AR and VR are similar or different is important for designing novel gestural interactions. To our knowledge, we are the first to examine this problem using gesture elicitation in AR and VR. The closest work compared real-world gestures with VR gestures using elicitation studies [40].

Our first research question (RQ_1) examines *What similarities and differences are found in hand gestures when the same elicitation study is performed in AR versus VR?* Our goal is to understand the influence of technology on users' gesture choices during an elicitation study. By identifying where user preferences converge and diverge across modalities, our findings provide theoretical support for designing unified XR interaction patterns, thereby serving as a stepping stone for exploring ways to reduce user learning costs during transitions between AR and VR. Our second research question (RQ_2) investigates *What gestures are commonly agreed upon by participants across VR and AR?*, aiming to assess the consistency of gestures identified in the elicitation study by comparing the most frequent gestures in both VR and AR. Finally, research question RQ_3 aims to understand users' perceptions of each technology and asks: *How would users perceive the gestures they generated after using AR and VR?*.

The main contributions of this work are as follows: (1) A key novelty of our experimental design is its direct comparison of gesture elicitation in AR and VR, which, to the best of our knowl-

*e-mail: jacinda.li@colostate.edu
†e-mail: zahra.borhani@colostate.edu
‡e-mail: buckmanm@colostate.edu
§e-mail: adam6457@gmail.com
¶e-mail: yanlin.guo@colostate.edu
‖e-mail: mbarrera@ucalgary.ca
**e-mail: andreea.danielescu@wordly.ai
††e-mail: F.Ortega@colostate.edu

edge, has not been explored in prior work. (2) By analyzing semi-structured interviews (using thematic analysis), this study investigated the factors influencing users' perspectives on gesture design when using AR vs VR. This helped us to understand participants' primary requirements for gesture design across both technologies from their perspective. (3) A discussion of future design opportunities for technology-adaptive gestures, along with design recommendations.

2 RELATED WORK

2.1 Gesture Preference

In the context of the elicitation methodology, several factors can influence the gestures users create for the same action or referent, including cultural background [29], physical ability or motor impairments [2], age [14], number of objects [54], and the specific technology or interface being used [61]. Cultural similarities and differences can play a role in user-defined gestures, e.g., a study across nine countries found a generally high level of agreement in the use of gestures on small, handheld touchscreen interfaces across cultures for 28 common actions like "zoom" and "copy" [29]. Another study found that wheelchair users preferred on-body and in-air gestures over on-wheelchair gestures, often involving touch input on the body or mid-air hand poses [2].

Elicitation studies have found that users' gesture preferences vary across interaction modes, such as desktop interaction [61], controller-based interaction [18], and mid-air interaction [19]. In *AR*, freehand gestures are widely employed, and users prefer one-handed gestures over two-handed gestures for single-object manipulations, and both hands for multi-object interactions [37, 38, 54]. Wobbrock et al. showed that users rarely cared about the number of fingers they used and generally preferred one hand over two for tabletop interactions [61]. On the other hand, in *VR*, natural body-based movements are widely employed to enable immersive experiences. However, many input techniques still rely on handheld controllers due to the greater fatigue caused by prolonged mid-air hand gestures [19]. We did not find prior work that directly compares AR- and VR-elicited gestures in the same environment.

2.2 Gestures Elicitation

In HCI, elicitation studies [51, 58, 61] have been extensively studied (e.g., [7, 50]), including in the domain of education [64] to understand user-centered interaction scenarios. This method, introduced by Wobbrock et al. [61], involves collecting and analyzing data on user behavior and interaction preferences. To achieve consensus, participants are asked to propose spontaneous gestures that correspond to the referents (i.e., actions) presented to them. The referents can be in text form [56], animation form [30, 57], and image form [5, 43], among others. For example, the referents can be prompts such as "move the object upward" or "delete the object" in text form. The primary advantage of playing animations is that they provide users with additional information, enhancing their understanding of the referents [31].

Elicitation studies can be used in various systems, such as mobile phones [44], televisions [52], and bicycles [48], among others [30]. Elicitation studies also aid in understanding user interaction preferences across different environments, such as full-body interactions for children [50], nose micro-gestures [42], and foot gestures [11]. Hence, this method can be used to explore suitable designs for users and ensure that these designs are more intuitive and natural than those designed without user input [61]. In addition, gesture designs in elicitation studies are also diverse, such as microgestures [4, 46] and freehand gestures [28, 62].

2.3 Comparisons between AR and VR

Past work has identified differences between AR and VR based on users' experiences. Comparisons between AR and VR have been explored across various fields, such as STEM education [1] and exposure therapy [49]. According to the Reality-Virtuality Continuum proposed by Milgram [33], AR and VR lie along a spectrum ranging from the physical environment to a fully virtual one. AR enriches the physical world by overlaying digital information [16]. In contrast, VR provides an immersive virtual space that isolates users from their physical surroundings [9, 59]. Starting from Ellis's early discussions on the interaction between AR and VR [8], studies have shown that the two technologies still have certain similarities in information-seeking tasks [23] and software usability [22].

In the field of interaction research, Perera et al. compared studies on mid-air gestures for smart-home control commands in real and virtual environments, and found that the two environments were generally similar in terms of usability and user experience [40]. However, participants in the VR condition reported a higher sense of immersion [40]. Li et al. pointed out that differences in immersive experience, interaction forms, and other factors between AR and VR can influence users' feelings of interaction [26]. For example, users in AR pay attention to their presence in the real world; however, VR provides a more immersive experience [26]. Moreover, Gaffary et al. showed that participants exhibited different biases in object perception in AR than in VR [12]. Specifically, participants perceived objects in the AR environment as softer to the touch than in the VR environment [12]. Although prior work has demonstrated differences in user experience between AR and VR, and in gesture elicitation across real and virtual environments, it remains unclear whether AR and VR exhibit similar or distinct patterns of gesture-based interaction in elicitation studies.

3 ELICITATION STUDY

The study was approved by Institutional Review Board (IRB).

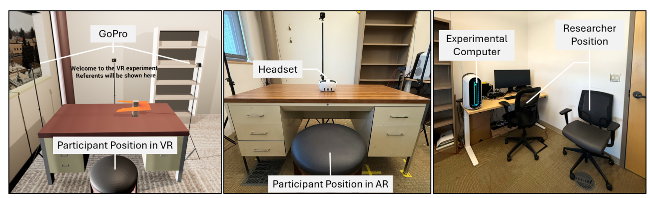

Figure 2: Experimental setup: Left: Participant Area in VR. Middle: Participant Area in AR. Right: Experimenter Area.

3.1 Methodology

Experiment Design The study employed a within-subjects gesture elicitation design with a single independent variable: AR or VR modality. Video see-through AR technology was used for the AR condition. In both conditions, participants proposed gestures for 33 referents within a controlled experimental environment. Controlled within-subjects designs are standard practice in gesture elicitation research [40, 56, 63]. To mitigate potential order effects, participants were randomly assigned to one of two counterbalanced groups (AR-first or VR-first). In addition, the evaluation tasks were organized into four referent sets (RS1–RS4). RS1–RS4 are counterbalanced referent order sets used in a Latin-square design to distribute the 33 referents and reduce order and learning effects.

Participants We recruited 16 participants (eight females, eight males, no other gender reported) for the experiment from undergraduate and graduate students, faculty, or alumni of Colorado State University. We estimated their age at 27.6 years using interval midpoints of age intervals [10]. All participants reported normal or corrected-to-normal vision. When recruiting participants, we did not require them to have any specific domain knowledge. Fourteen participants had some knowledge of VR, while two had never used VR devices before. Among participants with prior VR experience, 85.7% had used a VR headset 2–3 times per year, 7.1% every 2-3 months, and 7.1% 1–3 times per week. Regarding AR experience, five participants had some knowledge, eight were unfamiliar, and three reported no experience. Among the participants with prior AR experience, 80% had used an AR headset 2–3 times per year, while 20% had used it 1–3 times per week. Upon completing the study, participants received either a $20 US gift card or a 1% related class credit as compensation. Some participants were students who opted to receive course credit.

Environment We conducted the study in a dedicated laboratory space, divided into separate areas for the participant (Participant Area) and the experimenter (Experimenter Area). A desk (approximately 5×2.5 feet), a chair, and 3 GoPro cameras with tripods were set up in the participant area (approximately 12×10 feet). To capture the user's gestural data and interview content during the experiment, three stationary GoPros were set up at the front, left, and right sides of the desk. The experimenter controlled the application scenario, compared the data, and conducted the interviews in the experimenter area. Further details regarding the spatial layout are provided in Figure 2. Note that the idea was to create a digital twin of the experimental lab space in VR, so users could have a very similar environment when viewing AR alongside the real desks.

Apparatus We developed both AR and VR environments in Unreal Engine 5.2.1 using Unreal Meta XR Plugin 57.0 and the Blueprint scripting language. These tools were used to implement hand tracking and AR passthrough on the Meta Quest 3 headset[1]. Finally, we used a desktop computer with an Intel(R) i9-11900F processor and an NVIDIA GeForce RTX 3090 GPU for development and running the user study.

As shown in Figure 3, which is a screenshot of a participant's view during the VR and AR conditions, both modalities presented a digital wind simulation with text-based referents. This simulation included virtual buildings, smoke, and emitters. We developed the 3D wind simulation model using an in-house dataset collected from real-world wind conditions. In Unreal Engine 5.2.1, 15 emitters generated flowing particles to simulate the visual effect of smoke. The model supported dynamic visualization and physics-based obstacle interaction in both VR and AR. For example, when participants moved a virtual building, the direction of smoke updated in real time. The main difference between VR and AR conditions was how the surrounding environment and participants' hands represented. In VR, we recreated the physical room from the real environment using Blender 3.6 and included left and right virtual hands. In AR, participants viewed the physical environment through video passthrough and interacted using their real hands. In addition, the software collects the following data in the background: 4K video of the 3D gestures from three viewpoints (i.e., front, left, and right), and the trajectory of finger joints in the gesture data space.

The decision for using a wind simulation environment was to go beyond elicitation studies that have used either a single object (e.g., a cube) or multiple objects [56, 63]. We designed this environment (see Figure 3) to be richer in interaction and usable later as an application, thereby increasing the study's ecological validity. While our research project did not consider the educational aspects of this environment, the application we used may allow people to understand wind forces.

The wind simulation provides participants with a dynamic and intuitive representation of object movement. By observing changes in wind force and the resulting movement of the object, participants can clearly understand the expected outcome of gesture-based manipulation [36]. The continuous visual cues in the scene allow participants to rely on their understanding when designing gestures. The direction and color of the wind also change according to the

[1] https://www.meta.com/quest/quest-3/

Figure 3: The view the participant sees during the study. The left half is the AR environment, while the right half is the VR environment. The environments include a building, emitters, smoke, and, for the VR scenario, a virtual hand inside the virtual environment.

object's state. For example, when the wind interacts with the object, its color shifts to red Figure 3. Rather than manipulating a single object, the simulation presents the entire process of wind forces and object motion in a simplified environment. This helps participants to develop a clearer understanding of the task and infer the action goal from the overall context. These features make the wind simulation particularly suitable as a referent for elicitation gesture-design tasks.

Referents We elicited 33 referents (i.e., commands or actions). Each referent's name and a sample animation graph are shown in Figure 4. The references were selected for their potential to facilitate manipulation and understanding of the wind simulation environment. These referents include canonical manipulation referents (i.e., selection, translation, rotation, and scaling) [20, 25, 56], as well as non-canonical referents needed for the wind simulation system, necessitating adaptations to the rotation and selection referents. For example, rotating the building and emitters clockwise and selecting the three emitters closest to the participants. Moreover, to expand the types of referents, we included referents that support multi-object manipulation. For example, "Scale Building and Emitter Down" and "Yaw Building and Emitter Clockwise". Furthermore, we included abstract referents based on realistic interaction scenarios, such as "Change the Color", "Open the Chart", and "Close the Chart".

Procedure The study required each participant to complete two experimental sessions across two visits: Session 1 (S1) on Day 1, and Session 2 (S2) on Day 4. Day 2 and 3 served as rest days, during which participants followed their everyday routines. The gap in experimental conditions was explicitly introduced to reduce the impact of the elicitation provided in the first environment on the second. Each elicitation session lasted between 1.5 and 2 hours.To mitigate potential order effects, participants were randomly assigned to one of two counterbalanced groups: AR-first or VR-first.

In the first experimental session (day 1), participants gave informed consent and completed a demographic questionnaire. Then, the researcher explained the details of the user study and assisted them in wearing the VR Head-Mounted Display (HMD). Afterward, participants did two practice tasks to familiarize themselves with the application environment, gesture interaction, animation flow, and reference display format. During practice, participants were encouraged to use the "think-aloud" protocol to choose their preferred gestures from the ones suggested. After participants became familiar with the prototype, they proceeded to the elicitation phase, where they provided gestures for all 33 referents in response to animation and text prompts. Each animation played five times in a loop, but participants could request more loops if needed. During this phase, the researcher refrained from interfering with the participants' elicitation, e.g., asking to explain a gesture or choose

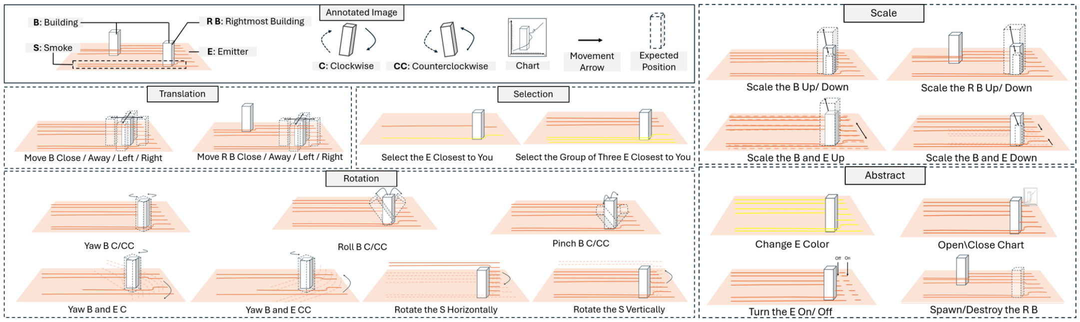

Figure 4: Example referents. R: Rightmost, B: Building, S: Smoke, E: Emitter, C: Clockwise, CC: Counter-Clockwise.

one. There was also no time limit, as participants requested the next referent using words (e.g., "next", "done").

In the second experimental session (day 4), participants completed a procedure similar to the first session, but used the other modality to visualize the wind simulation. After the elicitation phase, participants completed a 7-point Likert survey on system usage, which included evaluations of their comfort, understanding, and clarity regarding the experimental content in the two environmental scenarios (AR/VR). Participants also completed a recorded semi-structured interview in which they shared their thoughts and perspectives on the gestures they designed as they reviewed them. We divided the interview into two parts. The first part involved reviewing gestures in a single scene, where participants only needed to review either the content of the first or second experimental day to assess the impact of gesture design (such as animation, environment, experience). The second part involved analyzing and discussing specific gestures (i.e., different or novel gestures) across both experimental days. Participants could review gestures by watching videos or listening to verbal descriptions provided by researchers. They discussed the reasons for specific gestures and the factors influencing their design.

3.2 Gesture Coding

For gesture coding, we first selected the gestures by determining the final gesture for each referent provided by each participant. If a participant proposed only one gesture for a referent, we accepted it directly. For participants who provided multiple gestures for a single referent, we selected the gesture they verbally chose during think-aloud. This could allow us to capture their gesture choice directly. If they did not select any gesture, we chose the first completed gesture and ignored the others, including repeated or incomplete gestures. We followed the approach used in prior elicitation work [7], selecting the first gesture to simplify the analysis, as it conveys participants' intentions better. We do acknowledge that Morris et al. proposed options to reduce legacy [34], yet Williams et al. have shown that there is limited benefit in production for reducing it [55], which was another reason to follow our approach of taking the first gestures for those that didn't indicate otherwise. It was also very important that users not be interrupted at this moment, as we had asked them to think aloud.

After completing the gesture selection, we coded each gesture. To reduce subjectivity in gesture coding and ensure the reliability and effectiveness of the qualitative coding process, three researchers employed a consensus-based coding approach [7, 32]. Two researchers coded the gesture data, and a third researcher led them through the gesture classification process, including discussions and adjustments to the codebook. The two researchers independently coded all gesture images for 16 participants. After completing the coding for two participants, they would discuss and compare similarities and differences in coding. If disagreements

Table 1: Representative categories from the gesture codebook. We identified 29 hand-gesture types; but only six of them are added to the table as examples.

Category	Values
Hand gesture	Tap, Grab, Hold, Pinch, Swipe, Point, ...
Active hand	Single hand, Both hands
Movement	Move (6 orientations), Rotate (6 orientations)

arose during the discussion, we reviewed the data and revised as needed, refined the coding method, or developed a new gesture code. Based on the agreed-upon coding results, a code book was created and iteratively refined. The third researcher tracked and reviewed the entire process.

We collected 1,056 gestures from 16 participants who performed 33 reference gestures in both VR and AR scenarios (33 gestures × 2 days × 16 participants). Each gesture was independently coded, and a gesture coding manual was compiled (as shown in Table 1). If multiple gestures were consistent in terms of hand gesture, active hand, and movement characteristics, they were considered as the same gesture. We identified 178 unique gestures, following a common coding approach used in elicitation studies [61, 63] (see Figure 5 for details). During the discussion, ambiguous gesture names were clarified. For example, the difference between gestures "hold" and "grab" lies in whether the fingers present dynamic convergence. Since this paper does not delve into the differences between dominant and non-dominant hands in gestures, only single-handed or two-handed operations were considered in the active hand. The seven directions of orientation were named based on the gesture's location. For movement, the participant had hand or arm movement in space when performing the gesture.

3.3 Data Analysis

Agreement Rate Calculation We utilized the AGATe 2.0 tool (Agreement Analysis Toolkit)[2] to compute agreement among participants, including the agreement rate $\mathscr{AR}$ and (V_{rd}) significance test. The agreement rate, $\mathscr{AR}$, signifies the fraction of participant pairs that concur with each other out of all potential pairs that could possibly agree. For a single referent r, this is formally defined according to Equation 3.3, where P denotes all proposals for referent r and P_i represents subsets of equivalent proposals from P. We followed Vatavu et al.'s marginal agreement rates and interpretations to evaluate the level of consensus among participants for each referent [51]. An agreement rate above 0.30 represents high agreement, below 0.10 indicates low agreement, and between 0.10 and 0.30 is defined as medium agreement. Figure 6 presents the comparison

[2] http://depts.washington.edu/acelab/proj/dollar/agate.html

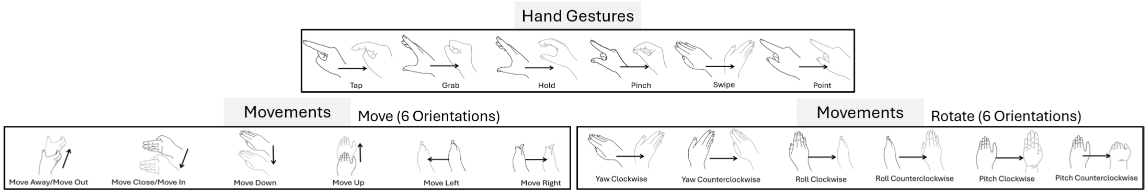

Figure 5: Hand gesture, gesture orientation, and gesture movement example diagram from code book.

of gesture agreement rates for AR and VR conditions using gesture data. Two dotted lines in this figure indicate the agreement rate [51].

$$\mathscr{AR}_r = \frac{\sum_{P_i \subseteq P} \frac{1}{2}|P_i|(|P_i|-1)}{\frac{1}{2}|P|(|P|-1)}$$

Statistical Analysis To analyze the results of a 7-point Likert survey, Cools et al. (2021) used the Wilcoxon Signed-Rank test to compare user experience between AR and VR across 5 dimensions [6]. A one-way analysis of variance (ANOVA) was used to analyze the scene sequences in the experiment [47]. To analyze the gesture agreement rate results, we used the Shapiro-Wilk (SW: $\mathscr{W}$) test to assess whether the data were normally distributed. For normally distributed data (translation and abstract categories), we used a t-test to assess whether the environment affected the gesture agreement rate. For non-normal data (all, scale, and rotation categories), we used ART (Aligned Rank Transform) ANOVA from ARTool (Align-and-rank data for nonparametric factorial ANOVA)[3], calculated using R version 4.5.0 to evaluate whether there were statistically significant effects [60]. Finally, we removed the selection category due to its limited sample size (n=2), which falls below the minimum threshold required for valid ANOVA testing (>=5 data points per group).

4 Results

4.1 System Usage Analysis

A Wilcoxon Signed-Rank test was conducted on the 7-point Likert survey for each dimension in both environments. Participants reported consistency across dimensions such as usability, comfort, animation quality, task clarity, and motion mapping (e.g., real hand in AR and virtual hand in VR) during the AR and VR experiments. All of the dimensions show no significant difference between the AR and VR environments. 62.5% of the participants believed that the virtual hand in VR could follow the real hand's movements well. 75% of the participants indicated that they could clearly see their real hands during the AR experiment. In addition, over 80% of participants thought the animations and task instructions were clear and understandable in both AR and VR environments.

4.2 AR and VR Condition Order Effect

In 16 participants, 8 (i.e., group 1) completed the AR experiment first, followed by the VR experiment, while the other 8 (i.e., group 2) completed the experiments in reverse order. To investigate whether completing the AR or VR condition first influenced the gesture produced, a one-way ANOVA was conducted on gesture consistency between the two groups. The measure of gesture consistency for assessment is the number of identical gestures across AR and VR for the 33 referents each participant performed. The analysis revealed no significant difference between the two groups ($\mathscr{F}(1,14) = 1.03$, $p = 0.33$).

[3] https://depts.washington.edu/acelab/proj/art/

Table 2: The columns labeled $\mathscr{AR}_{AR}$ and $\mathscr{AR}_{VR}$ show the average gesture agreement rates for each referent category in AR and VR. The final column reports the results of the ART ANOVA for non-normally distributed data and the t-test for normally distributed data. Statistically significant results are shown in bold.

Referent Category	$\mathscr{AR}_{AR}$	$\mathscr{AR}_{VR}$	AR vs. VR
All Categories	0.20	0.18	$\mathscr{F}(1,64) = 0.640,\ p = 0.43$
Rotation	0.26	0.24	$\mathscr{F}(1,18) = 0.680,\ p = 0.42$
Scale	0.09	0.10	$\mathscr{F}(1,10) = 0.006,\ p = 0.94$
Translation	0.34	0.25	$t(14) = 2.26,\ \mathbf{p < 0.05}$
Abstract	0.07	0.06	$t(12) = 0.28,\ p = 0.79$

4.3 Gesture Agreement Across AR and VR

On average, the agreement rate was slightly higher in AR ($\mathscr{AR}_{AR}$=0.20) than in VR ($\mathscr{AR}_{VR}$ = 0.18) (as shown in Figure 6). The influence of the referent type on the $\mathscr{AR}$ was significant for each technology (VR: $V_{rd(32,N=528)} = 661.527$, $p < 0.001$, AR: $V_{rd(32,N=528)} = 155.295$, $p = 0.001$). When comparing both modalities, we observed a non-significant difference in $\mathscr{AR}$ ($\mathscr{AR}_{AR}$ = 0.20 versus $\mathscr{AR}_{VR}$ = 0.18, see Table 2). A further analysis of the different categories (translation, rotation, scaling, selection, and abstract) showed only a significant difference in translation. The $\mathscr{AR}$ for translation gestures shows that gestures elicited in AR differ from those elicited in VR.

4.4 Winning Gestures

We identified the winning gestures for each referent in both the AR and VR environments. For each referent, every participant contributed one primary gesture (as described in subsection 3.2). The winning gesture (Top gesture) was defined as the gesture proposed most frequently among participants for each referent, representing the highest level of consensus. The Top-2 gesture was the second most frequently chosen gesture. Table 3 shows the winning gestures and their agreement rates for both environments.

Our analysis showed that 73% (52% for Top-2) of referents had matching top gestures across AR and VR. These trends are visualized in the highlighted table shown in Figure 7, where each cell indicates whether the Top or Top-2 gesture for a given referent matched across AR and VR (1 = match, 0 = no match). To further explore individual consistency, we calculated the average number of participants (68.2%) who used the same gesture for each referent across both conditions. These results are visualized in Figure 8.

4.5 Hand Usage in AR and VR

To investigate differences in finger and hand usage between AR and VR, we analyzed the most frequently used hand configurations among participants across different environments. Among hand gestures, the "hold" (28% in VR, 32% in AR), "grab" (14.4% in VR, 14.4% in AR), and "push" (9.5% in VR, 8.5% in AR) are the three most frequently used in both AR and VR. For hand use, single-hand and whole-hand operations remained the main modes in both AR and VR.

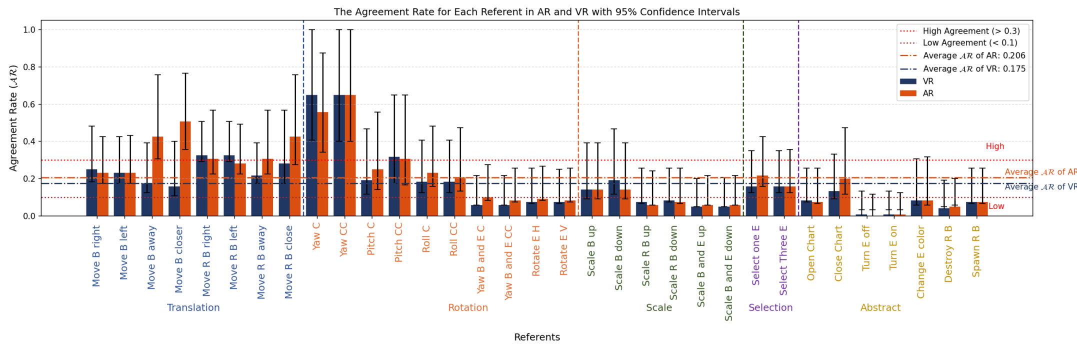

Figure 6: Agreement Rate in Virtual Reality and Augmented Reality for Each Referent with 95% Confidence Intervals. $\mathscr{AR}$: Agreement Rate, B: Building, E: Emitter, S: Smoke, R: Rightmost, C: Clockwise, CC: Counter Clockwise, V: Vertically, H: Horizontally.

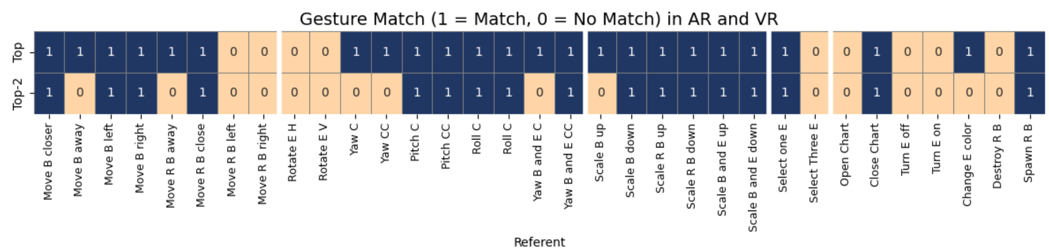

Figure 7: Summary of gesture matches across referents, where "Top" refers to the winning gesture and "Top-2" indicates the second most frequently selected gesture. B: Building, E: Emitter, S: Smoke, R: Rightmost, C: Clockwise, CC: Counter Clockwise, V: Vertically, H: Horizontally.

4.6 Interview Findings

Semi-structured interviews were conducted at the end of the experiment (Day 4), yielding 446 sentences of participant feedback on gesture design across AR and VR. Two authors independently performed a thematic analysis [3] using the Affinity Diagramming method [27]. Each author reviewed the transcripts, identified recurring patterns, and organized the data into preliminary themes to ensure unbiased interpretations. After completing their individual analyses, they compared their findings and collaboratively refined the themes through deliberation.

The analysis showed that participants' gesture design decisions in both AR and VR were shaped by multiple factors: uniqueness (50%), simplicity (44%), naturalness (19%), comfort (13%), animation task (94%), and prior interaction experience (37%). In addition, participants shared the sources of their gesture design inspiration. For example, integration with motion effects, device interaction experience, and familiar real-world gestures.

Uniqueness Half of the participants emphasized the importance of designing distinct gestures for different tasks. They often modified gestures based on task-specific nuances, such as using one hand to select a single object versus two hands to select multiple objects. This aligns with findings from previous studies [63, 39]. Conversely, one participant emphasized the value of consistency for similar tasks, explaining: "For each different task, I was looking for correlations between gestures and tasks. For example, if you rotate pitch up and down, I would consider using similar gestures" (P7).

Memorability Nearly half of the participants (44%) considered gesture simplicity an important factor in their design process, valuing it for both ease of execution and memorability. As P14 indicated, a simple gesture is easier to remember. Furthermore, simple gestures may help manage high-demand tasks (i.e., those requiring multiple steps or actions [56]). For example, P9 mentioned that "*I tried to think about how to select all of them at the same time.*"

Naturalness Three participants stated that when considering gesture design, they would like to use some natural gestures. These natural gestures can help them understand and complete some abstract referents. For example, most participants mentioned that it was hard to design a unique gesture for the referents of turning the emitters on and off. So some of them pushed the desk as if they were pressing a button to turn the emitters on or off. P4 mentioned, "*I think the gesture I did today, like using a button to turn on and off, ... is more natural.*"

Comfort Although gesture comfort was considered in prior work as a factor in gesture interaction [24, 39], only two participants reported considering gesture comfort during the design process. For instance, P16 mentioned that optimizing hand comfort is the primary consideration.

Referent Type The majority of participants (94%) reported that their gesture design was primarily influenced by the task requirements, particularly the need to demonstrate animations. For example, P4 emphasized the importance of the animation, stating,"*..., I can see the building movement,..., If the animation changes [even if there is a minor change], my gesture will change too.* " Additionally, 25% of participants mentioned that they focused more on interpreting the meaning of the animation rather than strictly following its motion. As P6 explained, "*I think I can move or do anything I want [after getting the meaning of animation].*"

Prior Interaction Experience Two participants indicated they would use their experience with physical manipulation. For example, if the object is very large or looks like a cup with a handle, the participants will adjust their gestures to match its shape. This has been validated previously by Zhou et al., who showed that when there is more than one object (e.g., a cube vs. a more complex environment), users rely on physical manipulations [63]. For instance, they will hold the handle rather than just the cup. P11 mentioned this during the interview; "*The shape and perceived sturdiness of the objects it affect the gestures, at lease the movement gestures.*"

Table 3: Agreement Rates in VR and AR. $\mathscr{AR}$: Agreement Rate, R: Rightmost, B: Building, E: Emitter, C: Clockwise, CC: Counter-clockwise, H: Horizontally, V: Vertically.

Referent/Mode	VR Gestures	$\mathscr{AR}$	AR Gestures	$\mathscr{AR}$
Move the B Close	hold+move in	0.158	hold+move in	0.508
Move the B Away	push away\hold+move out	0.175	hold+move out	0.425
Move the B Left	push left	0.233	push left	0.233
Move the B Right	push right	0.250	push right	0.233
Change the E Color	tap	0.083	tap	0.083
Rotate the E H	grab+rotate pitch	0.075	hold+rotate pitch	0.092
Rotate the E V	grab+rotate pitch	0.075	hold+rotate pitch	0.083
Select One E	pinch in	0.158	tap\pinch in	0.258
Select Three E	pinch in	0.158	grab	0.158
Yaw C	hold+rotate yaw	0.650	hold+rotate yaw	0.558
Pitch CC	hold+rotate pitch	0.317	hold+rotate pitch	0.308
Roll CC	hold+rotate roll	0.183	hold+rotate roll	0.208
Scale the B Up	pinch out	0.142	pinch out	0.142
Turn the E Off	pinch in	0.008	pinch in\others (15 gestures)	0
Turn the E On	pinch in	0.008	swipe right	0.008
Scale the R B Up	pinch out	0.075	pinch out	0.058
Move the R B Away	push away\hold+move out	0.217	hold+move out	0.308
Move the R B Close	hold+move in	0.283	hold+move in	0.425
Move the R B Left	push left	0.325	hold+move left	0.283
Move the B Right	push right	0.325	hold+move right	0.308
Scale the R B Down	pinch in	0.083	pinch in	0.075
Destroy the R B	push down	0.042	clap	0.050
Spawn the R B	pinch out	0.075	pinch out	0.075
Yaw CC	hold+rotate yaw	0.650	hold+rotate yaw	0.650
Pitch C	hold+rotate pitch	0.192	hold+rotate pitch	0.250
Roll C	hold+rotate roll	0.183	hold+rotate roll	0.233
Scale the B Down	pinch in	0.192	pinch in	0.142
Open Chart	tap	0.083	pinch out	0.075
Close Chart	tap	0.133	tap	0.200
Scale B and E Up	two hand grab+zoom out \two hand hold+zoom out	0.050	two hand grab+zoom out \two hand hold+zoom out	0.058
Scale B and E Down	two hand grab+zoom in \two hand hold+zoom in	0.058	pinch in\two hand grab+zoom in \two hand hold+zoom in	0.075
Yaw B and E C	hold+rotate yaw	0.067	hold+rotate yaw	0.117
Yaw B and E CC	hold+rotate yaw	0.067	hold+rotate yaw	0.100

Or they would like to find some inspiration from the real world. P7 created a ”taking photo” hand to complete the scale-up task. P10 and P14 mentioned that they considered the gestures they used in their life when designing gestures in VR.

In addition, when designing gestures for abstract tasks that may not have direct physical manipulations for the participant to draw from, such as changing color, participants turned to metaphorical actions or imagined tools [58], such as swiping as on TikTok or using virtual menus. P5 mentioned that “*to change the color, I used a swipe gesture, just like swiping videos on TikTok. In this way, I can view the next color.*” In addition, they might resort to imagined tools to help them complete the tasks. P11 mentioned that “*when I don't know what kind of gesture it should be, I will use the menu. ... kind of like an abstract property.*”

4.6.1 The Influence of AR and VR on Gesture Design

Environment A total of 75% of participants mentioned that the environment influenced the gestures they designed. Three participants (P2, P4, and P6) felt that VR is more like a computer game due to its fully digital nature. P4, P6, and P8 mentioned that this digital environment not only brings them fun but also makes them feel they can design gestures more easily than in AR. P14 mentioned that compared to AR, designing the gesture in VR made“*it easier to immerse* ” themselves in the environment. P8 echoed this sentiment but contrasted it with their experience in AR, stating: “*[When designing the gestures in AR,] I'm thinking more about how the gesture would be for all three of us (i.e., the participant and two researchers) rather than just for myself.*” For the AR, P1, P3, P5, and P15 described AR as more natural than VR, with P8 emphasizing how AR's connection to the physical world made them reflect more deeply on the source of their gestures: “*I can think more about whether the gesture is from reality or from myself.*” Overall, gestures were often grounded in how participants felt within each environment. Although participants reported that the environment affected their design decisions, this is not the only factor impacting their gesture design.

Hand Representation To understand whether participants' experiences with hand interactions, through either real or virtual hands, influenced how they approached gesture design, we asked them to share their opinion. For some, the mapping between their hand movements and the digital environment felt unfamiliar. P16 mentioned, “*[In VR,] I noticed that the interaction is very close to me. But I can still feel disembodied from the action I was doing.*” Three participants had uncomfortable feelings at the beginning of the use of the virtual hand. P16 described this as, “*the [virtual] hand movement is like a relearning movement.*”

For nearly half of the participants (56%), interacting with real hands felt more comfortable than interacting with virtual hands. As P4 mentioned that when using the real hand to “*interact with virtual objects, it feels very strange.*” P6 often used the palm in AR, but used the index finger or pinch (i.e., Thumb-and-index gesture) in VR. In contrast, participants described interactions in VR as more seamless. With both the hand and objects existing in the same digital space, the coordination felt more fluid. P14 described this as, “*I felt like it [using the virtual hand to grab a virtual object] was more tunable [than using a real hand to grab a virtual object].*”

5 DISCUSSION

We conducted a user study with 16 participants to examine differences in gestures elicited in VR and AR. Our participants elicited

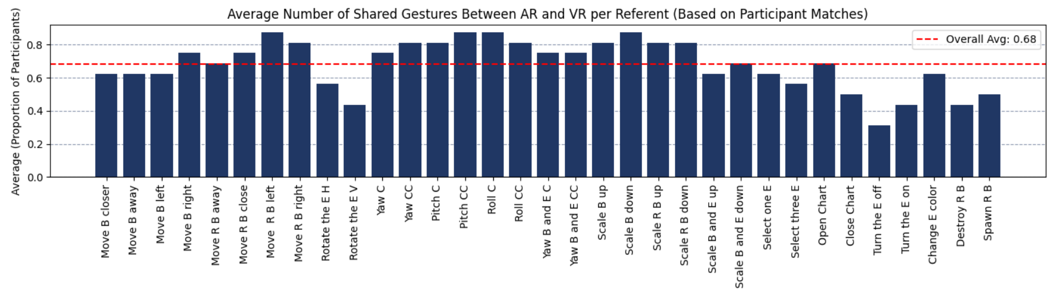

Figure 8: Gesture match summary across participants for each referent. B: Building, E: Emitter, S: Smoke, R: Rightmost, C: Clockwise, CC: Counter Clockwise, V: Vertically, H: Horizontally.

33 referents using a wind simulation. In this section, we discuss our three research questions based on our study results, focusing on gesture design across AR and VR.

RQ_1 *What are the similarities and differences found in hand gestures when the same elicitation study is performed in AR versus VR?* The findings from section 4.3 revealed a medium level of agreement for both AR and VR, with no statistically significant difference between the gestures created in these two modalities. Interestingly, only the translation category showed a significant difference between AR and VR in five referent categories. We also observed that participants maintained a shared understanding of gesture meaning regardless of the reality. For example, referents like“Yaw” and “Move”, had high agreement rate, while more abstract referents, such as “Turn E off,” and “Change E color” showed low agreement rates across both modalities. When we investigated the winning gestures among AR and VR (see section 4.4), our findings showed that 73% of referents had matching Top gestures and showed consistency (52%) in the second most frequent (Top-2) gesture across AR and VR, reinforcing the idea of gesture availability and consistency across realities. The findings suggest a significant overlap between AR and VR modalities in the proposed user gestures, implying that gesture sets elicited in one modality (e.g., AR) can be highly representative of the other (e.g., VR).

The agreement rate between AR and VR The similarity in average agreement rate between AR and VR might indicate that the gesture design for manipulation is less influenced by the environment. This implies that gesture design is possibly primarily influenced by the semantics of the task itself. This point was verified in the semi-structured interviews about animation tasks(see section 4.6. However, the slight differences in the average agreement rate between AR and VR in section 4.3 show that, although the gestures designed by participants could be used in both environments, this does not directly indicate that participants' cognitive processes for gestures are entirely environment-dependent. This is also discussed in the book by William and Ortega about distributed cognition (DCog) in the field of HCI [58]. Multiple factors might influence participants' gesture design process [58]. Beyond semantic understanding, object layout in spatial position and cultural differences between subjects also impact gesture design [58]. In future research, the influence of DCog in AR and VR modality interaction can be further explored to determine whether there is collaboration between semantics and the environment.

The specificity of the translation category In our study, translation was the only category that showed clear differences between AR and VR. Piumsomboon et al. noted that, in AR environments, whether gestures rely on physical surfaces (such as a desktop) is an important consideration for users when designing gestures [41]. Translation gestures are mostly object-centric and involve moving virtual objects to spatial positions [41], which may make them more susceptible to differences between AR and VR. Several factors may contribute to this difference. First, participants' preception of their spatial position in AR may have differed from that perceived in VR, potentially because they could not see their own trunk in the VR environment. However, translation tasks required participants to manipulate objects relative to their body and surrounding space. This might also explain why the translation category, which relies on spatial position for orientation, is the only one that shows differences across AR and VR interaction. Second, prior work has shown that VR affords larger, unconstrained arm movements in open space [13], whereas AR participants often preferred physical surfaces as body-anchored interfaces [15]. These differences in spatial properties and movement may explain why the translation gesture results were different from those of other categories across realities. Future research could examine this hypothesis by further reducing differences in body visibility and spatial perception between AR and VR conditions.

The specificity of the abstract category in winning gestures These findings revealed a distinction in gesture consistency based on the nature of the referent, which has also been found in prior studies [38, 56, 63]. Abstract referents such as “Turn E on/off”, “Destroy Rightmost Building”, and “Open Chart” demonstrated low agreement across realities, with no matching gestures observed in either the Top or Top-2 rows. One possible explanation is that these actions lack strong physical metaphors, leading to more subjective interpretations and diverse gesture proposals across realities [58]. In contrast, referents associated with spatially grounded or physically intuitive actions, such as “Move Building closer/away/left/right”, “Rotate”, “Pitch”, “Yaw”, and “Scale”, exhibited much higher consistency in gesture selection. These referents are closely aligned with real-world physical interactions, making them easier for participants to conceptualize similarly across both AR and VR environments. This suggests that physically grounded tasks benefit from a more universally understood gesture vocabulary, which enhances consistency across AR and VR modalities. This pattern was further supported by our participant-level matching data, which showed higher consistency for these physically grounded referents across AR and VR sessions.

RQ_2. *What gestures are commonly agreed upon by participants across VR and AR?* In both AR and VR, our result shows that “hold”, “grab” and “push” are the top three most popular gestures. In addition, single-hand gestures and whole-hand gestures are the most frequently used types from section 4.5. This high consistency, as shown in the comparison between AR and VR modalities, indicates that participants are less influenced by environmental factors when performing manipulation tasks. This situation was confirmed during the semi-structured interviews conducted after the experiment. A possible explanation is that, when designing gestures, users' primary consideration is convenience. These findings align with prior gesture elicitation research [7, 38, 50, 56, 61, 63].

Although previous studies have reported the use of gestures in a single-interaction environment, single-hand gestures have also been reported to be more convenient and comfortable to use. In addition, this high consistency in hand usage across modalities might inform the future design of gestures for AR and VR interaction.

From a future design perspective, these results suggest that interactions across AR and VR can use universally understood gestures, enabling users to perform the same gesture to move fluidly between the two modalities. Although this result on gesture usage shows strong agreement for most referents, the referents for abstract actions (e.g., changing the color) and multi-object manipulation (e.g., yawing the building and the emitters clockwise) might still require specific gestures. This need further investigation by making multi-objects manipulation and abstract referents.

RQ_3 How would users perceive the elicited gestures they generated after the fact when using AR and VR? Based on the interview analysis, participants reported that their considerations of the uniqueness, memorability, comfort, and naturalness of gestures influenced their design in both modalities. Previous studies confirm in user-based gesture design, simplicity and comfort are the primary factors [24, 56, 61]. From a future design perspective, this result suggests that supporting users' main requirements is significant in both modalities.

The difference in immersive experience between AR and VR may influenced gesture proposals as reported by participants. Li et al. found that differences in the immersive experience could affect participants' feelings of their interactions [26]. In our study, some participants mentioned that in the VR, since the hand model is in the same virtual dimension as the environment, participants were able to more easily associate the gestures with digital game interaction environment; while in the AR, because they can use their real hands and see the real world environment, participants tended to adopt interaction patterns that more closely match real-world behaviors. These findings align with the theory of distributed cognition in HCI [17]. In our study, hand proprioception, the visual experience of the hand, and the differences between digital and real sensations across two modalities are influencing factors in gesture design between AR and VR. However, participants still showed agreement in their gesture design between AR and VR. Those influences reported by participants may not have been reflected clearly in the analysis of gesture agreement. A possible explanation is that the factors reported by participants may apply only to the gestures they performed differently, rather than to the gestures that were identical in both AR and VR.

6 LIMITATIONS OF THE STUDY

The study has several limitations. First, most participants (76%) were between 20 and 29 years old, which may limit the generalizability of the findings to other age groups, although this distribution reflects the simulator's target audience. In addition, the sample included only 16 participants; however, prior gesture-elicitation studies suggest that saturation can occur with approximately 8–10 participants [7, 53]. Also, prior experience was uneven across modalities, with more participants reporting VR than AR experience. However, since most were infrequent users, this imbalance was unlikely to have substantially affected the results. Second, current VR hand tracking often struggles to capture subtle gestures (i.e., microgestures) and does not accurately reflect finger movements [24], potentially influencing participants' gesture proposals. However, based on the responses to the survey question, participants mostly reported the virtual hands closely reflected their real hand movements.

Third, participants' familiarity with conventional interfaces, particularly WIMP (Windows, Icons, Menus, and Pointing) systems [20] may have introduce legacy bias and influenced their gesture proposals. Although this legacy bias exists in elicitation studies (e.g., phone use), it has been shown to be effective in some cases [37]; its influence on the conclusions appears to be limited. Fourth, our within-subject design included a two-day interval between sessions to reduce gesture recall from the first session. While memory effects may persist, it will be fairly effective, as a large part of what they did would have been forgotten according to Ebbinghaus' forgetting curve [35], indicating that after 48 hours, 77% would have been forgotten. Despite this possibility, the within-subject design enables consistent comparison across conditions while controlling for individual differences between participants. Our analysis focuses on overall gesture patterns (i.e., agreement rates) between AR and VR. This within-subject design is commonly used in gesture elicitation studies, as reported in prior work [40, 56, 63]. Finally, the study focused on a single AR/VR wind-simulation scenario, which may limit transferability to other domains. Although, the referents included both general interaction referents (e.g., move, rotate, select) and application-specific ones (e.g., turn the emulator on). In addition, the goal of this study was not broad generalization. By analyzing participants' gesture preferences and their contextual considerations within this single scenario, we offer valuable support to inform future research and provide gesture-based interaction design in more complex scenarios.

7 CONCLUSION

In this study, we presented a repeated-measures gesture elicitation study across two sessions (i.e., AR and VR), aiming to examine whether participants agree on gesture type and behavior when engaged with an educational application in AR and VR. At the end of the experiment, a semi-structured interview was included, during which researchers asked participants to explain the reasoning behind their gesture choices. Overall, the gestures performed by participants in AR and VR modalities were highly similar, reinforcing the idea of gesture availability and consistency across realities. Furthermore, participants showed a strong preference for single-handed and hold gestures across both modalities.

Participants demonstrated diverse understandings when generating specific gestures across different interaction environments. However, it has no observable impact on the design of the gestures themselves. Nevertheless, we identified differences in gesture usage across modalities within the translation category. This also suggests that the design of gestures in response to specific referent characteristics may be influenced by different modalities. This could be explored in future work. Finally, it is crucial to establish a standardized method for coding gestures. While there are many paths to follow, one of the most important future directions for elicitation studies is to automate the analysis process, thereby making these studies more pervasive and easier to access for practitioners.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) awards NSF 2439474, NSF 2444132, NSF 2327569 NSF 2238313, and NSF 2223432. We would like to acknowledge Brendan Kelley for supporting the development of several virtual objects in VR environment design.

REFERENCES

[1] A. M. Al-Ansi, M. Jaboob, A. Garad, and A. Al-Ansi. Analyzing augmented reality (ar) and virtual reality (vr) recent development in education. *Social Sciences & Humanities Open*, 8(1):100532, 2023. doi: 10.1016/j.ssaho.2023.100532 2

[2] L.-B. Bilius, O.-C. Ungurean, and R.-D. Vatavu. Understanding wheelchair users' preferences for on-body, in-air, and on-wheelchair gestures. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2023. 2

[3] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. doi: 10.1191/1478088706qp063oa 6

[4] E. Chan, T. Seyed, W. Stuerzlinger, X.-D. Yang, and F. Maurer. User elicitation on single-hand microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 3403–3414. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858589 2

[5] Z. Chen, X. Ma, Z. Peng, Y. Zhou, M. Yao, Z. Ma, C. Wang, Z. Gao, and M. Shen. User-defined gestures for gestural interaction: Extending from hands to other body parts. *International Journal of Human–Computer Interaction*, 34(3):238–250, 2018. doi: 10.1080/10447318.2017.1342943 2

[6] R. Cools, J. Han, and A. L. Simeone. Selectvisar: Selective visualisation of virtual environments in augmented reality. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, DIS '21, p. 275–282. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3461778.3462096 5

[7] A. Danielescu and D. Piorkowski. Iterative design of gestures during elicitation: Understanding the role of increased production. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3501962 1, 2, 4, 8, 9

[8] S. Ellis, F. Breant, B. Manges, R. Jacoby, and B. Adelstein. Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency. In *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*, pp. 138–145, 1997. doi: 10.1109/VRAIS.1997.583063 2

[9] B. Familoni. Augmented and virtual reality in u.s. education: A review: Analyzing the impact, effectiveness, and future prospects of ar/vr tools in enhancing learning experiences. *International Journal of Applied Research in Social Sciences*, 6:642–663, 04 2024. doi: 10.51594/ijarss.v6i4.1043 2

[10] M. P. Fay. Estimating age conditional probability of developing disease from surveillance data. *Population Health Metrics*, 2(1):6, 2004. doi: 10.1186/1478-7954-2-6 2

[11] Y. Felberbaum and J. Lanir. Better understanding of foot gestures: An elicitation study. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173908 2

[12] Y. Gaffary, B. Le Gouis, M. Marchal, F. Argelaguet, B. Arnaldi, and A. Lécuyer. Ar feels "softer" than vr: Haptic perception of stiffness in augmented versus virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(11):2372–2377, Nov. 2017. doi: 10.1109/TVCG.2017.2735078 2

[13] N. M. Gamage, D. Ishtaweera, M. Weigel, and A. Withana. So predictable! continuous 3d hand trajectory prediction in virtual reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, p. 332–343. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474753 8

[14] Q. Gao and Q. Sun. Examining the usability of touch screen gestures for older and younger adults. *Human factors*, 57(5):835–863, 2015. 2

[15] J. Hartbrich, S. A. Arboleda, S. Goring, and A. Raake. The Effect of Hand Visibility in AR: Comparing Dexterity and Interaction with Virtual and Real Objects . *IEEE Transactions on Visualization & Computer Graphics*, 31(11):10026–10034, Nov. 2025. doi: 10.1109/TVCG.2025.3616868 8

[16] R. Hidayat and Y. Wardat. A systematic review of augmented reality in science, technology, engineering and mathematics education. *Education and Information Technologies*, 29(8):9257–9282, Sept. 2023. doi: 10.1007/s10639-023-12157-x 2

[17] J. Hollan, E. Hutchins, and D. Kirsh. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.*, 7(2):174–196, June 2000. doi: 10.1145/353485.353487 9

[18] L. Hong, C.-C. Hsu, I.-C. Chen, and Y.-Y. Kuo. Is holding better? a gesture-elicitation study with controllers in virtual reality. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '25. Association for Computing Machinery, New York, NY, USA, 2025. doi: 10.1145/3706599.3719703 1, 2

[19] P. Jain. Navigating in vr using free-hand gestures and embodied controllers: A comparative evaluation. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–2, 2023. 2

[20] J. L. Jr., E. Kruijff, D. Bowman, R. McMahan, and I. Poupyrev. *3D User Interfaces: Theory and Practice (Usability) 2nd Edition*. Pearson Education, 2017. 3, 9

[21] E. Kaiser, A. Olwal, D. McGee, H. Benko, A. Corradini, X. Li, P. Cohen, and S. Feiner. Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th International Conference on Multimodal Interfaces*, ICMI '03, p. 12–19. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/958432.958438 1

[22] J. Lacoche, E. Villain, and A. Foulonneau. Evaluating usability and user experience of ar applications in vr simulation. *Frontiers in Virtual Reality*, Volume 3 - 2022, 2022. doi: 10.3389/frvir.2022.881318 1, 2

[23] C. Lee, G. A. Rincon, G. Meyer, T. Höllerer, and D. A. Bowman. The effects of visual realism on search tasks in mixed reality simulation. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):547–556, 2013. doi: 10.1109/TVCG.2013.41 1, 2

[24] G. Li, D. Rempel, Y. Liu, W. Song, and C. H. Adamson. Design of 3d microgestures for commands in virtual reality or augmented reality. *Applied Sciences*, 11(14), 2021. doi: 10.3390/app11146375 6, 9

[25] J. Li, A. S. Coler, Z. Borhani, and F. R. Ortega. Collecting and analyzing the mid-air gestures data in augmented reality and user preferences in closed elicitation study. In J. Y. C. Chen and G. Fragomeni, eds., *Virtual, Augmented and Mixed Reality*, pp. 201–215. Springer Nature Switzerland, Cham, 2024. 1, 3

[26] X. Li, B. Xu, Y. Teng, Y.-t. Ren, and Z.-m. Hu. Comparative research of ar and vr technology based on user experience. In *2014 International Conference on Management Science & Engineering 21th Annual Conference Proceedings*, pp. 1820–1827, 2014. doi: 10.1109/ICMSE.2014.6930456 1, 2, 9

[27] A. Lucero. Using affinity diagrams to evaluate interactive prototypes. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2015*, pp. 231–248. Springer International Publishing, Cham, 2015. 6

[28] M. N. Lystbæk, P. Rosenberg, K. Pfeuffer, J. E. Grønbæk, and H. Gellersen. Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality. *Proc. ACM Hum.-Comput. Interact.*, 6(ETRA), May 2022. doi: 10.1145/3530886 2

[29] D. Mauney, J. Howarth, A. Wirtanen, and M. Capra. Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, p. 4015–4020. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1753846.1754095 2

[30] K. R. May, T. M. Gable, and B. N. Walker. Designing an in-vehicle air gesture set using elicitation methods. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '17, p. 74–83. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3122986.3123015 2

[31] E. McAweeney, H. Zhang, and M. Nebeling. User-driven design principles for gesture representations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3174121 2

[32] N. McDonald, S. Schoenebeck, and A. Forte. Reliability and interrater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), Nov. 2019. doi: 10.1145/3359174 4

[33] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994. 2

[34] M. R. Morris, A. Danielescu, S. Drucker, D. Fisher, B. Lee, m. c. schraefel, and J. O. Wobbrock. Reducing legacy bias in gesture elicitation studies. *Interactions*, 21(3):40–45, May 2014. doi: 10.1145/

2591689 4

[35] J. M. J. Murre and J. Dros. Replication and analysis of ebbinghaus' forgetting curve. *PLOS ONE*, 10(7):1–23, 07 2015. doi: 10.1371/journal.pone.0120644 9

[36] Y.-S. Oh and J.-H. Kim. Handwind-sim: Real-time-web-based cloth simulation with gesture-driven wind interaction. *SoftwareX*, 32:102442, 2025. doi: 10.1016/j.softx.2025.102442 3

[37] F. R. Ortega, A. Galvan, K. Tarre, A. Barreto, N. Rishe, J. Bernal, R. Balcazar, and J.-L. Thomas. Gesture elicitation for 3d travel via multi-touch and mid-air systems for procedurally generated pseudo-universe. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 144–153, 2017. doi: 10.1109/3DUI.2017.7893331 2, 9

[38] F. R. Ortega, K. Tarre, M. Kress, A. S. Williams, A. B. Barreto, and N. D. Rishe. Selection and manipulation whole-body gesture elicitation study in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1723–1728, 2019. doi: 10.1109/VR.2019.8798105 1, 2, 8

[39] S. Pei, A. Chen, J. Lee, and Y. Zhang. Hand interfaces: Using hands to imitate objects in ar/vr for expressive interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3501898 6

[40] M. Perera, T. D. Gedeon, M. Adcock, and A. Haller. Vr & in-lab gess: an analysis of empirical research designs for gesture elicitation in in-lab and virtual reality settings. *Virtual Reality*, 29, 2025. 1, 2, 9

[41] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. User-defined gestures for augmented reality. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2013*, pp. 282–299. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. 8

[42] J.-L. Pérez-Medina, S. Villarreal, and J. Vanderdonckt. A gesture elicitation study of nose-based gestures. *Sensors*, 20(24), 2020. doi: 10.3390/s20247118 2

[43] G. A. Rovelo Ruiz, D. Vanacken, K. Luyten, F. Abad, and E. Camahort. Multi-viewer gesture-based interaction for omni-directional video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 4077–4086. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2557113 2

[44] J. Ruiz, Y. Li, and E. Lank. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 197–206. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978971 2

[45] N. Samimi, S. von der Au, F. Weidner, and W. Broll. Ar in tv: Design and evaluation of mid-air gestures for moderators to control augmented reality applications in tv. In *Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia*, MUM '21, p. 137–147. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3490632.3490668 1

[46] A. Sharma, J. S. Roo, and J. Steimle. Grasping microgestures: Eliciting single-hand microgestures for handheld objects. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300632 2

[47] G. Sullivan and A. Artino. Analyzing and interpreting data from likert-type scales. *Journal of graduate medical education*, 5:541–2, 12 2013. doi: 10.4300/JGME-5-4-18 5

[48] Y. Tan, S. H. Yoon, and K. Ramani. Bikegesture: User elicitation and performance of micro hand gesture as input for cycling. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, p. 2147–2154. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3027063.3053075 2

[49] C.-F. Tsai, S.-C. Yeh, Y. Huang, Z. Wu, J. Cui, and L. Zheng. The effect of augmented reality and virtual reality on inducing anxiety for exposure therapy: A comparison using heart rate variability. *Journal of Healthcare Engineering*, 2018(1):6357351, 2018. doi: 10.1155/2018/6357351 2

[50] R.-D. Vatavu. The dissimilarity-consensus approach to agreement analysis in gesture elicitation studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300454 1, 2, 8

[51] R.-D. Vatavu and J. O. Wobbrock. 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, p. 1325–1334. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2702123.2702223 2, 4, 5

[52] R.-D. Vatavu and I.-A. Zaiti. 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, p. 131–138. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2602299.2602316 2

[53] S. Villarreal-Narvaez, J. Vanderdonckt, R.-D. Vatavu, and J. O. Wobbrock. 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, p. 855–872. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3357236.3395511 1, 9

[54] P. Vogiatzidakis and P. Koutsabasis. 'address and command': Two-handed mid-air interactions with multiple home devices. *International Journal of Human-Computer Studies*, 159:102755, 2022. 2

[55] A. S. Williams, J. Garcia, F. De Zayas, F. Hernandez, J. Sharp, and F. R. Ortega. The cost of production in elicitation studies and the legacy bias-consensus trade off. *Multimodal Technologies and Interaction*, 4(4), 2020. doi: 10.3390/mti4040088 4

[56] A. S. Williams, J. Garcia, and F. Ortega. Understanding multimodal user gesture and speech behavior for object manipulation in augmented reality using elicitation. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3479–3489, 2020. doi: 10.1109/TVCG.2020.3023566 2, 3, 6, 8, 9

[57] A. S. Williams and F. R. Ortega. Understanding gesture and speech multimodal interactions for manipulation tasks in augmented reality using unconstrained elicitation. *Proc. ACM Hum.-Comput. Interact.*, 4(ISS), Nov. 2020. doi: 10.1145/3427330 2

[58] A. S. Williams and F. R. Ortega. A concise guide to elicitation methodology. *CoRR*, abs/2105.12865, 2021. 1, 2, 7, 8

[59] A. S. Williams, X. Zhou, A. U. Batmaz, M. Pahud, and F. Ortega. A pilot study comparing user interactions between augmented and virtual reality. In G. Bebis, G. Ghiasi, Y. Fang, A. Sharf, Y. Dong, C. Weaver, Z. Leo, J. J. LaViola Jr., and L. Kohli, eds., *Advances in Visual Computing*, pp. 3–14. Springer Nature Switzerland, Cham, 2023. 2

[60] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963 5

[61] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, p. 1083–1092. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1518701.1518866 1, 2, 4, 8, 9

[62] W. Zhou, T. Li, S. Li, and Y. Zhu. Color-in-fist: a metaphor for color selection with mid-air interaction. *J. Vis.*, 25(1):207–218, Feb. 2022. doi: 10.1007/s12650-021-00792-w 2

[63] X. Zhou, A. S. Williams, and F. R. Ortega. Eliciting multimodal gesture+speech interactions in a multi-object augmented reality environment. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, VRST '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3562939.3565637 2, 3, 4, 6, 8, 9

[64] A. Şimşek Tontuş and S. İpek Kuru Gönen. Teachers' gestures in synchronous online language classrooms: embodied elicitation strategies for student participation. *Social Semiotics*, 0(0):1–23, 2025. doi: 10.1080/10350330.2024.2448007 2