2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)

# User-Defined Gesture Interactions for VR Museums: An Elicitation Study

Qianru Liu
School of Advanced Technology
Xi'an Jiaotong-Liverpool University
Suzhou, China

Yue Li *
School of Advanced Technology
Xi'an Jiaotong-Liverpool University
Suzhou, China

Bingqing Chen
School of Advanced Technology
Xi'an Jiaotong-Liverpool University
Suzhou, China

Huiyue Wu
School of Journalism and Communication
Sun Yat-sen University
Guangzhou, China

Hai-Ning Liang
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

## Abstract

Recognizing the potential of freehand gestures for interacting with virtual objects in virtual environments, our research introduces a user-defined freehand gesture set of *typical* referents in virtual reality (VR), focusing on a *specific* scenario: VR museums. We conducted a comprehensive elicitation study with two experiments to define and refine the gesture set. Meanwhile, we demonstrated an enhanced real-time *Wizard of Oz* approach that facilitated users' understanding of referents in VR and their gesture design. Our findings revealed significant improvements in gesture consistency and user agreement through two experiments, with an average agreement score of first-choice and second-choice advancing from 0.211 and 0.160 to 0.412 and 0.284, respectively. By offering a consistent user-centered gesture set, this work contributes to guiding museum curators toward creating more immersive user experiences in VR museums. The gestures can also be extended to other VR applications that necessitate *travel*, *selection and manipulation*, and *system control* tasks.

**Index Terms:** Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies

## 1 Introduction

Virtual Reality (VR) has emerged as an increasingly prevalent technology, finding applications across various domains such as museum tours [48], education [18], training [6], therapy [10], and entertainment [1]. In addition to the immersion experiences that VR technology brings to users, it enables diverse forms of interaction within virtual environments, such as the manipulation of museum artifacts in ways that would be impossible or restricted in physical museums, thereby providing users with a real-time and rich interactive experience. Currently, handheld controllers are the most widely used input device for VR interactions. However, these devices sometimes face ergonomic issues and limit the naturalness of user interactions within virtual environments [24]. Integrating hand-tracking technology can overcome these limitations by capturing and interpreting the human hand's movements and positions in real-time [20]. This advancement enables a more intuitive and seamless sense of control of the digital environment, aligning user interactions more closely with natural human behavior. Freehand gesture interaction is a method of Human-Computer Interaction (HCI) facilitated by hand-tracking technology, that offers touchless manipulation of digital content, enhancing the user experience through intuitive and natural movements of the hands [35].

Gesture elicitation studies (GESs) have emerged as a critical methodology for capturing users' requirements and expectations, thereby facilitating the creation of intuitive gestures for interactions. Although GESs have proven instrumental in the development of freehand gesture vocabulary across a variety of fields [38], a recent systematic review [43] highlighted a significant gap in existing GESs, that the most influential GESs today remain concentrated in non-immersive platforms such as mobile devices [16, 37], 2D interfaces [12, 27, 28, 51], and smart homes [19]. Although these studies have proposed task-specific gesture sets for various referents (i.e., the task or action that a gesture is meant to represent), obtained gestures are not seamlessly transferable to the 3D VR environment [43]. The review results pointed out a critical gap in the lack of exploration of freehand gestural interactions within 3D VR platforms, highlighting the urgent need for focused research on this interaction modality. Particularly, there is a need to develop a standardized and universally accepted vocabulary of freehand gestures specifically for immersive environments.

Following the research gap identified earlier, it is challenging to establish and validate a comprehensive and universal gesture set for VR in one study. However, it is practical to examine a *specific* scenario with *typical* referents in VR, such as travel, selection, manipulation, and system control [2]. Gesture elicited for these typical referents are more likely to be generalizable to other VR environments. Specifically, VR museums present an ideal context for investigating freehand gestural interactions for these typical VR referents. This involves user movements in the virtual environment, the selection and manipulation of museum artifacts, and system controls of multimedia, such as text panels and videos.

Although there have been museum-related studies adopting gesture control in the past [17,25,47], the gestures were not derived from elicitation studies. As such, there is a lack of consideration of user preferences and expectations. To address this gap, our study explores freehand gestures performed by users for interacting with a range of typical referents in VR museums, recording their first-person perspectives in a VR Head-Mounted Display (HMD). Specifically, we summarized eight groups of typical referents in an immersive VR environment within three categories outlined in [2]: (I) *Travel*: steer forward, teleport, turn (left / right); (II) *Selection and manipulation*: select an item, pick an item, scale an item (enlarge / shrink); (III) *System control*: panel (open / close), video (play / pause).

A two-phase gesture elicitation study proposed by Wu et al. [53] was employed as a strategy to examine gestural interactions in a VR museum setting. This approach was found effective in improving user agreement in gesture selection. A total number of 87 gestures for the eight groups of referents were obtained through the first elicitation study with 27 participants. In particular, we used an enhanced *Wizard of Oz (WoZ)* method that provides real-time feedback

*Corresponding author: yue.li@xjtlu.edu.cn

2473-0726/24/$31.00 ©2024 IEEE
DOI 10.1109/ISMAR62088.2024.00078

to support participants in visualizing the impact of their gestures from a first-person perspective. In contrast to the standard *WoZ* approach with only verbal instructions and third-person perspective animations on a 2D display, we asked participants to wear the HMD and provided responsive feedback when they were proposing and articulating their gestures. The qualitative feedback from participants showed that this approach helped to facilitate an effective understanding of gesture effects and enhanced the accuracy and relevance of gesture selection. Based on the gesture set collected in the first experiment, the same participants took part in the second experiment to refine the gesture set.

In summary, the paper makes the following contributions across different stakeholders. (1) *For museum curators and digital experience designers*, we introduce a set of intuitive user-centered freehand gestures that can be readily adopted in VR museum scenarios. The gesture set has been validated through an iterative choice-based user study that ensured its consistency and usability. (2) *For practitioners such as XR developers*, we present the gestures used for the eight groups of typical referents in VR from existing studies and show our results. The information and findings provide design alternatives and implications for their system implementation. (3) *For HCI researchers*, we detail that the enhanced *Wizard of Oz* approach with real-time responsive feedback has been found effective in refining the traditional GESs. Future work should consider adopting this approach to facilitate users' understanding of their gestural inputs and effects.

## 2 Related Work

### 2.1 Gesture Elicitation Studies in Virtual Environments

Gesture elicitation studies (GESs) represent a pivotal research method in human-computer interaction. It has been widely applied to design intuitive gesture commands for interactive systems and devices. This approach facilitates a comprehensive understanding of end-user behaviors, habits, expectations, and needs, enabling the creation of gestures that align more closely with user perceptions and operational preferences [44]. GESs often involve presenting users with referents, namely the tasks or actions that a gesture is meant to represent. These are often communicated to users through verbal descriptions, pictures, videos, or physical objects. The traditional GESs ask users to propose a gesture for each referent, and a set of gestures is then derived based on the frequency of the corresponding referents. However, this method usually results in low consensus on gestures and legacy bias. Users tend to provide gestures based on their prior knowledge, experiences, and cultural background. A revised approach to GESs is choice-based GESs, which asks participants to make selections of gestures based on their preferences from a series of predefined choices. Allowing people to select their preferred choices can reduce the chance of designers introducing gestures that are too complicated [26] and improve the consensus. Wu et al. [53] proposed a practical method for deriving more consistent gestures with improved consensus among users. This method consists of two experiments to define and refine the gestures. The first experiment takes the traditional approach and requires participants to propose two gestures for each referent, defining an initial gesture set. Based on the results, the second experiment presents the gestures, and participants are asked to select the top two gestures among the choices, resulting in a refined gesture set. The method was shown effective in alleviating the legacy bias issue and improving the consensus among participants.

During the gesture elicitation process, it is critical for users to understand their performed action and its effect on the system, namely, the input and the output. The *Wizard of Oz* technique [9] is an effective way to simulate interactions between users and a system that appears to be fully automated but is actually being operated by a human (the 'wizard') behind the scenes. This technique helps simulate the effect of users' gestural inputs and facilitates users' understanding of the system feedback from their actions. The recent formulation of the end-user elicitation method specified six detailed steps [42], where the *WoZ* method significantly impacts several stages. It aids users in forming and instantiating mental models of system effects and commands corresponding to referents and supports actual gesture proposals (steps 1-3). Additionally, this method helps researchers use devices like video cameras to capture clear user actions (step 4), facilitating more effective classification in subsequent steps. While the technique was referenced in previous GESs, most studies (95.7%) took a rather *reversed Wizard of Oz* approach, based on a review study [44]. The effect of users' gestural interaction was communicated to them *before* they propose any gestures, instead of showing real-time responsive feedback. In addition, the 'wizard' was apparent to users as they saw the moderator presenting the referents. However, one advantage of using the *Wizard of Oz* technique in VR is the ability to conceal the moderator from the users' view.

GESs have been widely conducted for eliciting user requirements and expectations in various contexts such as unconstrained object manipulation environments [3,29,31,32,34,50], shopping [53], smart home [45], virtual historical site visiting [22], virtual travel [13], and public installation visiting [46]. However, the related GESs in 3D environments (e.g., virtual reality and augmented reality) were comparatively sparse [43]. Even within the limited VR GESs, the referents studied varied significantly and the results sometimes show contradictory findings. In addition, a few typical referents allowed in VR hand-held controllers (e.g., turning viewing perspective) seemed missing in previous GESs in virtual environments. Bowman et al.'s 3D user interface design indicated three types of typical tasks [2]: (1) travel, (2) selection and manipulation, and (3) system control. To better understand the appropriate gesture inputs for the typical interactions in VR, we examined previous GESs involving 3D interactions outlined in a recent systematic review [43], and conducted some complementary searches to enrich our understanding of specific referents. The results are summarized in Table 1.

### 2.2 Freehand Gesture Interactions for VR Museums

The examination of previous GESs concerning 3D interactions revealed the issue of generalizability of gesture sets across different modalities and different contexts. Only a few gestures are seamlessly transferable across modalities. For example, for the referent *select an item*, the same gesture of *touch with the index finger* was used in immersive VR [50, 53], projection VR [22], and AR environments [34]. Similarly, AR and VR had used the same gesture for the referent *scale an item* [34, 50]; AR and smart home devices also used the same gesture for video controls [34, 45]. Aside from these examples, previous works showed that the gestures elicited for the same referent varied for different modalities. For example, previous research suggested the gesture of *five fingers clenched to a fist and drag* to *pick an item* in immersive VR [53], but the gesture of *pinch using thumb and index finger and drag* was used in AR [32]. Gestures used for travel-related referents, such as *steering* and *perspective turning* also varied significantly in mobile devices and immersive VR [13, 46].

In addition, even for the same modality, such as VR, various gestures were suggested for the same referent in previous works. This potentially violates the design principle of external consistency and indicates a significant learning cost for users. For instance, five different gestures were suggested for the referent *select an item* within different contexts. This indicates the current lack of universal design guidelines and established practices of freehand gestures in VR.

While a great body of work about GESs was focused on 2D user interfaces, elicitation works addressing 3D interactions were comparatively fewer and would require more in-depth examinations [43]. This is a challenging research gap because the typical tasks in dif-

Table 1: Typical referents in VR and the corresponding gestures used in previous works.

| Category | Referent | Gesture Description | Modality / Context | Reference |
|---|---|---|---|---|
| Travel | Steer | Torso leaning | Immersive VR | [13] |
| | | Cyclic movement of the right hand as calling someone | | |
| | | Bike riding with both hands | | |
| | | Tap and drag virtual joystick up | Mobile devices | [46] |
| | Teleport | Use palm ray for destination selection and hold for 1.5s to initiate teleportation | Immersive VR | [32] |
| | | Use finger ray for destination selection and teleport automatically | Immersive VR | [3] |
| | Turn (left / right) | Tap and drag virtual joystick to the left / right | Mobile devices | [46] |
| | | Rotate device to the left / right | | |
| | | Five fingers closed and drag to the left / right | Projection VR | [22] |
| Selection and Manipulation | Select an item | Point with index finger | AR (HMD), Immersive VR, Projection VR | [22, 34, 50, 53] |
| | | Elbow flexion | Immersive VR | [29] |
| | | Shoulder flexion | | |
| | | Palm forward and push | Immersive VR | [31] |
| | | Five fingers clenched to a fist (grab) | Immersive VR | [53] |
| | Pick an item | Five fingers clenched to a fist and drag | Immersive VR | [53] |
| | | Pinch using thumb and index finger and drag | AR (HMD) | [32] |
| | Scale an item (enlarge / shrink) | Two hands grab and move apart / together along X-axis | AR (HMD) | [34] |
| | | Two hands pinch each diagonal corner of the target and move apart / together along XY-plane | AR (HMD), VR | [34, 50] |
| | | Two hands grab and do vertical accordion (apart / together) | Immersive VR | [31] |
| | | Two palms move from the center middle to the outer left and right / the reversed direction | Immersive VR | [53] |
| | | One hand pinch out / in using thumb and index finger | | |
| System Control | Panel (open / close) | Swipe out (right) / in (left) | AR (HMD) | [34] |
| | | Pull up / push down | | |
| | | Spray / regroup all fingers | | |
| | | Pull down / up | Immersive VR | [53] |
| | | Tap twice / swipe away | | |
| | Video (play / pause) | Point with index finger / palm forward | AR (HMD), Home devices | [34, 45] |
| | | Point with index finger / five fingers clenched to a fist | Home devices | [45] |

ferent virtual environments often vary significantly, which means a gesture set derived from existing works may not be comprehensive enough to provide consistent gesture options for a specific context of use. Therefore, conducting context-specific GESs is essential for creating a consistent and intuitively user-defined gesture set. Through elicitation experiments for a specific context, it is possible to design and generate gestures based on the specific referent requirements and user needs, thereby enhancing the applicability and effectiveness of gestures. This process can stimulate innovative thinking and potentially uncover entirely new gesture patterns beyond those existing ones, thus providing a more intuitive interactive experience.

VR museums are digital entities that simulate and augment the characteristics of physical ones, allowing people to have rich real-time engagement with digital artifacts in an immersive environment [8, 15]. This digital format offers a more accessible and intuitive means for users to navigate the virtual museum environment and interact with virtual artifacts, typically digital twins of physical counterparts. For example, users can grab an artifact to examine details, view descriptions on information panels, and access multimedia such as videos [36]. Compared to controller interactions, freehand gesture interactions are considered more natural, intuitive, and easy to learn [35].

Therefore, freehand gestures have been explored for online virtual museums, which can provide users with a more satisfying and natural experience [21]. For example, Manghisi [23] developed an web-based interface and derived five gestures for users to explore a heritage site, including 'move the pointer on the screen', 'zoom in', 'zoom out', 'change gaze direction', and 'select items'. Freehand gestures are beneficial for cultural heritage explorations, facilitating immersion and engagement by allowing users to take a close look at and naturally interact with the artifacts [23]. In particular, VR technology has been adopted to create a more immersive experience than WIMP interfaces. For example, by capturing visitors' movements and positions with the Kinect, the Etruscanning project [33] allowed users to interact with VR installations inside the museum.

Existing studies incorporating gesture inputs in VR museums have been primarily driven by designers' decisions. These studies risk overlooking crucial insights for enhancing the user experience and thus may fail to meet users' expectations and preferences. To the best of our knowledge, there has been no GES study conducted for VR museums. It is missing design guidelines for gestural control in VR museums. Thus, there is a need for empirical studies that suggest appropriate gestures from the end user perspective. Despite that previous GESs could provide some design implications, our review of previous work has pointed out the issue in the generalizability of gesture sets in various contexts and modalities, and the conflicts of gesture use for different referents. Therefore, it is essential to adopt a user-centered approach to define a set of intuitive freehand gestures tailored for VR museum scenarios, thereby augmenting the immersion and the interactive experience of end users.

## 3 Experiment 1: Defining a Gesture Set

The first experiment aimed to understand and collect participants' gestural behavior in the eight groups of referents. Users were instructed to perform two gestures that they found appropriate for the referent and rank their first and second choices, as suggested in [53].

### 3.1 Referents for VR Museums

Based on the review results, we considered the following typical referents, and included two directions of actions where possible: (I) Travel: steer (forward / backward), teleport, turn (left / right); (II) Selection and manipulation: select / unselect an item, pick / unpick an item, scale an item (enlarge / shrink); (III) System control: panel (open / close), video (play / pause). We conducted a pilot study with end users to check user understanding of the referents, after which some referents were removed as they were found unnecessary or obtuse. For example, users found it more intuitive to change the direction first and then steer forward, rather than performing gestures for *steering backward*. In addition, users found it difficult to propose multiple gestures for unselecting and unpicking an item. They all agreed that a natural release action would do the job. As a result, we examined the following eight groups of referents for the virtual museum scenario, categorized into three groups based on Bowman's taxonomy of 3D interaction techniques [2]:

- $T_1$ Steer forward: Move forward continuously.
- $T_2$ Teleport: Move to a destination discretely.
- $T_2$ Turn (left / right): Change the first-person perspective to the left / right.
- $SM_1$ Select an item: Select an item (e.g., a museum artifact) of interest.
- $SM_2$ Pick an item: Pick up an item (e.g., a museum artifact) in hand.
- $SM_3$ Scale an item (enlarge / shrink): Change the size of an item (e.g., a museum artifact) uniformly.
- $SC_1$ Panel (open / close): View or dismiss an information panel about an item (e.g., a museum artifact).
- $SC_2$ Video (play / pause): Play or pause a video.

### 3.2 Participants and Settings

This experiment had a total of 27 participants (12 females, 15 males) aged 19 to 29 years ($M = 22.93, SD = 2.352$). All participants were right-handed. Most of them ($N = 22$) had prior experience with VR equipment. More than half of them ($N = 18$) had never experienced hand-tracking technology in VR before. On a scale from 1 (1 = not familiar at all) to 5, participants were familiar with VR ($M = 3.15, SD = 1.10$). No participant asked to stop the experiment early. The experiment was conducted in a quiet room with a one-to-one setup between each participant and the moderator to avoid any interference with users' choices. The experiment equipment included a laptop with an NVIDIA GeForce RTX 3070 graphics card and 8GB RAM, and a Meta Quest 2 VR HMD. The laptop hosted a VR museum system developed for processing freehand gesture inputs. The study was identified as low-risk research and approved by the ethics review board of our university.

### 3.3 Moderator's *Wizard of Oz* Control

During the experiment, the moderator could observe the participants' first-person perspective on the laptop screen. In addition, the moderator could provide real-time *Wizard of Oz* feedback to participants through the keyboard control on the laptop. By pressing the number key `0`, the experimental scene is reset with a default camera position and orientation. Aside from this birth point, we set up another four teleport points in the environment, which are controlled by pressing the number keys `6/7/8/9`. The keyboard controls on the moderator side are summarized below.

- $T_1$: press the up arrow key ($\uparrow$) to move forward.
- $T_2$: press the number keys `0/6/7/8/9` to teleport to one of the destinations.
- $T_3$: press the left arrow key ($\leftarrow$) to turn left, and the right arrow key ($\rightarrow$) to turn right.
- $SM_1$: press the number key `4` to highlight the outline of the selected artifact.
- $SM_2$: users can directly grab or pinch the artifact using the virtual hands.
- $SM_3$: press the plus key ($+$) to enlarge the artifact; press the minus key ($-$) to shrink the artifact.
- $SC_1$: press the number key `5` to open or close a panel.
- $SC_2$: press the `space` key to play or pause the video.

### 3.4 Procedure

Before commencing the experiment, participants were briefed on the study purpose, procedure, duration, and data collection methods, with screen recording utilized for data collection. After the participants confirmed a clear understanding of the experiment, they were invited to sign an informed consent form. The moderator then helped participants wear the VR HMD to ensure a comfortable fit. During the experiment, participants were tasked with proposing two gestures for each identified referent to reduce legacy bias [53]. A total number of 16 gestures per participant were collected. A Latin square design was applied to the sequence of the eight groups of referents to mitigate the order effects. Immediately after the participant performed a gesture, the moderator provided the corresponding feedback to simulate the effect in the VR museum environment. We also used the *think-aloud* protocol [11] to facilitate the collection of design insights. At the end of the experiment, a semi-structured interview was conducted to further understand users' gestural preferences. Two questions were asked in the semi-structured interview: (1) Do these gestures cover all your intended actions in a VR museum? If not, what else would you expect to have? (2) How do you feel about the feedback for your proposed gestures?

### 3.5 Results

We processed the user-defined gestures by classifying and grouping duplicated ones based on a gesture taxonomy defined from previous works. In addition, the agreement scores were calculated for the first- and second-choice gestures proposed by participants.

#### 3.5.1 Data Processing

The study design could lead to 432 (27 participants $\times$ 2 gestures $\times$ 8 referents) groups of gestures. However, some participants found it difficult to propose a second gesture for some referents, so we obtained a total number of 424 gestures. Participants' verbal explanations were audio-recorded and transcribed; their gestures were video-recorded and documented using screenshots.

Previous research has suggested gesture taxonomy based on the hand count (single or double), the orientation of the hand(s), and hand postures [7, 55]. Wu et al. [53] also mentioned the distinction of gestures in their forms (static or dynamic) and trajectory. Thus, we incorporated these five dimensions in our categorization of the gestures. The distribution of gestures is shown in Figure 2. Compared to the classification in previous works [30, 34, 54], our study showed a higher percentage of the 'Point' posture, accounting for 48.1% of the proposed gestures. In relation to this, unilateral gestures appeared to be the dominant orientation, which is 32.8% higher than that in [30]. We will further discuss this in Sec. 5.1.

Based on this taxonomy, two researchers examined the gesture data separately and coded the gestures. Specifically, screenshots of the gestures were reviewed to obtain a direct visual understanding of the gestures. In addition, researchers examined the video recordings and participants' verbal explanations to understand their mental models when articulating the gestures. Researchers then compared their coding results and reached an agreement on the categorization. As a result, we obtained 87 different groups of gestures in total.

#### 3.5.2 Agreement Scores

To quantify the degree of consensus and preference among participants regarding the specific gestures for the given referents, we calculated the agreement score (ranging from 0 to 1) for each referent using the revised consensus formula proposed by Vatavu and Wobbrock [40].

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

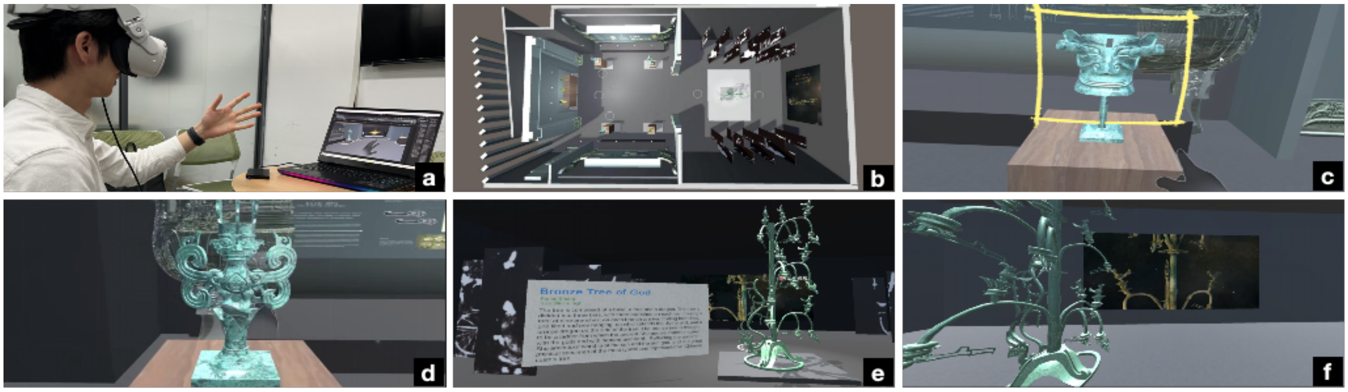


Figure 1: Experimental setting and the VR museum environment for our elicitation study: (a) a participant performing gestural control during the experiment; (b) the virtual museum environment with eight museum artifacts and five teleport points for *travel* gesture elicitation; (c) a museum artifact being highlighted in *selection* gesture elicitation; (d) front view of a museum artifact for *manipulation* (pick and scale) gesture elicitation; (e) a museum artifact with an information panel; (f) a museum artifact with a video description.

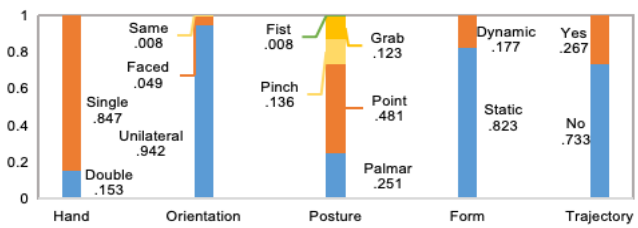


Figure 2: Five taxonomy dimensions and distribution of the gestures that users performed in this study. *Hand*: whether participants perform the gesture with both hands or one hand. *Orientation*: the orientation of the hand palm when performing the gesture. Unilateral indicates single hand facing in one direction (forward, back, left, right, up, down). *Posture*: the hand shape. *Form*: whether the hand shape changed when participants performed the gesture. *Trajectory*: whether the hand position changed when participants performed the gesture.

where $P$ represents the set of all gestures proposed for a referent $r$, $|P|$ denotes the size of the gesture set, and $P_i$ is the subset of identical gestures within $P$. The agreement score quantifies the probability of end users selecting the same gesture for a given referent. A higher agreement score indicates a greater likelihood of uniform gesture selection. Table 2 shows the agreement score for each referent.

The average agreement score for first- and second-choice gestures across eight groups of referents was 0.211 ($SD = 0.130$) and 0.160 ($SD = 0.067$). Previous studies reported an average score of 0.221 for mobile interaction [37], 0.242 for surface computing [51], 0.362 for TV control [39], 0.417 for AR [34], and 0.430 for 3D object manipulation [4]. Our results for both choices indicated a comparatively lower consensus among participants. The screenshots of the top two user-defined gestures are shown in Figure 3 (in blue).

For first-choice gestures, $SM_2$: *Video (play / pause)* had the highest agreement score of 0.363. Over half of the participants ($n = 16$) used the gesture, *tap the video with index finger*, out of nine proposed freehand gestures. On the other hand, participants showed comparatively low agreement scores for *Travel* referents, with the lowest agreement scores of 0.084 for $T_3$: *Teleport*. Over 10 unique gestures were proposed for each *Travel* referent.

For second-choice gestures, the highest agreement score of 0.246 was recorded for $SM_1$: *Select an item*. With 13 participants using the gesture, *point with index finger*, out of nine proposed freehand gestures. $T_1$: *Steer* had the lowest agreement score of 0.061. With a total of 13 freehand gestures identified for this referent, the top

Table 2: The agreement score (the number of user-defined gestures $|P|$) for each referent, and the total number of defined gestures in Experiment 1.

| Referent | First Choice | Second Choice | No. of Gestures |
|---|---|---|---|
| $T_1$ | 0.121 (12) | 0.061 (13) | 15 |
| $T_2$ | 0.087 (12) | 0.084 (8) | 14 |
| $T_3$ | 0.084 (13) | 0.188 (13) | 15 |
| $SM_1$ | 0.212 (5) | **0.246 (9)** | 9 |
| $SM_2$ | 0.340 (5) | 0.206 (7) | 8 |
| $SM_3$ | 0.184 (8) | 0.152 (7) | 8 |
| $SC_1$ | 0.151 (6) | 0.120 (9) | 9 |
| $SC_2$ | **0.363 (7)** | 0.224 (6) | 9 |
| Average | 0.211 | 0.160 | 11.875 |

gesture was only agreed upon by four participants.

Overall, participants had lower agreement scores for *Travel* referents than *Selection and manipulation* and *System control* referents.

#### 3.5.3 Interview Findings

The interview recordings were transcribed and analyzed using content analysis.

**Typical referents for VR museum visiting.** About 60% of the participants (16/27) felt that the eight groups of typical referents were sufficient for VR museum visiting. The remaining participants suggested additional functionalities. For *travel* related actions, P5 suggested incorporating more visual cues for navigation, such as illuminated pathways leading to the next artifact. Regarding *selection and manipulation*, P1 and P3 expressed interest in seeing realistic effects of physics and collision detection when tossing items. In terms of *system control*, P3 expected enhanced video functionalities, including fast-forwarding, rewinding, resizing the video panel, and relocating its position. Besides these three categories of referents, six participants mentioned the need for a virtual guide with voice explanation, which can help them to *'reduce the workload of reading texts'* and *'better understand the history'*. Furthermore, two participants would like to take photos during the visit. One of them proposed to have a personal page to store and browse these photos.

**Impact of real-time feedback on gesture elicitation.** Our results showed that 78% (21/27) participants endorsed the real-time feedback from the *Wizard of Oz* technique with specific positive feedback. No one reported any negative comments. About half of them were not aware that their gestural controls were performed by the moderator until they were told so. They found the visual

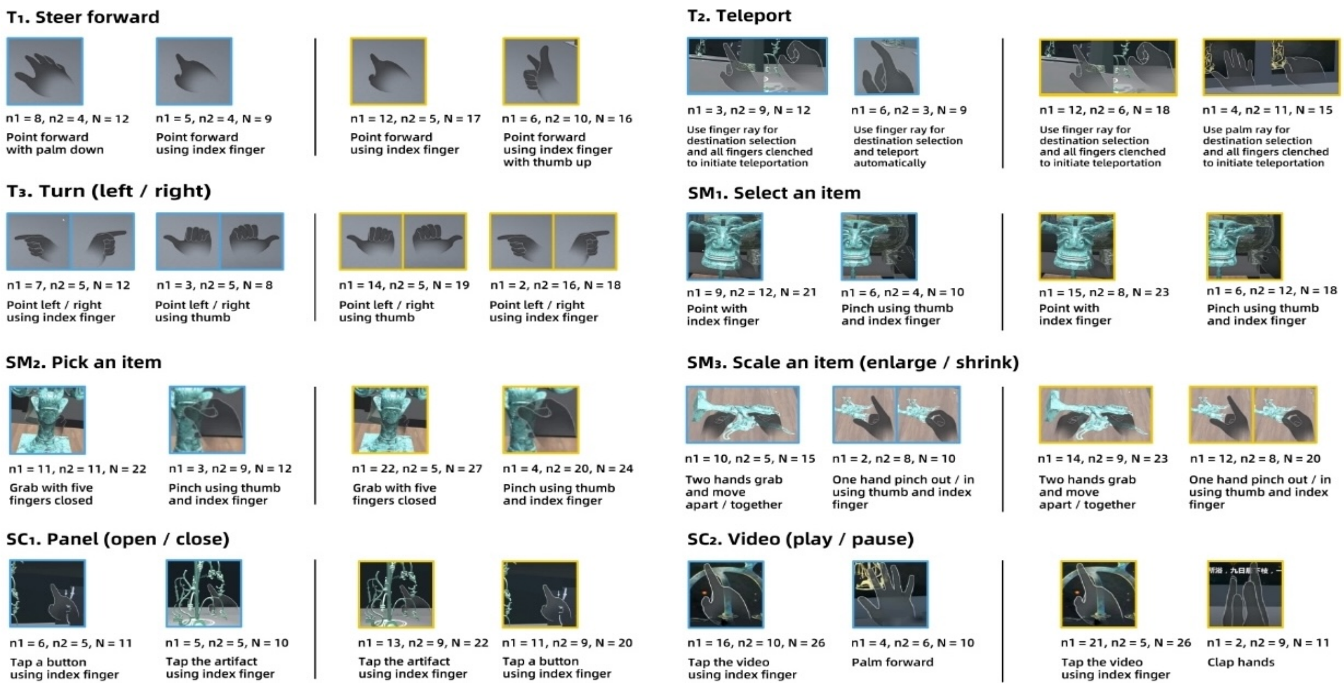


Figure 3: Top two user-defined gestures collected from Experiments 1 (in blue) and Experiment 2 (in yellow). The numbers *n1* and *n2* refer to the number of participants selecting the gesture as the first- and second-choice, respectively. *N = n1 + n2*, indicating the total selection count.

representations of gestures and the real-time feedback in the VR system helped visualize their input actions and understand the gesture effects. For example, P5 mentioned the system enabled him to *'immerse as if I am actually visiting a museum'*. P4, a VR novice user, was unfamiliar with the referent *'Teleport'*. She was confused at the beginning with the verbal explanations, but after receiving feedback of a performed gesture, she said, *'now I see what it means'*. Later in the interview, P4 commented that the real-time feedback helped her come up with more sensible gestures for the referents.

## 4 Experiment 2: Refining the Gesture Set

Experiment 2 took place two weeks after Experiment 1. The main objective was to further verify the usability and consistency of the gestures in Experiment 1, and to form a user-defined gesture set for VR museums.

### 4.1 Participants and Setting

The same 27 participants were re-invited to conduct a choice-based elicitation study. Building on the gesture set identified in Experiment 1, sample images of all gestures were extracted from the recordings and compiled into a slideshow for participants to rate the top 2 gestures for each referent. Same as Experiment 1, the experiment was conducted in a quiet room. To prevent any influence on participants' choices, the experiment maintained a one-on-one setting, with a participant and a moderator. The gesture sets were shown to participants on a PC monitor, and the Meta Quest 2 VR HMD was made available for participants to re-visit the VR museum environment.

### 4.2 Procedure

To begin with, participants were briefed about the experiment and provided their consent. Participants were instructed to wear the VR HMD to help them recall the experimental environment. Slides with screenshots showcasing gestures collected from Experiment 1 were presented to the participants. Following this, they were invited to select the two most appropriate gestures among the gesture candidates for each referent. During this process, they were advised to wear the VR HMD to experience the gestures and simulated feedback. The *think-aloud* protocol was employed to capture the rationale behind their selections. No hints were provided to participants.

### 4.3 Results

In this experiment, 87 groups of gestures identified from Experiment 1 were presented to participants. Each participant provided their top two selections for each referent, resulting in 432 (27 participants × 2 gestures × 8 referents) gesture selections in total.

#### 4.3.1 Agreement Scores

The agreement scores for the eight groups of referents were calculated using the same formula as detailed in Experiment 1. The results are demonstrated in Table 3. The average agreement scores for the first- and second-choice gestures for the eight referents were 0.412 ($SD = 0.153$) and 0.284 ($SD = 0.132$), respectively. The results of this experiment showed comparable agreement score magnitudes with the previous elicitation studies.

For the first-choice gestures, the highest agreement score for first-choice was 0.674 for $SM_2$: *Pick an item*, with the predominant gesture, *grab with five fingers closed*, agreed by 22 participants. Similar to the Experiment 1 results, $T_2$: *Teleport* had the lowest score of 0.252, where the top ranked gesture (*finger ray pointing*) was chosen by only 12 participants.

For the second-choice gesture, the lowest score was another *Travel* referent, $T_1$: *Steer forward*, with only 10 participants choosing *point forward* as the top gesture. The highest agreement score was 0.571 for $SM_2$, *Pick an item*, with the predominant gesture, *pinch using thumb and index finger*, selected by 20 participants.

A comparison of agreement scores between the two experiments revealed an overall increase for both choices, suggesting that the two-phase elicitation studies effectively elevated coherence for gestures.

#### 4.3.2 User-Defined Gesture Set

By summarizing gestures collected from and voted by participants in the two experiments, a refined set of user-defined freehand gestures

Table 3: The agreement score (the number of user-refined gestures $|P|$) for each referent, and the total number of refined gestures in Experiment 2.

| Referent | First Choice | Second Choice | No. of Gestures |
|---|---|---|---|
| $T_1$ | 0.261 (5) | 0.175 (9) | 10 |
| $T_2$ | 0.252 (6) | 0.235 (6) | 8 |
| $T_3$ | 0.345 (6) | 0.379 (7) | 7 |
| $SM_1$ | 0.342 (6) | 0.278 (6) | 7 |
| $SM_2$ | **0.674 (3)** | **0.571 (3)** | 4 |
| $SM_3$ | 0.446 (3) | 0.218 (6) | 6 |
| $SC_1$ | 0.376 (4) | 0.216 (6) | 6 |
| $SC_2$ | 0.601 (4) | 0.201 (6) | 6 |
| Average | 0.412 | 0.284 | 7.125 |

specific to VR museum visiting was established, as shown in Figure 3. For the eight groups of referents, 81.25% (13/16) of gestures obtained from Experiment 2 remained the same as that from Experiment 1. Gestures derived in Experiment 2 showed higher agreement scores, indicating stronger consensus.

#### 4.3.3 Differences Between Novices and Experienced Users

Given that Experiment 1 showed that some novice users were unfamiliar with VR controls, we thus conducted an analysis of these participant characteristics to examine if there is any difference in their proposed gestures. Specifically, we examined the demographic questionnaire and categorized the sample into two subgroups: novice ($n = 18$) and experienced ($n = 9$). Novice users are those who reported no previous experience with hand-tracking technology in VR. For $T_2$: *Teleport*, the palm-based gesture was mostly selected by novices ($n = 12$), while only three experienced users chose palm-based gesture as their second-top gesture. In addition, for $SC_1$: *Panel (open / close)*, compared to tapping the artifact ($n = 12$), novices preferred to imagine a button beside the panel for clicking ($n = 16$). In contrast, experienced users preferred to operate on the item directly ($n = 8$) over having a button ($n = 6$).

## 5 Discussion

Our two experiments defined and refined a gesture set for VR museum visiting, achieving a high agreement score of 0.412. The results provide some implications for the design of gestural control, the design of VR museums, and research on gesture elicitation.

### 5.1 Design Implications

#### 5.1.1 Mental models in 2D UIs transferred to 3D UIs

According to the user-defined gesture set we obtained, the gesture *Point with index finger* was found to frequently appear in the top two gestures among five typical groups of referents, including *steer forward*, *teleport*, *select an item*, *panel (open / close)*, and *video (play / pause)*. Given that these referents are typical in commercial VR systems, and that most participants ($N = 22$) have used VR, their familiarity with referents had minimal influence on their input. We speculate that this is related to users' mental models developed in 2D user interfaces. *Point and select* is arguably one of the most widely known and adopted interaction methods in both 2D and 3D user interfaces, thus it is reasonable that users are accustomed to its use for various purposes. Especially for system control referents, participants demonstrated many gestural controls and behaviors from 2D interactions, such as pointing to select, clicking on buttons to trigger information panels, tapping to play and pause, as well as the suggested enhanced video functionalities (see Sec. 3.5.3). Additionally, novices to hand-tracking technology constitute the majority of the participants (18/27), which is more likely to introduce legacy bias from 2D into 3D VR platforms. When articulating the gestures, participants indicated to point to different objects among referents (e.g., the floor, museum artifact, video, etc.). However, the repetitive use of such gestures may lead to system conflicts and confusion among end users. In such case, we suggest retaining the *Point with index finger* gesture for the *system control* category to align with the mental models of most users instead of overwriting it. For *Travel* tasks such as *steer forward* and *teleport*, there were other gestures suggested with a high number of votes, such as palm-based gestures. Regarding the referent *select an item*, pointing and dwelling can be a good alternative, as suggested in [32]. This could help distinguish it from gestures adopted in the *system control* category. In addition, the *pinch* gesture emerged as the second top gesture for this referent, which can be another acceptable alternative for item selection.

#### 5.1.2 Travel in VR with gestural control

Our study showed a comparatively lower agreement score for *travel* referents than *selection and manipulation* and *system control* referents. Thus, it is important to elaborate on our research findings in this aspect. Aside from the three gestures shown in the top two findings, some extremely creative gestures were proposed by participants. For example, participants proposed to *pull an imaginary rope*, *extend both arms forward as swimming*, and *simulate walking using index and middle fingers* for the *steer forward* referent. To *teleport*, some participants imitated Spiderman's web-slinging pose to choose a destination for teleport. These imaginative gestures were further endorsed by some other participants in Experiment 2, despite not ranking in the top two. Locomotion embodied in *travel* referents is unique and significant for VR. It affects users' perceived presence and simulator sickness. Our work endeavored to explore appropriate gestural controls for locomotion in VR. The results also pointed out possibilities for innovative designs in future work.

#### 5.1.3 Interaction metaphors in VR

Through our observations during users' gesture articulation, it showed that the virtual hand metaphor required less learning cost than the raycasting metaphor. In addition, we found that participants favored *indirect selection* but *direct manipulation*. Virtual hands are often found intuitive and easy to learn [5]. Our study showed that participants could easily adopt virtual hands for interactions in VR. They used the *grabbing* gesture to *pick* and *scale an item*, with their hands colliding with the target object. The intuitiveness of virtual hand controls could have contributed to the high agreement scores for *selection and manipulation* referents. On the other hand, some participants encountered challenges in mapping the movement actions to hand gestures. These *travel* referents could necessitate the raycasting metaphor, with a ray cast from the finger or palm for destination selection. In particular, participants indicated the need to visualize the parabola when selecting the destination. Although raycasting was also used in *system control* referents, participants did not ask the rays to be visualized. Instead, they expressed the need for feedback from the target, such as highlighting the object being selected. We found that less than half (42.11%, 8/19) of the user-defined gestures are supported by open-source projects designed for off-the-shelf HMD devices, including the XR Interaction Toolkit[1], the Microsoft Mixed Reality Toolkit[2] and Oculus Interaction SDK[3]. This revealed a significant design space for VR gesture control.

#### 5.1.4 Enrich the gesture set with secondary choices

Comparing the results in the two experiments, three out of the 16 gestures refined in Experiment 2 appeared to be 'new', indicating that choice-based experiments can help identify gesture options that users might not have encountered before but still find natural and easy to use once presented to them. Additionally, we found that novice participants were more likely to change their minds.

[1] https://github.com/Unity-Technologies/XR-Interaction-Toolkit-Examples
[2] https://github.com/microsoft/MixedRealityToolkit
[3] https://github.com/XRBootcamp/InteractionSDK

For example, for the *Video (play / pause)* referent, 8 out of 18 novices changed their minds and selected *clap hands* as their second choice. P17 commented that the clapping hand gesture can increase enjoyment and fun. Our results showed that 57.89% (11/19) gestures in the set are distinct from previously summarized ones in Table 1. This finding underscores the necessity of end-user elicitation studies to define gestures that meet users' actual needs and preferences.

#### 5.1.5 An enhanced *Wizard of Oz* technique in VR GESs

Our research findings indicated that the enhanced real-time *Wizard of Oz* technique demonstrated positive effects on users' understanding and gesture design. Compared to the traditional approach where the moderator observes participant behaviors from a third-person perspective, our approach allowed us to record participants' first-perspective views during their gesture articulation. Screenshots extracted from participants' video recordings could offer a more precise representation of user gesture behaviors compared to illustrations relying on moderators' memory recall. In contrast to a *reversed Wizard of Oz* approach adopted in previous GESs [44], the simulation of feedback resembled realistic gesture interactions in VR. Engaging in gestures related to events within the virtual environment could produce more ecologically valid outcomes compared to relying solely on verbal descriptions and participants' imaginations. This enhanced approach also allowed us to discover the effect of user demographics on their gesture selections. For novices with limited hand-tracking or VR technology experience, it can be challenging for them to understand some referents and propose gestures for them. For example, some novice users encountered difficulty at the beginning when proposing gestures for the referent *teleport* as they were unfamiliar with discrete movements and the *raycasting* metaphor. This is likely a significant factor contributing to the low consensus on this referent. With the enhanced approach and the simulation of effects, participants were able to articulate reasonable gestures. Our approach was found intuitive and natural by participants. Thus, we suggest researchers adopt this technology in their studies to bridge the gap in understanding.

#### 5.1.6 Standardized descriptions for user-defined gestures

Gestures were generally explained using images accompanied by text descriptions. However, we found significant discrepancies in the descriptions of the same gestures across different works, which posed difficulties in synthesizing the works. In particular, some descriptions of gestures were too brief (e.g., grab). Aligned with Vatavu and Wobbrock's suggestions regarding the standardized description of GES [42], we suggest incorporating taxonomy dimensions to standardize gesture descriptions, detailing specific body parts, postures, orientations, forms, and trajectories. This will facilitate better comparisons among multiple GESs.

### 5.2 Examples of Potential Extensions

The user-centered freehand gesture sets we provided encompass fundamental and typical referents in virtual environments, catering to most interaction and navigation needs. In addition to VR museums, these gestures can also be adopted in other scenarios with immersive environment simulations, object-centered interactions, and multimedia panels, such as VR galleries, historical reconstructions, and archaeological sites. Furthermore, our referent set can be easily expanded to fit various contexts, providing a relatively solid foundation for systems adopting gesture controls. For example, in VR educational applications, referents such as answering multiple choice quiz questions can be added, by combining *panel (open / close)* and *select an item* referents. In addition, the gestures for *selection and manipulation* can be used to manipulate virtual models and explore various spatial arrangements in architectural visualization and design. The natural gestures could also support patients to perform prescribed exercises for mobility training and physical rehabilitation. These extensions demonstrate the versatility and adaptability of the obtained freehand gesture sets across various VR applications.

### 5.3 Limitations and Future Work

Our study has some limitations. First, the same group of participants were involved in two experiments. While this approach allowed us to study the change of gesture proposals and improve agreement [53], it might have resulted in a net advantage [41]. Also, the effect of the enhanced *WoZ* method was observed in the interview, missing a quantitative perspective from a baseline group. Furthermore, the *codebook* method we adopted for agreement analysis [42] relies on ratings against predefined categories, which may introduce bias. To address this issue, Vatavu and Wobbrock [42] suggested that using *computer* models could improve efficiency and replicability. Recent work has introduced tools such as GEStory to facilitate the analysis [14]. One potential direction for future study is to allow end users to customize their own freehand gestures for different visiting referents, enabling a more personalized and intuitive interaction experience. However, our current study only examined hand gestures, but not other body parts. Future work could expand the body parts involved in the control. In particular, interactions for *travel* referents could be more natural if users' lower body parts could be tracked. Our first experiment suggested some additional referents required in VR museum settings, which future work could explore further. Some suggested referents could benefit from some existing works, such as the gestural control for enhanced video functionalities [52]. Furthermore, participants recruited for this study consisted of undergraduate and graduate students in China. In addition, many of them (18/27) were not familiar with gesture control. Future research could benefit from recruiting participants from broader age groups, cultural backgrounds, and technological proficiency to test the generalizability of our findings. Investigating left-handed users can also provide insights into accommodating diverse user needs. Finally, eye-tracking technology could be integrated with freehand gestures to support multimodal interactions [49]. Future work could aim to gain deeper insights into user behavior and preferences of these interactions, enabling more precise and responsive interactions within VR environments.

## 6 Conclusion

We present a gesture elicitation study (GES) examining appropriate gestures for *typical* referents in VR: travel, selection and manipulation, and system control, focusing on the VR museum scenario. We conducted the first experiment to define a gesture set, followed by the second one that refined the set, showing an agreement score of 0.412. We found users tended to transfer their mental models from 2D user interfaces to 3D spaces, showing a heavy use of the *pointing* interaction in various referents, such as steering, teleporting, selecting, and system controls. We also discussed the imaginative gestures for *travel* referents proposed by our participants and highlighted the under-explored design space of using gesture controls for locomotion in VR. Also, we demonstrated the applicability of an enhanced *Wizard of Oz* technique in VR GESs, showing the great benefits of real-time feedback in facilitating user understanding of gestures. Our work contributes a gesture set that can be readily adopted for VR museums, which can also be extended to other scenarios, such as education, architectural design, and healthcare.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (62207022 and 62272500). We thank our participants for their time and the reviewers for their constructive comments.

## References


[1] J. Bates. Virtual Reality, Art, and Entertainment. *Presence: Teleoperators and Virtual Environments*, 1(1):133–138, 02 1992. doi: 10.1162/pres.1992.1.1.133

[2] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., USA, 2004.

[3] E. Bozgeyikli, A. Raij, S. Katkoori, and R. Dubey. Point & teleport locomotion technique for virtual reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '16, p. 205–216. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2967934.2968105

[4] S. Buchanan, B. Floyd, W. Holderness, and J. J. LaViola. Towards user-defined multi-touch gestures for 3D objects. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '13, p. 231–240. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2512349.2512825

[5] G. Buckingham. Hand tracking for immersive virtual reality: Opportunities and challenges. *Frontiers in Virtual Reality*, 2, 2021. doi: 10.3389/frvir.2021.728461

[6] D. W. Carruth. Virtual reality for education and workforce training. In *2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, pp. 1–6, 2017. doi: 10.1109/ICETA.2017.8102472

[7] E. Choi, H. Kim, and M. K. Chung. A taxonomy and notation method for three-dimensional hand gestures. *International Journal of Industrial Ergonomics*, 44(1):171–188, 2014. doi: 10.1016/j.ergon.2013.10.011

[8] F. R. Cristobal, M. Dodge, B. Noll, N. Rosenberg, J. Burns, J. Sanchez, D. Gotshalk, K. Pascoe, and A. Runyan. Exploration of Coral Reefs in Hawai'i through Virtual Reality: Hawaiian Coral Reef Museum VR. In *Practice and Experience in Advanced Research Computing*, PEARC '20, p. 545–546. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3311790.3404542

[9] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of Oz studies — why and how. *Knowledge-Based Systems*, 6(4):258–266, 1993. Special Issue: Intelligent User Interfaces. doi: 10.1016/0950-7051(93)90017-N

[10] P. M. Emmelkamp and K. Meyerbröker. Virtual reality therapy in mental health. *Annual Review of Clinical Psychology*, 17(Volume 17, 2021):495–519, 2021. doi: 10.1146/annurev-clinpsy-081219-115923

[11] M. E. Fonteyn, B. Kuipers, and S. J. Grobe. A description of think aloud method and protocol analysis. *Qualitative Health Research*, 3(4):430–441, 1993. doi: 10.1177/104973239300300403

[12] M. Frisch, J. Heydekorn, and R. Dachselt. Investigating multi-touch and pen gestures for diagram editing on interactive surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, p. 149–156. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1731903.1731933

[13] P. Ganapathi and K. Sorathia. Elicitation Study of Body Gestures for Locomotion in HMD-VR Interfaces in a Sitting-Position. MIG '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3359566.3360059

[14] B.-F. Gheran, S. Villarreal-Narvaez, R.-D. Vatavu, and J. Vanderdonckt. RepliGES and GEStory: Visual Tools for Systematizing and Consolidating Knowledge on User-Defined Gestures. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, AVI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3531073.3531112

[15] W. Hürst, X. J. Tan, and F. de Coninck. Using Digital Extensions to Create New VR Museum Experiences. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, ACE '16. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/3001773.3001822

[16] S. K. Kane, J. O. Wobbrock, and R. E. Ladner. Usable gestures for blind people: understanding preference and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 413–422. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1979001

[17] A. Kargas, D. Varoutas, and G. Loumos. Augmented and virtual reality technologies in cultural sector: Exploring their usefulness and the perceived ease of use. 14:307–322, 09 2018. doi: 10.17349/jmc118223

[18] S. Kavanagh, A. Luxton-Reilly, B. Wuensche, and B. Plimmer. A systematic review of virtual reality in education. *Themes in Science and Technology Education*, 10(2):85–119, December 2017.

[19] C. Kühnel, T. Westermann, F. Hemmert, S. Kratz, A. Müller, and S. Möller. I'm home: Defining and evaluating a gesture set for smart-home control. *International Journal of Human-Computer Studies*, 69(11):693–704, 2011. doi: 10.1016/j.ijhcs.2011.04.005

[20] H. Y. Leng, N. M. Norowi, and A. H. Jantan. A user-defined gesture set for music interaction in immersive virtual environment. In *Proceedings of the 3rd International Conference on Human-Computer Interaction and User Experience in Indonesia*, CHIuXiD '17, p. 44–51. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3077343.3077348

[21] X. Li, D. Guan, M. Zhou, J. Zhang, X. Liu, and S. Li. Comparison on user experience of mid-air gesture interaction and traditional remotes control. In *Proceedings of the Seventh International Symposium of Chinese CHI*, Chinese CHI '19, p. 16–22. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3332169.3333570

[22] V. M. Manghisi, A. E. Uva, M. Fiorentino, M. Gattullo, A. Boccaccio, and G. Monno. Enhancing user engagement through the user centric design of a mid-air gesture-based interface for the navigation of virtual-tours in cultural heritage expositions. *Journal of Cultural Heritage*, 32:186–197, 2018. doi: 10.1016/j.culher.2018.02.014

[23] V. M. Manghisi, A. E. Uva, M. Fiorentino, M. Gattullo, A. Boccaccio, and G. Monno. Enhancing user engagement through the user centric design of a mid-air gesture-based interface for the navigation of virtual-tours in cultural heritage expositions. *Journal of Cultural Heritage*, 32:186–197, 2018. doi: 10.1016/j.culher.2018.02.014

[24] A. Masurovsky, P. Chojecki, D. Runde, M. Lafci, D. Przewozny, and M. Gaebler. Controller-free hand tracking for grab-and-place tasks in immersive virtual reality: Design elements and their empirical study. *Multimodal Technologies and Interaction*, 4(4), 2020. doi: 10.3390/mti4040091

[25] N. F. Mohd Noor Shah and M. Ghazali. A systematic review on digital technology for enhancing user experience in museums. In N. Abdullah, W. A. Wan Adnan, and M. Foth, eds., *User Science and Engineering*, pp. 35–46. Springer Singapore, Singapore, 2018.

[26] 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

[27] M. R. Morris, J. O. Wobbrock, and A. D. Wilson. Understanding users' preferences for surface gestures. In *Proceedings of Graphics Interface 2010*, GI '10, p. 261–268. Canadian Information Processing Society, CAN, 2010.

[28] M. A. Nacenta, Y. Kamber, Y. Qiang, and P. O. Kristensson. Memorability of pre-designed and user-defined gesture sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 1099–1108. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2470654.2466142

[29] V. Nanjappan, H.-N. Liang, F. Lu, K. Papangelis, Y. Yue, and K. L. Man. User-elicited dual-hand interactions for manipulating 3D objects in virtual reality environments. *Human-centric Computing and Information Sciences*, 8, 2018. doi: 10.1186/s13673-018-0154-5

[30] V. Nanjappan, R. Shi, H.-N. Liang, H. Xiao, K. K.-T. Lau, and K. Hasan. Design of interactions for handheld augmented reality devices using wearable smart textiles: Findings from a user elicitation study. *Applied Sciences*, 9(15), 2019. doi: 10.3390/app9153177

[31] 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

[32] T. Pham, J. Vermeulen, A. Tang, and L. MacDonald Vermeulen. Scale Impacts Elicited Gestures for Manipulating Holograms: Implications for AR Gesture Design. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, p. 227–240. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/

3196709.3196719

[33] E. Pietroni, C. Ray, C. Rufa, D. Pletinckx, and I. Van Kampen. Natural interaction in VR environments for Cultural Heritage and its impact inside museums: The Etruscanning project. In *2012 18th International Conference on Virtual Systems and Multimedia*, pp. 339–346, 2012. doi: 10.1109/VSMM.2012.6365943

[34] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. User-defined gestures for augmented reality. pp. 955–960, 04 2013. doi: 10.1145/2468356.2468527

[35] D.-M. Popovici, D. Iordache, R. Comes, C. G. D. Neamțu, and E. Băutu. Interactive exploration of virtual heritage by means of natural gestures. *Applied Sciences*, 12(9), 2022. doi: 10.3390/app12094452

[36] G. Ramsay. Hands-on, hands-off: The personal, social and physical context of interactives in museums. In *International Cultural Heritage Informatics Meeting*, 1999.

[37] 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

[38] Y. Tao. A VR/AR-based Display System for Arts and Crafts Museum. In *2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pp. 124–129, 2020. doi: 10.1109/CyberC49757.2020.00029

[39] R.-D. Vatavu. User-defined gestures for free-hand TV control. In *Proceedings of the 10th European Conference on Interactive TV and Video*, EuroITV '12, p. 45–48. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2325616.2325626

[40] 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

[41] R.-D. Vatavu and J. O. Wobbrock. Between-subjects elicitation studies: Formalization and tool support. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 3390–3402. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858228

[42] R.-D. Vatavu and J. O. Wobbrock. Clarifying agreement calculations and analysis for end-user elicitation studies. *ACM Trans. Comput.-Hum. Interact.*, 29(1), jan 2022. doi: 10.1145/3476101

[43] S. Villarreal-Narvaez, A. Sluÿters, J. Vanderdonckt, and R.-D. Vatavu. Brave New GES World: A Systematic Literature Review of Gestures and Referents in Gesture Elicitation Studies. *ACM Comput. Surv.*, 56(5), jan 2024. doi: 10.1145/3636458

[44] P. Vogiatzidakis and P. Koutsabasis. Gesture elicitation studies for mid-air interaction: A review. *Multimodal Technologies and Interaction*, 2(4), 2018. doi: 10.3390/mti2040065

[45] P. Vogiatzidakis and P. Koutsabasis. Mid-air gesture control of multiple home devices in spatial augmented reality prototype. *Multimodal Technologies and Interaction*, 4(3), 2020. doi: 10.3390/mti4030061

[46] S. Vosinakis and A. Gardeli. On the use of mobile devices as controllers for first-person navigation in public installations. *Information*, 10(7), 2019. doi: 10.3390/info10070238

[47] B. Wang and Y. Liu. The research on application of virtual reality technology in museums. *Journal of Physics: Conference Series*, 1302(4):042049, aug 2019. doi: 10.1088/1742-6596/1302/4/042049

[48] Y. Wang, Y. Li, and H.-N. Liang. Comparative Analysis of Artefact Interaction and Manipulation Techniques in VR Museums: A Study of Performance and User Experience. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 761–770. IEEE, 2023.

[49] Y. Wei, R. Shi, D. Yu, Y. Wang, Y. Li, L. Yu, and H.-N. Liang. Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2023.

[50] 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

[51] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. CHI '09, p. 1083–1092. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1518701.1518866

[52] C. Wu, Z. Tan, Z. Wang, and S. Yang. A Dataset for Exploring User Behaviors in VR Spherical Video Streaming. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, MMSys'17, p. 193–198. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3083187.3083210

[53] H. Wu, Y. Wang, J. Qiu, J. Liu, and X. L. Zhang. User-defined gesture interaction for immersive VR shopping applications. *Behaviour & Information Technology*, 38(7):726–741, 2019. doi: 10.1080/0144929X.2018.1552313

[54] Y. Yan, X. Xi, C. Yu, and Y. Shi. Gesture-based target acquisition in virtual and augmented reality. *Virtual Reality Intelligent Hardware*, 1(3):276–289, 2019. Human-computer interactions for virtual reality. doi: 10.3724/SP.J.2096-5796.2019.0007

[55] Y. Yan, C. Yu, X. Ma, X. Yi, K. Sun, and Y. Shi. VirtualGrasp: Leveraging Experience of Interacting with Physical Objects to Facilitate Digital Object Retrieval. 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: 10.1145/3173574.3173652