Understanding Gesture Performance for Object Selection in VR: Classification and Taxonomy of Gestures in HCI

Shimmila Bhowmick
Department of Design, Indian Institute of Technology (IIT) Guwahati
shimmila.bhowmick@gmail.com

Pratul Kalita
Department of Design, Indian Institute of Technology (IIT) Guwahati
pratulkalita@iitg.ac.in

Keyur Sorathia
Department of Design, Indian Institute of Technology (IIT) Guwahati
keyur@iitg.ac.in

ABSTRACT

Classification of body gestures is vital in HCI research as it benefits researchers with appropriate gestures considering user behavior, mental model, and constraints. This work presents the gesture classification of 87 unique gestures identified through an elicitation study for object selection in Virtual Reality (VR). In addition to adopting existing gesture taxonomies, we propose two new taxonomies based on hand dominance– (i) dominant hand only (ii) non-dominant hand first and (iii) equal and simultaneous hand dominance and multi-body part movement gestures – (i) simultaneous and (ii) sequential. We believe that the proposed classification and taxonomies will assist designers and researchers in designing effective full-body gestural interfaces, including VR interfaces.

KEYWORDS

Gesture taxonomy, Virtual reality, Gesture classification, Virtual environment, Gestures, Object selection

ACM Reference Format:
Shimmila Bhowmick, Pratul Kalita, and Keyur Sorathia. 2021. Understanding Gesture Performance for Object Selection in VR: Classification and Taxonomy of Gestures in HCI. In *India HCI 2021 (India HCI 2021), November 19–21, 2021, Virtual Event, India.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3506469.3506483

1 INTRODUCTION

Gestures are a well-studied and accepted interaction modality in Human-Computer Interaction (HCI). Gesture classification is vital in HCI research, given the diversity and complexity of the gestures used in different applications [7], contexts [11], and cultures [1]. Gesture designers, state-of-the-art practitioners and researchers can benefit from a well-structured classification of gestures [8]. It will provide them with variations of appropriate gestures that best fit their design-related purposes while considering actual user behaviors [9], mental models [5, 10] and constraints. Gesture classification and information of its user behavior and mental model have supported designing surface computing and motion gestures [6, 9]. In addition, a well-established taxonomy of gestures can help HCI researchers better comprehend and understand the different perspectives in the area of gestural interaction [10].

Karam and Schraefel [3] have presented the most recent gesture classification research in HCI. Wobbrock et al. [9] also excellently covered hand gesture classification based on human discourse for surface computing. Ruiz et al. [6] proposed two broad classes of taxonomy dimensions: gesture mapping and physical characteristics. To cover dimensionalities in Augmented Reality (AR), Piumsomboon in [4] developed two additional dimensions: symmetry and locale. Overall, the classification presented covers wide dimensionalities of gesture taxonomy; however, existing literature does not accommodate gesture classification studies conducted for VR platforms and excludes full-body gestures classification, including VR.

This work presents gesture taxonomies in gestural interactions in HCI, particularly for whole-body gestures performed for Virtual Reality (VR) interfaces. We classify our identified gestures from an elicitation study [2] and propose two new gesture taxonomies based on 87 unique gestures. This work makes two primary contributions (i) analysis and classification of the user-defined gesture set identified from an elicitation study and (ii) a proposal of 2 new gesture taxonomies.

Overall, the classification presented covers wide dimensionalities of gesture taxonomy; however, existing literature does not accommodate gesture classification studies conducted for VR platforms and excludes full-body gestures classification, including VR.

2 OVERVIEW OF GESTURE ELICITATION STUDY

In our previous study [2], we conducted a two-stage gesture elicitation study to identify natural and intuitive controller-less techniques while considering full-body interaction for nail size object selection in 4 different VEs. In study 1, we asked participants to propose the most natural and intuitive gestures for selecting a nail size object in 4 different VE conditions (i) VE1: dense VE with the target object placed at arm's length (within 50 cm motor space), (ii) VE2: dense VE with the target object placed at scaled length (5.2 m), (iii) VE3: dense VE with the occluded target object placed at arm's length (within 50 cm motor space) and (iv) VE4: dense VE with the occluded target placed at a scaled length (5.2 m). The targets were partially (>50%) occluded in VE3 and VE4. Once the participants performed the body gestures, they informed the moderator. After completing the task, participants were asked to rank their proposed gestures from best to worst. 40 participants participated in the study. 737 gestures were generated for all VEs. 87 unique gestures (22, 24, 17 and 24 unique gestures) were proposed for VE 1, 2, 3 and 4, respectively.

India HCI 2021, November 19–21, 2021, Virtual Event, India

ACM ISBN 978-1-4503-9607-3/21/11…$15.00
https://doi.org/10.1145/3506469.3506483

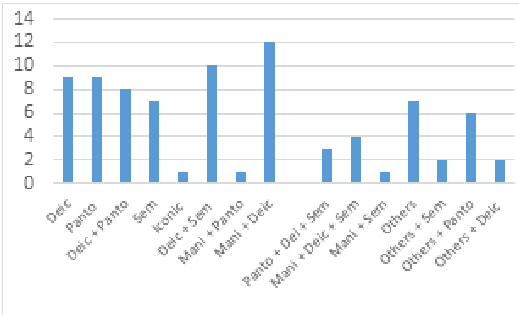

Figure 1: Number of unique gestures in each gesture category

The second study evaluated the extracted gestures to propose 3, 2, 3, and 3 gestures for 4 VEs, respectively. A new group of 40 participants performed the evaluation. The gestures were evaluated using ease of performance, gesture suitability, appropriateness, user preference, and effort. An overall score was calculated to evaluate and propose the final gesture set. The second study followed the same VEs and protocol as the first.

3 GESTURE CLASSIFICATION

We present classification for gestures extracted from the first study (figure 1). 737 gestures were performed that included 87 total unique gestures with 22, 24, 17, and 24 unique gestures for VE1-VE4, respectively. We refer to the taxonomies proposed by [3] due to its extensive coverage of taxonomies that covers most taxonomies presented in the literature and primarily focuses on full-body and in-air gestures. Overall, we observed a higher preference for upper body gestures (87%) than lower body (12%) and full-body gestures. Among the upper-body gestures, hand-based gestures (76%) were preferred the most in all VEs. Further, gesture classification revealed dominance of *manipulative and deictic* (14.3%) gestures followed by *pantomimic* (10.6%), *deictic* (10.6%), *deictic and pantomimic* (9.5%), and *semaphoric* (8.3%) gestures. We discuss the VE-specific classifications and elaborate on them in the following sections.

3.1 Gesture Classification for VE 1

We observed a higher preference for *pantomimic* (12 out of 22 gestures) for VE1, including the combination of *deictic* and *pantomimic* gestures. *Pantomimic* gestures enabled participants to perform either a static pose or a dynamic movement or a combination for object selection, i.e. *grab (to identify) and rotate (to confirm); grab with curved fingers and open palm; make a fist (static pose); punch (to identify) and wrist rotate (to confirm); swipe a hand over the target* were performed by the participants. We believe that the target size could be one reason for the higher preference for pantomimic gestures. Participants matched their hand size with the target size, where they could shrink and enlarge the curved fingers according to the target size (i.e., curving the finger to reduce the grabbing area for *grab and rotate* gesture). A participant added, "I can map the target size to the size of my curved fingers. It gives confidence that I have held the target properly for selection." We also observed the use of pantomimic gestures with gaze-based interactions, i.e., *gaze and grab; gaze and nod; and gaze and blink.*

Further, a series of gestures performed combined *pantomimic* gestures with *deictic* gestures, i.e., *point and bring hand closer to the body; palm point near body and thrust; point and draw a circle around the target.* It ensured that the selection was validated by signifying a confirmation gesture to ensure correct selection. We observed a two-stage strategy for performing *deictic* and *pantomimic* gestures. The participants performed *deictic* gestures to indicate the target, while the object selection was confirmed using *pantomimic* gestures.

Semaphoric gestures (3 out of 22) comprised dynamic semaphores, i.e., *call gesture using vertical palm, knuckle tap, and flip hand,* and *knee touch and tap with a leg.* While semaphoric gestures are not considered natural [4], we believe they were performed primarily due to eyes-free interactions.

The *remaining gestures consisted of various gestures that do not fall into categories suggested by [*4, 10*], i.e., dwell gaze, blowing the object,* etc. We did not observe any pattern among these gestures, but some were adopted from low-cost VR interactions, i.e., gaze.

3.2 Gesture Classification for VE 2

We observed the use of *manipulative and deictic* (5 out of 24), *pantomimic* (3 out of 24), *deictic and semaphoric* (4 out of 24) and *deictic* (2 out of 24) gestures in VE 2. A combination of gesture categories was also seen performed, including the combination of pantomimic and deictic; *manipulative, pantomimic, deictic and semaphoric; others* and *manipulative; pantomimic, deictic, and semaphoric* gestures.

Overall, we observed a combination of various gesture categories due to challenges associated with distant objects placed in a dense VE. Participants wanted to bring target objects to visible space to ensure accurate and error-free object selection. For example, the *manipulative and deictic* gestures were performed primarily because they felt the need to manipulate the environment to bring it closer. Although manipulation demanded increased physical effort, they preferred to make objects appear closer for precise selection, i.e., *grab and bring environment near and point; scroll with the index finger to zoom and point; pinch-out VE and point and tap* etc. Such approach was seen for gestures performed under *deictic and semaphoric; pantomimic, deictic and semaphoric and manipulative; pantomimic, deictic and semaphoric categories.*

3.3 Gesture classification for VE 3

The target object was partially occluded and kept at arm's reach in the VE 3. We observed the use of *semaphoric gestures* (3 of 17), *others* (3 of 17), *deictic* (3 of 17), *deictic and semaphoric* (2 of 17) gestures in VE3. A combination of gesture categories was also seen performed by the participants, including the combination of *Pantomimic and others, deictic and pantomimic, manipulative semaphoric and deictic, pantomimic, and semaphoric, semaphoric and others* and *pantomimic deictic and semaphoric* gestures. These combinations were performed once by the participants.

Overall, we observed a comprehensive combination of various gesture categories due to challenges associated with the target being partially occluded and placed in a dense VE. Participants performed simple gestures usually performed in real-life to avoid error

Table 1: Taxonomy of gestures in virtual reality

				Definition
Hand dominance	Dominant hand only	Single movement gestures	Static hand pose	A gesture is performed by the dominant hand and has a single movement. Hand pose is held static. i.e., *Extend hand and*
			Dynamic hand pose	The gesture is performed by the dominant hand and has multiple movements. Hand pose is dynamic. i.e., *Point and bring hand closer to body*
		Multi-movement gestures		The gesture is performed by a dominant hand that has multiple movements i.e., *pinch-out VE and point and tap*
	Non-dominant hand			A gesture is performed by a non-dominant hand (to perform a secondary task) and followed by a dominant hand (to perform a primary task). i.e., *pinch-out and rotate the VE (non-dominant hand) and point and tap (dominant hand)*
	Equal hand dominance			A gesture that involves simultaneous movement of both hands. No specific preference is given to any hand for the task. i.e., *point (with the index finger of both hands) and tap*
Multi-body part movement gestures	Simultaneous			Gestures that occur simultaneously to complete a task. i.e., *lift leg and hand together and tap*
	Sequential			Gestures that involve a definite sequence to complete a task. i.e., *walk (to the target object) and rotate the environment and point and tap*

while selecting the occluded target. For example, *semaphoric* gestures were performed because participants felt these were symbolic gestures frequently used in real-life.. e. g. *grab, grab (both target and occluding object) and grab and select the target, tap on occluding sphere (it disappears) then tap on target.* Similar approaches were seen for gestures performed under *deictic, others,* and *deictic and semaphoric* gestures.

3.4 Gesture classification for VE 4

The target object was distant and partially occluded in VE 4. Similar to the findings observed in VE 2, gestures performed by the participants were dominantly *manipulative and deictic* (30.2 %). The participants preferred to remove the occlusion to increase the visibility of the target objects, assuming that it would result in error-free object selection. Gestures such as *pinch-out*, *leaning-in, and tap using leg,* etc. were performed to zoom the VE and increase target object visibility and reachability. The VE was scaled using the non-dominant hand, whereas the dominant hand selected the target object. A similar approach was seen using *manipulative, deictic and semaphoric; pantomimic and deictic; and deictic and semaphoric* gestures.

A different approach was seen while performing *deictic* gestures for object selection in VE 4. Here, participants imagined a graphical visualization while performing deictic gestures such as scaled visualization of the highlighted area using pointing; *point to teleport and point; and point and bend ray.*

Other combinations such as *manipulative, deictic, and others; manipulative and semaphoric; iconic; deictic pantomimic and semaphoric* were performed once each by the participants.

4 TAXONOMY OF GESTURES IN VR

We propose the gesture taxonomies from 87 unique gestures extracted from the gesture elicitation study. Gestures were classified contextually based on the hand dominance in performing the task and the motion performed by hands. We also observed the posture and the sequence of gestures to propose the taxonomies. We introduce two taxonomies: *hand-dominance* and *multiple body-part movement gesture*, each with sub-categories. These have been defined in table 1 and explained in detail below. The focus of the work is on full-body gestures; hence, gestures performed by facial/micro expressions are out of the scope of the taxonomy.

The first dimension we developed, *hand dominance,* allowed classification of gestures depending on whether the gestures were performed using a *dominant hand* or a *non-dominant* hand. We group the gestures into *dominant hand only, non-dominant hand first,* and *equal hand dominance* category.

As the name indicates, a dominant hand performs the gesture under the *dominant hand only* category. This category comprised 46% (40 out of 87 unique gestures) of the total unique gestures performed in all VEs. It is further sub-divided into *single-movement,* and *multi-movement* gestures that require one movement and multiple hand movement by the dominant hand, respectively. For the *single-movement dominant hand only* category, the gestures can have the final hand pose held static (i.e., extend hand and *point*) or dynamic (i.e., point and *bring hand closer to the body*). The single-movement dominant hand comprises 39% (16 out of 40 gestures) of the *dominant hand only* category. The multi-movement gestures have more than one movement performed by the dominant hand (i.e., point and bring hand closer to the body, palm point, and thrust). Although there is no limitation to the number of hand movements, our observation suggests a maximum of 3-4 dominant hand movements performed as it may increase physical effort for the participants. The gesture performed under the *dominant hand only* category is helpful in a multi-tasking environment where the non-dominant hand is responsible for performing secondary tasks, i.e., use of non-dominant hand to virtually navigate and dominant hand to pick up virtual shopping items in a VR shopping mall.

We propose a *non-dominant hand first* taxonomy under the hand dominance category. Here, the gesture is performed with the non-dominant hand followed by the dominant hand to complete the gesture. The non-dominant hand can perform a secondary or a supporting activity to accomplish an actual task. For example, *pinching and rotate* gesture is performed to ensure increased visibility and access to the target object for *pinch-out & rotate the VE (with non-dominant hand) and point and tap* gesture. *Point and tap* is performed to select the target object accurately and performed with a dominant hand.

The third taxonomy under the *hand dominance* category is equal hand dominance. It proposes equal weightage given to both hands to perform the primary task. For example, *pointing with the index finger of both hands and tap* to select uses both hands to perform the object selection independent of hand preference and dominance.

We propose multi-body part movement gestures where more than one body part performs the gesture. We further sub-categorize them based on *simultaneous* and *sequential* multi-body-part movements performed for accomplishing the task. For example, *walk towards the object, rotate the environment and point and tap* required multiple body movements (i.e., legs and hand) in which walking, rotation and pointing were performed sequentially (17.2%, 15 out of 87 unique gestures). *The simultaneous multi-body part* gesture is performed in unison to confirm the selection of the target object, i.e., *lift leg and hand together and tap; gaze and point; nod and tap.*

5 CONCLUSION

This paper presented a classification of extracted gestures identified through a gesture elicitation study for object selection task on a VR interface. We also proposed two new gesture taxonomies based on hand dominance and the involvement of body parts in performing the gestures. These taxonomies were identified after analyzing 87 unique gestures performed during the elicitation study. The hand dominance was further detailed through a dominant hand only, non-dominant hand first, and equal dominance taxonomies. Dominant hand only taxonomy was further categorized into single and multiple hand movements to indicate the use of one and multiple hand movements in accomplishing a task. We also propose multi-body part movement gestures where different body parts perform sequentially or simultaneously to accomplish a task. We believe that these new taxonomies add to the existing literature on gesture classifications, which is vital in understanding user behaviors, mental models and user performances for different tasks and technology platforms.

REFERENCES

[1] Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindbauer, Alexandra Ion, Shengdong Zhao, and J. T. K. V. Koh. "Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci." Microsoft Research TechReport MSR-TR-2012-111 2 (2012): 30
[2] Shimmila Bhowmick, Pratul Kalita, and Keyur Sorathia. 2020. "A Gesture Elicitation Study for Selection of Nail Size Objects in a Dense and Occluded Dense HMD-VR." In IndiaHCI'20: Proceedings of the 11th Indian Conference on Human-Computer Interaction, pp. 12-23
[3] Maria Karam. 2005 "A taxonomy of gestures in human computer interactions."
[4] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. "User-defined gestures for augmented reality." In IFIP Conference on Human-Computer Interaction, pp. 282-299. Springer, Berlin, Heidelberg, 2013
[5] Isabella Poggi. 2001. "From a typology of gestures to a procedure for gesture production." In International Gesture Workshop, pp. 158-168. Springer, Berlin, Heidelberg
[6] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined Motion Gestures for Mobile Interaction (CHI' 11). ACM, New York, NY, USA, 197–206. DOI: http://dx.doi.org/10.1145/1978942.1978971
[7] Feng Tian, Fei Lyu, Xiaolong Zhang, Xiangshi Ren, and Hongan Wang. 2017. "An empirical study on the interaction capability of arm stretching." International Journal of Human–Computer Interaction 33, no. 7: 565-575
[8] Fereydoon Vafaei. 2013. "Taxonomy of gestures in human computer interaction."
[9] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing (CHI' 09). ACM, New York, NY, USA, 1083–1092. DOI: http://dx.doi.org/10.1145/1518701.1518866
[10] Huiyue Wu, Yu Wang, Jiali Qiu, Jiayi Liu, and Xiaolong Zhang. "User-defined gesture interaction for immersive VR shopping applications." Behaviour & Information Technology 38, no. 7 (2019): 726-741