What feels natural?: exploring the use of gesture based interactions on a distant 2D screen in virtual reality

Yangfan Chen
5194652
Utrecht University
y.chen40@students.uu.nl

Dongcao Lu
2218259
Utrecht University
d.lu2@students.uu.nl

Yu Su
5268672
Utrecht University
y.su3@students.uu.nl

Axel van Buren
7960522
Utrecht University
a.w.p.vanburen@students.uu.nl

Tim Goedhuys
6231926
Utrecht University
t.h.j.goedhuys@students.uu.nl

Artur Salek
6790615
Utrecht University
a.t.salek@students.uu.nl

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

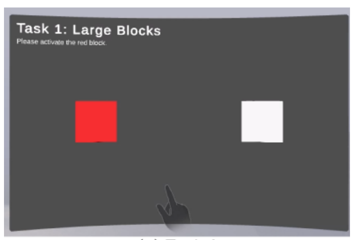
Abstract
With the increasing accessibility of virtual reality, it has become apparent that the various amounts of approaches to interacting in a VR space differ quite. As hand-tracking technology is only increasing in its importance this research pursues a goal state where an universal standard for gestures is created. For this a 4 stage research is done, where stages 1 and 2 concern an elicitation study focusing on the initial gesture made when interacting with a distant 2D screen in virtual reality performing various tasks. Stages 3 and 4 form an evaluative study on newly implemented gestures. Following these 4 stages, it can be concluded that to create the feeling of natural gestures a spectrum of gestures needs to be created. Based on the findings of the 4 stages factors, the gestures should be ranging from the smallest to largest gesture forms being applicable to the same task. And evaluated on the found themes. Using these findings the feeling of natural gestures in the future implementation for VR can be enhanced.

Author Keywords
Virtual Reality; Hand-tracking; Gestures; Natural interactions; Intuitiveness; Qualitative Study

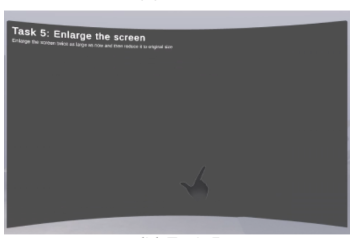
CCS Concepts
•Human-centered computing → User studies; User centered design;

Nr.	Task Description
1	Activate two large blocks on opposite sides of the screen.
2	Activate two small blocks on opposite sides of the screen.
3	Adjust a slider from 0 to 75%.
4	Center a paragraph marked in red by scrolling.
5	Zoom in the screen to 200% and return to original size.
6	Move the screen closer by 50% and back to original.

Table 1: list of tasks for elicitation study



(a) Task 1



(b) Task 5

Figure 1: sample of tasks, captured from an elicitation study session

Introduction

Hand-tracking-based gesture interaction is currently experiencing a surge in popularity. For example, Meta's Quest 2 and Quest 3 devices, allow users to interact with the virtual world solely using their hands. This interaction method leverages hands as the primary part of us to interact with objects, frees us from additional hardware such as controllers, thus offering a more natural and immersive experience in many use cases[2]. Despite its advantages, hand-tracking interaction is not without its flaws, such as lower precision and the absence of tactile feedback[2]. Thus, further research about how to design gestures for specific tasks is necessary. In this study, we focus on researching natural gestures for interacting with distant 2D screens within virtual reality (VR) environments.

Previous studies on gestures have yielded many meaningful conclusions for designing gestures. Studies by Xu et al. [7] found that the hand dwell gesture outperforms the hand pinch in efficiency and accuracy for task operation. Schjerlund et al.[5] introduced a gesture that maps the user's hand to a matrix of virtual hands in VR, decreasing the distance to targets. The world-in-miniature method, which compresses the space into an arm's reach map, facilitates the manipulation of distant objects [2]. Meanwhile, Maslych et al. [3] integrated cone-casting [2], world-in-miniature, and the grasping metaphor to develop an interactive disocclusion mini-map, enabling the selection of long-distance and occluded objects.

As for the definition of natural gestures, Vatavu et al. [6] describe them as actions influenced by users' educational and cultural backgrounds, as well as the specific situations they encounter. Nielsen et al. [4] outline principles for developing intuitive, ergonomic gesture interfaces that are easy to use and remember. LaViola et al. [2] delve into natural decision-making, highlighting it as the process individuals use to navigate complex tasks, shaped by experiences and societal influences.

The contribution of this research is two-fold. First, w

Building on the aforementioned insights into natural gestures and decision-making processes, we propose our research question as follows:

RQ: For users with no VR experience, what is perceived as the most natural way to interact with a 2D screen beyond arm's reach in a VR environment using hand-tracking?

Elicitation Study Design

This research initiated an elicitation session to collect initial data on the natural gestures people use for tasks involving interacting with a 2D screen in VR beyond arm's reach.

For our study, we recruited 8 students from Utrecht University's Multimodal Interaction course, ensuring they had no previous experience with VR hand-tracking to maintain unbiased perceptions of natural interaction. The study's elicitation sessions were conducted in a classroom setup, where participants were seated. They engaged in a virtual environment created with Unity on a Meta Quest 3 device, connected to a Windows laptop via USB. From about 1 meter away, researchers operated the laptop and observed as participants performed a series of tasks (see Table 2.) using intuitive gestures.

To be able to capture every possible gesture, the virtual environment did not actually respond to the user's gestures. Instead, we applied the Wizard of Oz method [1] to provide participants with feedback. Participants did not interact with the VR world directly. When a participant was asked to complete a task, the response of the operation (e.g., in task

Operation times	First gesture	Second gesture
Pointing	12	5
Pointing + other movement (pointing +)	7	1
Pinching	5	3
Pinching + other movement (pinching +)	10	5
Grabbing	0	2
Grabbing + other movement (Grabbing +)	6	0
Palm spreading	3	4
Others	1	2
Sum	44	22

Table 2: list of tasks for elicitation study

1, a block becoming red) was manipulated by the research conductors based on their observation of the participant's behaviour. Tasks were expected to be repeated three times. Two sample tasks can be found in Figure 1.

After each task, participants were asked about their initial gesture choice, its suitability for the task, its intuitiveness and ease, and physical comfort. Additionally, they were inquired if they had alternative gestures they wished to test, allowing for task repetition

Elicitation Study Results

For 6 tasks, we received 44 natural gesture choices, 30 natural gesture choice reasons, and 22 second gestures after data cleaning. The statistics of gestures is shown as Table 2. The reasons for natural gesture choices are concentrated in phone-like interaction, natural object interaction, and object size, see appendix A for the quotations.

Pointing and pinching as primary choices Across all tasks, "pointing" or "pointing +" as initial choice, and chose "pinching" or "pinching" as secondary gesture or reversed. These 4 kinds of gestures occupied 57.6% gestures. Besides, "Grabbing", "Grabbing +", and "Palm spreading" happened mostly in task 5 and task 6.

Object size influences the size of the gesture. Task 1 and 2 are simple selection tasks therefore "pointing" and "pointing +" were chosen 10 out of 14 times. Task 1 compared to task 2 is a larger object, leading to 6 participants to chose pointing. Task 2, requiring more accuracy, 5 participants opted for slower pointing, while others altered their gesture to create a circle to lock on the object.

Feedback influences gesture performance On complex tasks, like task 3 and task 6, requiring extra movement and moving the fingers to a specific location, it's more complicated and precise requiring the system to provide gradual feedback. Various altering gestures were used to finish the task, such as pinching, grabbing, and pushing. Among them, "pinching +" gestures were used 7 out of 14 times. "Grabbing" was also used 4 times here, as said by participants "feels natural to just move it back and forth".

The similarity of the experience also influences the gesture. Task 4 and 5 are interactions that the participants also operate on 2D screens. Thus most of their reactions are similar to 2D screen gesture interactions, such as "pointing and flicking up", "pinching and zooming", and "scrolling by pointing and moving". Participants' reasons for using such gestures were similar, mentioning "phone-like" 11 out of 14 times.

Implementation of the Selected Gestures

Based on the results of the elicitation study, gestures were implemented that positioned distant 2D screens through eyes and index finger top, as this method is dominantly used according to the elicitation study. During the implementation, we applied some key adjustments to make this positioning method usable (see Figure 2). First, binocular vision creates two images of hands when focusing on a distant screen, which causes ambiguity to the user. To resolve this, we introduced a hand model projection on the target screen, calculated from the eyes' midpoint and hand position to disambiguate the double image issue. Figure 2 Second, using the fingertip as a reference causes position drift during actions like tapping. Instead, we opted for the wrist as a reference point, with a manual position compensation from the wrist to the fingertip. Using this positioning method, we implemented two gestures: a direct index finger tap and a pinch. The tap gesture uses a simple algorithm to detect the index finger's forward and backward movements (see Figure 4). The pinch gesture is based on the relative

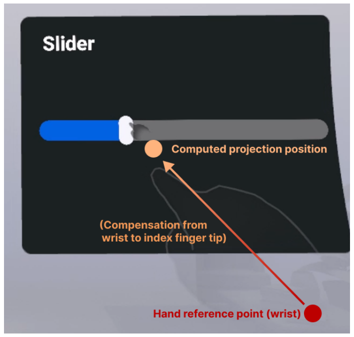


Figure 2: demonstration of hand projection positioning

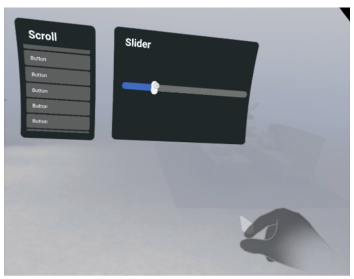


Figure 3: Meta Quest default ray gesture interaction

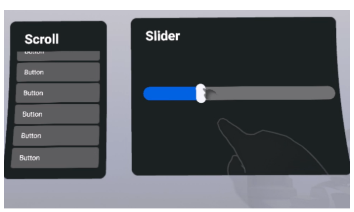


Figure 4: Implemented tap gesture

distance between the tips of the index finger and thumb (see Figure 5). The input from these gestures is then processed by the event handlers of the 2D screens.

Evaluation Study Design

The sequential step consists of a think-aloud evaluation study, where participants explore gestures on their own and report their thoughts and feelings to researchers.

The evaluation study sessions were conducted in one of the researchers' houses, with a setup similar to that of the elicitation sessions. Four master's students from diverse programs were recruited for the study.

A think-aloud protocol was used for the study. Before the experiment, participants were informed how to adjust the VR headset, the experiment's purpose, the gestures they would use, what the VR environment would look like, and how to do a think-aloud session. Participants then explored three gestures to interact with two distant VR 2D screens side by side. One screen allowed scrolling and browsing through a list, while the other had a slider for adjustment. The study excluded tasks involving direct screen manipulation (tasks 5 and 6 from the elicitation study) due to time constraints. Interruptions occurred if participants ceased to verbalize thoughts, needed elaboration, or struggled with gestures or technical issues. Researchers also observed and took notes for sequentially analysis.

Participants explored three gestures: Meta's default gesture, which fires a ray from the user's hand displaying a round cursor on the 2D screen, and the two previously mentioned and implemented gestures.

Evaluation Study Results

Several essential themes emerged from analyzing the study's transcriptions and video recordings. **Fatigue:** was a notable concern. Some participants explicitly reported relief and comfort when switching back to the Meta default gesture, while others reported fatigue when using both the implemented tap and pinch gestures. Additionally, The tap gesture led to larger movements and arm extension due to missing feedback, further contributing to fatigue.

Intuitiveness: In terms of the selection mechanism, participants displayed a clear preference for gestures that aligned with familiar interactions, such as tapping, which resembles the everyday use of smartphones and tablets. However, the pinch gesture was seen as more intuitive by some as it resembled manipulating real-life objects, indicating varied user perceptions regarding the naturalness of different gestures. Regarding the positioning method, some participants reported that the Meta default gesture was not intuitive to use. One participant, in particular, required a long period of adjustment and some instructions to eventually learn how to use this gesture. For the implemented gestures, all participants quickly grasped the method to use them and found them intuitive.

Feedback and Uncertainty: The tap gesture notably lacked feedback, leading to high levels of uncertainty among users. Researchers observed participants increasing their finger's travel distance and tapping multiple times quickly when the system failed to recognize their intent. In contrast, the pinch methods from both the Meta default and the implemented gestures met user expectations. Participants reported immediate system responses to the touch between two fingers, finding it satisfying.

Discussion

The main limitation of this research is time and consequently the subject pool, only containing Dutch students, is not varied enough to determine if there is a universal intu-

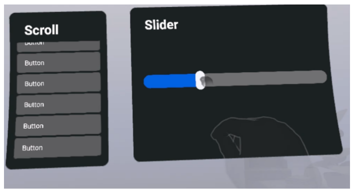


Figure 5: Implemented pinch gesture

itive consensus. However, based on the before-mentioned results it can be seen that there seems to be no real consensus for initial gesture usage. For the evaluation study, incorporating more varying backgrounds would increase the reliability of the results. Furthermore, this study tries to observe intuitive behaviour when it comes to gestures in VR, by applying the Wizard of Oz method, the freedom to make any gesture work is achieved. However, this does leave room for experimentation error where the participant interacts too fast or unexpectedly for the researcher to correctly influence the VR space. The results of the elicitation study showcased the factors: object size, feedback and similarity of the experience influence the gesture that is initially made.

Using this knowledge new gestures were implemented. From the evaluation study on these gestures, the results showed that the themes of fatigue, intuitiveness and uncertainty combined with feedback are most prominent when evaluating the performance of gestures. These results have revealed that for gestures in VR to feel natural and intuitive, a spectrum of gestures should be created. Many participants performed the initial gesture of other participants for their second gesture. A key phrase used by many participants was "I want this to also work" (see Appendix A), whilst performing their second gesture. Therefore the idea of creating a spectrum of gestures that allows for gestures ranging from the smallest to the largest, to work in the same way as manipulating the VR space, is appropriate. However, based on the evaluative study those 3 themes should be kept in mind whilst designing the outer ranges of these gestures, by performing more user evaluation on them. In a follow-up study, the main focus should lie on allowing different variations of gestures to achieve the same effect. Creating a system that works with hand-tracking gestures that for example range from pinching to full-on 2-handed grabbing. This would improve the natural feeling of gestures and enhance the idea of immersion. After creating such a spectrum of gestures this should be evaluated across varying participants improving the reliability of the spectrum encapsulating as many preferences as possible.

Conclusion

Looking at the vast amounts and differences between gestures and their implementations across VR platforms, the pursuit of finding the ideal state in which a universal standard for hand-tracking interactions in VR is established. However, a universal standard can only be achieved when the gestures are perceived as natural, which is achieved when they seem intuitive and/or shaped by personal experience. To address our research question we conducted a 4 stages study. In stages 1 and 2 the elicitation study was done to observe initial gesture behaviour. The results showed no real consensus for specific gestures, however, it did provide insights into 3 specific factors influencing gesture choice: object size, feedback and similarity of the experience. Sequentially, in stages 3 and 4 the evaluation study was done. By implementing new gestures focusing on the perception of the virtual world, to positively influence the factors of feedback and similarity of the experience these were evaluated. From the evaluation study, the following themes were found: fatigue, intuitiveness, feedback and uncertainty. Even though the subject pool was limited and the time allocated for this research was short it can be concluded that these results have led to the understanding that for gestures in VR to feel as natural and intuitive, a number of factors influence this. Our research provides a solution that requires creating a spectrum of gestures that allows for gestures ranging from the smallest to the largest, to work in the same way when manipulating the VR space. However, based on the evaluative study, those 3 themes should be kept in mind whilst designing the ranges of these gestures, by performing more user evaluation on them.

REFERENCES

[1] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces*. 193–200.

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

[3] Mykola Maslych, Yahya Hmaiti, Ryan Ghamandi, Paige Leber, Ravi Kiran Kattoju, Jacob Belga, and Joseph J. LaViola. 2023. Toward Intuitive Acquisition of Occluded VR Objects Through an Interactive Disocclusion Mini-map. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, Shanghai, China, 460–470. DOI: http://dx.doi.org/10.1109/VR55154.2023.00061

[4] Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. 2004. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction*, Gerhard Goos, Juris Hartmanis, Jan Van Leeuwen, Antonio Camurri, and Gualtiero Volpe (Eds.). Vol. 2915. Springer Berlin Heidelberg, Berlin, Heidelberg, 409–420. DOI: http://dx.doi.org/10.1007/978-3-540-24598-8_38 Series Title: Lecture Notes in Computer Science.

[5] Jonas Schjerlund, Kasper Hornbæk, and Joanna Bergström. 2021. Ninja Hands: Using Many Hands to Improve Target Selection in VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. DOI: http://dx.doi.org/10.1145/3411764.3445759

[6] Radu-Daniel Vatavu, Stefan Pentiuc, and Christophe Chaillou. 2005. On natural gestures for interacting in virtual environments. *Advances in Electrical and Computer Engineering* 24 (Jan. 2005).

[7] Wenge Xu, Xuanru Meng, Kangyou Yu, Sayan Sacar, and Hai-Ning Liang. 2022. Evaluation of Text Selection Techniques in Virtual Reality Head-Mounted Displays. (Oct. 2022). http://arxiv.org/abs/2209.06498 arXiv:2209.06498 [cs].

Appendix A

Task	Initial gestures (*nr. Of participants*)	Remarks
1	Pointing/tapping with 1 finger (6) Point with entire hand (1) Pinching with 2 fingers (1)	“Like a phone”/”Tapping feels natural”/”reaching is not comfortable” “Object is small so to select make sure to entirely cover it” “Feels like I need to confirm my selection”
2	Pointing/tapping with 1 finger (5) Pinch with 2 fingers to form selection circle (2) Pinching with 2 fingers (1)	“Need for accuracy”/”Like a phone”/”I move slower than before” “Too small so surround the object to select” “Feels natural and precise”
3	Grabbing with 2 fingers then drag (3) Point with 1 finger and then slide (3) Use entire hand to point and slide (1) Pinch on 1 spot and then the next (1)	“Want to grab the knob” “Like phone volume slider” “The button is very small so I want make sure I touch it” “Faster than just sliding it”
4	Point with 1 finger then flick up (4) Grab and place with 1 hand (1) Wave entire hand up and down (2) Pinch with 2 fingers and move it up (1)	“Scrolling like phone” “Feels natural to reposition like that” “Like a phone but big” “First need to grab it to move it”
5	Pinch and drag with both hands (4) Pinch zooming with 2/3 fingers (3) Clench and spread entire hand (1)	“Zooming like a big phone”/”Natural too stretch something like that” “Feels natural like a phone” “Zooming on a very big phone”
6	Pinching with 1 hand then drag with arm (5) Grab with 2 hands and then push and pull (1) One handed cup and drag to then throw back (1) Push and pull with 1 hand, no grabbing (1)	“Thin object is easy to grab with 2 fingers, like movies” “Large screen and want it up close like a newspaper” “Feels natural to just throw back like in the movies” “Feels natural to just move it back and forth”