# Exploring Visualizations for Precisely Guiding Bare Hand Gestures in Virtual Reality

Xizi Wang
University of Waterloo
Waterloo, Canada
l84wang@uwaterloo.ca

Ben Lafreniere
Reality Labs Research, Meta
Toronto, Canada
benlafreniere@fb.com

Jian Zhao
University of Waterloo
Waterloo, Canada
jianzhao@uwaterloo.ca

## ABSTRACT

Bare hand interaction in augmented or virtual reality (AR/VR) systems, while intuitive, often results in errors and frustration. However, existing methods, such as a static icon or a dynamic tutorial, can only inform simple and coarse hand gestures and lack corrective feedback. This paper explores various visualizations for enhancing precise hand interaction in VR. Through a comprehensive two-part formative study with 11 participants, we identified four types of essential information for visual guidance and designed different visualizations that manifest these information types. We further distilled four visual designs and conducted a controlled lab study with 15 participants to assess their effectiveness for various single- and double-handed gestures. Our results demonstrate that visual guidance significantly improved users' gesture performance, reducing time and workload while increasing confidence. Moreover, we found that the visualization did not disrupt most users' immersive VR experience or their perceptions of hand tracking and gesture recognition reliability.

## CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; **Empirical studies in visualization**; **Gestural input**.

## KEYWORDS

Virtual reality, visual guidance, error visualization, hand gesture recognition.

**ACM Reference Format:**
Xizi Wang, Ben Lafreniere, and Jian Zhao. 2024. Exploring Visualizations for Precisely Guiding Bare Hand Gestures in Virtual Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3613904.3642935

## 1 INTRODUCTION

Bare hand interaction is becoming increasingly common in augmented reality (AR) and virtual reality (VR) systems and has been integrated into many commercial AR/VR headsets (e.g., Meta Quest series and HoloLens series). It has emerged as a relatively new approach for achieving various tasks in AR/VR, such as object selection and manipulation, locomotion, game interactions, art creation, teaching and learning, and home device controls [22, 42, 47, 48]. Compared to traditional handheld controllers, bare hand interaction demonstrates a range of benefits. First, hand interaction is natural and intuitive to users. This is because, in our everyday lives, we use our real hands to interact with the surrounding environment, such as grasping, rotating, and moving real objects. Thus, this mode of interaction enhances the sense of realism, presence, and immersion, especially in VR systems [7]. In addition, hand interaction facilitates social interactions and communication, allowing AR/VR users to express feelings through their natural body language and providing richer interactions [50].

However, due to the limitations of current sensing and computer vision technologies, hand tracking and gesture recognition are not perfect and often result in errors. For example, when a user intentionally performs a grab gesture, the system may not recognize the input, having no response, or incorrectly recognize the gesture, executing unwanted commands. Both cases can ruin the user's experience in AR/VR [25]. On top of these tracking and recognition errors, AR/VR application designers also face a dilemma. That is, supporting more hand gestures, while providing a richer interaction vocabulary, could lead to a worse user experience, because users may forget a gesture due to non-consensus hand gestures [22], infrequent use of the application, and frequently switching back and forth between multiple applications [2].

One notable solution is to provide detailed visual descriptions of the gesture in the onboarding phase [8] and hope users can replicate it correctly when they perform hand gestures. Existing methods include displaying a static icon signifying the gesture or a dynamic 2D/3D demonstration of the gesture, such as in Meta's First Hand [32]. These methods are suitable for guiding users with simple hand gestures and at a *coarse level*, but fail to express more complex gestures that need guidance at a *micro level* and with higher accuracy. Another solution is increasing the threshold to make the recognition less sensitive; nonetheless, it limits user interaction and does not apply to certain applications. Precisely performing complex hand gestures in AR/VR plays a valuable role in scenarios like emotion expression in social communication (e.g., VRChat[1]), training that requires precise gesture learning [42], and games with rich interactions [7]. However, to the best of our knowledge, there currently exist no studies adequately exploring micro-level visual guidance for precise bare hand gestures in AR/VR.

To fill in this gap, we investigated different visualization designs that can facilitate users with precise bare hand interaction in immersive environments, more particularly, static pose gestures based on Hosseini et al.'s classification [22]. We based our study on VR and considered both single-handed (e.g., thumb up) and double-handed gestures (e.g., take photo). We first conducted a formative study

*CHI '24, May 11–16, 2024, Honolulu, HI, USA*


[1] https://hello.vrchat.com

with five VR users with various experiences in hand-gesture-based VR/AR applications. From the formative study, we discovered four types of main information for designing useful visual guidance, including: the error showing where the user fails (error), the target position that the user should move towards (target), the path to the target position (direction), and the difference between the current and target positions (difference). We also identified several principles for generating effective visualizations for gesture guidance, such as simplicity and universal symbolism. Next, we designed 15 visualizations that exhibit different combinations of the four main information types (e.g., error, error + target). Through an iterative design process along with an additional formative study, which involved six more participants, we carefully assessed each visualization and selected four of them, with each representing the most effective visualization at a different complexity level (i.e., containing different numbers of information types). Subsequently, we carried out a within-subjects controlled experiment with 15 participants to investigate the effectiveness of the four selected visualizations on performing 10 single-handed and 10 double-handed gestures, together with a baseline, as well as collected their perceptions and feedback about the visualizations. Our results indicate that micro visual guidance helped users faster and more precisely perform single-handed and double-handed gestures, as well as reduced the workload during hand gesture performing. Also, we found that the majority of users expressed positive perceptions of the visual guidance with regards to reliability, confidence, immersive experience, helpfulness, and usability. While our studies were conducted in VR, we believe the visualization designs and the obtained empirical knowledge can shed light on the cases of AR and other immersive scenarios.

In summary, our work contains the following key contributions:

- Design guidelines on creating effective visual guidance for precise hand interaction in VR, derived through a two-part formative study;
- An exploration of various visualizations for guiding both single-handed and double-handed gestures, grounded by an iterative design process;
- A controlled experiment that provides both quantitative and qualitative empirical knowledge about the four selected visualizations.

## 2 RELATED WORK

In a recent survey, Hosseini et al. [22] identified a total of six types of mid-air gestures: *static pose gestures* (that holds a pose without considering hand moving), *static pose gestures with path* (that hold a pose and at the same time moves hands), *dynamic pose gestures* (that change poses without considering the hand moving), *dynamic pose gestures with path* (that change poses while moving hands), *stroke gestures* (that consist of continuous motion or stroke using a finger or a stylus), and *multiple gestures* (that include bi-manual gestures mixed of any of above five gestures). In this work, as a first attempt, we are interested in investigating the design of visual guidance for users to perform *static pose gestures* (referred to as "gestures" henceforth). This is because such gestures have been widely used in different types of domains, including accessibility, education, and games. For example, static pose gestures account for over 50% in gaming [22].

In the following, we first review the literature on hand feature errors and guidance in AR/VR and then the general topic of AR/VR visualization techniques.

### 2.1 Hand Tracking and Gesture Recognition Errors

Due to the limitations of optical tracking systems, hand-tracking errors are unavoidable, which may lead to a hand gesture recognition error that ruins users' experience in AR/VR [25]. When the recognition fails, users get confused and frustrated. There exist two types of gesture recognition errors: false positive (FP) and false negative (FN) errors. FP input errors occur when the user does not intend to perform a gesture, but the system recognizes it, or when the user intends to perform one gesture, but the system recognizes it as another one. FN input errors often occur when users intentionally perform a gesture, but the system does not recognize the input. While Lafreniere et al. [25] showed that FP errors tend to be more frustrating, developers usually just fix FP errors by tightening the thresholds; however, this results in more FN errors [24]. The improvement of FN errors is relatively harder by merely manipulating the algorithms, as users' actions also play a crucial role.

Many reasons can cause FN errors, for example, the unfamiliarity of gestures or interactions [8], tight thresholds for recognizing the gestures (a method usually used to prevent FP errors), limitations of optical hand recognition technology (such as the occlusion of partial hands), and poor hand-tracking models. In this work, we mainly focus on exploring real-time visualizations in VR for understanding and correcting gesture errors caused by unfamiliarity and tight thresholds, which are the most predominant reasons that could always exist. Other causes, we believe, are more likely to be addressed in the future as computer vision and tracking technologies become more advanced.

The unfamiliarity often happens when a new gesture is introduced to the user [8] or their usage of the gesture recognition system is infrequent. Thus, users forget how to perform the gestures, encountering a legacy bias, and they tend to transfer gestures that have been performed in previous systems to the current system [2, 33] or try to perform a non-consensus hand gesture [22] but gets confused with a variant gesture that shares the same name. This is a "user error" that cannot be easily resolved by just improving the recognition technologies, which we choose to focus on.

On the thresholding aspect, previous research has investigated improving hand gesture recognition by using loose thresholds or applying bi-level thresholds to reduce FN recognition errors [24, 35]. Although these techniques can mediate the FN and FP errors [24], they do not work for richer and more complex hand gesture interactions that are essential in many application scenarios, for example, teaching and learning of the American Sign Language (ASL) as well as using AR/VR for training in critical domains like medical, engineering, and military. As a more concrete case, ASL involves many gestures that are very similar [26] and requires precise gesture performing because subtle changes may change the entire meaning of a word; the different replacement of the index finger distinguishes the letter "d" and number "1" [28]. As a

pressing requirement, we aim to explore visualizations to address this problem in our work.

## 2.2 Visual Gesture Guidance

Visual guidance is an effective means to facilitate users' gesture performance in different environments including AR/VR. For example, some systems use visual tutorials to assist users in understanding the details of a gesture, like displaying static hand poses in AR/VR [32, 49]. However, this method fails to help users learn and correct their gestures because of the absence of real-time feedback on where the error really occurs. Previous studies have also investigated visualizations to support *stroke gestures* and *dynamic hand pose gestures with path* on a 2D surface. For example, OctoPocus [5] and ShadowGuides [16] dynamically display possible future *stroke gestures* with annotations to help users understand and learn gesture sets. Gesture Heatmaps [45] visualizes *stroke gesture* interactions with color maps to help the user understand gesture variants.

Building upon these 2D techniques, researchers have investigated visualizations for *stroke gestures* and *static hand pose gestures with path* in VR or 3D environments. For instance, Delamare et al. proposed OctoPocus3D [9] that extends the idea of OctoPocus [5], verified its feasibility and adjust-ability in 3D and investigated the positive effect of feedback and feedforward on the performance. Later, Fennedy et al. presented OctoPocusVR [14] to bring the idea of origin OctoPocus [5] into VR and demonstrated its benefits during execution. Lilija et al. [29] explored a new type of guidance embedded in a user's avatar by showing corrected virtual hands in the target location. They found that this new technique improved the short-term accuracy during training and may reduce visual distraction, but not the accuracy over time or outperform ghost hands. Unlike the previous research focusing on facilitating single movements on paths with particular types of information (e.g., error and target), we studied multiple adjustments during pose correction as well as the effect of different levels of information complexity on users' performance and perceptions.

There have also been attempts to visualize a user's arm or body movements in 3D and AR/VR. Diller et al. [10] provided a comprehensive summary of 13 types of visual cues in motor skill training in mixed reality (MR). These cues include but are not limited to body outline [18], end positions [51], and rubber bands [51], that have been used in previous studies and inspired our visualization design for guiding bare hand gestures. Yu et al. [51] explored different perspectives and contrasted two different visual coding styles for body motions. They discovered that the guidance from a first-person perspective is more effective than that from a third-person view, and discrete arm motion guidance outperforms continuous streamer-ribbon-look guidance. This also highlights the importance of showing visualizations of joints in designing effective motion guidance.

Similarly, YouMove [1] uses green ribbons/lines to provide a future movement guide, red dots to indicate error joints, and a green skeleton to illustrate correct posture through an AR mirror to facilitate physical body movement learning. Moreover, LightGuide [43] projects arrows directly on users' hands to instruct mid-air body movements. For arm movement guidance, Han et al. [18] proposed AR-Arm, a coarse-grained egocentric guidance that uses ghost hands to instruct arm movements of Tai-Chi in AR. Along this line, Dürr et al. [11] investigated two ghost-hand-look visual appearances with different levels of fidelity and three paces (the way they guide a user) of coarse-grained egocentric guidance in arm movements and found that using a realistic visual shape and continuous guidance increased accuracy and tended to receive more positive perceptions.

While a large body of techniques has been proposed to instruct users with visual guidance in AR/VR and 3D scenarios, previous research has predominantly focused on guiding hand, arm, and body gesture movements at a coarse level. To the best of our knowledge, no adequate studies have been done on exploring visualizations to guide precise hand gesture interaction in AR/VR as feedback to support *static pose gesture* learning and adjustment, which is our focus in this paper.

## 2.3 Visualization Techniques in Augmented and Virtual Reality

Various visualizations and applications in AR/VR have been extensively explored, especially in the domains of training, analytics, and assessments. In the realm of sports training, Faure et al. [13] conducted a scoping review on the use of different visualizations in different kinds of ball sports training and assessment in VR. They synthesized previous works, including Vignais et al.'s work [46] that explored different visual cues (e.g., dots, wire-frames, and textured models) for representing players and balls. They also suggested using different visualizations with various graphic complexity for skilled and beginning sports players because rich graphic information tends to distract experienced users but has no negative effects on novices [17]. This insight aligned with our observations that VR users may prefer different visual feedback based on their experience and familiarity with hand gestures. Also, it inspired us to further explore and compare different visualizations with various levels of information complexity. Additionally, Lin et al. [30] investigated the effectiveness of a visualized ideal 3D shooting arc for MR basketball training, emphasizing the potential of real-time, fine-grained visual feedback in the development of free-ball-shot skills. This reinforced our commitment to developing dynamic, spatial, and precise visual guidance to facilitate gesture performance and learning in VR.

In the context of analytics and assessments, Lee et al. [27] investigated the transformation of data visualization between 2D and 3D in MR, providing a high-level design guideline for transformation design. Further, Luo et al. [31] proposed and evaluated a visual MR analytics toolkit with various types of visualizations to support explorations of human movement data, utilizing visual elements like lines and arrows for path and direction representation. Similarly, Nebeling et al. [34] introduced MRAT, an analytics toolkit designed for the collection, visualization, and analysis of user performance data in MR in diverse scenarios, including crisis informatics and nursing. The toolkit uses 3D arrows to indicate directions and abstract objects to represent individual users. Our design of the visual guidance, such as using arrows, spheres, bars, and lines, has been inspired by some visual encoding used in these systems.

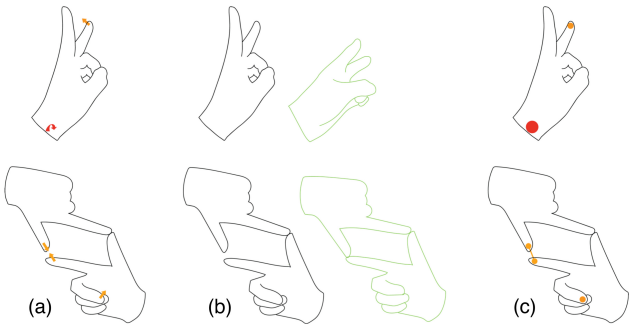


**Figure 1: Initial designs of visual guidance used during the co-design session of the formative study. (a) uses arrows to instruct users on finger movements and wrist rotation for gesture adjustment. (b) presents a ghost hand alongside the user's virtual hands, demonstrating the correct gesture. (c) utilizes spheres to signify where hand positioning may need correction. Graduated color shifts in (a) and (c) indicate the degree of deviation from the target position, with red indicating the larger deviation and orange signifying fewer errors.**

## 3 VISUAL GUIDANCE DESIGN

As there exist few studies comprehensively exploring visualization designs for guiding users to perform static hand gestures in AR/VR, we conducted a two-part formative study with 11 participants to investigate the design space and understand the users' needs.

### 3.1 Formative Study: Part I

The goal of the Part I formative study is to understand what information the users might need to understand gesture recognition errors and what properties a visual guidance should have to help users recover from these errors. In particular, as mentioned previously, we focus on false negative errors caused mainly by users' unfamiliarity with the gestures and high-precision requirement of the gesture performance (i.e., tight thresholds).

*3.1.1 Participants and Procedure.* We recruited five participants (1 female and 4 males) with different experiences in VR hand tracking and recognition applications, from mailing lists and through word of mouth. Three of them are in the age group of 20-29, and the other two are in 30-39. Two reported to have 3-5 years of experience in AR/VR, and three reported 1-3 years of experience. One participant is a CTO of a VR startup, and the other participants are graduate students. During the study, we first conducted a 30-minute semi-structured interview to collect their experience with hand or body tracking and gesture recognition errors, especially in AR/VR. We also asked participants to share their experience of existing tools that helped them when they failed to perform a gesture correctly. We then invited them to participate in a co-design session to explore potential visual guidance and criticize three initial visualization designs that were inspired by some visualizations in previous studies [11, 43, 49] using mock-ups and sketches (Figure 1). We asked them to provide their ideas of how to improve the visualizations or describe their ideal visual guidance designs. At the end of the study, each participant received $10 as remuneration. In the following, we use "S1-P[X]" to refer to the participants in the formative study.

*3.1.2 Results.* We transcribed the entire study session, the leading author conducted a thematic analysis of the interviews using an affinity diagram, and the rest two authors cross-checked the results. Results were organized around several themes (see Appendix Table 7). Overall, participants mentioned that they often got frustrated when they failed to perform a hand gesture in AR/VR applications, and expected someone to tell them why. *"I don't know if I forgot the motion or if my position is just changing, or if I was subconsciously changing something, the recognition accuracy dropped like crazy. And I felt so disappointed. It was working two minutes ago. And the problem is I don't know why it's not working. I don't know what's going on."* -S1-P5 This confirms the need for real-time visual guidance in AR/VR systems, especially for those requiring deliberate hand gestures.

From participants' comments, we distilled four types of information that an ideal guidance should contain:

1. **Error (mentioned by S1-P1, P4 and P5): What is wrong?** The errors showing why the users fail. For example, S1-P5 thought that *"showing what is wrong, like just showing errors, will help me fix the problems"*
2. **Target (mentioned by all participants): What is correct?** The information containing: a) the correct/target position from a micro perspective, and b) the overview of the correct gesture from a macro perspective. For example, S1-P2 thought *"(c) is like a mystery because it does not tell me what to do next and what the correct answer would be."*
3. **Direction (mentioned by all participants): Which way is it?** The path or direction indicating how the users can fix the errors. For example, S1-P1 thought that *"a combination of (a) and (c) would be more helpful because it shows both my current location and then correct location. Also, it gives me a path for how to move from the current location to the correct location."*
4. **Difference (mentioned by S1-P2, P3 and P5): How far is it?** The information indicating where the users are in recovering themselves to the correct/target position. For example, S1-P5 thought that *"having the numbers to know how far I am away from the threshold is helpful. And over time, I will naturally get better at meeting the threshold every single time I want to be."*

In addition, based on our thematic analysis, we summarized four important properties that effective visual guidance for static hand gestures should contain:

- **Simple.** Our participants believed that the visualization should be clear and informative and, at the same time, not overwhelming. Also, the gesture guidance, which is on top of the existing VR applications (e.g., a game), should minimize attention distraction and immersion breaking. *"The (a) on the top left, it doesn't make any sense to me. It's too complicated."* -S1-P2 *"If I see a ghost hand beside, and I know that's my target gesture, then I will just try to fit my hand into the ghost hand (b). And that's a lot easier than trying to adjust for six different arrows [in a]. So maybe the ghost hand one (b) is just better and straight up."* -S1-P5
- **Universal.** The participants expected that the visualization should use symbols that convey the same information to a broad audience. For example, we found that color gradients might not be a

good option to represent the differences because our participants had different interpretations of the meaning of the colors. *"I like (c) the least because the meaning of the color has lost me already. But I believe it will be painful in the first few times when I try to use (c). I will be constantly thinking, what does red mean? What is orange?"* -S1-P1. Additionally, our observations suggest that using arrows as an indicator to guide users in wrist rotation may not be the most effective approach. *"I don't know what the red flip arrow on the bottom mean. I thought I have to flip my hand, but I am not sure. So I have to find a way to make a consent about how to use the symbols and what those symbols mean in a way that we both agree."* -S1-P2

- **Spatial.** As VR applications are inherently 3D, our participants preferred a 3D guidance (e.g., a 3D hand gesture or image) over a simple text description or a 2D image (e.g., an icon). For example, S1-P4 commented, *"The ideal solution for me should be like a full 3d models that demonstrate how that could be."* Also, in 3D, the guidance could use transparent texture to avoid occlusion. *"In a TV game, it's okay to block maybe a third of the screen. But in a VR world, I would expect the pop-up error feedback UI to be partially transparent."* -S1-P1
- **Dynamic.** The visual guidance should provide real-time feedback to reveal the details of changes in hand poses for effectively guiding the users. As noted by our participants, *"There should be animations of the models and how they are doing. I think that could be helpful."* -S1-P4 *"Just from these three types, I would prefer (b), but realistically I want animations."* -S1-P1

From this formative study, we also found that participants expected to see visual guidance in two key scenarios: 1) during a tutorial session or the first time when the gesture is introduced (S1-P1, P2, P3, and P5), and 2) when the user intends to perform a gesture and the system fails to recognize the gesture (S1-P1, P3, P4, and P5). For example, S1-P1 mentioned *"if the system knows I want to do this gesture and I am not triggering it, this [guidance] will be extremely useful, especially when the system is teaching me how to use the gestures."*

Moreover, participants may expect a different style of guidance in the tutorial compared to a follow-up reminder. *"When the first time a gesture is introduced in the tutorial, you would want to show them the ghost hand (b) because it is more constrained and tells you more in a diagram way how you should do it so the user can visualize it. But then, after that, maybe it could move on to (a) where it's more of a suggestion, it's less force for everyday reminders."* -S1-P3 This also implies that experts and novices may prefer different types of guidance. Novices may require more extensive and detailed guidance to effectively adjust their gestures, compared to experienced users who have hand tracking experiences or are more familiar with the gesture. *"Clear guidance always is always a good thing. But how much you want to go into the details for the expert and for the newbies are so different."* -S1-P2 This observation also aligns with Fitts and Posner's three stages of learning [15] that beginning users during the *cognitive* stage can benefit more from detailed guidance. As they move toward the *associative* stage, they may need fewer visual cues, while experts during the *autonomous* stage may require minimum guidance (such as a small reminder).

In addition, we found that participants preferred different types of guidance for various amounts of deviations or during different phases of error recovery. *"If the difference in positions is subtle, you can't really tell the difference from (b)."* -S1-P2 For example, they may need an overview of the gesture at the beginning of error correction or when the difference between their current state and the correct state is large; and prefer a subtle guide for fine adjustments when they are close to the target position.

## 3.2 Exploration of Visual Guidance Designs

Based on the above four types of information, we designed a total of 15 visualizations, each containing a different combination of information types, altogether exhibiting different levels of complexity. Four of them contain one type of information, six contain two different types, and so forth ($\binom{4}{1} + \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 15$). We situated our visualization designs based on some existing AR/VR practices (e.g., the visualization of arm/hand skeletons [49, 51]) and MR visual cues (e.g., the rubber band and end position [10, 51]). We also refined our designs considering participants' feedback on our initial design mock-ups, such as the confusion over the use of color for conveying the information "difference" and preferences of the "rubber bands" design shown in our initial design (c). In specific, we used small spheres to present the position of hand joints/fingertips, and used the orientation and length of lines to present the direction and distance. As an example, Figure 2 illustrates all the 15 visualization design sketches for the "take photo" gesture in 2D forms. In Figure 2-V1, which just indicates the error information, the purple (representing keypoints which may be fingertips or joints) and orange (joints) spheres show which ones are wrong; whereas in Figure 2-V2, which just indicates the target information, dots are presented to exhibit the target positions of the corresponding joints/fingertips. In Figure 2-V3, which just shows the direction information, lines with an equal length are presented from the current positions of the joints towards the target positions; whereas in Figure 2-V4, lines with different lengths are shown to indicate the magnitudes of distances to the target positions but they are all in parallel with the hand skeleton bones (i.e., no directions indicated). In the case of Figure 2-V34, V234, and V1234, which encompass both direction and difference information, there is only one line extending from each joint or fingertip's current position towards the target position with changes of length and conveying both directional guidance and the deviation from the target position. Similar design ideas were employed for combining different types of information in the visualizations, as shown in Figure 2.

## 3.3 Formative Study: Part II

Based on the exhaustive list of visualizations, we moved on to develop high-fidelity prototypes and aimed to further assess the visual guidance designs in Part II of our formative study. We carefully reviewed all the visualizations and ruled out some designs based on the four key properties of effective visual guidance derived from Part I of the formative study. We filtered out V2, V3, V4, V23, V24, V34, and V234 because these visualizations exclude the "error" information (Figure 2). Without the error (i.e., the user does not know which part is incorrect), visualizations need to show visual indicators on every joint and fingertip of the hands, leading to an

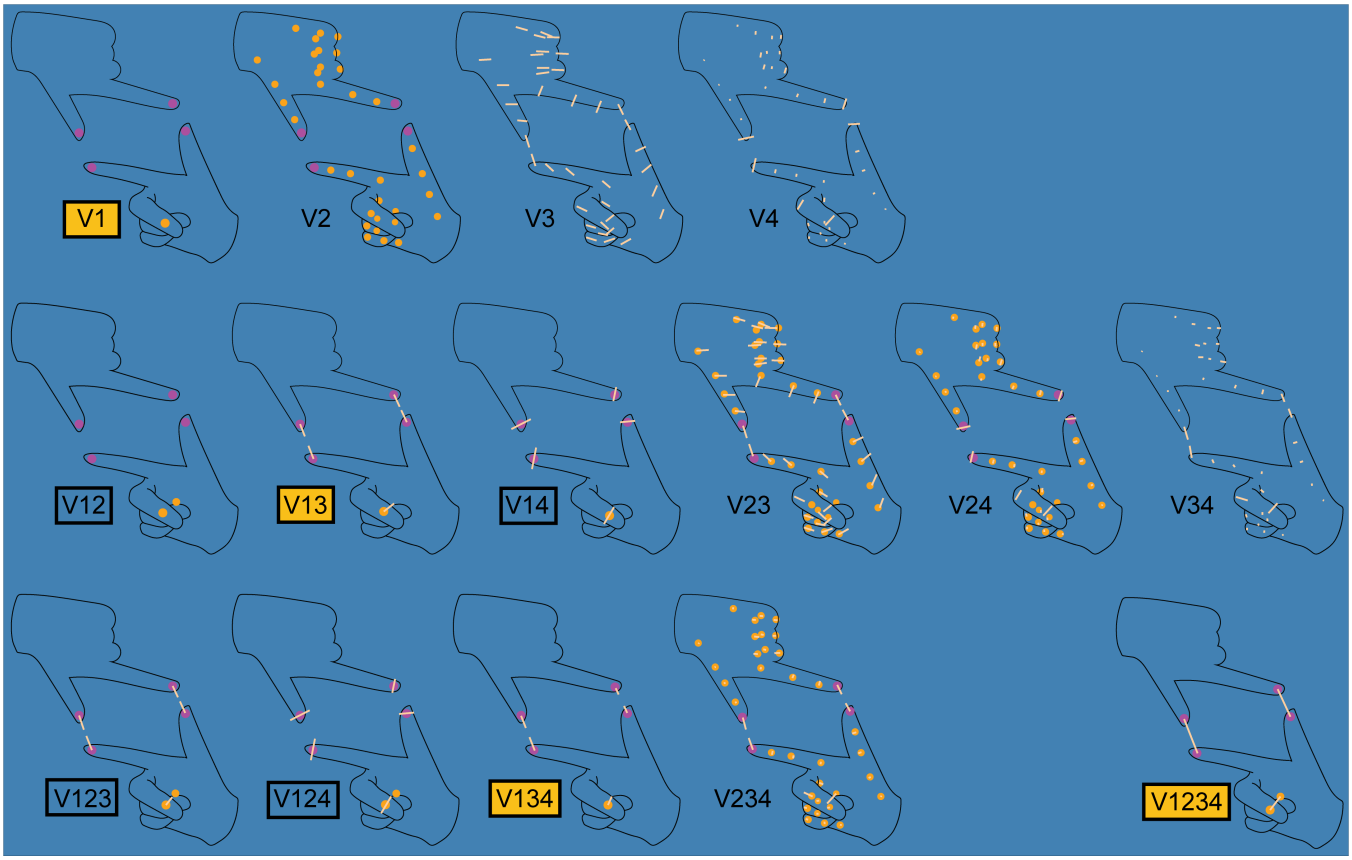


**Figure 2: A sketch of 15 types of visualizations on a pair of hands performing the "take photo" gesture. The naming of the sketch (V[X]) indicates the types of information the visualization contains, where 1 = error, 2 = target, 3 = difference, and 4 = difference (e.g., V12 is a visualization showing error and target information). Outlined designs indicate the visualizations selected to rank in Part II of the formative study. Orange-colored designs indicate the four top-ranked visualizations within their complexity level.**

overwhelming number of visual cues. This decision aligns with the essential property "simple," ensuring the guidance conveyed to the user is clear and minimally distracting.

*3.3.1 Participants and Procedure.* We recruited an additional six participants (2 females and 4 males) through the same means in Part I. All of them are students and in the age group of 20-29. Five reported have 0-1 year of AR/VR experience, and one reported 1-3 years of AR/VR experience. Three have no hand gesture experience, and the other three have used hand gestures as input in AR/VR or other systems. Using a low-cost evaluation approach introduced in [21], we developed initial prototypes for the selected eight visual guidance with the same "photo taking" gesture in VR, and created a video recording of a user performing the gesture for each visual guidance (Figure 3). During the study, we instructed the participants to watch the videos and rank these eight visualizations using an online questionnaire. We invited the participants to consider different aspects in their rankings, including intuitiveness, simplicity, clarity, cognitive workload, informativeness, and learnability. We also asked them to share their rationales for the ranking. Like

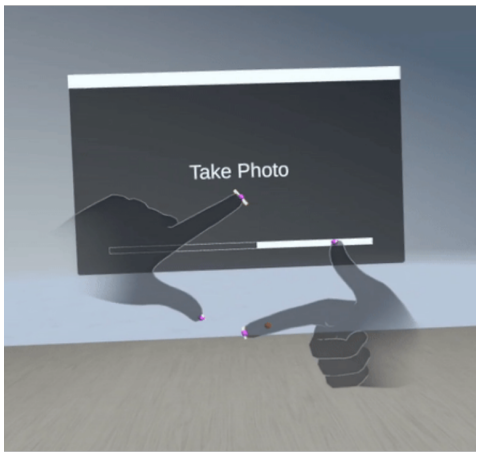


**Figure 3: A screenshot of a recorded video of performing a "take photo" hand gesture with the visual guidance design V13 (see Figure 2).**

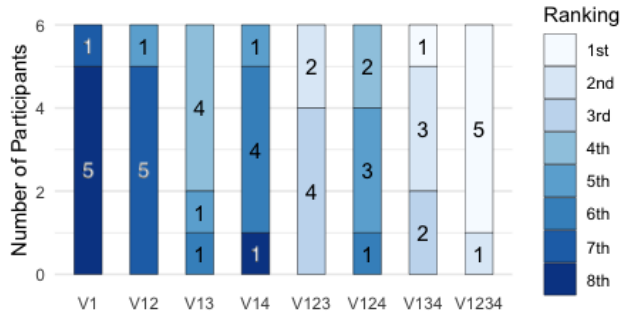


**Figure 4: Distribution of participants' rankings for the eight visualizations in Part II of our formative study (1 = error, 2 = target, 3 = difference, and 4 = difference). A lighter color indicates a higher ranking.**

in Part I, we continue labeling the participants with "S1-P[X]" for Part II.

*3.3.2 Results.* Figure 4 shows the distribution of rankings from the participants. We computed the averages of rankings and selected the lowest four (i.e., the top-ranked visualizations). We can see that V1234 and V134 were most participants' first or second favorite choices. With the visualizations containing two types of information, V13 had the highest average ranking. V1 was the least favorite visualization, which is not a surprise because it contains significantly less amount information. We also identified a trade-off between the richness of information and cognitive workload. *"I think I need the visualization which has the most feedback but is also as simple as possible. Showing the destination for all joints might be a little bit too confusing and hard to distinguish."* -S1-P10 Also, we confirmed our design choices of using lines, where participants thought lines could effectively guide them to move their hands or fingers and make adjustments. *"The line in [V1234] connects two dots, making it easier to know where to move."* -S1-P6

From this study, we also observed that when provided with more information, participants considered certain pieces of information more important. First, knowing which part is incorrect is the most crucial since this information filters out the unnecessary information and makes the user focus on the effective adjustment. The second most important piece of information is the direction, which explains that V13 is the most preferred among visualizations that contain exactly two types of information. This is because, with the direction information, participants could follow the guide without wandering around. *"The 'direction' information is more important than the 'difference' information. If there is no direction, I need to guess."* -S1-P7 The other two types of information, target and distance, were the least appreciated; However, we cannot assert which information is more important because we observed only a subtle difference in the ranking result between V123 and V134.

Based on participants' rankings, we selected four visualizations, one from each complexity level (based on the number of different information types included), including V1, V13, V134, and V1234 which will be labeled as VG1-4 hereforward. Also, based on the feedback, we further edited the visual guidance designs to improve clarity and learnability. We used green spheres to represent the target joint positions to distinguish those from the orange spheres that represent the current positions. We altered the texture of the purple spheres, making them hollow with a purple surface/outline. Particularly for double-handed gestures, we used purple lines instead of orange ones to guide users to move their two hands together. The final visual guidance designs are shown in Figure 5.

## 4 CONTROLLED EXPERIMENT

We conducted a controlled experiment to further investigate the four selected visualizations from our formative study. The goal is to gain both quantitative and qualitative knowledge about the usability and effectiveness of the visual guidance designs.

### 4.1 Gesture Set

We considered a set of 22 hand gestures, which consists of 11 single-hand gestures and 11 double-hand gestures, as shown in Figure 6. Among these gestures, S1-10 and D1-10 were used in the actual tasks, and D11 and S11 were used in the tutorials of the study, which will be described later in this section. Most gestures (20 out of 22) were selected from previous papers [2–4, 6, 22, 28, 37–39, 47], except that S10 and S11 were chosen from ASL [26, 28]. Both single-handed and double-handed gestures were recorded using Oculus Integration SDK's *Hand Grab Pose Recorder* feature by the same researcher in a bright environment.

### 4.2 Study System

To simulate the usage of hand gestures in a VR application, we developed a study system using Unity (2020.3.42f) with Oculus Integration SDK (version 50.0). We also edited a demo scene provided by the SDK, where users can see a virtual desk and a virtual information panel in the front surrounded by furniture sets (Figure 8). In the front virtual panel, users can find a blue progress bar at the top, a white timer bar at the bottom, and the name and image of the hand gesture they should perform in the middle. During the study, a desktop computer was used to run the study system and stream the VR scene into the Oculus Quest 2.

One key question is to define the criteria for a successfully performed gesture. Previous research [37, 39] employed the sum of joint offsets (between the current and target poses) with a threshold (e.g., 5 cm) to detect if a pose is correct. To pursue precise hand gesture interaction, in our system, we imposed a threshold for the offset of each individual joint. For double-handed gestures, we followed the method by Pei et al. [37] and also computed the distances between predefined pairs of keypoints, which can be joints or fingertips, with a threshold. In particular, we used the following measure with a threshold $T$:

$$S = \begin{cases} \bigwedge_{m=1}^{17}(\| t_m - c_m \| \leq T), & \text{if a single-handed gesture} \\ \bigwedge_{m=1}^{17}(\| t_m - c_m \| \leq T) \wedge \\ \quad \bigwedge_{n=1}^{17}(\| t_n - c_n \| \leq T) \wedge \\ \quad \bigwedge_{i=1}^{N}(\| c_{p_i^1} - c_{p_i^2} \| - k_i \leq T), & \text{if a double-handed gesture} \end{cases} \quad (1)$$

In this measure, $t_m$ or $t_n$ represents the target position of the $m$th (the dominate hand) or $n$th (the non-dominate hand) joint and $c_m$ or $c_n$ represents its current position; $c_{p_i^1}$ and $c_{p_i^2}$ represent the

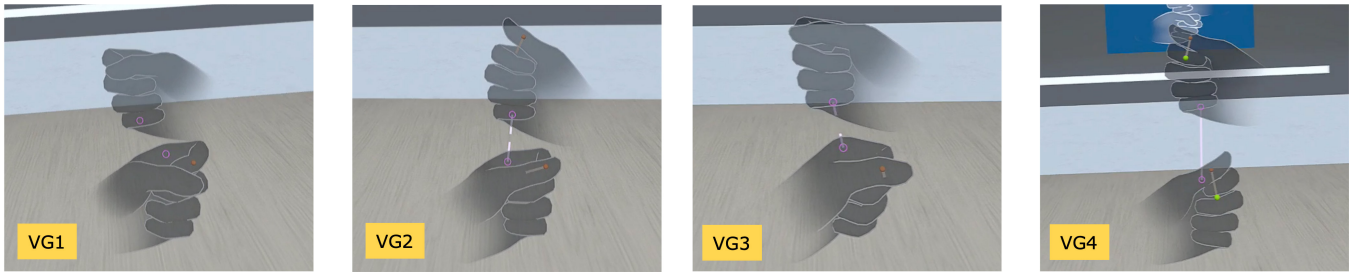


**Figure 5: The final visual guidance designs of the four top-ranked visualizations for performing a "make in ASL" hand gesture. From left to right: VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**


**Figure 6: The 22 hand gestures included in our controlled experiment, where S1-11 are single-handed gestures and D1-11 are double-handed gestures. All pictures above are for right-handed participants. Left-hand participants were provided with mirrored hand gesture pictures.**

current positions of the $i$th predefined pair of keypoints, and $k_i$ is an adjustment parameter taking the consideration of bone/skin thickness. Depending on the specific pair of keypoints required in different gestures, $k_i$ can be 0, 5 mm, or 10 mm. For instance, in the "book" gesture (see Figure 7), the $k_i$ for the pair of pinky distal phalange bones is 5 mm, and the $k_i$ for the pair of pinky proximal phalange bones is 10 mm, because these pairs of keypoints can not touch each other, and different bone/skin has different thickness. However, the $k_i$ for pairs of fingertips in the "Heart" gesture is 0, because there is no gap when two fingertips touch each other. To determine a suitable threshold, we ran a pilot study with three volunteer participants, and in the end, we chose $T = 15$ mm. In summary, for a single-handed gesture, all 17 joints of the hand need to be within the threshold of the correct positions to be recognized as a correct gesture, and for a double-handed gesture, the adjusted distances of all pairs of keypoints need to be within the threshold as well.

## 4.3 Participants

We recruited 15 participants, 6 females and 9 males, with an average age of 27.0 ($SD = 5.15$), via university mailing lists and word-of-mouth. Among these participants, 13 are right-handed, and two are left-handed. They have diverse AR/VR experiences and usage frequencies, as shown in Figure 9. Nine of the participants have previously used hand tracking in either AR/VR ($N = 4$), other systems ($N = 1$), or both ($N = 4$). All participants received 20 CAD as a reimbursement after successfully completing the study. In the following, we use "S2-P[X]" to refer to the participants in this experimental study.

## 4.4 Task and Design

As Figure 8 shows, with the study system, participants followed a series of steps to perform a gesture required by the experimental task. All participants completed the tasks sitting on an office chair at the same spot under the same lighting conditions across the whole

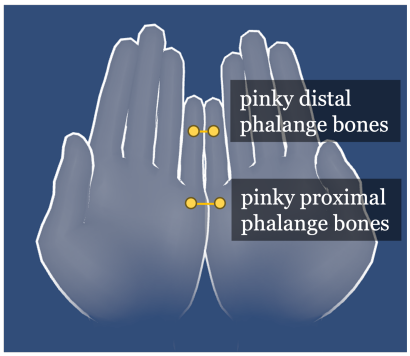


**Figure 7: A "book" gesture with two pairs of keypoints indicated with yellow dots: pinky distal phalange bones (top) and pinky proximal phalange bones (bottom). The yellow line connecting a pair of dots indicates the thickness of the bone/skin.**

study. Whenever the participant felt ready, they first put two hands on the virtual table and used their dominant hand to touch the green sphere to start performing a hand gesture. During each task, participants used their dominant hand to execute single-handed gestures and both hands to perform double-handed gestures. In order to prevent participants from accidentally triggering gestures, the system required them to hold the correct gesture for at least 1.5 seconds to complete the task. When the participant was holding the correct hand gesture, a green progress bar surrounding the gesture image would appear. Participants were given 60 seconds to complete each task. If they failed to hold a correct gesture for 1.5 seconds within 60 seconds, the system ended the trial and moved to the next one.

We employed a within-subjects design for our study, and the two independent variables were visualization and gesture type. In addition to the four selected visualizations (VG1-4) from our formative study, we considered a baseline condition without visualization (NoVG). For the baseline, we adopted an approach commonly used in existing commercial VR applications, i.e., displaying a static hand pose illustration. As mentioned earlier, for gesture type, we considered both single-handed and double-handed gestures. Each participant experienced all five visualizations (NoVG and VG1-4); and for each visualization, they performed four different gestures (including two singled-handed and two double-handed gestures) selected from our gesture set (S1-10 and D1-10). There were two repetitions for each task. Altogether, each participant completed 5 Visualizations × 4 Gestures × 2 Repetitions = 40 trials in the study. The combinations of visualization and gesture were generated for each participant by guaranteeing each gesture was only paired with one visualization (i.e., 20 gestures were required for five visualizations, and each gesture was only seen once), using a balanced Latin square design to counterbalance the visual conditions and gestures to eliminate the ordering effect. This design was to minimize the learning effect of gestures in the study. The order of the visualizations presented to the participants was also randomized.

**Table 1: Results of repeated-measure ANOVAs on the visualization and gesture type for task completion time, completion rate, average sum offset, and average task load. ($^{*}$ for $p < 0.05$, $^{**}$ for $p < 0.01$, and $^{***}$ for $p < 0.001$).**

| Factor | Metric | P-value |
|---|---|---|
| Visualization | Completion Time | $0.002^{**}$ |
| | Completion Rate | $0.01^{*}$ |
| | Avg. Sum Offset | 0.07 |
| | Avg. Task Load | $< 0.001^{***}$ |
| Gesture Type | Completion Time | $< 0.001^{***}$ |
| | Completion Rate | $< 0.001^{***}$ |
| | Avg. Sum Offset | $< 0.001^{***}$ |
| Visualization × Gesture Type | Completion Time | 0.108 |
| | Completion Rate | 0.06 |
| | Avg. Sum Offset | $0.038^{*}$ |

## 4.5 Procedure

At the beginning of the study, we collected participants' consent forms and used a pre-survey to gather their demographic information and previous experiences with AR/VR and hand-tracking systems. Next, participants were presented with five visualizations, one by one, in a random order. For each visualization condition, we first displayed and introduced the visualization, and then participants put on an Oculus Quest 2 to complete four sample tasks (two with gesture S11 and two with gesture D11) in a tutorial session to get familiar with the given visualization and the task. Each task in the tutorial session lasted up to 5 minutes, and they may skip a task in the tutorial if they think they are familiar with the visualization and the task. After the tutorial, participants performed the actual tasks with four different gestures using the same visualization, each having two repetitions. After completing the eight tasks, we helped participants take off the VR headset and asked them to fill out a post-survey regarding their experience and perception of the visualization. The post-survey included a NASA-TLX questionnaire [20] and some questions on their perceived benefits, helpfulness, and usability of the visual guidance, using 7-point Likert scales. Next, we asked them if they needed a break and proceeded to the next visualization. In the end, we conducted a short semi-structured interview with each participant to collect their feedback and rankings on the four visualizations. The entire study took about 60 minutes.

# 5 RESULTS

## 5.1 Quantitative Task Measures

From the study system logs, we obtained task completion time, task completion rate, and average sum of offsets for different conditions in our experiment. Here, we report the results of our statistical analyses on these quantitative measures.

*5.1.1 Completion Time.* As shown in Table 1, a two-way repeated-measure ANOVA indicates a significant effect of visualization ($p = 0.002^{**}$) and gesture type ($p < 0.001^{***}$) on task completion time. Descriptive statistics regarding the completion time for different visualizations and gesture types are shown in Table 2. Further, Figure 10 indicates the means and 95% CIs of the task completion time

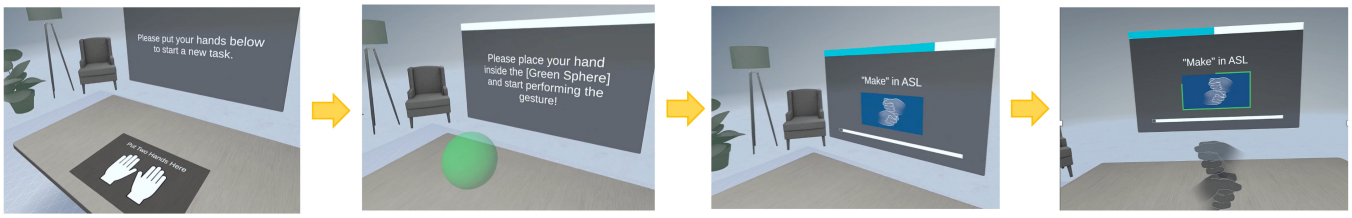


**Figure 8: Steps for completing a double-handed gesture task with our study system.**

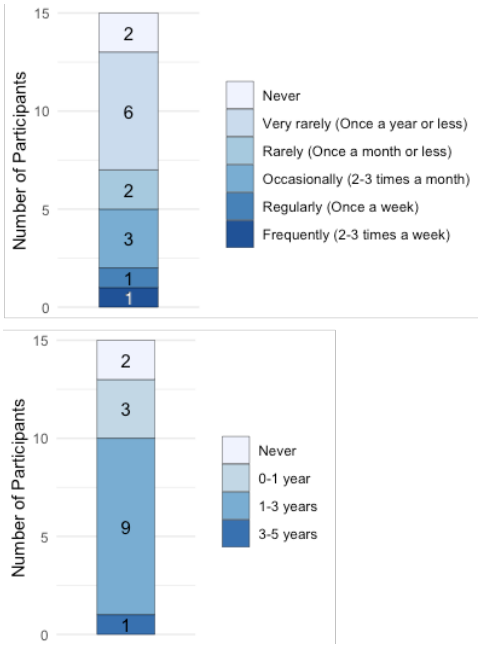


**Figure 9: Distribution of participants with various AR/VR usage frequencies (top) and experiences (bottom).**

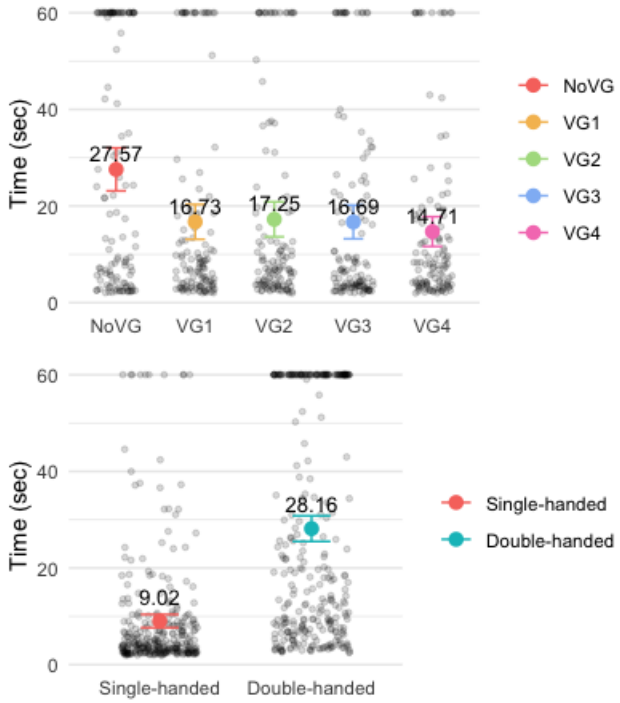


**Figure 10: Mean task completion time (with 95% Confidence Interval) for different visualizations (top) and gesture types (bottom). NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

**Table 2: Descriptive statistics of task completion time and completion rate for different visualization and gesture type conditions. NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

| Condition | Avg. Completion Time (seconds) | Avg. Completion Rate |
|---|---|---|
| NoVG | 27.6 (SD = 24.9) | 82/120 = 68.3% (SD = 0.467) |
| VG1 | 16.7 (SD = 20.2) | 101/120 = 84.2% (SD = 0.367) |
| VG2 | 17.3 (SD = 20.2) | 102/120 = 85% (SD = 0.359) |
| VG3 | 16.7 (SD = 19.5) | 104/120 = 86.7% (SD = 0.341) |
| VG4 | 14.7 (SD = 17.2) | 109/120 = 90.8% (SD = 0.29) |
| Single-handed | 9.02 (SD = 12.1) | 290/300 = 96.7% (SD = 0.18) |
| Double-handed | 28.2 (SD = 23.4) | 208/300 = 69.3% (SD = 0.462) |

for different visualization and gesture type conditions. Post-hoc comparisons (paired t-tests with the Bonferroni correction) further show that participants spent significantly less time performing the gestures with visualizations (i.e., VG1-4) than that with no visual guidance (NoVG), as shown in Table 3. This shows that visual guidance is effective in helping participants perform gestures faster than that without visualization, but encoding additional types of information in visualizations may not further enhance participants' overall performance.

Within our expectation, as shown in Table 1, participants completed single-handed gestures significantly faster than double-handed gestures ($p < 0.001^{***}$) because double-handed gestures are inherently more complex. Moreover, the data points of task completion time for single-handed gestures seem more concentrated, as shown in Table 2 and Figure 10. This indicates that participants were more consistent when performing single-handed gestures, whereas double-handed gesture tasks exhibit much larger variation in completion time.

**Table 3: Pairwise comparisons of different visualizations on task completion time, completion rate, and average task load using paired t-tests with the Bonferroni correction ($^{*}$ for $p < 0.05$, $^{**}$ for $p < 0.01$, and $^{***}$ for $p < 0.001$). NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

| Metric | Paired Conditions | P-value |
|---|---|---|
| Completion Time | NoVG-VG1 | $< 0.001^{***}$ |
| | NoVG-VG2 | $0.001^{**}$ |
| | NoVG-VG3 | $< 0.001^{***}$ |
| | NoVG-VG4 | $< 0.001^{***}$ |
| Completion Rate | NoVG-VG1 | $0.015^{*}$ |
| | NoVG-VG2 | $0.022^{*}$ |
| | NoVG-VG3 | $0.005^{**}$ |
| | NoVG-VG4 | $< 0.001^{***}$ |
| Avg. Task Load | NoVG-VG1 | 0.133 |
| | NoVG-VG2 | $0.013^{*}$ |
| | NoVG-VG3 | $0.045^{*}$ |
| | NoVG-VG4 | $0.013^{*}$ |

Moreover, we classified the 20 gestures into three difficulty levels based on their average difficulty ratings which will be reported in Section 5.2.7 (Figure 17): low ($rating < 1.67$), moderate ($1.67 \geq rating \geq 3.33$), and high ($rating > 3.33$). Notably, all single-handed gestures were categorized as "low difficulty". Subsequently, we conducted a one-way repeated-measure ANOVA within each group, and we ran Greenhouse-Geisser corrections for low-difficulty and moderate-difficulty groups where Mauchly's Tests show the sphericity assumption is violated. As shown in Table 4, the results reveal a significant impact of visualization on completion time for low-difficulty ($p = 0.017^{*}$), moderate-difficulty ($p < 0.001^{***}$) and high-difficulty ($p < 0.001^{***}$) gestures. A post-hoc test using paired t-tests with Bonferroni correction revealed a significant difference between VG1 and NoVG ($p = 0.04^{*}$) for low-difficulty gestures (Table 5); however, no significant differences were found between other pairs of visual conditions. Also, a post-hoc test using t-tests with pooled SD and Bonferroni correction for the group of moderate-difficulty gestures showed significant differences between VG1 and NoVG ($p < 0.001^{***}$), between VG2 and NoVG ($p < 0.001^{***}$), between VG3 and NoVG ($p < 0.001^{***}$), and between VG4 and NoVG ($p < 0.001^{***}$); however, no significant differences were observed among the four types of visual guidance (Table 5). The result indicates that all visual guidance types (i.e., VG1-4) enhanced performance speed for moderate-difficulty gestures compared to NoVG, yet they were similarly effective when compared with each other. In addition, the same post-hoc test revealed significant differences between VG4 and NoVG ($p < 0.001^{***}$) and between VG4 and VG1 ($p = 0.011^{*}$) for high-difficulty gestures (Table 5). The result also indicates that VG4 improved the speed of performing high-difficulty gestures compared to NoVG and VG1.

*5.1.2 Completion Rate.* A two-way repeated-measure ANOVA also indicates there was a significant effect on task completion rate for both visualization ($p = 0.01^{*}$) and gesture type ($p < 0.001^{***}$), as shown in Table 1. Descriptive statistics regarding the completion rates are further shown in Table 2. Similarly, paired t-tests (with the Bonferroni correction) indicate that participants completed significantly more tasks when they were provided with visualizations (i.e., VG1-4) than when they had no dynamic visual guidance (NoVG). This reveals that visual guidance could help participants recover from errors and lead them to more successful performance, but the effects of the richness of information in the visualization may not result in huge differences in completion rate. Moreover, participants completed significantly more single-handed gestures than double-handed gestures in the study (Table 1 and 2). The variance of task completion rate for single-handed gestures is also lower than that of double-handed gestures, as indicated in Table 2. This further confirms that participants were more consistent with single-handed gestures, and thus, an interesting future direction is to explore ways to reduce the variability of double-handed gesture performing tasks.

**Table 4: Results of one-way repeated-measure ANOVAs on visualisations for task completion time within each subgroup ($^{*}$ for $p < 0.05$, $^{**}$ for $p < 0.01$, and $^{***}$ for $p < 0.001$). $^{\dagger}$ indicates with Greenhouse-Geisser Correction.**

| Factor | Metric | Subgroup | P-value |
|---|---|---|---|
| Visualization | Completion Time | Low Difficulty | $p^{\dagger} = 0.017^{*}$ |
| | | Moderate Difficulty | $p^{\dagger} < 0.001^{***}$ |
| | | High Difficulty | $p < 0.001^{***}$ |

**Table 5: Pairwise comparisons of different visualizations on task completion time within each subgroup using the t-tests with Bonferroni correction ($^{*}$ for $p < 0.05$, $^{**}$ for $p < 0.01$, and $^{***}$ for $p < 0.001$). NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

| Subgroup | Paired Conditions | P-value |
|---|---|---|
| Low Difficulty | NoVG-VG1 | $p = 0.04^{*}$ |
| | NoVG-VG2 | $p = 0.22$ |
| | NoVG-VG3 | $p = 0.12$ |
| | NoVG-VG4 | $p = 1$ |
| Moderate Difficulty | NoVG-VG1 | $p < 0.001^{***}$ |
| | NoVG-VG2 | $p < 0.001^{***}$ |
| | NoVG-VG3 | $p < 0.001^{***}$ |
| | NoVG-VG4 | $p < 0.001^{***}$ |
| High Difficulty | NoVG-VG1 | $p = 1$ |
| | NoVG-VG2 | $p = 0.693$ |
| | NoVG-VG3 | $p = 0.166$ |
| | NoVG-VG4 | $p < 0.001^{***}$ |
| | VG1-VG4 | $p = 0.006^{**}$ |
| | VG2-VG4 | $p = 0.072$ |
| | VG3-VG4 | $p = 0.304$ |

*5.1.3 Offset.* To reflect the accuracy of gesture performance, we computed the average offset between participants' current hand positions and desired hand positions across different timestamps. At each timestamp, we calculated the sum of offsets based on all

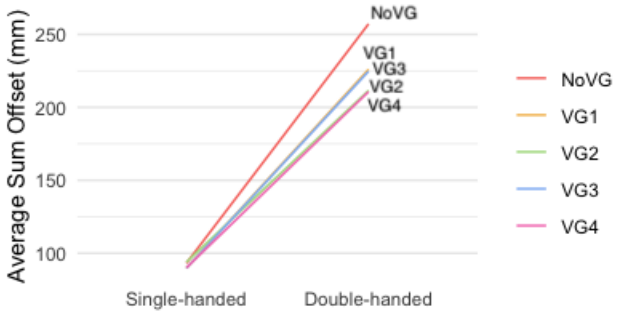


**Figure 11: Interaction plot of visualization and gesture type on the average sum of offsets.**

the joints and keypoints:

$$O_t = \begin{cases} \sum_{m=1}^{17}(\parallel t_m - c_m \parallel), & \text{if a single-handed gesture} \\ \sum_{m=1}^{17}(\parallel t_m - c_m \parallel) + \\ \quad \sum_{n=1}^{17}(\parallel t_n - c_n \parallel) + \\ \quad \sum_{i=1}^{N}(\parallel c_{p_i^1} - c_{p_i^2} \parallel - k_i), & \text{if a double-handed gesture} \end{cases} \tag{2}$$

In this equation, $\parallel t_m - c_m \parallel$ or $\parallel t_n - c_n \parallel$ represents the offset for an individual joint of the dominant hand or the non-dominant hand, while $\parallel c_{p_i^1} - c_{p_i^2} \parallel - k_i$ is the offset for $i$th pair of keypoints. For single-handed gestures, we summed the offsets of the 17 joints; for double-handed gestures, we totaled the offsets of the 34 joints (17 for each hand) and the offsets of $N$ pairs of keypoints. Then, we filtered out the outliers that are greater than a threshold to ignore the data captured when participants were resting or not actively performing a gesture. The threshold was set based on the sum of the maximum possible joint-wise and keypoint-wise offsets (i.e., 15 mm) that triggers the gesture recognition.

A two-way repeated-measure ANOVA showed a significant main effect of gesture type ($p < 0.001^{***}$) and a significant interaction effect ($p = 0.038^{*}$) on the average sum of offsets (Figure 11). Then, we ran a t-test within each of the five visualization conditions, and the results show significant simple effects of gesture types (all $p < 0.001^{***}$). Also, we conducted one-way repeated-measure ANOVAs within the single-handed and double-handed gesture groups. The results show a significant simple effect of visualizations under the double-handed gesture condition ($p = 0.035^{*}$). Further paired t-tests with the Bonferroni correction reveal significant differences between VG2 and NoVG ($p = 0.001^{**}$) and between VG4 and NoVG ($p = 0.003^{**}$). These results imply that VG2 and VG4 significantly helped participants perform a more precise double-handed gesture during the task, compared to the baseline (i.e., NoVG).

## 5.2 Subjective Perceptions

In addition to quantitative measures, we collected participants' perceptions of the visual guidance designs using questionnaires. Next, we report different aspects of the perception in detail.

*5.2.1 Task Load.* We collected participants' perceptions of their experience using the NASA-TLX questionnaire. As shown in Table 1, the results of a one-way repeated-measure ANOVA indicate that there was a significant effect on the average task load rating for visualization ($p < 0.001^{***}$). Paired t-tests with Bonferroni correction revealed that there were significant differences between NoVG and VG2 ($p = 0.013^{*}$), NoVG and VG3 ($p = 0.045^{*}$), as well as NoVG and VG4 ($p = 0.013^{*}$), as shown in Table 3 and Figure 12. This indicates that VG2-4 effectively reduced participants' overall workload during hand gesture performing, compared to no visual guidance; whereas VG1, which just contains the error information, was not perceived much differently from NoVG in terms of workload. However, we did not observe a significant difference between any pairs of visualizations on the average task load.

To comprehensively investigate the impact of visualization on task load, we performed one-way repeated-measure ANOVAs on six subscales. The results reveal significant effects of visualization on mental demanding ($p < 0.001^{***}$), physical ($p = 0.007^{**}$), performance ($p < 0.001^{***}$), effort ($p = 0.003^{**}$), and frustration ($p < 0.001^{***}$). Follow-up paired t-tests with Bonferroni correction demonstrate that, compared to NoVG, VG2 reduced participants' mental demanding ($p = 0.012^{*}$), effort ($p = 0.009^{**}$), and frustration ($p = 0.012^{*}$); VG3 decreased participants' mental demanding ($p = 0.025^{*}$) and frustration ($p = 0.034^{*}$), and at the same time, improved their perception of performance ($p = 0.033^{*}$); and VG4 decreased mental demanding ($p = 0.03^{*}$), effort ($p = 0.033^{*}$), and frustration ($p = 0.009^{**}$) while improving the perception of performance ($p = 0.042^{*}$). VG1 did not have any significant differences on any subscales compared to NoVG. This is plausible because only visualizing errors might not provide enough guidance for hand gesture performing. However, we did not observe that VG1-4 significantly reduced physical or temporal demands compared to NoVG. This could be because during the study, we gave sufficient time to complete each task, and the task setting required performing very precise mid-air hand gestures and holding the gesture for 1.5 seconds, which introduced the problem of "gorilla arm." [19]

*5.2.2 Reliability.* Based on the ratings in our post-survey, a Friedman test revealed a significant effect on the perceived reliability of the hand tracking and gesture recognition system ($p = 0.014^{*}$). Although we observed a large deviation of responses to VG1-4, as shown in Table 6, a Wilcoxon signed rank test plus the Bonferroni correction indicated that participants with VG4 thought the hand tracking/recognition system was more reliable than that without visual guidance ($p = 0.048^{*}$). This observation could be attributed to the performance improvement, including reduced completion time and task load, associated with VG4. While we did not find a significant difference in the response between NoVG and any of the other three visualizations (Table 6), participants' responses tended to be more positive when they were provided with the visual guidance, by observing the medians in Figure 13.

*5.2.3 Confidence.* We carried out a Friedman test on the ratings of participants' confidence over the visualizations, and the results revealed a significant effect ($p < 0.001^{***}$). With a later post-hoc test using Wilcoxon signed rank test with Bonferroni correction (Table 6), we found that participants felt significantly more confident in performing hand gestures with all types of visual guidance (i.e., VG1-4) than that without any visual help. We did not find any significant differences between any pairs of the visualizations. But from Figure 13, we observed that the responses for VG4 have a

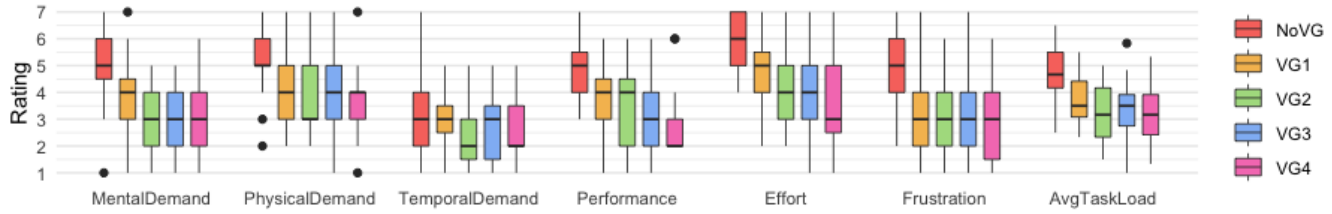


**Figure 12: Box-and-whisker diagrams of unweighted NASA-TLX ratings on a 7-point Likert scale (the lower the better). NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

**Table 6: Pairwise comparisons of different visualizations on participants' responses related to reliability and confidence using the Wilcoxon signed rank test with Bonferroni correction ($^*$ for $p < 0.05$, $^{**}$ for $p < 0.01$, and $^{***}$ for $p < 0.001$). NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

| Metric | Paired Conditions | P-value |
|---|---|---|
| Reliability | NoVG-VG1 | $p = 0.403$ |
| | NoVG-VG2 | $p = 0.084$ |
| | NoVG-VG3 | $p = 0.137$ |
| | NoVG-VG4 | $p = 0.048^*$ |
| Confidence | NoVG-VG1 | $p = 0.028^*$ |
| | NoVG-VG2 | $p = 0.015^*$ |
| | NoVG-VG3 | $p = 0.02^*$ |
| | NoVG-VG4 | $p = 0.015^*$ |

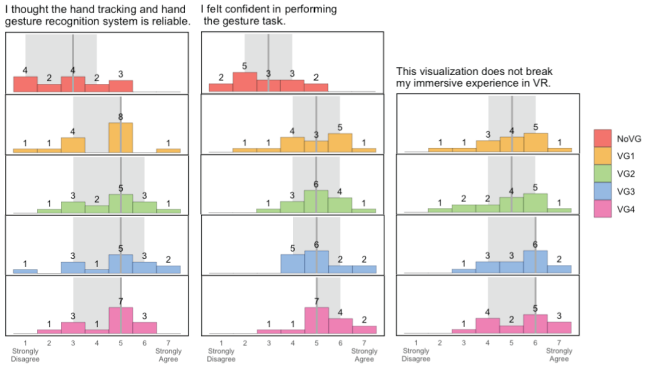


**Figure 13: Distributions of participants' responses related to reliability, confidence, and immersion of their experiences with different visualizations. The dark grey line indicates the median and light grey area indicate the interquartile range (IQR). NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

smaller IQR compared to the other three conditions. This might reflect that VG4 offered a more consistent impression.

*5.2.4 Immersion.* Overall, we found that visual guidance did not negatively affect participants' immersive experience in VR. As shown in Figure 13, few participants thought that the visualization broke their immersive VR experience. For VG1 and VG2, two to three participants disagreed or somewhat disagreed with the statement that providing visual guidance does not interfere with their immersive VR experience. For VG3 and VG4, the majority (14 out of 15) gave at least neutral responses, and 10 of 15 participants expressed positive responses to varying degrees, ending with a high median response ($Md = 6$). Further, a Friedman test did not show a significant difference in response related to immersive experience among these four visualizations ($p = 0.662$).

*5.2.5 Helpfulness.* By observing Figure 14, we found that most participants either kept neutral or agreed that the support of visual guidance has caused a positive impact on the precision and the speed of hand gesture performing as well as system failure understanding, with the median ratings all equalling or above 5. For example, participants tended to agree the most that VG4 helped them perform more precise gestures ($Md = 6$), where as VG1-3 had a slightly lower median rating (all $Md = 5$). Additionally, participants thought VG2, VG3, and VG4 helped them gain a better understanding of why their gestures failed (all $Md = 6$), compared to VG1 ($Md = 5$). We also collected participants' opinions on the helpfulness of the visualizations for different gesture types. As Figure 15 shows, the majority of participants thought the visual guidance was beneficial for both single-handed and double-handed gestures. The results of Friedman tests indicated no significant difference for the four visualizations on the responses to whether participants thought the visual guidance helped them perform gestures faster ($p = 0.825$) or more precisely ($p = 0.344$). Similarly, no significant effect was found in helping them better understand gesture recognition failures ($p = 0.5$). These results indicate that the four visual guidance designs were roughly equally helpful to the participants.

*5.2.6 Usability.* Figure 16 shows participants' responses to different aspects of the usability of the visual guidance. Overall, participants all agreed that the four visualizations were easy to understand, learn, and remember, with the medians all equalling 6. The results of Friedman tests revealed no statistical difference in the impact of the visualization on the above aspects, including ease of understanding ($p = 0.277$), learning ($p = 0.662$), and remembering ($p = 0.442$). While all visualizations were perceived useful, several participants thought VG1 was more challenging to understand and learn compared to others, exhibiting a larger IQR and Q1 (first quartile) towards the negative side. This can potentially explain

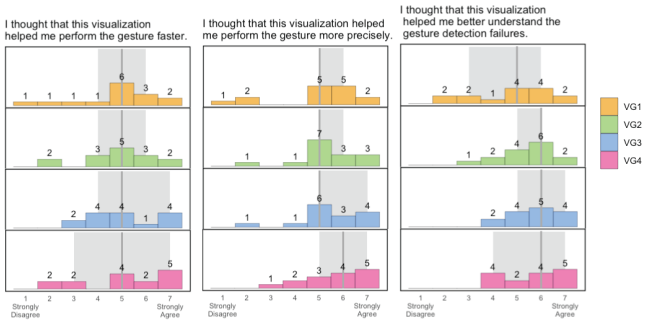


**Figure 14: The distributions of participants' responses related to the helpfulness of the visualizations. The dark grey line indicates the median and light grey area indicate the IQR. NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

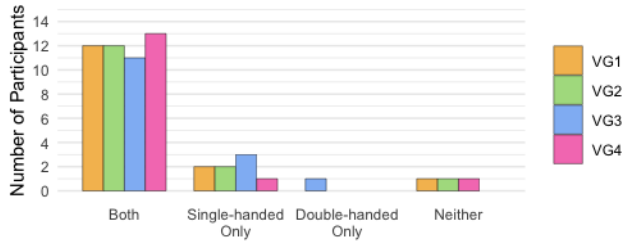


**Figure 15: Distribution of participants' subjective perceptions regarding whether the visual guidance is helpful for single-handed gestures only, double-handed gestures only, both or neither. NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

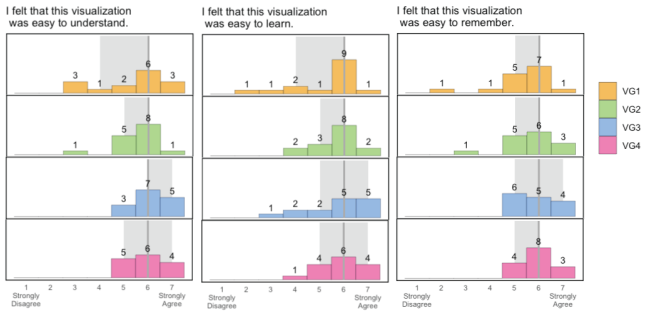


**Figure 16: Distributions of participants' responses related to the easiness to understand, learn and remember for the visualizations. The dark grey line indicates the median and light grey area indicate the IQR. NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

why there was no significant difference between VG1 and NoVG on the average task load (Table 3).

*5.2.7 Gestures.* We collected participants' perceptions of the difficulty of the 20 gestures they experienced in the study, as shown in Figure 17. Without a surprise, we can see that overall double-handed gestures were perceived more difficult than single-handed gestures. Overall, the actual task completion rates matched with participants' perceptions. A paired t-test verified this by showing a significant difference in difficulty rating for the gesture type ($t = -18.353, p < 0.001^{***}$). One thing to note is that most participants failed to complete the "take photo" gesture with a completion rate of 3%. By examining the system, we found that the Oculus Quest hand tracking algorithm often fails to accurately identify the degree of thumb flexion during this gesture performing. This flags the limitation of the current optical tracking system that can be enhanced in the future.

## 5.3 Qualitative Feedback

From the interviews with the participants, we obtained several insights into the design of effective visual guidance for bare hand gesture interaction.

**Necessity of visual guidance.** All participants expressed a need for visual guidance during precise hand gesture interaction. Participants felt the visualization is helpful and may make them *"psychologically feel better"* (S2-P6). *"With no visualization, it's pretty hard to find out what and which part of your gesture is incorrect. So you have to try all the ways that you can think of and hope one of them works. That's really frustrating."* -S2-P5 Also, participants felt frustrated (S2-P2, P3, P6, P7, P15), confused (S2-P10), tougher (S2-P1, P3, P4, P5, P11), and less effective (S2-P14) to perform a correct gesture when no visual guidance was provided, especially when they were performing double-handed gestures (S2-P8, P11). The above results confirm the significance of offering visual guidance for bare hand gesture performing as well as the importance of our investigation.

**Comparison of visualizations.** When comparing different visual guidance designs, participants mentioned that the lack of certain critical information often caused confusion and frustration. For example, S2-P11 and P14 felt that VG1 was confusing because they did not know how to adjust their gestures or move their joints. S2- P10 and P12 also reported that by just giving them the information of error and direction (i.e., VG2), they might over-shoot when adjusting their hand gestures. This is also related to the fact that six participants actively mentioned that VG1 was their least favorite visualization.

Moreover, several key findings from the formative study were confirmed. For example, VG3 and VG4 remained the top choices by the participants in the controlled study; eight (S2-P1, P3, P5-7, P10-12) considered VG4 as their top choice, while four (S2-P8, P13-15) favored VG3 the most. Among the remaining participants, S2-P4 expressed a preference for VG1, S2-P2 for VG2, and S2-P9 exhibited no particular preference. This makes sense because, with VG4, participants perceived a higher reliability of hand tracking and gesture recognition system, completed a higher number of tasks in an average shorter time, and had a better perception of task load, which was reported before.

Further, we verified that there is a trade-off between the amount of information and the workload in processing the information,

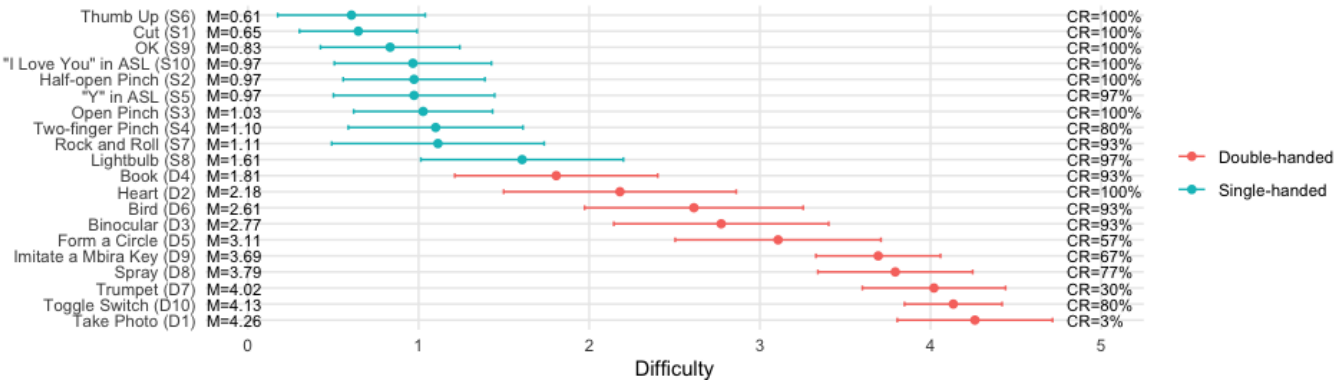


**Figure 17: Mean difficulty ratings (with 95% Confidence Interval) of all the 20 gestures (Figure 6) performed in the actual tasks (the higher the more difficult). The gestures are sorted in ascending order (top to bottom) based on the mean difficulty score where the score ranges from 0 to 5. The CR stands for completion rate.**

which aligned with the results of Chauvergne et al.'s work [8]. If the additional information significantly increases the cognitive load or causes too much distraction, participants tend to opt for simpler alternatives. For example, some participants thought VG4 *"was a little bit too much, like the green ball was a lot of information that was not necessary"* -S2-P13 while VG1-3 *"were better in terms of how much information they gave"* -S2-P8. This explains why four out of 15 participants preferred VG3 over VG4 because they perceived VG3 as a simpler option with reduced information overload and distraction. We also observed two participants show a strong bias toward simplicity. S2-P4 thought that *"the visualizations are absolutely necessary to accomplish some of those gestures. But telling me which way my fingers need to go, it can get pretty distracting and definitely immersion-breaking."* Similarly, S1-P2 liked VG2 the most because she thought the lines in VG3 were helpful but not clear and that VG4 *"becomes confusing when all the lines are showing, so I can't focus on one thing, and it just becomes really messy."* This was summarized nicely by S2-P9: *"less information is helpful, and more information is also insightful."* Overall, providing more information (e.g. directional guidance) is beneficial and reduces cognitive workload during gesture adjustment; meanwhile, it might introduce distractions and break the immersive experience for some users.

**General impression and suggestion.** Participants generally did not find visual encoding (e.g., spheres and lines) of the visualizations confusing and thought they were intuitive, which is supported by their ratings on the aspects of helpfulness and usability reported previously. We also received many inspirations to further enhance the visual guidance design. Most suggestions revolved around simplifying the visualization to reduce distraction and information overload. For instance, S2-P3 suggested removing spheres representing the current and target positions for VG4 because the line already conveyed the same information as two dots. S2-P1 suggested displaying fewer marks because *"having one guide for each joint becomes a bit overwhelming."* One potential solution could be showing indicators for each finger rather than for each joint. S2-P8 proposed disassembling the visual guidance to reduce the workload for double-handed gestures: presenting guidance for one hand first, followed by guidance for another hand, and lastly, showing guidance for the proximity errors. Also, two participants (S2-P5, P10) proposed a new concept of visual guidance, a 3D semi-transparent ghost hand displaying the correct hand gesture superimposed over the visual hands, which was discussed in the formative study.

# 6 DISCUSSION

## 6.1 Take-aways

Based on our study, we verified that visual guidance was useful and needed in precise bare hand gesture performance in VR, especially for double-handed or complex gestures. Figure 18 provides an overview of the key results. Compared with no visual guidance, participants completed gesture tasks faster with the selection of four visualizations (VG1-4) that encode different types of information (i.e., error, target, direction, and difference) at different levels of complexity. And they completed double-handed gesture tasks more precisely when provided with VG2 or VG4. Although the comparisons between two types of visual guidance on completion time and rate were not significant for overall gestures, there seems to be a tendency that employing more types of information (e.g., VG4) resulted in faster performance and a higher average completion rate (Table 2). In addition, our findings suggest that the effectiveness of visual guidance and the need for richer information vary depending on the difficulty of hand gestures. For single-handed or low-difficulty gestures, low-information-complexity visual guidance (i.e., VG1) enhance users' gesture performance, while visual guidance with richer information (i.e., VG2-4) may not show many benefits. However, their advantages, including improving speed and precision, became more apparent as the difficulty of gestures increased. It was also interesting to observe that, for challenging hand gestures, integrating additional information into the visual guidance seemed necessary for better assistance. This could be attributed to the significant difference between NoVG and VG4 and between VG1 and VG4 on completion time in the high-difficulty gesture group.

It is worth noting that we did not observe many significant differences among the four types of visual guidance on many measurements, such as the precision, speed, completion rate, and overall

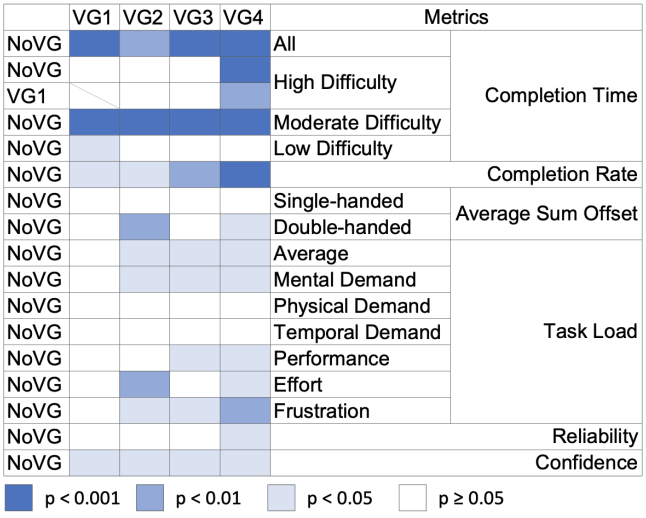


**Figure 18: An overview of found significant difference between five visual conditions. NoVG = no visual guidance, VG1 = error, VG2 = error + direction, VG3 = error + direction + difference, and VG4 = error + target + direction + difference.**

task load. This might due to the specific visual coding styles we utilized, and opens up future research possibilities for exploring other formats of visualizations that use various marks and channels for guidance and correction, such as a ghost hand [9–11, 18] and heatmap [45]. While we confirmed the significance of error indication in guiding simple gestures, we did not observe sufficient quantitative evidence to prove the necessity of other information types individually. We believe that the difficulty of the gesture may affect the need for more information. Specifically, there was a significant difference between VG1 and VG4 on completion time for high-difficulty gestures. Adding additional information about direction, difference, and target significantly improved the performance, as the gestures moved from low to high difficulty level.

On the subjective aspects, overall, visual guidance was perceived as helpful in understanding the system's failure and having good usability while not breaking the immersion experience. When participants used visual guidance with richer information (e.g., VG2-4), they also psychologically felt more confident, less frustrated, less mentally demanding, less needed effort and more successful in performance, compared to NoVG. Also, among VG1-4, VG4 outperformed NoVG on the most number of subscales of the task load, which implies that encoding more types of information could lead to a more comprehensive reduction of task loads. With the support of VG4, participants also thought that hand-tracking and gesture recognition were more reliable than that without any visual guidance.

Moreover, it is important to note that while providing more information can reduce confusion [44], it could raise the problem of overwhelming information and distraction [17]. Thus, we suggest that when considering adding pieces of information, designers should consider finding a balance between low cognitive workload and rich information, for example, assessing the difficulty of hand gestures and prioritizing the information of the error and direction, which are more important and helpful. For applications that require most single-handed or low-difficulty hand gesture interactions, designers could consider only showing users which parts are wrong (i.e., errors) to speed up the gesture performing. Conversely, when designing visual guidance for double-handed and especially challenging gesture interactions, designers should consider showing richer information (e.g., error, target, direction, and difference shown in VG4) to more comprehensively enhance gesture performance and overall user experience.

## 6.2 Limitations and Opportunities

Our study is not without limitations. Here, we discuss them and point to possible future research directions. First, we did not consider gesture orientation. For example, a thump up and a thump down gesture, if not considering the orientation, are the same gesture from the hand gesture recognition point of view. However, they have completely different semantic meanings. Thus, it is interesting to extend our study along this line to investigate the visualization design for orientation information as well as when the micro-level visual guidance is supported, how gesture orientation affects participants' performance and perception.

Second, there exist various thresholding methods for determining whether a gesture is successful or not. We used the offset of each individual joint and/or pair of keypoints to ensure a precise hand gesture performance. Other methods like the sum of offsets of all joints and using angles instead of distances (e.g., Oculus Interaction SDK [2]) exist in different applications. These different thresholding methods could impact users' experience and performance of hand gestures, and visualizations may need to be designed differently to accommodate them for users to better understand the errors and how/why the recognition system fails. Thus, future studies could be conducted to explore and evaluate visualizations for interpreting different thresholding methods.

Third, the timing of showing visual guidance could be further investigated. In our experiment, the system shows the visualization all the time, which could be more adaptive using various detection technologies. For example, previous research has demonstrated the possibility of using gazing responses and brain activities to detect and distinguish different types of errors, such as FP from FN errors of controller-based and pinch-gesture-based inputs [23, 36, 40]. Possible future work can be combining the intention detection technologies with micro-level visual guidance to explore a more integrated system.

Fourth, poor hand tracking may impact users' perceptions of the reliability of our visual guidance, which might affect our results. However, it is challenging to isolate this factor in an experiment due to the limitation of optical hand-tracking technologies of VR/AR headsets. Nine participants reported that handing tracking was not accurate, especially when they were performing double-handed gestures. This can occur when certain parts of a hand are occluded, parts of two hands overlap, or the system fails to accurately determine the depth. This implies that just employing visual guidance cannot address the whole problem of errors, and more research

[2]https://developer.oculus.com/documentation/unity/unity-isdk-hand-pose-detection/

should be devoted to reducing both user and system errors at the same time.

Last but not least, in this study we, focus on *static pose gestures* as a first attempt. There exist lots of scenarios where dynamic gestures are needed and employed. It is worth exploring visual guidance that is optimized for dynamic gestures by extending the knowledge we obtained in this study, such as the types of information, properties of visual guidance, etc. Also, future experiments should be conducted to evaluate the effects of different visualizations by considering the nature of gesture (i.e., static or dynamic) as a factor. This will further shed light on the design of visual guidance for precise hand interaction in a broader range of AR/VR applications. Further, the types of information to visualize in the guidance were obtained from a formative study with only five participants, which might introduce some biases, while many of the past studies [12, 41] also have a small sample pool. More future studies could be conducted to investigate whether other factors could be considered in visual guidance design for bare hand VR gestures.

## 7 CONCLUSION

We have presented a design exploration and evaluation of different visualizations for guiding users to perform precise hand interaction in VR. The design of such visual guidance has been informed by a two-part formative study in an iterative manner. We first identified four types of essential information to visualize and properties of effective visual guidance, and then explored different design alternatives and distilled four visualizations to assess later in our controlled experiment. By comparing with no visual guidance as a baseline, we report and discuss various quantitative and qualitative results, demonstrating that using visual guidance, including but not limited to increasing speed and precision of gesture performing, reducing the workload, and enhancing confidence. We envision our study results pave the way for fostering richer, more precise, and more complex gesture interactions in the domain of social communication, training, and entertainment in the future.

## ACKNOWLEDGMENTS

We would like to thank all our participants for their time and invaluable thoughts. Additionally, we express our gratitude to our reviewers whose insightful comments significantly enhanced the quality of this paper. This research was made possible, in part, with support from the NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Grant and sponsorship from Meta Platforms Technologies, LLC.

## REFERENCES

[1] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: Enhancing Movement Training with an Augmented Reality Mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. Association for Computing Machinery, New York, NY, USA, 311–320. https://doi.org/10.1145/2501988.2502045

[2] Emilie Maria Nybo Arendttorp, Kasper Rodil, Heike Winschiers-Theophilus, and Christof Magoath. 2022. Overcoming Legacy Bias: Re-Designing Gesture Interactions in Virtual Reality With a San Community in Namibia. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 555, 18 pages. https://doi.org/10.1145/3491102.3517549

[3] Emilie Maria Nybo Arendttorp, Heike Winschiers-Theophilus, Kasper Rodil, Freja B. K. Johansen, Mads Rosengreen Jørgensen, Thomas K. K. Kjeldsen, and Samkao Magot. 2023. Grab It, While You Can: A VR Gesture Evaluation of a Co-Designed Traditional Narrative by Indigenous People. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 308, 13 pages. https://doi.org/10.1145/3544548.3580894

[4] Rahul Arora, Rubaiat Habib Kazi, Danny M. Kaufman, Wilmot Li, and Karan Singh. 2019. MagicalHands: Mid-Air Hand Gestures for Animating in VR. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 463–477. https://doi.org/10.1145/3332165.3347942

[5] Olivier Bau and Wendy E. Mackay. 2008. OctoPocus: A Dynamic Guide for Learning Gesture-Based Command Sets. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology* (Monterey, CA, USA) *(UIST '08)*. Association for Computing Machinery, New York, NY, USA, 37–46. https://doi.org/10.1145/1449715.1449724

[6] Idil Bostan, Oğuz Turan Buruk, Mert Canat, Mustafa Ozan Tezcan, Celalettin Yurdakul, Tilbe Göksun, and Oğuzhan Özcan. 2017. Hands as a Controller: User Preferences for Hand Specific On-Skin Gestures. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) *(DIS '17)*. Association for Computing Machinery, New York, NY, USA, 1123–1134. https://doi.org/10.1145/3064663.3064766

[7] Gavin Buckingham. 2021. Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges. *Frontiers in Virtual Reality* 2 (2021). https://doi.org/10.3389/frvir.2021.728461

[8] Edwige Chauvergne, Martin Hachet, and Arnaud Prouzeau. 2023. User Onboarding in Virtual Reality: An Investigation of Current Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 711, 15 pages. https://doi.org/10.1145/3544548.3581211

[9] William Delamare, Thomas Janssoone, Céline Coutrix, and Laurence Nigay. 2016. Designing 3D Gesture Guidance: Visual Feedback and Feedforward Design Options. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (Bari, Italy) *(AVI '16)*. Association for Computing Machinery, New York, NY, USA, 152–159. https://doi.org/10.1145/2909132.2909260

[10] Florian Diller, Gerik Scheuermann, and Alexander Wiebel. 2022. Visual Cue Based Corrective Feedback for Motor Skill Training in Mixed Reality: A Survey. *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–14. https://doi.org/10.1109/TVCG.2022.3227999

[11] Maximilian Dürr, Rebecca Weber, Ulrike Pfeil, and Harald Reiterer. 2020. EGuide: Investigating different Visual Appearances and Guidance Techniques for Egocentric Guidance Visualizations. In *International Conference on Tangible, Embedded, and Embodied Interaction*. https://api.semanticscholar.org/CorpusID:211041359

[12] Mehrad Faridan, Bheesha Kumari, and Ryo Suzuki. 2023. ChameleonControl: Teleoperating Real Human Surrogates through Mixed Reality Gestural Guidance for Remote Hands-on Classrooms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 203, 13 pages. https://doi.org/10.1145/3544548.3581381

[13] Charles Faure, Annabelle Limballe, Benoit Bideau, and Richard Kulpa. 2020. Virtual reality to assess and train team ball sports performance: A scoping review. *Journal of Sports Sciences* 38, 2 (2020), 192–205. https://doi.org/10.1080/02640414.2019.1689807

[14] Katherine Fennedy, Jeremy Hartmann, Quentin Roy, Simon Tangi Perrault, and Daniel Vogel. 2021. OctoPocus in VR: Using a Dynamic Guide for 3D Mid-Air Gestures in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 27, 12 (2021), 4425–4438. https://doi.org/10.1109/TVCG.2021.3101854

[15] P.M. Fitts and M.I. Posner. 1967. *Human Performance*. Brooks/Cole Publishing Company. https://books.google.ca/books?id=XtFOAAAAMAAJ

[16] Dustin Freeman, Hrvoje Benko, Meredith Ringel Morris, and Daniel Wigdor. 2009. ShadowGuides: Visualizations for in-Situ Learning of Multi-Touch and Whole-Hand Gestures. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (Banff, Alberta, Canada) *(ITS '09)*. Association for Computing Machinery, New York, NY, USA, 165–172. https://doi.org/10.1145/1731903.1731935

[17] Kazunobu Fukuhara, Hirofumi Ida, Takahiro Ogata, Motonobu Ishii, and Takahiro Higuchi. 2017. The role of proximal body information on anticipatory judgment in tennis using graphical information richness. *PLoS ONE* 12, 7 (2017), 1–11. https://doi.org/10.1371/journal.pone.0180985

[18] Ping-Hsuan Han, Kuan-Wen Chen, Chen-Hsin Hsieh, Yu-Jie Huang, and Yi-Ping Hung. 2016. AR-Arm: Augmented Visualization for Guiding Arm Movement in the First-Person Perspective. In *Proceedings of the 7th Augmented Human International Conference 2016* (Geneva, Switzerland) *(AH '16)*. Association for Computing Machinery, New York, NY, USA, Article 31, 4 pages. https://doi.org/10.1145/2875194.2875237

[19] Jeffrey T. Hansberger, Chao Peng, Shannon L. Mathis, Vaidyanath Areyur Shanthakumar, Sarah C. Meacham, Lizhou Cao, and Victoria R. Blakely. 2017. Dispelling the Gorilla Arm Syndrome: The Viability of Prolonged Gesture Interactions. In *Virtual, Augmented and Mixed Reality*, Stephanie Lackey and Jessie Chen (Eds.). Springer International Publishing, Cham, 505–520.

[20] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[21] Jay Henderson, Tanya R. Jonker, Edward Lank, Daniel Wigdor, and Ben Lafreniere. 2022. Investigating Cross-Modal Approaches for Evaluating Error Acceptability of a Recognition-Based Input Technique. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 1, Article 15 (mar 2022), 24 pages. https://doi.org/10.1145/3517262

[22] Masoumehsadat Hosseini, Tjado Ihmels, Ziqian Chen, Marion Koelle, Heiko Müller, and Susanne Boll. 2023. Towards a Consensus Gesture Set: A Survey of Mid-Air Gestures in HCI for Maximized Agreement Across Domains. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 311, 24 pages. https://doi.org/10.1145/3544548.3581420

[23] Fotis P. Kalaganis, Elisavet Chatzilari, Spiros Nikolopoulos, Ioannis Kompatsiaris, and Nikos A. Laskaris. 2018. An error-aware gaze-based keyboard by means of a hybrid BCI system. *Scientific Reports* 8, 1 (2018), 1–11. https://doi.org/10.1038/s41598-018-31425-2

[24] Keiko Katsuragawa, Ankit Kamal, Qi Feng Liu, Matei Negulescu, and Edward Lank. 2019. Bi-Level Thresholding: Analyzing the Effect of Repeated Errors in Gesture Input. *ACM Trans. Interact. Intell. Syst.* 9, 2–3, Article 15 (apr 2019), 30 pages. https://doi.org/10.1145/3181672

[25] Ben Lafreniere, Tanya R. Jonker, Stephanie Santosa, Mark Parent, Michael Glueck, Tovi Grossman, Hrvoje Benko, and Daniel Wigdor. 2021. False Positives vs. False Negatives: The Effects of Recovery Time and Cognitive Costs on Input Error Preference. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 54–68. https://doi.org/10.1145/3472749.3474735

[26] Jolanta Lapiak. [n. d.]. Signs for make. https://www.handspeak.com/word/1331/

[27] Benjamin Lee, Maxime Cordeil, Arnaud Prouzeau, Bernhard Jenny, and Tim Dwyer. 2022. A Design Space For Data Visualisation Transformations Between 2D And 3D In Mixed-Reality Environments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 25, 14 pages. https://doi.org/10.1145/3491102.3501859

[28] Karen B. Lewis and Roxanne Henderson. 2001. *Sign language made simple*. Three Rivers Press.

[29] Klemen Lilija, Søren Kyllingsbæk, and Kasper Hornbæk. 2021. Correction of Avatar Hand Movements Supports Learning of a Motor Skill. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1–8. https://doi.org/10.1109/VR50410.2021.00069

[30] Tica Lin, Rishi Singh, Yalong Yang, Carolina Nobre, Johanna Beyer, Maurice A. Smith, and Hanspeter Pfister. 2021. Towards an Understanding of Situated AR Visualization for Basketball Free-Throw Training. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 461, 13 pages. https://doi.org/10.1145/3411764.3445649

[31] Weizhou Luo, Zhongyuan Yu, Rufat Rzayev, Marc Satkowski, Stefan Gumhold, Matthew McGinity, and Raimund Dachselt. 2023. Pearl: Physical Environment based Augmented Reality Lenses for In-Situ Human Movement Analysis. *Conference on Human Factors in Computing Systems - Proceedings* (2023). https://doi.org/10.1145/3544548.3580715

[32] Meta. 2022. Introducing 'First Hand,' Our Official Hand Tracking Demo Built With Presence Platform's Interaction SDK. https://developer.oculus.com/blog/introducing-first-hand/

[33] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing Legacy Bias in Gesture Elicitation Studies. *Interactions* 21, 3 (may 2014), 40–45. https://doi.org/10.1145/2591689

[34] Michael Nebeling, Maximilian Speicher, Xizi Wang, Shwetha Rajaram, Brian D. Hall, Zijian Xie, Alexander R. E. Raistrick, Michelle Aebersold, Edward G. Happ, Jiayin Wang, Yanan Sun, Lotus Zhang, Leah E. Ramsier, and Rhea Kulkarni. 2020. MRAT: The Mixed Reality Analytics Toolkit. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Honolulu</city>, <state>HI</state>, <country>USA</country>, </conf-loc>) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376330

[35] Matei Negulescu, Jaime Ruiz, and Edward Lank. 2012. A Recognition Safety Net: Bi-Level Threshold Recognition for Mobile Motion Gestures. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services* (San Francisco, California, USA) *(MobileHCI '12)*. Association for Computing Machinery, New York, NY, USA, 147–150. https://doi.org/10.1145/2371574.2371598

[36] Candace E. Peacock, Ben Lafreniere, Ting Zhang, Stephanie Santosa, Hrvoje Benko, and Tanya R. Jonker. 2022. Gaze as an Indicator of Input Recognition Errors. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 142 (may 2022), 18 pages. https://doi.org/10.1145/3530883

[37] Siyou Pei, Alexander Chen, Jaewook Lee, and Yang Zhang. 2022. Hand Interfaces: Using Hands to Imitate Objects in AR/VR for Expressive Interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 429, 16 pages. https://doi.org/10.1145/3491102.3501898

[38] Tran Pham, Jo Vermeulen, Anthony Tang, and Lindsay MacDonald Vermeulen. 2018. Scale Impacts Elicited Gestures for Manipulating Holograms: Implications for AR Gesture Design. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 227–240. https://doi.org/10.1145/3196709.3196719

[39] Alexander Schäfer, Gerd Reis, and Didier Stricker. 2022. AnyGesture: Arbitrary One-Handed Gestures for Augmented, Virtual, and Mixed Reality Applications. *Applied Sciences* 12, 4 (2022). https://doi.org/10.3390/app12041888

[40] Naveen Sendhilnathan, Ting Zhang, Ben Lafreniere, Tovi Grossman, and Tanya R. Jonker. 2022. Detecting Input Recognition Errors and User Errors Using Gaze Dynamics in Virtual Reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 38, 19 pages. https://doi.org/10.1145/3526113.3545628

[41] Xinyu Shi, Ziqi Zhou, Jing Wen Zhang, Ali Neshati, Anjul Kumar Tyagi, Ryan Rossi, Shunan Guo, Fan Du, and Jian Zhao. 2023. De-Stijl: Facilitating Graphics Design with Interactive 2D Color Palette Recommendation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 122, 19 pages. https://doi.org/10.1145/3544548.3581070

[42] Jungpil Shin, Akitaka Matsuoka, Md. Al Mehedi Hasan, and Azmain Yakin Srizon. 2021. American Sign Language Alphabet Recognition by Extracting Feature from Hand Pose Estimation. *Sensors* 21, 17 (2021). https://doi.org/10.3390/s21175856

[43] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: Projected Visualizations for Hand Movement Guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 179–188. https://doi.org/10.1145/2207676.2207702

[44] Chase Stokes, Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti A. Hearst. 2023. Striking a Balance: Reader Takeaways and Preferences when Integrating Text and Charts. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 1233–1243. https://doi.org/10.1109/TVCG.2022.3209383

[45] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2014. Gesture Heatmaps: Understanding Gesture Performance with Colorful Visualizations. In *Proceedings of the 16th International Conference on Multimodal Interaction* (Istanbul, Turkey) *(ICMI '14)*. Association for Computing Machinery, New York, NY, USA, 172–179. https://doi.org/10.1145/2663204.2663256

[46] Nicolas Vignais, Benoit Bideau, Cathy Craig, Sébastien Brault, Franck Multon, Paul Delamarche, and Richard Kulpa1. 2009. Does the Level of Graphical Detail of a Virtual Handball Thrower Influence a Goalkeeper's Motor Response? *Journal of sports science & medicine* 4 (2009).

[47] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. 2020. Mid-Air Gesture Control of Multiple Home Devices in Spatial Augmented Reality Prototype. *Multimodal Technologies and Interaction* 4, 3 (2020). https://doi.org/10.3390/mti4030061

[48] Robert Wang, Sylvain Paris, and Jovan Popović. 2011. 6D Hands: Markerless Hand-Tracking for Computer Aided Design. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 549–558. https://doi.org/10.1145/2047196.2047269

[49] Tianyi Wang, Xun Qian, Fengming He, Xiyun Hu, Yuanzhi Cao, and Karthik Ramani. 2021. GesturAR: An Authoring System for Creating Freehand Interactive Augmented Reality Applications. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 552–567. https://doi.org/10.1145/3472749.3474769

[50] Xiaoying Wei, Xiaofu Jin, and Mingming Fan. 2022. Communication in Immersive Social Virtual Reality: A Systematic Review of 10 Years' Studies. arXiv:2210.01365 [cs.HC]

[51] Xingyao Yu, Katrin Angerbauer, Peter Mohr, Denis Kalkofen, and Michael Sedlmair. 2020. Perspective Matters: Design Implications for Motion Guidance in Mixed Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 577–587. https://doi.org/10.1109/ISMAR50242.2020.00085

# A APPENDIX

**Table 7: An overview of results in Part I of our formative study.**

| Themes | Explanations |
|---|---|
| Types of Hand Tracking and Gesture Detection Errors That Users Encountered | This theme summarizes four types of errors shared by interviewees:<br>• Failed to recognize (FN errors)<br>• Triggered without users' intentions (FP errors)<br>• Latency<br>• Drifting (hands disappear and reappear somewhere else suddenly)<br>Also, these users express a low tolerance in errors. If they fail few times or tracking is poor, they will switch back to the controller that they think is a faster and more precise input. |
| Causation of Gesture Detection Errors | Interviewees think that gesture detection errors are caused by the following reasons:<br>• Environment problems (e.g., a dark or complex environment)<br>• Occlusions of hands (e.g., due to the complexity of gestures or the orientation of gestures)<br>• Wrong gestures (e.g., caused by forgetting the gesture or not pushing fingers down far enough)<br>• Hands not detected or hands drifting when moving the headset/hands too fast<br>• Not interactable (e.g., the interviewee guessed that it was caused by changes of the system or the interaction was not implemented.)<br>• The recognition algorithm is not trained well |
| Types of Information Users Expect to Receive | These VR users think their ideal guidance should contain four types of information:<br>• Error (What is wrong?)<br>• Target (What is correct?)<br>• Direction (Which way is it?)<br>• Difference (How far is it?) |
| Visual Characteristics of the Guidance That Users Anticipated | This theme summarizes four properties of a good guidance from interviewees' comments:<br>• Simple<br>• Universal<br>• Spatial<br>• Dynamic |
| When to Show Visual Guidance | Interviewees expected to see the guidance:<br>• During tutorial sessions or when the system first introduces a gesture<br>• When the user intends to perform a gesture and fails |
| Preferences of Types and the Amount of Guidance | This theme highlights interviewees' different preference of guidance in different scenarios:<br>• Experts and novices may prefer different types of guidance.<br>• Users may expect to be guided differently in the tutorial when the gesture is first introduced and in a following-up reminder.<br>• Users may prefer different types of guidance for varying degrees of errors or during different phases of correction. |
| Design Suggestions | Interviewees proposed different types of visual guidance, including displaying a ghost hand that overlays the virtual hand, showing all possible gestures in a panel demonstration, combining two initial designs, and more. |