The Effects of Hand Tracking on User Performance: an experimental study of an object selection based memory game

Nima Jamalian*
Goldsmiths, University of London

Marco Gillies†
Goldsmiths, University of London

Frederic Fol Leymarie‡
Goldsmiths, University of London

Xueni Pan §
Goldsmiths, University of London

Figure 1: We designed and carried out an experimental study using a memory task (c) to understand the effect of different interaction input systems such as hand tracking (a) and controller tracking (both with or without haptics) (b), has on user performance.

Abstract— Until recently, Virtual Reality (VR) applications relied on controllers to enable user interaction in virtual environments. With advances in tracking technology, HMDs are now able to track the movements of users' hands in real-time with significantly greater accuracy, allowing us to interact with the digital world directly with our hands. However, it is not entirely clear how hand tracking affects users' performance. In this study, we investigate user performance using an in-game analytics-based assessment methodology for a VR memory puzzle task. We conducted a within-subjects experiment with 30 participants in three conditions: 1- Hand-tracking, 2- Controller Without Haptics, and 3- Controller With Haptics. In all our measurements (correct order and pattern, correct pattern only, and trial completion) except for the initial selection time, participants performed best with hand tracking. The use of controllers with haptics did not outperform controllers without haptics in most measures, possibly because other feedback cues compensated for the lack of haptics. This study helps us better understand the three selected interactivity methods when used in VR, as well as the importance of naturalistic experience in interaction design.

Index Terms—Virtual Reality, Hand Tracking, Cognition, Haptic, Multisensory Feedback

1 Introduction

Recent years have seen the growth of Virtual Reality (VR) applications with rapid technological advances and new features for head-mounted displays (HMDs). One such feature is hand tracking using monochrome cameras that are part of the HMD inside-out positional tracking system providing high accuracy even for peripheral vision. It is now commonplace to use a artificial neural network architecture to detect hands and estimate key-points locations, enabling a high tracking frequency of 60Hz [3]. Since its first introduction to the consumer market in mid-2020, modern hand-tracking is rapidly becoming available across brands with a growing user community, allowing VR users to more naturally interact with objects and navigate in a virtual world hand-free. As demonstrated in previous research, the use of controllers could lower the perception of naturalness and the level of immersion [23]. Other than gaming and entertainment, VR is being used in many fields such as learning [25], medicine [24] and manufacturing [26]. Precision, hand movement, and realistic representation are essential for tasks in many of these areas. Hand tracking could play a significant role and open new interaction paradigms in VR. However, few scientific studies have been conducted to examine the impact current VR hand tracking technology has on user experience and performance. Here we attempt to address this gap.

The most apparent benefit of hand tracking is the additional degrees of freedom, allowing users to use their hands and fingers to grab object or to gesture more flexibly. There are also possible cognitive benefits. Since hand tracking allows for more natural interactions, it may help reduce the cognitive load when compared to using controllers. In particular, hand tracking is likely to reduce significantly the mismatch between visual feedback and proprioception.

In this work, we are interested in the impact of hand tracking on user psychology and performance, in comparison to using controllers. As one potential disadvantage of hand-free interactions could be the lack of haptic feedback, we also compared two different conditions using controllers: with and without haptic feedback. We designed and implemented a puzzle memory VR game, which allowed us to

*e-mail: n.jamalian@gold.ac.uk
†e-mail: m.gillies@gold.ac.uk
‡e-mail: ffl@gold.ac.uk
§e-mail: x.pan@gold.ac.uk

978-1-6654-5325-7/22/$31.00 ©2022 IEEE
DOI 10.1109/ISMAR55827.2022.00095

test user performance in three conditions. Furthermore, as part of our game design, we utilised multi-sensory input (sound and visual) to compensate for the lack of haptic feedback. In the following section, we review the most relevant related work, organised in four main areas: the psychological impact of hand-tracking, cognitive load and working memory, haptic feedback, and 3D user interaction.

2 Related work

2.1 The Psychological impact of Hand-tracking

2.1.1 The Virtual Body

The concept of body ownership refers to the feeling that we own a physical body which allows us to experience and manipulate the environment in which we live. We can regard our hands as our own and as a part of our body when we believe that they are our own [30]. It has been demonstrated in cognitive neuroscience that we can also experience this feeling towards artificial bodies, which is referred to as a body ownership illusion [20].

Hand tracking allows us to visualize user hand movement and hand poses more accurately. Using controllers, we can also see our hands, but the model representing our hands will often not accurately reflect our hand pose nor allow for natural hand poses (figure 2). The relationship between our body in everyday interaction and related thought process is a complex subject which involves cognitive science, neuroscience, and philosophy [2]. Previous work has shown that interaction techniques can be more successful if they make more use of motor skills and common ways of interaction that we have learned through our normal interactions in the real world [11, 19, 21, 22]. Hand tracking, by being more similar to how we use our hands in the real world, can better leverage these prior learned skills.

Virtual hands form an important part of the virtual body of a participant, or a self-avatar, and research conducted on the impact of self-avatars on cognitive load in VR has shown that they can have a positive effect on presence, interaction and perception of space. The ability to represent an accurate self-avatar has been a quest in VR since its inception. In 1995, Slater et al. [33] demonstrated that the virtual body had a significant impact on self-reports of presence during locomotion. Other research has also shown that the virtual body can be a crucial factor of presence generating experience [36]. More recently, Steed et al. showed that having a self-avatar could aid participants' cognitive process in VR [34]. Another study showed that using pure hand tracking provides a more positive and less arousing experience in comparison to relying on controllers while performing a typing task [37]. Overall, prior work suggests that the representation of self-avatar is a significant factor for motor-related tasks and general interaction in an immersive environment. Such work on embodiment is a strong indicator that visual feedback of a self-avatar can positively affect user experience. Overall, it appears that hand tracking will remain a key feature of immersive virtual reality (even though the technology has not reached its optimal stage yet). Many aspects of our social, teaching, and learning worlds could benefit from effective and engaging interpersonal communication and more formal presentations in a remote setting [8]. In order to develop the next generation of VR hardware and immersive experiences, it is important to understand the challenges that accompany these opportunities. There will be an in-depth discussion of design and challenges of hand tracking in the section 3.

2.1.2 Effect of Gesture on Cognition

People gesture while speaking, particularly while explaining something, and this appears to support the process of cognition [34]. Previous research conducted by Hostetrer et al. has shown that in a task where users are asked to describe ambiguous dot patterns, the more complex the pattern resulted in more complex gestures from participants [18]. Another research conducted by Goldin-Meadow et al. explored the role of gesture and memory [15]. They asked adults and children to remember letter sequences while explaining how they solved mathematical problems. The results showed that when participants were prevented from gesturing, they recalled significantly fewer letter sequences, suggesting that gesture appears to significantly affect cognitive effort.

2.2 Cognitive Load and Working Memory

Cognitive load theory was developed by John Sweller in the late 1980's out of a study on problem solving [35]. Cognitive load refers to how much information our working memory can hold at one time. According to Sweller, instructional methods should not overload working memory with activities that will not directly contribute to learning. Depending on how physical activity is implemented, it can go from being beneficial to detrimental for learning [32].

Paas and Sweller proposed that cognitive load theory should be updated to include insights from embodied cognition theory [27]. A study of autobiographical memory also examined the effect of embodied cognition on memory performance by studying the influence of body position on ease of recall [9, 10]. When asked to recall a previously remembered event, participants were instructed to take positions compatible or incompatible with the original body position. Researchers have found that participants given compatible body positions recalled memories more quickly than those given incompatible body positions, indicating that body position facilitates access to autobiographical memories [9]. Thus, we can experience fluctuations in our cognitive load as a consequence of our body movement or posture.

It has been suggested that Embodied Cognitive Load Theory can be used to predict the utility of interactive features in learning environments [31]. According to this theory, embodied modes of interaction can be more effective when their benefits (such as easier cognitive processing) outweigh their cognitive costs (such as motor coordination).

Studies in different fields and through different tasks have shown that memory and embodied cognition are related [39]. Research on embodied cognition and memory typically examines how manipulations of the body influence memory performance or, a contrary, how manipulations of memory tasks result in bodily changes [10]. Glaenberg, for example, examined the relationship between memory and action from an embodied cognition viewpoint, establishing that memory can be described as a sequence of coordinated actions confined by the body [14]. According to Glaenberg, memory, action, and perception have a reciprocal relationship. As a consequence, manipulations of the body or movement can alter memory [10], [14].

In addition, new perspectives have been provided on neural processes and the neural structure underlying embodied cognition, episodic memory, recall, and recognition [38], [39]. Neuronal states, which are produced by action, perception, and introspection systems, can be reenacted as experiences. Sensory perception includes sensory elements, motor perception includes motor movements, and introspective perception includes emotional, mental, and motivational aspects. Collectively, these modes contribute to shaping our experiences in different ways. Cognitive processes that involve memory thus facilitate the choice of the appropriate action in a given situation, not by remembering exactly what the situation is, but rather by remembering how the action relates to the circumstance [39]. As an example, remembering and identifying a party attended the day before is said to be related to the body, since sensory-motor aspects of the recalled event are being reconstructed, as well as details of actions that took place [38], [5].

2.3 Haptic Feedback

The term "haptic technology" is often used interchangeably with "kinaesthetic communication" or 3D touch, as it refers to any technology that can create an experience of touch by applying forces, vibrations, or motion to the user [6] to indicate vibrate tactile. It is becoming increasingly common for VR systems to include haptics, which adds the sense of touch to previously audio-visual only interfaces. An example of a haptic device would be HMD controllers with built-in vibrate motor components or haptic gloves [7]. With bespoke devices, a previous study examined the role of haptic feedback in virtual embodiment through the use of a drawing task that had users color in certain shapes using three interaction modes: with force feedback, vibrotactile feedback, or no haptic feedback at all [3]. Results showed that force feedback was more ecological in the sense that it was more analogous to reality, while vibrotactile feedback was more symbolic. Force feedback was significantly more effective regarding embodiment than a lack of haptic feedback and significantly more effective than the other

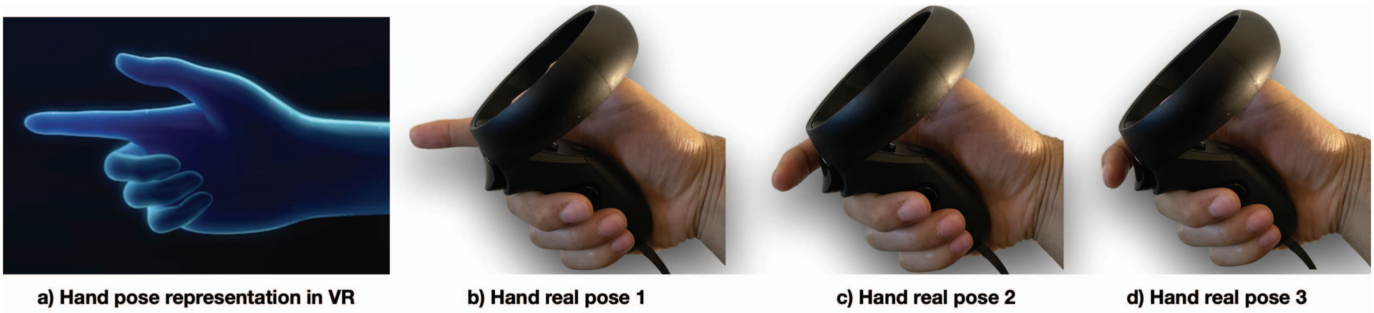

Figure 2: Controller Hand Pose Mismatch: a) Visualize user hands in VR (pointing pose). As an example of b,c,d, show various hand positions in real life while using controllers that result in pointing pose (a) in VR. This illustrates the possible mismatch between hand pose in real life and those represented in VR while using controllers as an input system.

two modes in terms of subjective performance. These results suggest that ecological feedback is more appropriate for eliciting embodiment during fine manipulation tasks [3].

Gibbs et al. [13] evaluated the psychological impact of haptic only, visual only, and haptic plus visual feedback and found that bi-modal feedback enhances presence compared with uni-modal, but visual alone does not provide a better sense of presence than haptic alone. Richard et al. [28] also found that force feedback outperformed the lack of haptics when it comes to embodiment and that force feedback was significantly better than no haptic or vibrotactile feedback for perceived performance (measured by subjective questionnaires). However, when it comes to the *actual* performance (degree of completeness and precision of the task), no significant differences were found. In general, prior research has confirmed that haptic feedback, be it force feedback or vibrotactile, can enhance the user's sense of presence and the level of embodiment, while in terms of actual performance, little difference has been found.

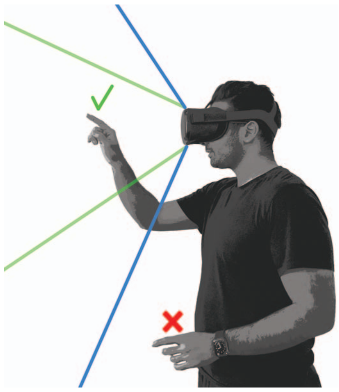

Figure 3: Hand Tracking Volume: The green lines indicate the display field of view (DFOV), which is the optimal hand tracking area, while the blue lines indicate the larger tracking volume. The user is unlikely to be able to see a rendered hand in VR if it is within the tracking volume but outside the DFOV, despite the fact that the hand is being tracked. It is thus important to try to keep the user activity within the DFOV.

3 User Interaction Design

3.1 Current Hand Tracking Technology in the Consumer Market

Many latest consumer HMD devices come equipped with hand-tracking technology integrated into the device, without the need for external tracking hardware (e.g. Meta Quest [16]), using fish-eye cameras ensure a tracking volume that is larger than users' field-of-view (Figure 3) and machine learning algorithms to teach head-mounted cameras to determine the orientation of our hands in real-time [16]. Currently, Quest considers several additional factors to ensure high-quality output. First, tracking latency is maintained at a very low level, ensuring a high level of embodiment. Second, jittering is reduced: the key points of the reconstructed hand no longer vary or oscillate significantly from frame to frame. Previously, the tracking methods used Keypoint estimation without considering history, which resulted in high levels of hand jittering. This problem is solved by integrating history into Keypoint estimation to help handle complex occlusions [1, 16]. Third, Quest hand tracking works under different lighting conditions, from dim to bright lighting. This has the advantage of allowing an extensive range of possible interfaces and interactions using hand tracking in VR, such as using a virtual keyboard or interacting with a user interface, or interacting with virtual objects using physics.

3.2 Hand Tracking Application Design and Limitation

Our application is built within the Unity engine using the Meta integration SDK. We developed tasks including object manipulation, object selection, user interface interaction and gestural interaction, using Meta Quest and Quest 2. We present here an overview of the state of hand tracking and its limitations during our experiment since it is rapidly developing. Two limitations were identified. The first one is that hand-hand interactions do not work well with the current hand tracking system. For instance, if users try to use hand tracking to simulate washing their hands, the action will not perform as expected. Regardless of recent improvements in computer vision, handling overlaps of the hand remains a challenge [16]. Therefore in this study we focus on designing interactions that can be executed with just one hand. The second limitation is the tracking volume. In our application, we designed our tasks which keep users from reaching outside of the tracking area so that their relevant movements can always be tracked. In our pilot study, we also noted that users perform better with tasks involving interactions they are familiar with in real life. For example, pinch gestures are used for grabbing objects in many hand-tracking toolkits. However, many users found it difficult as they do not typically perform pinch gestures in the real world counterpart actions. We propose that it is more beneficial to use interaction paradigms which users are already familiar with via real-world interaction since it supports a more natural, easier to achieve experience and leads to better VR immersion results.

We designed a task that can be completed with either one or two hands, while ensuring two hands never overlap. The main interaction area was designed to be located at the centre of the screen, ensuring that no interaction is out of range. Additionally, we incorporated only natural interaction users are familiar with.

3.3 Multi-sensory Interaction Design

One of the key elements in both object selection and manipulation in VR is the mechanism for providing feedback to compensate for the lack of real-world force feedback we would expect. We reviewed the multi-sensory feedback in the top 30 played VR games on the market [12], including: Pistol Wipe, Robo Recall, Beat Saber, Angry Birds VR, and others. We have also included applications specifically designed to introduce VR to new users, such as Meta's First Steps

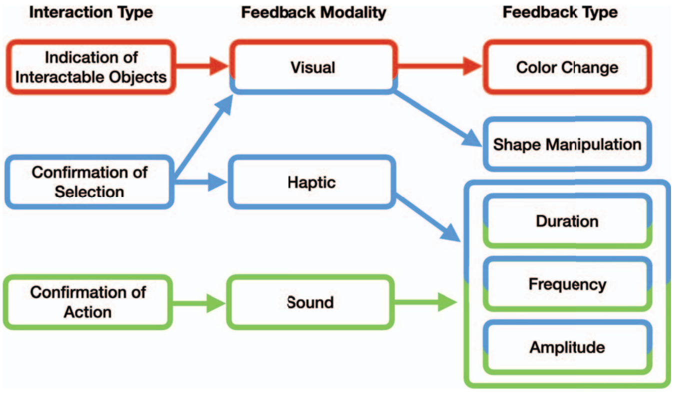

Figure 4: Multi-sensory Design Principle: Interaction type in VR mapped to feedback modality and type.

and First Contact. As an example, in the First Step application, which is designed to introduce users to VR using an interactive experience, interactive objects are highlighted, giving the user an indication that the object can be interacted with. Upon selecting the object, the user gets a visual (i.e., the object's position changes) and haptics feedback. User's action (such as button pressing) are normally confirmed through sound feedback. We examined the multi-sensory feedback approaches used by each application, identified common design principles, and developed a multi-sensory feedback design concept (Figure 4). Based on the analysis of different interactions, three categories were identified. 1: Indication of Interactable Objects: providing users with clues (feedback modalities), in this case, a visual cue to inform them that a particular object is intractable. 2: Confirmation of Selection: providing visual and haptic feedback to users to indicate whether or not the user has selected an interactable object. 3: Confirmation of action: Confirmation that an action has been successfully performed by means of a sound.

4 Methods

Hand-tracking reduces the mismatch between users' real hand pose and the virtual ones, resulting in increased immersion and a more natural and accurate self-avatar with better gestural representations. Here we compare user performance in a memory puzzle task using hand tracking and controller input systems.

Accordingly, based on the theory of cognitive load, embodied cognition, their relationship with memory load, and how action affects our memory, we designed a task where users need to remember a series of patterns while performing actions based on what has been pattern shown to them. In this task, interactions (actions) are designed to mimic real-world interactions as much as possible. By providing VR actions that are closer to real-world experiences, which are actions that users are already familiar with, we are interested in whether this method can reduce cognitive load on the users and improve recall. The task is a 3D version of Corsi's block-tapping test [4], which assesses short-term working memory in visuo-spatial domains, in which the difficulty is increased incrementally.

Comparing the effect between hand-tracking and using controllers inevitably involves the question of haptic feedback. This is because with current consumer HMDs, hand tracking mode, or "hands-free" mode, comes with the price of having to sacrifice vibrotactile feedback when users put their controllers down. Here, based on the literature, hand tracking limitation and multi-sensory design principles, we created a simple memory task which helped us understand the impact of performance of the hand-free mode, which on one hand could improve performance due to reduced cognitive load, but on the other hand reduce performance because of lack of haptic feedback.

4.1 Hypotheses

First of all, the paper examines whether using hand tracking can reduce the user cognitive load during a memory puzzle task, resulting in user's performance being improved. Our first hypothesis is:

> **Hypothesis 1 (H1):** Compared to using controllers, hand tracking can reduce cognitive load by offering a more natural method of interaction with less mismatch of hand presentation. Thus, participants will perform better in our memory game which relies on working memory and cognitive load.

Our next research question relates to the use of the dominant hand. We expect that users will use both hands more equally when performing the task in a hand tracking mode than when using a controller, where a dominant hand operating a controller is more likely. In real life, we tend to use both hands actively when we type (by pressing a button) or interact with objects surrounding us. However, when it comes to tasks that are more complex or require more precision, we tend to use our dominant hand more often as it is known to be stronger, faster, and better at dexterity [29]. Therefore, we hypothesize that in the controller condition, users would predominately use their dominant hand; whereas in the hand tracking, they will be using both hands more equally:

> **Hypothesis 2 (H2):** Since hand tracking mode is more natural, participants will use both hands equally, whereas using controllers participants will have greater reliance on the dominant hand.

When using multi-sensory feedback, we expect the haptic mode to provide the user with richer experiences, improving users' performance. Hence, we formulate the following third hypothesis.

> **Hypothesis 3 (H3):** Participants would perform better using a controller with haptic feedback than one without.

4.2 Experimental design

A within-group study was designed to examine the hypotheses in which participants performed a memory puzzle task in three conditions:

- **Hand:** Hand Tracking Condition
- **Controller:** Controller Without Haptics Condition
- **Haptic:** Controller With Haptics Condition

The order of conditions was randomly assigned to the participants, so their familiarity with the task would not affect the results. Participants were presented with a 3 x 3 board with 9 virtual physics-based buttons as shown in Figure 5.a. Here, a virtual physics-based button refers to a button that functions in a virtual environment in a manner that is similar to a physical button in real life (Section 4.3). The participant was required to apply physical force in order to operate the virtual button by moving their hand toward the button and pushing/pressing on the button.

In each of the conditions of the memory puzzle task, participants were required to complete 12 trials. In each trial, the buttons on the board were highlighted randomly one by one to form a pattern (Figure 5.a). When the entire pattern was highlighted, the buttons would return to their original colours, becoming indistinguishable from the other buttons. Participants then needed to reproduce the pattern they just saw, in the same order, by pressing the buttons one at a time (Figure 5.b). Upon completing the selection, the user was presented with the results of their selection. If the button was selected in the correct position and order, the button would turn green. If a selected button had the correct position but in the wrong order, only the base of the button would turn green while the button itself would turn red (see figure 5.c). Any buttons that were selected in the incorrect position, both the base and the button would be colored red.

The application would log multiple data points throughout the activity while participants performed the task via custom-made code. Initially, the application would record the puzzle pattern, in this case referred to as Puzzle Selection, into a list and also record the pattern the participant selected into a list. Subsequently, it would compare these two lists in order to determine if they contained the same elements, in which case it would mean that the participant had correctly identified the overall pattern of the puzzle. In the event that the participant had got

a) Puzzle shown to the user on the board   b) User selecting their input using buttons   c) Result of user selection gets display to user

Figure 5: Experiment footage: (a) Puzzle Board, (b) User Input, (c) Result

the overall pattern correctly, the application would check if the order of the elements appearing in the list matches. If this was the case, it meant that the participant also correctly replicated the order of the puzzle. The application would record this as a Correct Order and Pattern data point. Cases where the overall pattern was correct but not the order would be recorded as a **Correct Pattern Only** data point.

In each condition, there were 12 puzzle trials with different puzzle complexities to provide a range of stimuli. Puzzle complexity refers to the pattern that was displayed to the user as well as how many buttons the user must memorize. Trial 1-3 had puzzle complexity 3, trial 4-7 puzzle complexity was 4, trial 8-12 puzzle complexity was 5. The program would generate a random puzzle combination based on the puzzle complexity number; for example, if the difficulty level of the puzzle is 3, the program would choose three buttons on the board to be highlighted in random order, and the user would need to repeat this pattern.

4.3 Implementation

The application was developed using the Unity game engine and C# programming language. It uses both the Oculus Integration SDK and Unity XR Plug-in Management package. Custom scripts were added to facilitate transitions between hand tracking and controller tracking. Scripts have been added to allow the detection of user hand poses when the user is in controller mode. The application was created using Unity's in-built rendering pipelines since our testing showed that it achieved the best performance with Meta Quest. This application has been optimized to maintain a constant frame rate of 72, which is the maximum refresh rate for Meta Quest.

4.3.1 Gesture and Interaction Design

The experiment was designed by considering both interaction input systems (hand and controller tracking). The user interface elements associated with the task were all positioned on the user's right side. Early testing indicated that this would be less distracting for the user. The user interface provides detailed instructions of what to expect, how to proceed to the next phase, and the current stage of the experiment. In addition, a small UI element at the top of the board identifies the input system which informs the players whether or not to hold the controller. In order to prevent misreading the buttons, the board positions and sizes were carefully adjusted and tested. We also designed two versions of the application, one for seated use and the other for standing. We chose the standing version for the final study based on feedback from pilots as it was the preferred option allowing better mobility.

Due to the fact that users would be continuously using their hands throughout the experiment, we considered economic factors when positioning elements in the experiment. The design of interactable elements in VR can directly influence users' muscular fatigue. When the user stays in strenuous positions for extended periods of time, their arm may become fatigued, or even start to hurt, causing the 'gorilla arm' effect [17]. Although the distance to the object would vary depending on the length of the user's arm, the recommendation states the object should be placed within 35 centimetres of the user's abdomen [17]. We took this into consideration in our design and positioned the main interaction element within 35cm from the user's abdomen.

In order to reduce the possibility of accidental selection, a hand pose system was programmed and introduced to the participant. They would be told to place their thumb on the thumbstick of the controller and move their index finger away from the grip button to perform a pointing hand pose in VR (Figures 2 and 6.b) which is designed based on the standard Meta Quest interaction paradigm. Participants could also perform hand pointing pose using real-time hand tracking (Figure 6.a).

4.3.2 Physics-based Button System

For the core system of interaction within the application, we chose a physics-based button system since it was the most similar to the interaction in the real world. The interaction with buttons is also enhanced by the inclusion of multi-sensory feedback. We tested our implementation with different set-ups ourselves and through pilots. The final version used is one we believe replicated manipulation of virtual buttons in a way that corresponds to the state-of-the-art found in common VR experiences by replicating real-world interaction physics. The physics-based button system was implemented in the Unity game engine using the Spring Joint component as well as our own physics code. The spring functions as a piece of elastic that attempts to bring the two anchor points together to the same location. With force per unit of distance set by the Spring property, the strength of the pull is proportional to the current distance between the points. It is possible to adjust the Damper value in order to prevent the spring from oscillating indefinitely. In proportion to the relative speed between the two objects, the spring force will be reduced. If the value is higher, the oscillation will abate more quickly. Through the use of this method, the VR button would simulate the physical properties of an actual spring button. The button functions like a physical button, even though it can only be pressed with a movement finger (and not actual force). It means that if a user simply touches the button, it will not register as a click. They must push the button after they have applied pressure to the button to a certain point; this will then register as a click/selection. Following the pressing of the button, the user would receive visual feedback — the colour of the button would change, as well as feedback through sound. In the process of pressing hand pose, the participant can observe a 3D object being manipulated, at which point the button begins to move inside. Manipulating objects have been used as another method of providing multi-sensory feedback in order to inform the user that they are performing the task correctly.

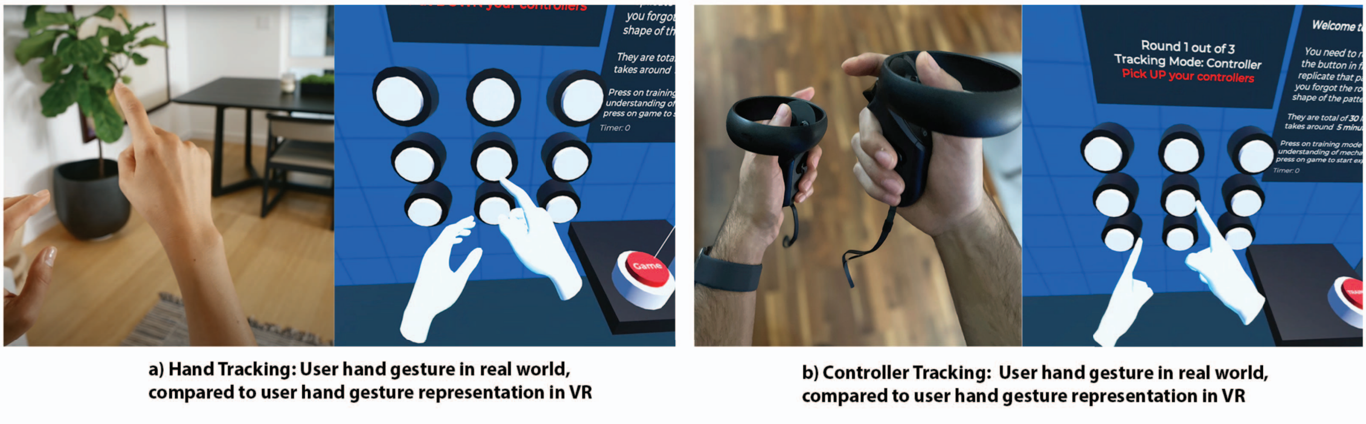

Figure 6: Accuracy of hand pose representation in hand tracking and controller tracking

4.4 Participants

Thirty-three participants were recruited for our final study via emails sent to different mailing lists at our university. Three of the participants' data were excluded due to technical failures (battery in controller dying), thus they were unable to complete the experiment. The remaining participants were comprised of 14 males, 14 females, and two non-binary individuals. The average age of the participants was 27.6, and the range was 21 to 43. The majority of participants were university students in their twenties. They were asked at the beginning of the experiment to identify their dominant hand, and out of the 30 participants, 24 were right-handed and six left-handed. The majority of participants had prior experience with VR and the use of controllers.

4.5 Procedure

Participants arrived one at a time. They were given the information about the research, instructions about the task, and also the duration of the experience and the number of rounds to be conducted. We then demonstrated to participants how to hold the controllers before they put on their HMDs. Following this, participants were asked to put on the Meta Quest headset provided by us. We helped them adjust it to a comfortable position. We would then ask them to launch the application from Meta Quest. In the tutorial section, they can also see images that explain how to hold the controller in VR. Participants were instructed to place their thumb on the joystick button and to move their index finger away from the trigger button to create a hand pointing pose. Meta Quest stand-alone version was used for the entire experiment; however, the device was connected to a PC via a USB port, and the participant view was streamed into the computer. This was done in order to guide the participant in case any problems should occur, as well as to observe participant behaviour during the experience.

Inside VR, we ask the participant to enter their age to confirm they are over 18 years old, they agree to our data collection policy, and that they give consent to participate. This also serves as part of the application's on-boarding process. Next, participants were shown again how to use the hand tracking and controller mode using text, images and video. We asked participants to complete a simple test (Pressing on a button) in order to validate that they understood the basics of interaction systems for hand tracking and controllers. After the participant successfully completed the test, they were asked to select their dominant hand and confirm it. Following that, the participant would go through a training session where they could play through three rounds of puzzles with three levels of complexity to gain an understanding of the task. After this, they would perform the actual task, which consists 3 sessions (i.e., the three conditions) with 12 trials each. Each participant took approximately 7-8 minutes to complete the study.

4.6 Measurements

First of all, in order to test our hypothesis that hand tracking reduces cognitive load, we collected behavioural results related to their overall

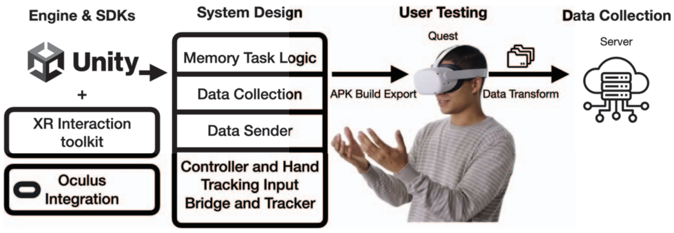

Figure 7: Application build pipeline overview.

performance in terms of both accuracy and speed (**H1**). Our primary measure is the percentage of winning for each round (*Correct Order and Pattern*): i.e., when participants managed to repeat the pattern in the correct order. Our secondary measure is the percentage of getting the overall pattern right, regardless of the order (*Correct Pattern Only*). In terms of speed, we measure both the first selection time (the time it takes the user to select the first button of the puzzle) and trial completion time (the time it takes the user to complete each trial). For each participant, an average of their first selection times were calculated for each session across the 12 trials as their *First Selection Time*, the same for *Trial Completion Time*. We use the same performance measurements to also investigate **H3**, related to haptic feedback.

Additionally, each participant was asked to identify his or her dominant hand and the application would monitor both hands of the participant while performing the task. The application would record which hand was used to select the buttons on the puzzle board. For each participant, a database is created that allows us to track how many times the participant has used the left or right hand to select buttons and solve the puzzle. Data points from these statistics were used to calculate a percentage of left-handed and right-handed use. We have labeled this as *Dominant Hand* in order to investigate **H2**.

4.7 Ethics

The Research Computing Department Ethics Committee of Goldsmiths University of London has approved this research. Since the experiment occurred during the recent pandemic, the ethics committee requested that a sanitized environment be created for the user, as well as the HMD device to be cleaned between each participant's participation. In additional, we also provided a sterile VR mask to the participants to ensure their safety. The experiment was conducted with one participant at a time, complying the rules of social distancing.

5 Results

In the analysis, we refer to the three conditions as: *Hand* (Hand Tracking), *Controller* (Controller Without Haptics), and *Haptic* (Controller with Haptics). Also, see Figure 8 for the box-plots of all behavioural

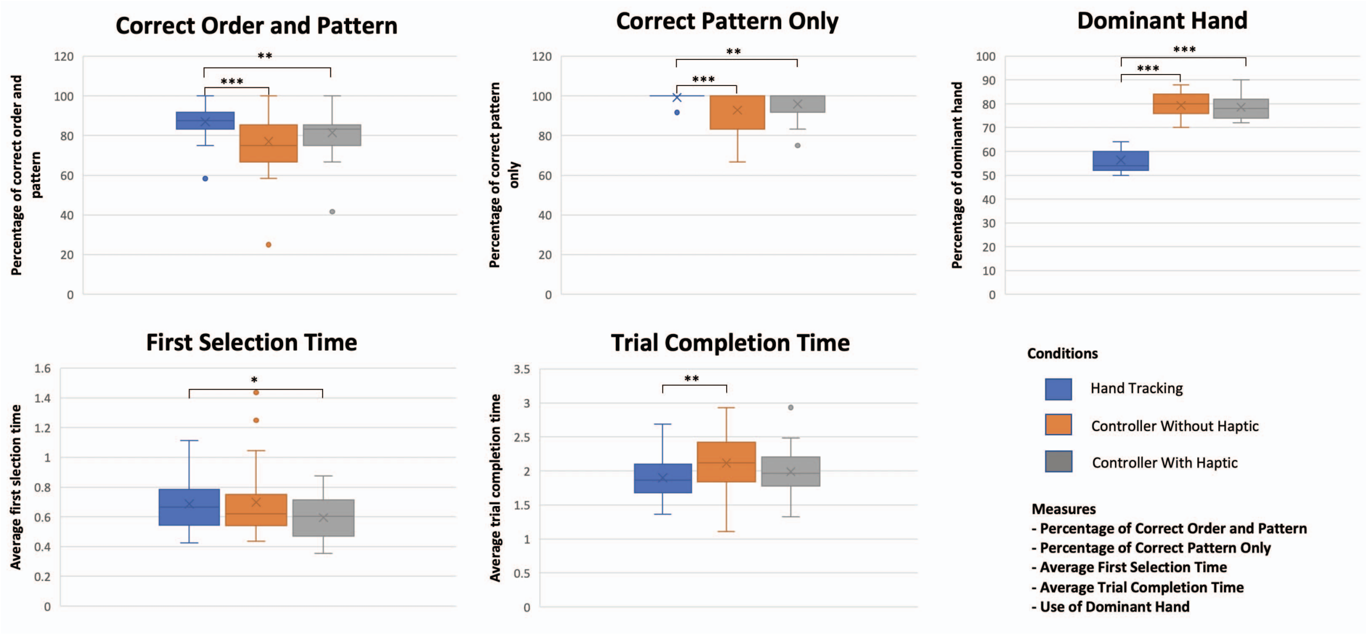

Figure 8: Behavioural Results: percentage of correct order and pattern, correct pattern only, dominant hand, and time for first selection and completion for all three conditions. ***: p = 0.000, **: p <0.01, * p < 0.02

data. We performed Repeated Measure One-Way ANOVA comparing the three conditions (*Hand*, *Controller*, *Haptic*) and found that there is significant difference for percentage of correct order and pattern $F(1,29) = 11.8$, **p=0.002**, $\eta^2 = 0.29$, percentage of correct pattern only $F(1,29) = 12.4$, **p=0.001**, $\eta^2 = 0.30$, average time of reaching at the first button $F(1,29) = 9.4$, **p=0.005**, $\eta^2 = 0.25$, and Percentage of using Dominant Hand $F(1,29) = 73.8$, **p=0.000**, $\eta^2 = 0.72$. No significant difference was found for average time of completing each trial $F(1,29) = 3.4, p = 0.074, \eta^2 = 0.12$. In the following, we present pairwise comparison for all measurements.

5.1 Percentage of Correct Order and Pattern

In terms of percentage of correct order and pattern, as we predicted, participants performed much better using *Hand*, achieving an average of 87%, as compared to 77% with *Controller* and 81% *Haptic*. Pairwise comparison reveals a significant difference between *Hand* and *Controller* **p = 0.000** as well as *Hand* and *Haptic* **p = 0.005**. However, there was no significant difference between *Controller* and *Haptic* $p = 0.063$. As test of normality was rejected (Hand: $p = 0.005$; Controller $p = 0.002$; Haptics $p = 0.000$), we performed Wilcoxon Signed Rank test which confirmed our results (Hand and Controller: $p = 0.000$, Hand and Haptics: $p = 0.007$).

5.2 Percentage of Correct Pattern Only

Similarly, in terms of percentage of getting at least the correct pattern, again participants performed much better using *Hand*, achieving an average of 99%, as compared to 93% with *Controller* and 96% *Haptic*. Pairwise comparison reveals a significant difference between *Hand* and *Controller* **p = 0.000** as well as *Hand* and *Haptic* **p = 0.004**. However there was no significant difference between *Controller* and *Haptic* $p = 0.075$. As test of normality was rejected (Hand: $p = 0.000$; Controller $p = 0.000$; Haptics $p = 0.000$), we performed Wilcoxon Signed Rank test which confirmed our results (Hand and Controller: $p = 0.001$, Hand and Haptics: $p = 0.004$).

5.3 First Selection Time

We also recorded the time it took for participants to reach the first button for each trial. On average, participants were the quickest to reach the first button with *Haptic* (0.60*s*), and slower with *Hand* (0.69*s*) or *Controller* (0.70*s*). Pairwise comparison reveals a difference between *Hand* and *Haptic* **p = 0.014**, with *Haptic* being significantly quicker than *Hand*. No other significant difference was found. As test of normality was rejected for the Controller condition (Hand: $p = 0.187$; Controller $p = 0.000$; Haptics $p = 0.330$), we performed Wilcoxon Signed Rank test which confirmed our results (Hand and Haptics: $p = 0.003$).

5.4 Trial Completion Time

When it comes to trial completion time, *Hand* was the quickest (1.9*s*) compared to *Controller* (2.1*s*) or *Haptic* (2.0*s*). Pairwise comparison reveals a significant difference between *Hand* and *Controller* **p = 0.006**, this time with *Hand* significantly better than *Controller*. No other significant difference was found. Test of normality was not rejected (Hand: $p = 0.562$; Controller $p = 0.941$; Haptics $p = 0.467$).

5.5 Percentage of Dominant Hand

It is interesting to observe that, as we predicted, only with *Hand* participants seem to be using both hands more equally (dominant hand percentage 59% on average), whilst with the other two sessions, they are using mainly their dominant hands (both *Controller* and *Haptic* have an average of 75%). Pairwise comparison indicated that there is indeed a significant difference between *Hand* and the other two sessions (**p = 0.000** for both *Controller* and *Haptic*). As test of normality was rejected for the Haptics condition (Hand: $p = 0.512$; Controller $p = 0.526$; Haptics $p = 0.000$), we performed Wilcoxon Signed Rank test which confirmed our results ($p = 0.000$ for both Controllers and Haptic).

6 Discussion

Most significantly, our performance result suggested that participants were able to have better accuracy with hand-tracking (better overall accuracy and better pattern correct). As shown in Figure 8, majority of participants got all of the pattern correct in the hand-tracking mode. This provides strong evidence to support **H1** that participants performed better in the cognitive load task with the hand-tracking mode. One possible explanation could be that, in the controller conditions, the mismatch between perceived and actual hand pose, coupled with having to hold the controllers, were the cause for the increase in cognitive load. This is supported by participants' comments in the interview. Additionally, this is consistent with previous research in which it has been shown that the use of pure hand tracking provides a more positive and less arousing experience than relying on controllers while performing selection tasks [37]

Our data also supported **H2** as participants tended to use both of their hands more equally in hand-tracking mode compared to controller conditions. We conclude that this is a direct outcome of hand tracking being a more natural interaction paradigm. This was also highlighted in the interviews conducted after the study, as participants commented that the hand tracking interaction felt more natural to them. This supports further that natural interaction is the cause for more active use of both hands while in hand tracking. In real world interactions, we tend to use both of our hands simultaneously, and we tend to use our dominant hand while performing tasks that require more precision. In controller tracking conditions, participants tended to rely on their dominant hand more. This is indicative of controller interaction being unnatural, such that users felt they needed to assert with more control and precision.

In terms of performance speed, participants were able to press their first button significantly faster in the controller with haptics condition than with hand tracking. However, since our results indicated that hand-tracking reduces cognitive load, we think this effect is a direct result of tracking latency. Although hand-tracking latency has improved dramatically recently, it is still slower compared with controller tracking, which is marker-based with built-in infrared diodes. The use of haptic cues has been shown in previous research to enhance accuracy for manual tasks [8], suggesting that participants may have been able to locate the buttons more quickly due to the haptic cues. Interestingly, the initial delay is then compensated when looking at the trial completion time, where participants were significantly faster in completing the trial in hand-tracking mode in comparison to the controller without haptics mode. This could be a result of reduced cognitive load. It could also be a direct result of the fact that participants were using both hands in hand-tracking mode, rather than just their dominant hand. Or, rather, being able to use both hands efficiently may well be the result of a lower cognitive load.

When it comes to the effect of haptics, our data indicated that for first selection time, controller with haptics is significantly quicker than hand tracking, whereas there is no difference between controller no haptic and hand tracking. This indicated improved speed with haptics. Similarly, for trial completion time, there is only a significant difference between hand tracking mode and controller no haptic mode, but no difference between hand tracking mode and haptic mode. This, again, indicates better speed in completing the task with haptics. Therefore, our result partially supported **H3**. Furthermore, another explanation could be the result of us implementing a multi-sensory feedback, which included: three forms of feedback: Sound, Visual (Color and object animation), and Controller Haptic. We compared the condition in which we would only take away the Controller haptic feedback and keep the remaining multi-sensory feedback. We think our results show that it is possible to use other forms of feedback combining sound and vision to make up for the lack of haptics, and that this is the underlying reason why there was no direct significant difference between the conditions of controller with and without haptics.

7 Limitations and Future Research

Overall our research shows that the use of hand tracking can reduce user cognitive load leading to improved performance in VR, and that our participants tended to behave more naturally while using hand tracking. This further supports that hand tracking has great potential for use in VR and mixed realities in general as it offers a more intuitive interaction method.

The biggest limitation of our study is that we were not able to include a condition where both haptics and hand-tracking were enabled. Initially designed during the lockdown in 2020, the whole experiment relied on consumer VR devices so that we were prepared to run it both in person and remotely. In future, hand-tracking can be combined with devices such as the Arduino board which can be mounted on user's hand or single vibrotactile motor on their finger tips. Further, In this study, we logged performance data during the user interaction in VR. However, we did not log users' hand position in VR, which we now think could yield interesting results. Last but not least, although hand-tracking clearly had advantages in our memory game task, it has obvious limitations and thus not suitable for all types of more sophisticated interactions. In comparison to controller tracking, hand-tracking has limited tracking volume, lower tracking accuracy, latency issues, lack of haptic feedback, and occlusion problems. For instance, it is not possible to perform overlapping hands gestures. Moreover, certain interactions, such as teleportation, rely on the use of the buttons and triggers on the controllers which provide reliable and quick input. Future studies are needed to test how to re-design those controller-based interactions to make them intuitive and robust. In the future, these limitations are likely to be resolved through advancements in hand tracking technology, which shall open up new possibilities for a 3D interaction paradigm.

Acknowledgments

Xueni Pan is funded by the AHRC Project "Immersive, Innovative, and Interactive Experience" AH/T011416/1.

References

[1] V. Athitsos and S. Sclaroff. Estimating 3d hand pose from a cluttered image. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, pp. II–432, 2003. doi: 10.1109/CVPR.2003.1211500

[2] L. W. Barsalou. Grounded cognition. *Annu. Rev. Psychol.*, 59:617–645, 2008.

[3] H. Benko, C. Holz, M. Sinclair, and E. Ofek. Normaltouch and texture-touch: High-fidelity 3d haptic shape rendering on handheld virtual reality controllers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, p. 717–728. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2984511.2984526

[4] D. B. Berch, R. Krikorian, and E. M. Huha. The corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition*, 38(3):317–338, 1998. doi: 10.1006/brcg.1998.1039

[5] L. M. Bietti. Towards a cognitive pragmatics of collective remembering. *Pragmatics & Cognition*, 20(1):32–61, 2012.

[6] S. Biswas and Y. Visell. Emerging material technologies for haptics. *Advanced Materials Technologies*, 4(4):1900042, 2019.

[7] J. Blake and H. B. Gurocak. Haptic glove with mr brakes for virtual reality. *IEEE/ASME Transactions On Mechatronics*, 14(5):606–615, 2009.

[8] G. Buckingham. Hand tracking for immersive virtual reality: opportunities and challenges. *CoRR*, abs/2103.14853, 2021.

[9] K. Dijkstra, M. P. Kaschak, and R. A. Zwaan. Body posture facilitates retrieval of autobiographical memories. *Cognition*, 102(1):139–149, 2007. doi: 10.1016/j.cognition.2005.12.009

[10] K. Dijkstra and R. A. Zwaan. Memory and action. *The Routledge handbook of embodied cognition*, pp. 314–323, 2014.

[11] P. Dourish. *Where the action is: the foundations of embodied interaction*. MIT press, 2004.

[12] J. Feltham. Best vr games of all-time: 25 titles to play now. https://uploadvr.com/best-vr-games/, 09 2022. (Accessed on 08/11/2022).

[13] J. K. Gibbs, M. Gillies, and X. Pan. A comparison of the effects of haptic and visual feedback on presence in virtual reality. *International Journal of Human-Computer Studies*, 157:102717, 2022. doi: 10.1016/j.ijhcs.2021.102717

[14] A. M. Glenberg. What memory is for. *Behavioral and Brain Sciences*, 20(1):1–19, 1997. doi: 10.1017/S0140525X97000010

[15] S. Goldin-Meadow, H. Nusbaum, S. D. Kelly, and S. Wagner. Explaining math: Gesturing lightens the load. *Psychological science*, 12(6):516–522, 2001.

[16] S. Han, B. Liu, R. Cabezas, C. D. Twigg, P. Zhang, J. Petkau, T.-H. Yu, C.-J. Tai, M. Akbay, Z. Wang, et al. Megatrack: monochrome egocentric articulated hand-tracking for virtual reality. *ACM Transactions on Graphics (TOG)*, 39(4):87–1, 2020.

[17] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1063–1072, 2014.

[18] A. B. Hostetter, M. W. Alibali, and S. Kita. I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and cognitive processes*, 22(3):313–336, 2007.

[19] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-based interaction: a framework for post-wimp interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 201–210, 2008.

[20] K. Kilteni, A. Maselli, K. P. Kording, and M. Slater. Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in human neuroscience*, 9:141, 2015.

[21] D. Kirsh. Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(1):1–30, 2013.

[22] S. R. Klemmer, B. Hartmann, and L. Takayama. How bodies matter: five themes for interaction design. In *Proceedings of the 6th conference on Designing Interactive systems*, pp. 140–149, 2006.

[23] R. McGloin, K. Farrar, and M. Krcmar. Video games, immersion, and cognitive aggression: does the controller matter? *Media psychology*, 16(1):65–87, 2013.

[24] J. L. McGrath, J. M. Taekman, P. Dev, D. R. Danforth, D. Mohan, N. Kman, A. Crichlow, W. F. Bond, S. Riker, A. Lemheney, et al. Using virtual reality simulation environments to assess competence for emergency medicine learners. *Academic Emergency Medicine*, 25(2):186–195, 2018.

[25] T. Monahan, G. McArdle, and M. Bertolotto. Virtual reality for collaborative e-learning. *Computers & Education*, 50(4):1339–1353, 2008.

[26] A. Nee and S. Ong. Virtual and augmented reality applications in manufacturing. *IFAC Proceedings Volumes*, 46(9):15–26, 2013. 7th IFAC Conference on Manufacturing Modelling, Management, and Control. doi: 10.3182/20130619-3-RU-3018.00637

[27] F. Paas and J. Sweller. An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review*, 24(1):27–45, 2012.

[28] G. Richard, T. Pietrzak, F. Argelaguet, A. Lécuyer, and G. Casiez. Studying the role of haptic feedback on virtual embodiment in a drawing task. *Frontiers in Virtual Reality*, 1:28, 2021.

[29] S. M. Scharoun and P. J. Bryden. Hand preference, performance abilities, and hand selection in children, Feb 2014.

[30] S. Seinfeld, T. Feuchtner, J. Pinzek, and J. Müller. Impact of information placement and user representations in vr on performance and embodiment. *IEEE transactions on visualization and computer graphics*, pp. 1–13, 2020.

[31] A. Skulmowski, S. Pradel, T. Kühnert, G. Brunnett, and G. D. Rey. Embodied learning using a tangible user interface: The effects of haptic perception and selective pointing on a spatial learning task. *Computers & Education*, 92:64–75, 2016.

[32] A. Skulmowski and G. D. Rey. Embodied learning: introducing a taxonomy based on bodily engagement and task integration. *Cognitive research: principles and implications*, 3(1):1–10, 2018.

[33] M. Slater, M. Usoh, and A. Steed. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2(3):201–219, 1995.

[34] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)*, pp. 67–76, 2016. doi: 10.1109/VR.2016.7504689

[35] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988.

[36] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr. Walking¿ walking-in-place¿ flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 359–364, 1999.

[37] J.-N. Voigt-Antons, T. Kojic, D. Ali, and S. Möller. Influence of hand tracking as a way of interaction in virtual reality on user experience. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–4, 2020. doi: 10.1109/QoMEX48832.2020.9123085

[38] M. Wilson. The case for sensorimotor coding in working memory. *Psychonomic bulletin & review*, 8(1):44–57, 2001.

[39] M. Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4):625–636, 2002.