Do It Fast, Forget It Fast: How Timing and Limb Visualizations Affect First-Person Augmented Reality Instructions

Clara Sayffaerth
LMU Munich
Munich, Germany
clara.sayffaerth@ifi.lmu.de

Ehbal Ablimit
LMU Munich
Munich, Germany
e.ablimit@campus.lmu.de

Annika Köhler
University Hospital Würzburg
Würzburg, Germany
koehler_a6@ukw.de

Jonas Wombacher
TU Darmstadt
Darmstadt, Germany
jonas.wombacher@tu-darmstadt.de

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Florian Müller
TU Darmstadt
Darmstadt, Germany
florian.mueller@tu-darmstadt.de

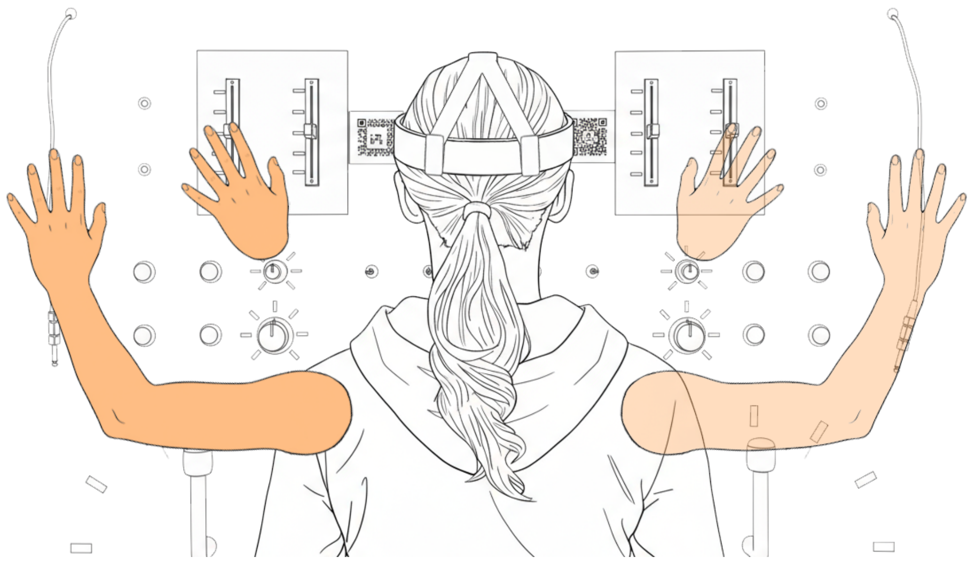


Figure 1: In this paper, we evaluate the influence of imitation timing and limb visualizations on efficiency, embodiment, social factors, and user comfort during first-person two-hand task instructions in AR.

Abstract

Acquiring tacit knowledge and practical skills often depends on direct observation and in situ training. AR offers an alternative by overlaying first-person step-by-step instructions that guide users through tasks such as assembly and repair. Previous work demonstrates the effectiveness of AR instruction for specific applications. In our experimental work, we systematically explore aspects of the broader design space. We conducted a controlled experiment (n = 40) to investigate three key factors identified in learning theory and XR embodiment research: imitation timing (parallel vs. sequential), limb visualization (hand vs. full arm), and limb visibility (opaque vs. semi-transparent). Across all conditions, participants followed AR instructions and afterward repeated the tasks from memory. We assessed performance, user experience, and retention. Our results show that parallel imitation is faster and increases embodiment, whereas sequential imitation enhances memory retention and comfort. Our findings provide guidance for the temporal and visual design of first-person AR tutorials.

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**;
• **Applied computing → Computer-assisted instruction**.

Keywords

Extended Reality, Instructions, First-Person, Machine Interface

ACM Reference Format:
Clara Sayffaerth, Ehbal Ablimit, Annika Köhler, Jonas Wombacher, Albrecht Schmidt, and Florian Müller. 2026. Do It Fast, Forget It Fast: How Timing and Limb Visualizations Affect First-Person Augmented Reality Instructions. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3772318.3791471

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791471

1 Introduction

In our everyday lives, we encounter tasks that are unfamiliar to us. We operate new devices, repair systems, perform maintenance tasks, and assemble instruments. For such tasks, we often rely on others for guidance or turn to video tutorials. These approaches are valid and widely used, but have limitations. While asking another person is not always possible, traditional video tutorials can miss critical aspects of spatial relationships or context-sensitive actions [11].

Extended Reality (XR) tutorials offer an alternative. They allow recording and replaying expert demonstrations or including Artificial Intelligence (AI)-generated instructions to provide spatial and contextual step-by-step guidance on demand [32, 61] directly within the user's environment. By situating the instructions in the physical workspace, XR systems preserve spatial context, reduce cognitive load, and offer a more immersive learning experience, allowing the user to move intuitively with less visual obstruction and occlusion [63]. This approach has proven particularly valuable for procedural and body-coordinated tasks, especially when the instructor is visually represented [11]. Previous work indicated that shifting from a third-person to a first-person perspective can improve task performance, foster embodiment, and enhance social connectedness with the virtual instructor [55].

Despite these promising results, however, fundamental questions remain about how to best design asynchronous first-person Augmented Reality (AR) instructions, where the real body is still visible without overlays or manipulations. A central challenge is fostering a strong sense of embodiment and making the instructions feel like the user's own actions rather than those of an external expert: Prior work shows that seeing oneself performing a skill at a level not yet achieved can improve performance beyond watching an instructor [19]. This suggests that strengthening embodiment in AR instruction could further enhance learning outcomes. Techniques such as synchronizing users' movements with virtual limbs can reinforce body ownership [60] and even social connectedness with the instructor [64]. However, it remains unclear how synchronizing with instructions by performing them in parallel rather than observing them first affects performance, memory retention, and user perception in asynchronous AR contexts.

A connected design dimension is limb representation and visualization. Virtual limbs can either be connected to the user's shoulders to create a full-arm representation, which may strengthen the sense of ownership and embodiment, or be visualized as hands only, which could reduce visual clutter. Similarly, limb transparency is frequently employed to prevent occlusion of the task workspace [65] and could therefore contribute to improving performance and memorization, yet it also can lead to dehumanization [51]. The effects of these design dimensions on embodiment, memory, and performance remain largely unexplored in this context.

In this work, we address this gap by systematically investigating the effects of the imitation *timing* (*parallel*, *sequential*), the visualization of the *limbs* (*hand*, *arm*), and their *visibility* (*opaque*, *transparent*) on performance (accuracy and task completion time), retention, mental load, embodiment, social factors (social closeness and presence) as well as on the comfort regarding the visualizations during two-handed tasks. For this, we designed and conducted a user study ($n = 40$) in which the participants followed AR instructions involving step-by-step manipulation of manual controls in a predefined order while varying the aforementioned variables, and afterward repeated them without guidance to evaluate their influence on memory. We found that *parallel* imitation was faster, increased embodiment and perceived co-presence, whereas *sequential* interaction lowered mental load, felt easier, and yielded better immediate retention of order, values, and hand use. *Hands*-only visualizations further boosted order retention. However, *visibility* did not influence the measured variables. With these findings, our work offers clear design levers for speed-vs-retention trade-offs in the design of XR instructions.

The contribution of this paper consists of:

(1) An experimental study of the effects of *timing*, *limb* visualization, and *visibility* during a two-handed manual task and the results on performance, user experience, and immediate retention ($n = 40$).
(2) Based on the findings, we contribute design guidelines for future XR tutorial systems.

2 Related Work

This paper is grounded on related work in the areas of learning by imitation, XR instructions, and body representations. This section will go into detail about the approaches influencing this work.

2.1 Learning by Imitation

Practical skills often rely on spatial information and coordinated movement, making the choice of instructional medium particularly important [8]. From early childhood, humans acquire new skills through imitation, a process that builds on observing and reproducing the behavior of others [3]. This form of social learning emphasizes how the actions of a model and their interconnected consequences form the foundation for knowledge. Further, demonstrations by an expert represent a crucial first step in skill acquisition, typically followed by imitation, repeated practice, and the development of effective and active learning strategies [10, 26]. But it does not always need another person to act as a model, seeing yourself performing a task on a skill level that has not yet been achieved can further increase performance [19]. Therefore, factors such as embodiment and social connectedness have been shown to positively influence learning efficiency and overall experience [2, 24]. Importantly, embodiment is not static, but can shift depending on the learner's perspective. These perspective changes can also strengthen social connectedness [70], further shaping how instructions are experienced and internalized. Since qualified people are not always available to provide guidance, and traditional methods make it difficult to achieve a shift in perspective during instruction, we need to find alternative ways to transfer practical skills. XR technologies promise, through their contextual and spatial abilities, to solve this problem. In the following, we will look at related research in this area.

2.2 XR Instructions

XR applications have been increasingly employed to teach knowledge and practical skills, utilizing their three-dimensional nature to achieve more accurate task demonstrations compared to traditional

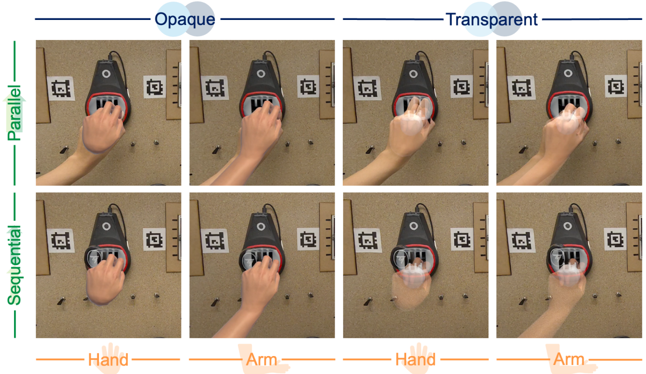


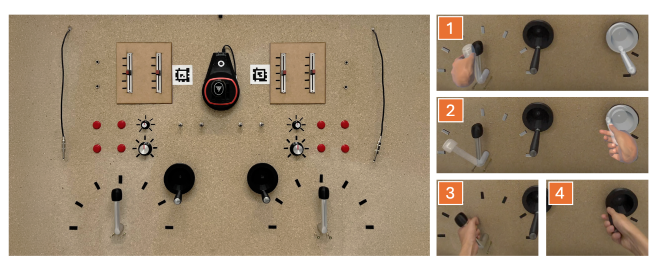
Figure 2: Timing (parallel, sequential) × limbs (hand, arm) × visibility (opaque, transparent) conditions during the machine task instructions in AR.

manuals [45], 2D tutorial videos [11, 15, 65], and, in some cases, even in-person training [69]. These instructional approaches can be broadly categorized into synchronous and asynchronous solutions.

Synchronous systems allow users and experts to interact in real time, bridging large physical distances and enabling training scenarios that would otherwise not be possible [56]. Examples include systems that let users adopt another person's perspective [13], experience co-embodiment for improved imitation [25, 38, 62], or benefit from enhanced communication through annotations and spatial cues [18, 23, 28, 66]. However, synchronous training can become impractical when experts are required to instruct many learners simultaneously, leading to cognitive overload and scalability issues [49]. Asynchronous solutions address these limitations by decoupling the guidance from real-time expert availability. Such systems can be based on human recordings, AI-generated instructions, or a combination of both [16, 32, 45, 61]. Asynchronous XR instructions are particularly effective for procedural and body-coordinated tasks, especially when the instructor is visually represented [11]. Most existing systems use third-person visualizations of the instructor [11], while a smaller number implemented first-person views [29, 55], or allow participants to select their preferred perspective [41, 45, 65]. First-person visualizations have been shown to outperform third-person ones in terms of task performance, social connectedness, and embodiment [55]. Despite these promising findings, research on asynchronous first-person XR instruction remains limited. Moving in synchrony with a virtual body can enhance embodiment and social connectedness, which are factors known to support learning [60, 64], but few studies have explicitly investigated these effects in asynchronous settings. Timing appears to be an important design factor: while communication delays can negatively affect collaboration and performance [35], well-structured imitation timing can improve speed imitation in non-avatar-based systems [12]. Yang and Kim [69] compared synchronous, semi-transparent, parallel first-person instructions to real-world, third-person sequential training, finding that performance was maintained, and in some cases improved, in the first-person parallel condition. However, it is difficult to determine whether these improvements are a result of the perspective shift, imitation timing, or a combination of these factors, and how these findings generalize to asynchronous settings. In addition to timing and perspective, body visualization plays a critical role. In AR, when the user's own physical body remains visible, mismatched or additional limb representations can induce discomfort or even a sense of creepiness [55, 57, 68]. Over time, such effects may not only hinder task performance and embodiment but also impact user well-being. Consequently, there is a strong need, besides instruction timing, to understand how different body representations influence performance, memory retention, embodiment, and comfort during asynchronous first-person AR instructions.

2.3 Body Representations

Because every human body is unique and applications differ, XR researchers have explored various ways to depict participants' bodies within immersive environments. When looking at realistic representations, most studies adapt the virtual avatar's gender to the participant, but often neglect to adjust skin tone or exclude participants based on gender or skin color [4]. This omission can negatively impact the user experience and study results, as dissimilar skin tone has been shown to decrease the sense of embodiment [21]. It can even lead to behavioral changes in motor tasks [34], underlining the importance of similarity between virtual and real body features. Recent work argues that matching skin color, not only gender, is essential, both to prevent these influences and to promote ethical inclusion [4, 40]. As an alternative, researchers have experimented with arm and hand representations besides realistic human arms, ranging from cartoonish to robotic, abstract, or even invisible visualizations [57] and investigated abstract versus realistic guiding arms for assistance tasks [22]. While some studies show the full arm of an instructor for the limbs [13], others only visualize the hands [45, 55], part of the arm [22], or adapt it depending on the proximity [41]. Such changes increase but also decrease performance based on the task, and can be perceived as uncanny or uncomfortable [22, 55, 57, 68]. Transparency has emerged as another design variable: semi-transparent hands can make tasks feel easier [65] and, when paired with interpenetrable interaction, improve efficiency [67]. A work by Zhang et al. [71] explored how blue ghost and realistic opaque full arm visualizations affect ownership, agency, and social presence of one's



(a) **(b)**

Figure 3: The machine interface (a) and an exemplary imitation step in the (*sequential, hand, opaque*) condition (b). The interface with its individual elements consists of the same number of buttons, switches, pins, knobs, sliders, wheels, and levers on both sides, as well as a shift and an additional button in the middle of the board.

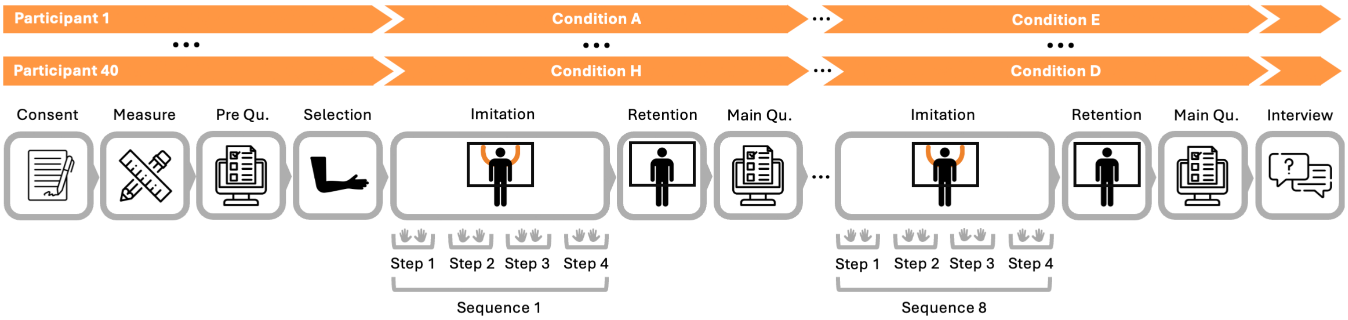


Figure 4: The different phases of the study showing the varied conditions for each participant on the top and the fixed steps of the procedure on the bottom.

own versus the instructor's hand in Virtual Reality (VR) using controllers instead of the participants' hands. The results showed that realistic hand visualizations improved ownership, while transparency helped differentiate the different hands from each other. However, transparency may also dehumanize avatars and influence visibility differently depending on skin tone, due to current display technology limitations [51]. Beyond visual fidelity, research has compared collocated and interactive hand representations [20, 39]. Other systems omit body visualization altogether, leaving the user without a virtual proxy while only showing the tools to interact with [15] or only cues [12]. While this seems like the easiest way to visualize interactions and include all kinds of participants, it can negatively impact performance compared to visualizing them with body parts [11]. In general, this mismatch between the virtual body and the participant's real body can have safety implications. Studies show that dissimilar avatars can reduce body awareness [36, 37], diminish body ownership [52], alter pain perception [43, 44], and even increase the risk of injury [14], which can be fatal when using devices and tools during manual instructions. Most studies focused on changing one's own body representations, especially in VR environments. Therefore, it is still unclear how limb guidance from a first-person perspective needs to be designed not only to improve factors such as performance and retention, but also embodiment, social factors, and comfort.

3 Methodology

Prior research demonstrates that XR, particularly first-person AR instructions, offers clear advantages over traditional manuals, videos, and even some in-person training methods [45, 65, 69]. Studies indicate that both the body representation and timing of information delivery shape how effectively users fulfill tasks [12, 57]. However, these factors have never been systematically examined for asynchronous first-person AR instructions. As a result, fundamental questions remain about how timing and body visualization affect not only task performance and retention, but also critical aspects of user experience such as embodiment, social connectedness, and comfort. Consequently, addressing this gap is essential for designing future AR guidance that seamlessly integrates with real-world interactions. With ethics committee approval, we conducted a controlled user study to investigate these factors and answer the following research questions:

RQ1 How does the *timing* of first-person instructions affect performance, immediate retention, and user experience during two-handed tasks?
RQ2 How does the *limbs* visualization of first-person instructions affect performance, immediate retention, and user experience during two-handed tasks?
RQ3 How does the *visibility* visualization of first-person instructions affect performance, immediate retention, and user experience during two-handed tasks?

3.1 Design

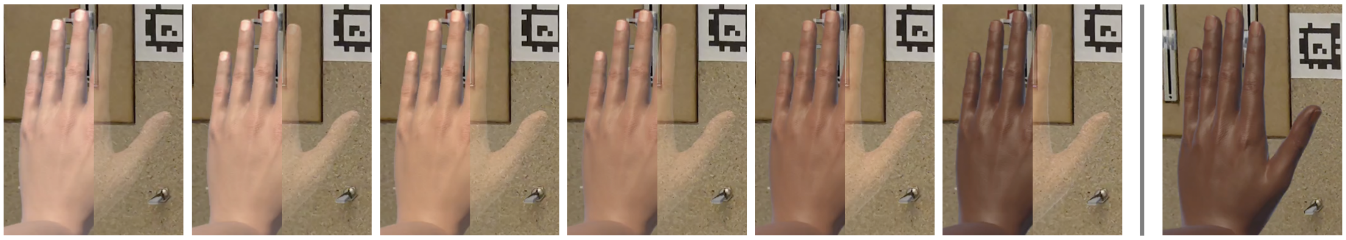
In order to answer the research questions, we designed a user study varying the *timing*, *limbs*, and *visibility* as our three independent variables, each with two levels:

Timing of imitating the instructions, doing them in *parallel* (at the same time) or *sequential* (watch-then-do).
Limbs visualization during the instructions, consisting of either only the *hand* or full *arm* representations.
Visibility of the limbs differentiated by being displayed *opaque* or semi-*transparent*.

We varied all independent variables in a within-subjects design with a total of $2 \times 2 \times 2 = 8$ conditions and counterbalanced the order using a balanced Latin square (see Figure 2).

3.2 Task

To precisely evaluate the performance and immediate memory retention, we designed a two-handed task inspired by prior work on asynchronous XR instructions with machine interfaces [11, 42, 55]. We selected a two-handed task that required Latin square balanced use of both hands to minimize potential effects of handedness [33]. Throughout the study, participants observed two virtual limbs from a first-person perspective, guiding them solely visually through actions on a machine interface. Depending on the condition, participants either imitated the movements simultaneously with the animation (*parallel*) or first observed and then imitated them afterward (*sequential*). Further, we varied the virtual limbs, depicting only the *hand* or the full *arm*, and to appear either completely visible (*opaque*) or see-through (*transparent*). After completing all imitation steps in a given condition, participants immediately repeated the entire sequence from memory without visual guidance



(a) #FFFFFF (α : 0.22) **(b) #F2EDE4** (α : 0.28) **(c) #ECDFC9** (α : 0.30) **(d) #CCBCA9** (α : 0.30) **(e) #AB947E** (α : 0.62) **(f) #735A43** (α : 0.99) **(g) #735A43** (α : 1.00)

Figure 5: The different skin tones of the limbs and their transparent counterpart, including the hexadecimal code and alpha value (α). (a-f) show limb shape 1 and (g) shape 2.

to assess retention. To ensure consistency, they stood in a fixed position marked on the floor during both phases.

We designed the task to remain challenging but feasible by following the guideline of 7 ± 2 information chunks from Miller [46], which defines the typical capacity of human working memory. This resulted in 8 imitations per condition, divided into four sub-steps with one left-hand and one right-hand interaction each, to improve perception and memorization of the steps [65]. As the same order of elements could have led to learning effects, we implemented 8 different element types, each only used once per sequence, while varying the specific element and value. We counterbalanced the element types with a Latin square together with the hand usage. We mounted the 8 different element types mirrored on a vertical wooden panel: 8 buttons, 4 switches, 4 sliders, 2 levers, 4 rotary knobs, 4 pin sockets, including 2 pins, 2 wheels, and 1 shift in the middle to improve reachability. We added another button in the top middle of the panel to distinguish the actions of the participant and the study supervisor. The machine interface is shown in Figure 3a. With 8 conditions and 8 imitations per sequence, this results in a total of $8 \times 8 = 64$ combinations. While only varying the order of our 8 conditions, the 64 combinations of sequences, hand usage, values, and elements stayed the same for every participant. For instance, the first step involves using the left lever and then turning the right wheel. One participant receives these instructions in the (*parallel*, *hand*, *opaque*) condition, while the next completes the same first sequence in the (*sequential*, *hand*, *opaque*) condition. This exemplary first step during the *sequential* condition is shown in Figure 3b, and the whole procedure in Figure 4.

3.3 Dependent Variables

As part of the study, we collect data using the loggers of the machine interface and application, videos, quantitative, and qualitative questionnaires, as well as conducting interviews and taking notes, resulting in the above-mentioned dependent variables:

3.3.1 Performance. was evaluated through correctness and task completion time of the instructions imitation in the first phase. We calculated the **correctness** by logging the electronic parts of the machine interface and video recordings. Three authors validated all measured values independently through the recorded videos, which we then compared. We calculated 4 boolean values for correctness out of the data for each step: correct value adjustment, correct element order, correct hand usage, and the overall correct imitation. For elements with continuous values, we measured the boundary between two setting options, resulting in reference areas. We estimated the correct hand usage solely through the video recordings independently by two authors, and then compared the data. Further, we measured the **task completion time** through the logger data. We calculated the time from the beginning of the animation to the last interaction with the interface (overall time) as well as the time between the first and last interaction with the interface (interaction time).

3.3.2 Retention. was evaluated by comparing the participants' recall from memory in the second phase against the initial instructions and their imitation in the first phase. Specifically, we measured the amount of correct **value** imitations, elements correctly following one another (**order**), and **hand usage** similar to the first phase.

3.3.3 User Experience. collected using questionnaires. We evaluated the **mental load** using the *RAW (NASA-)TLX* [30, 31] (21-point Likert scale, 0: Very Low, 20: Very High), the *Single Ease Question (SEQ)* to record the feeling of task difficulty [54] (7-point Likert scale, 1: Very Difficult, 7: Very Easy) as well as our own question regarding the participants' perceived performance (7-point Likert scale, 1: Strongly disagree, 7: Strongly agree). Further, we obtained the information about the **embodiment** of the artificial limbs using the short *Avatar Embodiment* questionnaire [50] (7-point Likert scale, 1: Strongly disagree, 7: Strongly agree). For the **social factors**, we measured the closeness towards the virtual expert using the *Inclusion of Other and Self (IOS)* [1] (7-point Likert scale visualized through circles with 1 being not at all close and 7 being extremely close) and the Co-Presence using the *Perception of self* sub-questionnaire of the *Networked Minds Social Presence Inventory (NMSPI)* [5, 6] (7-point Likert scale, 1: Strongly disagree, 7: Strongly agree). The feelings towards the **instructions**, including the comfort regarding the limb visualization, were assessed with our own questions (7-point Likert scale, 1: Strongly disagree, 7: Strongly agree)

As qualitative measurements, we wrote notes during the study and also allowed participants to leave written comments at the end of the questionnaire after each condition. The semi-structured interview questions at the end of the study addressed the experience during the instructions, the timing, limbs, visibility, ownership of the limbs, and general suggestions to get insight into the feelings of the participants by using a mixed-methods approach.

3.4 Apparatus

To conduct the study, we built a machine interface and a corresponding AR application. We crafted the interface using a combination of 3D printing, laser cutting, milling, drilling, jigsaw cutting, soldering, and hot glue assembly. To capture user interactions with all elements except the shift control, we integrated sensors including buttons, switches, sliders, potentiometers, and rotary encoders. These sensors were connected to an Arduino Mega ADK, which communicated with a PC. The industrially manufactured shift was connected directly to the PC. We implemented data logging and event handling using the Arduino IDE with Python scripts.

The AR application was developed in Unity and on a Meta Quest 3 headset, which remained connected to the PC throughout the study to enable real-time control, marker tracking, and pass-through streaming. To avoid battery drain, we alternated between two identical headsets. For each participant, we positioned the interface on an adjustable table at the beginning of the study so that it aligned with the participant's shoulder level. This ensured that the virtual arms appeared consistent with the participant's body, as we rendered the limbs up to the shoulder in the full-*arm* condition using a similar representation as the one Lampen et al. [41] implemented for their proximity condition. We used Character Creator 4 to generate two limb shapes, each available in six skin tones based on the Hafnia Hands dataset [52]. Participants could select their preferred limb shape and tone at the beginning of the study. To manipulate opacity for each skin tone, we adjusted the alpha values to achieve a consistent level of transparency for all visualizations, using the darkest shade as the baseline [51] (see Figure 5). For visual alignment, we created white ghost digital twins of all interface elements and edited them with the recorded avatar animations. Digital twins are commonly used in AR to enhance realism and provide clear, spatially aligned guidance [48].

To measure performance, we logged the timestamp (in milliseconds), element identifier, and new value whenever a machine element's state changed. We additionally recorded key events such as application start, animation start and end, sound cue playback, and condition completion. A video camera continuously recorded the interface during the study, allowing us to verify correct hand usage and cross-check logged interaction data.

3.5 Procedure

After welcoming the participants, we introduced the procedure whereby we framed both the task and interface in a neutral way without a specific purpose to prevent any unintended bias. Further, we obtained informed consent, measured their height, and fitted the adjustable table with the interface on top to their shoulder height. Next, we asked them to complete a pre-questionnaire capturing demographic information and prior knowledge. The participants then put on the Head-Mounted Display (HMD), and we launched the application. We aligned the virtual machine elements with their real-world counterparts through QR codes and adapted them via a video stream. Then the participants chose a preferred limb shape and color from the prepared set of models. Before each condition, we told the participants whether they should perform the instructions simultaneously with the instructor or after the demonstration. Once they were ready, we started the first condition. We played a sound cue to signal when they should begin imitating, and the participants verbally indicated when they had completed the step, followed by us triggering the next animation using a keyboard command. This was an intentional decision not to interfere with the memory of the participants by pressing an additional button between the animations. The participants did this process four times. Afterward, they turned around so that they could not watch us resetting the interface. Before and after resetting, we pressed a button on top of the interface so that we could differentiate between our actions and the participants' actions. This took approximately 30 seconds. They then repeated the full instruction sequence from memory, this time without visual guidance, before pressing the button on top of the interface to signal that they are finished. Next, they removed the HMD and completed a questionnaire on a computer about the current condition. This inter-condition pause of approximately 4 minutes mitigated simulator sickness and enabled reconfiguration of the interface for the subsequent condition. We repeated this procedure with different conditions, elements, and values until the participants had completed all eight conditions. Finally, we conducted an audio-recorded interview to capture their impressions. The entire experiment took about 75 minutes per participant.

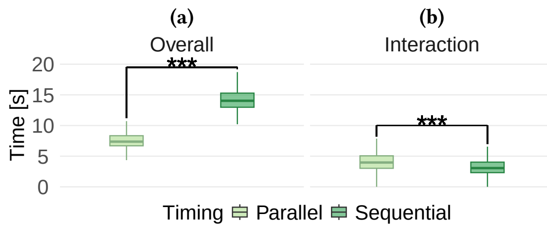


Figure 6: Box plots of the overall (a) and interaction (b) task completion time by *timing*, visualizing an increase of the overall time and a decrease of the interaction time for the *sequential* compared to the *parallel* condition.

3.6 Participants

For the study, we recruited 40 participants (22 women, 18 men), aged between 21 and 64 (*Mean* = 27.95, *SD* = 7.49). 38 participants were right-handed, and 2 were left-handed, with the height varying between 1.59 m and 1.91 m (*Mean* = 172.23, *SD* = 8.54). The participants consisted of 24 students, 4 researchers, 2 healthcare workers, 2 designers, 2 trainees, 1 physiotherapist, 1 consultant, 1 mechanic, 1 graduate, 1 unemployed person, and 1 language correspondent. All participants voluntarily took part in the study and got reimbursed 12€ per hour or study points.

3.7 Analysis

We analyzed our data using Linear and Generalized Mixed Effects Models (LMM/GLMM). For ordinal questionnaire responses (e.g., IOS), we used `clmm` with a logit link fitted using the `ordinal` R package as suggested by Christensen [17]. We employed GLMMs (logit link) with a binomial family fitted using the `glmmTMB` [9] package for binary variables (e.g., correct value). For count data like the retention measures, we followed a similar approach but used a Poisson family. We used (Gaussian) linear models fitted using `lme4` for continuous data such as the task-completion time. All models

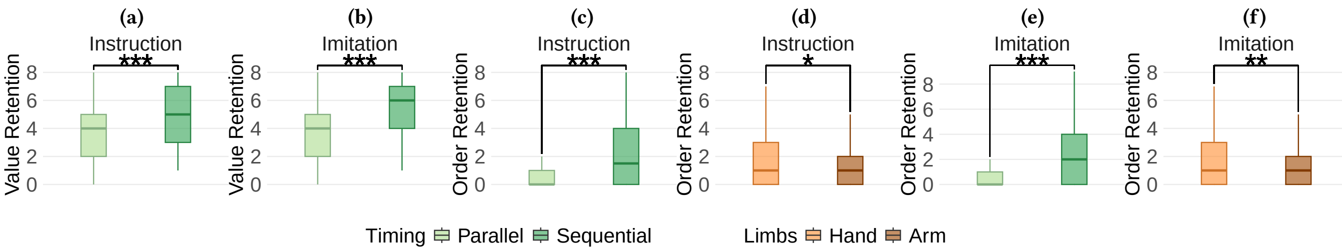


Figure 7: Box plots of the value (a-b) and order (c-f) retention compared to the initial instructions (a, c, d) and the inputs of the participants in the imitation phase (b, e, f) by *timing* (green) and *limbs* (orange) showing increased retention in the *sequential* and *hand* condition.

included our independent variables: *timing* (*parallel*, *sequential*), *limbs* (*hand*, *arm*), and *visibility* (*opaque*, *transparent*), and their interactions as fixed effects. We further added a random intercept per participant and tested fixed effects with Type-III Wald tests. We assessed model assumptions via residual diagnostics using the `performance` package. To estimate the mean response and perform post-hoc contrasts, we used Bonferroni-corrected Estimated Marginal Means (EMMs) as proposed by Searle et al. [58] using the `emmmeans` package. Due to loss of data recordings, we excluded 6 of the 320 questionnaires from 4 different conditions and 52 of the 1,280 runs from all 8 conditions. The Linear Mixed Effects Models (LMM) fully accounted for this missing data, as it is well-suited for handling incomplete datasets. We highlighted significant differences (p<.001: $***$, p<.01: $**$, p<.05: $*$) for all box plots.

4 Results

In the following, we report the results of our controlled quantitative and qualitative user study with a focus on the variables that showed significance. In the pre-questionnaire, the participants reported on a 7-point Likert scale (1: Strongly disagree, 7: Strongly agree) their prior experience with AR ($Mean = 4.12$, $SD = 1.73$), HMD ($Mean = 3.42$, $SD = 2.07$), and machine interfaces ($Mean = 3.42$, $SD = 1.84$). 21 participants picked limb shape 1 (Figure 5a-5f) and 19 picked shape 2 (Figure 5g) as a representation at the beginning of the study. 9 chose (a), 6 (b), 16 (c), 6 (d), 2 (e), and 1 (f) from Figure 5 as the skin tone. The Mean and standard deviation (SD) values are listed in Table 1.

4.1 Performance

We assessed the performance of the first imitation phase by analyzing the correctness per step as Boolean values and the task completion time in milliseconds.

4.1.1 Correctness. In order to analyze the correctness, we calculated the correct value adjustment, action order, hand usage, and the resulting overall correct execution. Thereby, we could not find any significant effects for the correct value ($Mean = 0.78$, $SD = 0.42$), order ($Mean = 0.93$, $SD = 0.26$), hand ($Mean = 0.94$, $SD = 0.24$) or overall correctness ($Mean = 0.71$, $SD = 0.45$).

4.1.2 Task Completion Time. We determined the task completion time in two ways: the beginning of the animation to the last interaction with the interface (overall time) and the time from the first to last interaction with the interface (interaction time). The analysis shows that the *timing* significantly affects the overall time ($\chi^2(1) = 2150.81,\ p < .001$) with *sequentially* taking longer than *parallel* (Figure 6a), while for the interaction time ($\chi^2(1) = 84.05,\ p < .001$) *parallel* is more time intensive than *sequential* (Figure 6b).

4.2 Retention

For the retention, we compared the value adjustments, action order, and hand usage of each retention phase with the initial instruction sequence, as well as the imitated actions of the participants in the first round, as count values.

4.2.1 Value. When looking at how the participants recalled the initial instruction sequence, the results show that a significant effect on *timing* value adjustment ($\chi^2(1) = 28.95,\ p < .001$) with *sequential* being more accurate (Figure 7a). In addition, we compared the actions of the participants in the imitation phase to the retention in the repetition phase. The results indicate a significant difference for *timing* ($\chi^2(1) = 42.09,\ p < .001$) with *parallel* performing worse than *sequential* (Figure 7b).

4.2.2 Order. For the recall of the initial instruction sequence, the results show that the *timing* ($\chi^2(1) = 83.80,\ p < .001$) significantly affects the order, with *sequential* being more accurate (Figure 7c). Further, the analysis reveals that the *hand* leads to a higher order recall than the *arm limbs* ($\chi^2(1) = 5.73,\ p < .05$) (Figure 7d). When comparing the imitation of the order from the first phase with the recall in the second phase the analysis shows significant effects for *timing* ($\chi^2(1) = 81.97,\ p < .001$) and *limbs* ($\chi^2(1) = 9.31,\ p < .01$) with *sequential* (Figure 7e) and *hand* (Figure 7f) leading to better results than their counterpart.

4.2.3 Hand usage. The analysis of the correct hand usage for *timing* shows significant effects both for recalling the initial imitation sequence ($\chi^2(1) = 22.22,\ p < .001$) (Figure 8a) as well as the recall of the participants actions in the first phase ($\chi^2(1) = 21.09,\ p < .001$) (Figure 8b) with *sequential* leading in both cases to improved hand usage recall.

4.3 User Experience

In addition to the performance and retention measurements, we collected the user experience through questionnaires regarding mental load, embodiment, social factors, and the instructions, including comfort towards the limbs.

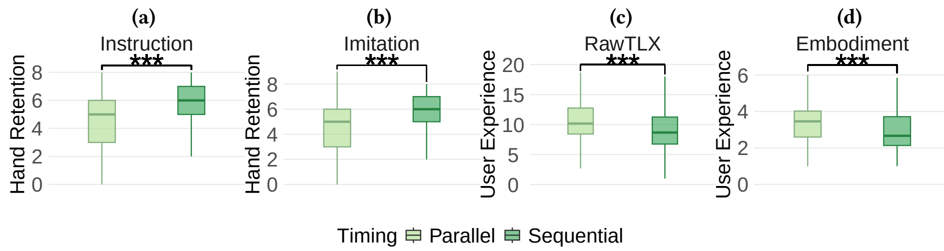


Figure 8: Box plots of the hand retention compared to the initial instructions (a) and the inputs of the participants in the imitation phase (b), as well as the Raw TLX (c) and embodiment (d) from the user experience questionnaire by *timing*, showing an increase for the order retention and a decrease for Raw TLX and embodiment for *sequential* compared to *parallel*.

4.3.1 Mental Load. To measure mental load, we used the Raw TLX showing a significant effect for *timing* ($\chi^2(1) = 38.25,\ p < .001$) with *parallel* leading to a higher rating (Figure 8c). Furthermore, we found significant effects regarding *timing* for the SEQ evaluating perceived task difficulty ($\chi^2(1) = 46.09,\ p < .001$) and our own question about successful task completion ($\chi^2(1) = 40.01,\ p < .001$) both with *sequential* resulting in an increased rating (Figure 9a).

4.3.2 Embodiment. We determined the embodiment with the short Avatar Embodiment questionnaire. The analysis shows significant effects for *timing* ($\chi^2(1) = 35.58,\ p < .001$) with *parallel* resulting in a higher embodiment (Figure 8d). Moreover, the results showed an interaction effect for the limbs and visibility ($\chi^2(1) = 6.20,\ p < .05$) with the post-hoc test showing no significance.

4.3.3 Social Factors. We assessed the social closeness using the IOS scale and the co-presence with part of the NMSPI. The results showed significant effects for *timing* for the IOS scale ($\chi^2(1) = 13.43,\ p < .001$) and the question if you feel like being in the same room as the instructor ($\chi^2(1) = 4.71,\ p < .05$) from the NMSPI both with *parallel* leading to higher ratings compared to *sequential* (Figure 9b). The data also showed an interaction effect regarding *timing* and *limb* for the question if the participant had the feeling of being in a different place than the instructor ($\chi^2(1) = 6.19,\ p < .05$) with the post-hoc test revealing no significance ($Mean = 2.84$, $SD = 1.52$). In addition, the other co-presence questions regarding awareness of ($Mean = 4.75$, $SD = 1.71$) or hardly noticing the instructor ($Mean = 3.19$, $SD = 1.81$) showed no significant effects.

4.3.4 Instructions. Further, we asked our own questions regarding the instructions. The analysis shows significant effects regarding *timing* for the questions if the participants liked ($\chi^2(1) = 12.69,\ p < .001$) and understood the instructions ($\chi^2(1) = 8.78,\ p < .01$) as well as if the instructor performed well ($\chi^2(1) = 12.80,\ p < .001$) all with *sequential* resulting in higher ratings than *parallel* (Figure 10a). Moreover, the analysis showed significant effects regarding the discomfort towards the virtual limbs both for *timing* ($\chi^2(1) = 6.69,\ p < .01$) and *limbs* ($\chi^2(1) = 7.48,\ p < .01$) with *parallel* and *arm* achieving higher ratings (Figure 10b).

4.4 Interviews and Qualitative Notes

We asked the participants about their experience in semi-structured interviews, which we recorded. After transcribing the audio, we conducted a thematic analysis [7], with two researchers independently coding 20% of the transcripts in ATLAS.ti. We then reviewed, discussed, and refined the codes. We continued to code the data and iteratively discussed code groups and whether codes should be added. This resulted in 175 codes summarized in the following.

When we asked the participants what they ***liked***, 37.5% found the experience interesting, fun (20%), good (37.5%), natural (10%), and better than traditional instructions (7.5%). 20% enjoyed the new experience in XR with 5% stating that it felt like *"playing a game"* (P11). Of the participants, 17.5% liked the machine interface, 5% found it good, that the instructions were clear (5%), the imitation task easy (7.5%), with some participants liking the challenge (7.5%).

The participants ***disliked*** that it was difficult to remember the instructions (67.5%), with 22.5% finding them too fast, sometimes too slow or too fast (7.5%), repetitive (7.5%), and lacking meaning (5%). 5% stated that they did not know where the instructions started, with 42.5% answering that the experience was difficult. P6 stated: *"It was quite fun, but also partly demanding, at least for me, because I really tried my best for the memorizing part."* 10% expressed that the repetition was challenging, the elements hard to differentiate (7.5%), sometimes not visible (10%), with some disliking specific haptic or virtual elements (5%). For 30%, the experience got more challenging over time, with 15% noting that there were too many rounds. Of the participants, 10% disliked the HMD, as it was uncomfortable (5%). 30% were disappointed in their own performance, and 17.5% pointed out that *"there is nothing really to dislike"* (P33).

Regarding the ***timing***, 80% of the participants preferred *sequential*, 15% *parallel*, and 5% were indifferent. For *parallel*, 10% of the participants liked that it was easier to remember with less time to forget the instructions (10%) and that it felt natural (10%). 20% disliked that they were reflecting less about their actions, as *parallel* felt like only doing motor activity. P17 mentioned: *"When I just follow along, I'm concentrating on following, so I don't really know what I'm doing."* The participants liked that *sequential* has advantages in terms of remembering (70%), more time (40%), overview (32.5%), and visibility (5%) of the instructions. Further, they perceived *sequential* as easier (10%), and not as stressful (17.5%). Still, P6 said regarding the *sequential* condition: *"My body was kind of acting on its own. It's like I couldn't wait."* 5% of the participants noticed no difference in memorization between *sequential* and *parallel*.

For the ***limbs*** 22.5% preferred *hand*, 30% *arm*, and 47.5% noticed no difference. On the one side, the participants preferred the *hand* visualization, as the instructions were better visible (5%). On the other side, the *hand* representation felt uncomfortable (15%) and distracting (7.5%). P9 mentioned: *"You suddenly see some hands, and*

Table 1: Mean and SD of all variables divided by variables and levels. All values that show significance are highlighted (black).

	Timing		Limbs		Visibility	
	Parallel Mean (SD)	Sequential Mean (SD)	Hand Mean (SD)	Arm Mean (SD)	Opaque Mean (SD)	Transparent Mean (SD)
Performance						
Correct Value	0.76 (0.43)	0.78 (0.41)	0.79 (0.41)	0.76 (0.43)	0.78 (0.41)	0.76 (0.43)
Correct Order	0.92 (0.27)	0.94 (0.24)	0.94 (0.24)	0.92 (0.27)	0.93 (0.25)	0.93 (0.26)
Correct Hand	0.95 (0.23)	0.93 (0.26)	0.93 (0.25)	0.94 (0.24)	0.94 (0.24)	0.93 (0.25)
Correct Overall	0.71 (0.45)	0.71 (0.45)	0.73 (0.44)	0.70 (0.46)	0.72 (0.45)	0.70 (0.46)
Time Overall	7.57 (1.39)	14.25 (2.02)	10.91 (3.80)	11.00 (3.78)	11.00 (3.80)	10.92 (3.78)
Time Interaction	4.08 (1.55)	3.39 (1.68)	3.70 (1.57)	3.77 (1.73)	3.82 (1.75)	3.65 (1.55)
Retention						
Value: Instruction	3.70 (1.77)	5.00 (1.86)	4.50 (1.92)	4.21 (1.93)	4.39 (1.88)	4.33 (1.98)
Value: Imitation	3.89 (1.88)	5.49 (1.83)	4.77 (2.08)	4.61 (1.97)	4.73 (1.98)	4.65 (2.07)
Order: Instruction	0.92 (1.43)	2.32 (2.38)	1.78 (2.10)	1.49 (2.07)	1.53 (1.95)	1.73 (2.22)
Order: Imitation	1.00 (1.61)	2.42 (2.47)	1.93 (2.28)	1.51 (2.12)	1.63 (2.14)	1.81 (2.28)
Hand: Instruction	4.63 (1.91)	5.88 (1.74)	5.27 (2.01)	5.26 (1.86)	5.28 (1.91)	5.25 (1.96)
Hand: Imitation	4.75 (1.89)	5.99 (1.75)	5.39 (2.01)	5.36 (1.84)	5.42 (1.89)	5.34 (1.96)
User Experience						
Raw TLX	10.6 (3.33)	8.92 (3.51)	9.68 (3.56)	9.83 (3.49)	9.68 (3.57)	9.82 (3.48)
SEQ	2.96 (1.47)	3.86 (1.60)	3.42 (1.61)	3.42 (1.59)	3.41 (1.59)	3.43 (1.61)
Task Completion	3.30 (1.54)	4.21 (1.74)	3.85 (1.71)	3.67 (1.70)	3.87 (1.68)	3.66 (1.72)
Embodiment	3.44 (1.13)	2.94 (1.15)	3.19 (1.14)	3.18 (1.19)	3.20 (1.14)	3.17 (1.19)
IOS	3.91 (1.83)	3.39 (1.91)	3.64 (1.77)	3.65 (1.99)	3.59 (1.91)	3.70 (1.86)
NMSPI: Same Room	5.00 (1.58)	4.80 (1.65)	5.01 (1.49)	4.78 (1.73)	4.92 (1.59)	4.88 (1.65)
NMSPI: Different Place	2.88 (1.53)	2.81 (1.51)	2.74 (1.43)	2.94 (1.60)	2.80 (1.55)	2.89 (1.49)
NMSPI: Awareness	4.85 (1.66)	4.66 (1.76)	4.83 (1.66)	4.68 (1.76)	4.79 (1.74)	4.72 (1.69)
NMSPI: Noticing	3.22 (1.84)	3.16 (1.78)	3.08 (1.72)	3.29 (1.90)	3.22 (1.84)	3.15 (1.78)
Instructions: Liked	4.66 (1.56)	5.13 (1.35)	4.91 (1.44)	4.89 (1.51)	4.90 (1.49)	4.90 (1.47)
Instructions: Understood	5.87 (1.10)	6.10 (0.92)	6.06 (0.93)	5.91 (1.10)	5.97 (1.02)	6.01 (1.02)
Instructions: Performance	4.73 (1.43)	5.19 (1.34)	5.06 (1.35)	4.87 (1.46)	4.94 (1.40)	4.99 (1.41)
Instructions: Discomfort	2.63 (1.62)	2.30 (1.50)	2.33 (1.49)	2.59 (1.63)	2.37 (1.45)	2.56 (1.67)

you don't know where they come from. But if you see the whole thing, then you know it's from a person." According to the participants with *arm*, they could remember better (7.5%), felt more like their own limbs (17.5%), more realistic (10%), and could more easily perceive the movement (10%). But the *arm* also felt uncomfortable (15%) with an out-of-the-body experience (5%). Further, 12.5% found it distracting, and obstructing the view (5%). P26, for example, explained: *"The whole arm basically makes it more confusing, because then it feels like it's not me."* In general, 5% said that the *limbs* were making it hard to see the position of the elements, and better when they imagined the *limbs* as an extension of themselves (5%). For the people who did not see a difference between the *limbs*, 5% mentioned they only saw the *arm*, and 15% only the *hand*.

Looking at the ***visibility***, 17.5% prefer *opaque*, 12.5% prefer *transparent*, 65% did not notice a difference, and 5% had no preference. 12.5% of the participants mentioned that *opaque* lead to clearer instructions as they *"knew where to focus"* (P38). 7.5% found it more realistic, better to blend with your body (7.5%), but also irritating (7.5%). With *transparent* visualization, it was easier to follow (5%), and improved visibility of instructions (15%): *"It doesn't feel like an obstruction of my sight, and it is clearer. I have fuller vision, I guess."* (P8). The participants also mentioned that *transparent* felt like a disconnection between view and reality (5%).

Of the participants, 12.5% said it felt like ***another person***, 70% thought it was not another person, and 17.5% had the feeling in specific moments. The participants perceived the instructions as a presence (5%), a navigation guideline (10%), or explicitly as virtual (25%). P1 explained: *"I know the instructions are from a computer, not from you or from my head."* 5% only felt like there was another person at the beginning, or 5% during *sequential* in combination with *opaque*. 25% stated they mostly focused on the task.

When we asked if the visualizations felt like ***themselves***, 42.5% of the participants did not feel like the *limbs* belonged to them, 50% felt only in specific moments like the instructor, 5% always felt, and 2.5% nearly felt like the instructor. 5% felt like they could control the instructor. On the contrary, they said feeling like the instructions are yourself is weird (7.5%). Of the participants, 32.5% only felt like the instructor in the *parallel* condition. P3 forgot to move in the *parallel* condition due to this feeling: *"That part, when you reminded me I should do it alongside, that was the part when I thought I was already doing it."* 5% felt like the instructor during longer interactions and 7.5% only at the beginning.

We also asked the participants for ***general suggestions***. 5% recommended improving where to look, and 5% to slow down the speed of the parallel instructions. Of the participants, 10% said that

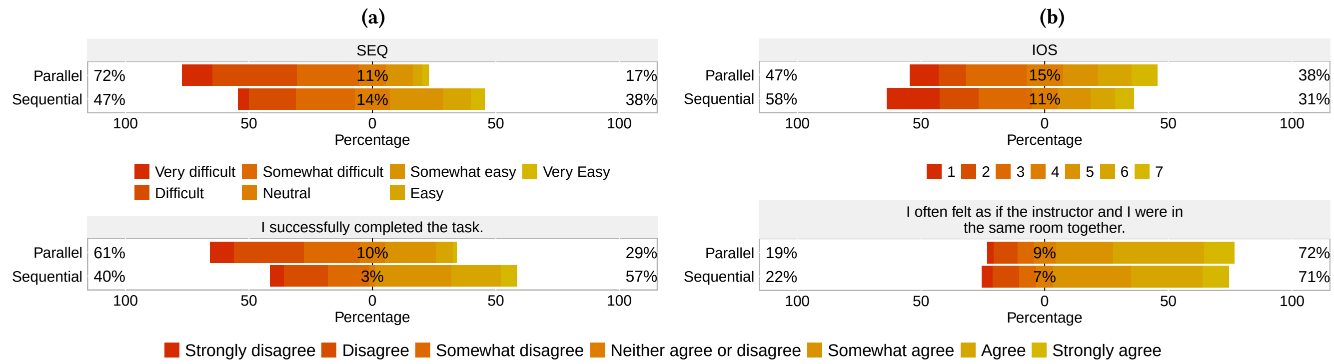


Figure 9: Likert scales of the SEQ and perceived performance (a) as well as the IOS and NMSPI subquestion (b) by *timing* showing an increase in results for SEQ and perceived performance for sequential compared to parallel as well as a decrease in results for the IOS and NMSPI subquestion.

it would be great to use the instructions in real life: for companies and *"for people to do things by themselves"* (P7).

Additionally, we also gave the participants the option to write comments at the end of each questionnaire part and took notes during the study. In the questionnaire, the participants also mentioned that they sometimes had to tilt their head to see specific instructions, and that they felt disconnected when the movement of the instructions was different from their own in the *parallel* condition. Despite the sound cue, we sometimes had to remind participants during the study to either start directly in the *parallel* or wait in the *sequential* condition. Thereby, it was visible that some participants had to actively stop themselves from starting immediately during the *sequential* condition. In 4 rounds, we encountered headset issues between conditions, which led to longer calibration periods.

5 Discussion and Guidelines

Our study results provide consistent evidence that the *timing* and *limb* visualizations, but not the *visibility*, influence performance, immediate retention, and user experience during first-person AR instructions. Regarding RQ1, the findings show that while *parallel timing* leads to higher embodiment and is overall faster, *sequential* improves comfort and memorization. For RQ2, we observed that the *hand* only *limbs* representation can improve comfort as well as correct order retention. Still, *arm* is more preferred among the people who noticed a difference between the visualizations. About RQ3, the empirical results show that *visibility* did not influence the measured variables in our study design. In the following, we will discuss our findings in detail.

5.1 Faster is not Always Better

Our results show that timing did not influence correctness during the first imitation, confirming that participants can complete a task accurately regardless of whether instructions are given in *parallel* or *sequentially*. However, when we consider the impact on time, immediate memory, and user experience, a more nuanced picture emerges. As participants had to watch the full instructions before starting the task in the *sequential* condition, *parallel* instructions resulted in a shorter total task duration. By contrast, *sequential* instructions were faster during the interface interaction itself. We attribute this finding to the participants needing to actively process, remember, and plan their actions while watching, rather than focusing passively and solely on copying in real time. In addition, as the participants first had to watch and then act in the *sequential* condition, they already had to recall the instructions once, leading to further memory reinforcement and therefore better retention. From a cognitive perspective, *sequential* instructions appeared to be easier and less mentally demanding. Participants reported a better understanding of the instructions and greater satisfaction with how they completed the task. This was reflected in performance: *sequential* instructions led to more accurate repetition from memory of order, values, and hand usage, again suggesting stronger memory consolidation. This is consistent with the assumption that memorization and, in the long run, learning take time [27, 59]. Interestingly, while the participants preferred *sequential* imitations, we still observed impatience with the interface in this condition, an observation that aligns with previous findings on first-person *sequential* instructions [55]. Therefore, it is still unclear how to effectively time instructions to prevent users from unintentionally starting the interaction too early. *Parallel* instructions, on the other hand, elicited a stronger sense of embodiment and co-presence, as participants felt socially closer to the instructor and more in the same room. However, this enhanced embodiment did not translate into better immediate retention and may even have contributed to increased discomfort. Qualitative feedback suggests that this effect might stem from moments of confusion about whether the virtual limbs were perceived as their own or as belonging to another person. They stated that the experience felt not entirely natural, which may have caused uneasiness. They also perceived the instructor performing worse in the *parallel* condition and expressed a preference for *sequential* instructions in the future. As we created a rather difficult task, some participants said in the interviews that they were disappointed in their own performance. This, combined with the blurred distinction between self and instructor in the *parallel* condition, may have caused participants to project their own performance onto the observed instructions.

These findings raise questions about the relationship between embodiment, social factors, memory, and task difficulty. While *parallel* instructions may increase embodiment and social factors, they

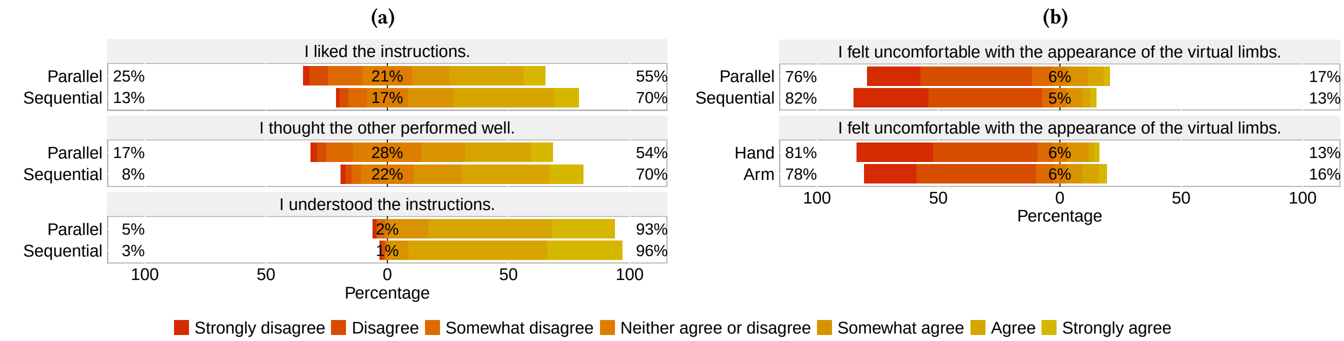


Figure 10: Likert scales of the questions regarding the quality of the instructions (a) and discomfort (b) by *timing* and *limbs* with *sequential* leading to higher ratings for the instructions questions as well as *parallel* and *arm* resulting in increased discomfort ratings.

do not necessarily support retention or a positive user experience. From a design perspective, our results indicate that the choice between *parallel* and *sequential* instructions should depend on the instructional goal. For tasks that must be completed quickly and where memorization is less critical, *parallel* instructions may be preferable. Conversely, when retention, accuracy, and comfort are priorities, *sequential* instructions appear more effective. As concurrent guidance optimizes immediate execution by encouraging cognitive offloading, delayed imitation requires and encourages users to memorize and recall, which supports retention. Based on these observations, we define the following design guideline:

Use *parallel timing* when speed, embodiment, social closeness, and co-presence matter more than retention, and *sequential* when memorization, and comfort are priorities.

5.2 Hands Support Retention

Our results show that *limb* representation had no significant effect on direct imitation performance, mental load, or instruction quality, suggesting that the dimensions of the virtual *limbs* did not interfere with participants' ability to accurately reproduce the demonstrated steps directly. However, we observed more nuanced effects when considering short-term memory-based performance. Specifically, *hand* representations improved participants' ability to recall the order of the instructions. This effect extended beyond the instructional sequence to the reproduction of participants' own prior actions, indicating that *hand*-based representations may facilitate stronger motor memory consolidation and self-referential encoding [53]. We attribute these effects to the fact that *hand* representations focus attention more precisely on the action being performed while minimizing obstruction and improving the view to the interface [11, 55]. Although participants were able to imitate actions equally well under both conditions, memorizing the sequence order later proved to be more challenging when using *arm* representations. This may relate to the embodiment experience: some participants reported that *arm* representations felt less comfortable and did not feel as if the limbs belonged to them. This reduced sense of ownership may have interfered with encoding and later recall. In some cases, participants even perceived the arms as belonging to another person, while at the same time leading to no improvement in social closeness. This could have created a subtle feeling of intrusion into their personal space and contributed to discomfort [41, 47]. Interestingly, while approximately half of the participants reported not noticing a difference between limb representations, those who did notice often preferred *arms* over *hands* during interviews. This may be explained by the *arms*' greater visual realism, better guiding properties, and increased visibility, which some participants found helpful for following along during the task.

Taken together, in our study setup, *hand* visualizations improved immediate order retention and comfort. Still, many participants preferred the *arm* representation in the interviews, suggesting that limb representation design should be adaptable. Different users may benefit from different representations depending on whether the goal is memorization, comfort, or preference. Allowing users to choose between *hand*- or *arm*-based representations or dynamically adjusting representations based on task demands could optimize both user experience and retention outcomes. Such customization could also accommodate individual differences in embodiment perception, ensuring that virtual limbs are perceived either as part of the user's own body or as those of another person, depending on the user and desired interaction context. These findings lead to the following design recommendation:

***Hands*-only representations aid immediate order retention and comfort, while some users may prefer *arms* for guidance.**

5.3 Invisible Visibility

Our findings show that limb visibility did not significantly influence performance, immediate retention, embodiment, or user experience in our setup. This outcome stands in contrast to prior research suggesting that reducing visual occlusion can improve task execution [11, 55], that *transparent* limbs can make tasks feel easier [65, 67], and that *opaque* visualizations may enhance body ownership [20]. A plausible explanation lies in the nature of the stepwise procedural motor task, which required participants to focus on and remember interface elements. In this setting, differences in limb *visibility* may not have carried enough perceptual weight to affect behavior. Interview data support this interpretation: although *transparent visibility* improved the view of interface elements, this benefit did not lead to measurable improvements and even created confusion between virtual and physical elements. Conversely, while *opaque* limbs sometimes covered parts of the interface, they appeared to support focus for some participants. This suggests that improved *visibility* alone does not guarantee more effective guidance. The lack of an embodiment effect also indicates that, while avatar transparency can cause dehumanization [51] in first-person AR instruction contexts, ownership is shaped more by spatial alignment, motion synchrony, and viewpoint consistency than by the transparency of the limbs. This implies that realistic *transparent* limb visualizations can remain valuable without compromising body ownership. Importantly, over half of the participants did not notice the *visibility* manipulation at all, aligning with the phenomenon of attention tunneling during cognitively demanding tasks [63]. When cognitive load is high and instructional demands are immediate, peripheral visual properties may be overlooked. Among those who did notice a difference, preferences were approximately evenly split with only a slight leaning towards *opaque visibility*, indicating no clear tendency among users.

In summary, these results suggest that limb *visibility* alone is unlikely to affect performance, immediate retention, or user experience in tightly guided first-person AR instruction scenarios. Instead, *visibility* appears to be a matter of personal preference rather than functional necessity. Consequently, systems should allow users to tailor transparency to their own comfort and perceptual needs, leading to the following design implication:

***Visibility* should be individually adaptable, even though it may not influence performance, immediate retention, or user experience.**

6 Limitations and Future Work

Our study provides valuable insights into the effects of instruction timing, limb representation, and visualization style on task performance, immediate retention, and memory. In the following, we qualify the scope of our claims and the limitations that result from this. Technical advances will take some limitations away. Several participants noted that the limited field of view of the HMD made it difficult to focus simultaneously on the task interface and the displayed limbs. New headset technology, particularly wider field-of-view devices, could improve the visibility of virtual limbs and reduce cognitive load during task execution.

6.1 One Task does not Fit All

To ensure reproducibility and increase internal validity, we intentionally conducted the study in a fixed laboratory setup, focusing on procedural motor memory based on a repetitive two-hand task. The aim was to reveal differences between conditions while using a counterbalanced within-subject design. With this well-defined apparatus that is described in detail, we ease the recreation of the experiment. The stationary setting allowed us to study the effects on short-term memory and performance experimentally, providing high internal validity. We expect that our results transfer to other settings and tasks, but our empirical findings are limited to this setup and study design, constraining the external validity.

6.2 Every Body is Different

To optimize embodiment and reduce potential biases, we allowed participants to select a limb shape and skin tone from a predefined set that matched their own appearance to a large extent. This approach was informed by prior work on embodiment and inclusion [4, 21, 34], and was intended to minimize behavioral changes and perceptual distraction caused by visual dissimilarity. Using the predefined set comes with limitations with regard to the realism of the virtual limbs. In our setting, we could not recreate participants' individual physical characteristics, such as tattoos, body hair, jewelry, and age-related features, precisely. We also could not recreate the movement behavior that exactly matched their movement. These factors could potentially limit the identification with the avatar and, in turn, the embodiment effects.

6.3 Towards Personal Helping Hands

Going beyond the highly controlled task in our study, future work could extend the experimental work to evaluate tasks with varying difficulty levels, including more naturalistic settings where distractions, spatial constraints, locomotion, and environmental variability are present. This could be combined with further exploration on how pacing can be communicated and enforced. Since our work focused on procedural learning, examining whether the same patterns hold for declarative learning could provide a richer understanding of how different memory systems respond to AR-based instructions. Incorporating more personalized avatar generation or 3D scans could further improve future systems through more realistic avatar representations. Varying transparency systematically beyond the two levels could reveal optimal levels for task focus, limb shape, and interface visibility. A bigger research challenge, that goes beyond instructional XR systems, is to experimentally explore how embodiment interacts with memory and user satisfaction.

7 Conclusion

In this work, we systematically investigated the effects of instruction *timing*, *limb* representation, and *visibility* on task performance, immediate memory retention, and user experience during a two-handed manual task. To explore these factors, we developed a custom AR application and a physical machine interface that showed first-person instructional animations. In a controlled experiment, 40 participants imitated the XR instructions and subsequently repeated the full sequence from memory. Our findings provide the following key insights: While none of the manipulated variables significantly affected the accuracy of the first imitation, performing the instructions in *parallel* with the animation resulted in faster overall task completion, a stronger sense of social closeness, co-presence, and embodiment. In contrast, performing them *sequentially* enhanced memory retention and increased subjective comfort. Representing the *limbs* with *hands* only positively influenced immediate order retention and comfort, although many participants reported a preference for the full-*arm* visualization. Furthermore, *visibility* did not significantly affect the measured variables. Together, these results demonstrate that the design of first-person AR instruction systems must carefully balance efficiency, memory retention, and user experience. We encourage future work in this area to analyze different use cases and the visualization degrees of the measured variables to further generalize these findings. We are convinced that this study represents an important step toward ubiquitous first-person AR instruction systems that not only enable users to operate devices autonomously but also support memorization.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 517713394 and the LOEWE initiative (Hesse, Germany) within the emergenCITY center [LOEWE/1/12/519/03/05.001(0016)/72].

References

[1] Arthur Aron, Elaine N. Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology* 63, 4 (1992), 596–612. doi:10.1037/0022-3514.63.4.596

[2] Domna Banakou, Sameer Kishore, and Mel Slater. 2018. Virtually Being Einstein Results in an Improvement in Cognitive Task Performance and a Decrease in Age Bias. *Frontiers in Psychology* 9 (June 2018). doi:10.3389/fpsyg.2018.00917

[3] Albert Bandura. 1969. Social-Learning Theory of Identificatory Processes. In *Handbook of Socialization Theory and Research*. 262.

[4] Kristin A. Bartlett and Claudia Krogmeier. 2025. Observations on Virtual Reality Avatar Alignment with Research Participants' Skin Tone and Gender. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 288–295. doi:10.1109/VR59515.2025.00053

[5] Frank Biocca and Chad Harms. 2003. Guide to the Networked Minds Social Presence Inventory v. 1.2. (2003).

[6] Frank Biocca, Chad Harms, and Judee K Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators & virtual environments* 12, 5 (2003), 456–480. doi:10.1162/105474603322761270

[7] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[8] Rudy Bretz. 1971. *The Selection of Appropriate Communication Media for Instruction: A Guide for Designers of Air Force Technical Training Programs.* Technical Report. Communications Department, Rand, 1700 Main Street, Santa Monica, California 90406.

[9] Mollie E. Brooks, Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Machler, and Benjamin M. Bolker. 2017. glmmTMB Balances Speed and Flexibility among Packages for Zero-Inflated Generalized Linear Mixed Modeling. *The R journal* 9, 2 (Dec. 2017), 378–400. doi:10.32614/RJ-2017-066

[10] Jerome Bruner. 1999. Folk Pedagogies. In *Learners & Pedagogy*. SAGE, 4–20.

[11] Yuanzhi Cao, Xun Qian, Tianyi Wang, Rachel Lee, Ke Huo, and Karthik Ramani. 2020. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. doi:10.1145/3313831.3376688

[12] Jeroen Ceyssens, Bram van Deurzen, Gustavo Rovelo Ruiz, Kris Luyten, and Fabian Di Fiore. 2024. The Art of Timing: Effects of AR Guidance Timing on Speed Control. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 31–40. doi:10.1109/ISMAR62088.2024.00017

[13] Morgan Le Chenechal, Thierry Duval, Valerie Gouranton, Jerome Royan, and Bruno Arnaldi. 2016. Vishnu: Virtual Immersive Support for HelpiNg Users an Interaction Paradigm for Collaborative Remote Guiding in Mixed Reality. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*. IEEE, Greenville, SC, USA, 9–12. doi:10.1109/3DCVE.2016.7563559

[14] Antonin Cheymol, Rebecca Fribourg, Anatole Lécuyer, Jean-Marie Normand, and Ferran Argelaguet. 2023. Beyond My Real Body: Characterization, Impacts, Applications and Perspectives of "Dissimilar" Avatars in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (Nov. 2023), 4426–4437. doi:10.1109/TVCG.2023.3320209

[15] Subramanian Chidambaram, Hank Huang, Fengming He, Xun Qian, Ana M Villanueva, Thomas S Redick, Wolfgang Stuerzlinger, and Karthik Ramani. 2021. ProcessAR: An Augmented Reality-Based Tool to Create in-Situ Procedural 2D/3D AR Instructions. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 234–249. doi:10.1145/3461778.3462126

[16] Subramanian Chidambaram, Sai Swarup Reddy, Matthew Rumple, Ananya Ipsita, Ana Villanueva, Thomas Redick, Wolfgang Stuerzlinger, and Karthik Ramani. 2022. EditAR: A Digital Twin Authoring Environment for Creation of AR/VR and Video Instructions from a Single Demonstration. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Singapore, Singapore, 326–335. doi:10.1109/ISMAR55827.2022.00048

[17] Rune Haubo B. Christensen. 2018. Cumulative Link Models for Ordinal Regression with the R Package Ordinal. *Submitted in J. Stat. Software* 35 (2018), 1–46.

[18] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch. 2016. Workspace Awareness in Collaborative AR Using HMDs: A User Study Comparing Audio and Visual Notifications. In *Proceedings of the 7th Augmented Human International Conference 2016 (AH '16)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/2875194.2875204

[19] Christopher Clarke, Jingnan Xu, Ye Zhu, Karan Dharamshi, Harry McGill, Stephen Black, and Christof Lutteroth. 2023. FakeForward: Using Deepfake Technology for Feedforward Learning. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. doi:10.1145/3544548.3581100

[20] Diane Dewez, Ludovic Hoyet, Anatole Lecuyer, and Ferran Argelaguet. 2022. Do You Need Another Hand? Investigating Dual Body Representations During Anisomorphic 3D Manipulation. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (May 2022), 2047–2057. doi:10.1109/TVCG.2022.3150501

[21] Tiffany D. Do, Juanita Benjamin, Camille Isabella Protko, and Ryan P. McMahan. 2024. Cultural Reflections in Virtual Reality: The Effects of User Ethnicity in Avatar Matching Experiences on Sense of Embodiment. *IEEE Transactions on Visualization and Computer Graphics* 30, 11 (Nov. 2024), 7408–7418. arXiv:2407.10412 [cs] doi:10.1109/TVCG.2024.3456196

[22] Maximilian Dürr, Rebecca Weber, Ulrike Pfeil, and Harald Reiterer. 2020. EGuide: Investigating Different Visual Appearances and Guidance Techniques for Egocentric Guidance Visualizations. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, Sydney NSW Australia, 311–322. doi:10.1145/3374920.3374945

[23] Omid Fakourfar, Kevin Ta, Richard Tang, Scott Bateman, and Anthony Tang. 2016. Stabilized Annotations for Mobile Remote Assistance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1548–1560. doi:10.1145/2858036.2858171

[24] Shihui Feng, Shuming Qiu, David Gibson, and Dirk Ifenthaler. 2023. The Effect of Social Closeness on Perceived Satisfaction of Collaborative Learning. In *Open and Inclusive Educational Practice in the Digital World*, Dirk Ifenthaler, Demetrios G. Sampson, and Pedro Isaías (Eds.). Springer International Publishing, Cham, 101–113. doi:10.1007/978-3-031-18512-0_7

[25] Rebecca Fribourg, Nami Ogawa, Ludovic Hoyet, Ferran Argelaguet, Takuji Narumi, Michitaka Hirose, and Anatole Lécuyer. 2021. Virtual Co-Embodiment: Evaluation of the Sense of Agency While Sharing the Control of a Virtual Body among Two Individuals. *IEEE Transactions on Visualization and Computer Graphics* 27, 10 (Oct. 2021), 4023–4038. arXiv:1911.03166 [cs] doi:10.1109/TVCG.2020.2999197

[26] Jason Friedman and Maria Korman. 2019. Observation of an Expert Model Induces a Skilled Movement Coordination Pattern in a Single Session of Intermittent Practice. *Scientific Reports* 9, 1 (March 2019), 4609. doi:10.1038/s41598-019-40924-9

[27] Gökhan Gönül, Nike Tsalas, and Markus Paulus. 2021. The Effect of Time Pressure on Metacognitive Control: Developmental Changes in Self-Regulation and Efficiency during Learning. *Metacognition and Learning* 16, 2 (Aug. 2021), 459–483. doi:10.1007/s11409-021-09262-y

[28] Sebastian Günther, Sven Kratz, Daniel Avrahami, and Max Mühlhäuser. 2018. Exploring Audio, Visual, and Tactile Cues for Synchronous Remote Assistance. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference (PETRA '18)*. Association for Computing Machinery, New York, NY, USA, 339–344. doi:10.1145/3197768.3201568

[29] 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*. ACM, Geneva Switzerland, 1–4. doi:10.1145/2875194.2875237

[30] 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 (Oct. 2006), 904–908. doi:10.1177/154193120605000909

[31] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, 139–183. doi:10.1016/S0166-4115(08)62386-9

[32] Keiichi Ihara, Kyzyl Monteiro, Mehrad Faridan, Rubaiat Habib Kazi, and Ryo Suzuki. 2025. Video2MR: Automatically Generating Mixed Reality 3D Instructions by Augmenting Extracted Motion from 2D Videos. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 1548–1563. doi:10.1145/3708359.3712159

[33] Rachel L. Kelly and Lewis Wheaton. 2013. Differential Mechanisms of Action Understanding in Left and Right Handed Subjects: The Role of Perspective and Handedness. *Frontiers in Psychology* 4 (Dec. 2013). doi:10.3389/fpsyg.2013.00957

[34] Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. 2013. Drumming in Immersive Virtual Reality: The Body Shapes the Way We Play. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (April 2013), 597–605. doi:10.1109/TVCG.2013.29

[35] Eishi Kim, Vsevolod Peysakhovich, and Raphaëlle N. Roy. 2021. Impact of Communication Delay and Temporal Sensitivity on Perceived Workload and Teleoperation Performance. In *ACM Symposium on Applied Perception 2021*. ACM, Virtual Event France, 1–8. doi:10.1145/3474451.3476233

[36] Martin Kocur, Florian Habler, Valentin Schwind, Paweł W. Woźniak, Christian Wolff, and Niels Henze. 2021. Physiological and Perceptual Responses to Athletic Avatars While Cycling in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–18. doi:10.1145/3411764.3445160
[37] Martin Kocur, Lukas Jackermeier, Valentin Schwind, and Niels Henze. 2023. The Effects of Avatar and Environment on Thermal Perception and Skin Temperature in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3544548.3580668
[38] Daiki Kodama, Takato Mizuho, Yuji Hatada, Takuji Narumi, and Michitaka Hirose. 2023. Effects of Collaborative Training Using Virtual Co-embodiment on Motor Skill Learning. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (May 2023), 2304–2314. doi:10.1109/TVCG.2023.3247112
[39] Matthew Lakier and Daniel Vogel. 2022. Investigating the Effect of Superpower Priming and Representation on Virtual Reality Interaction. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction (SUI '22)*. Association for Computing Machinery, New York, NY, USA, 1–2. doi:10.1145/3565970.3567704
[40] Nathanaël Lambert and Matthew Lakier. 2024. Towards an Avatar Customization System for Semi-realistic Ethnically-diverse Virtual Reality Avatars. In *Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology (VRST '24)*. Association for Computing Machinery, New York, NY, USA, 1–2. doi:10.1145/3641825.3689497
[41] Eva Lampen, Jannes Lehwald, and Thies Pfeiffer. 2020. A Context-Aware Assistance Framework for Implicit Interaction with an Augmented Human. In *Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications*, Jessie Y. C. Chen and Gino Fragomeni (Eds.). Vol. 12191. Springer International Publishing, Cham, 91–110. doi:10.1007/978-3-030-49698-2_7
[42] Ziyi Liu, Zhengzhe Zhu, Enze Jiang, Feichi Huang, Ana M Villanueva, Xun Qian, Tianyi Wang, and Karthik Ramani. 2023. InstruMentAR: Auto-Generation of Augmented Reality Tutorials for Operating Digital Instruments Through Recording Embodied Demonstration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. doi:10.1145/3544548.3581442
[43] Matteo Martini, Daniel Perez-Marcos, and Maria V Sanchez-Vives. 2013. What Color Is My Arm? Changes in Skin Color of an Embodied Virtual Arm Modulates Pain Threshold. *Frontiers in human neuroscience* 7 (2013), 438. doi:10.3389/fnhum.2013.00438
[44] Marta Matamala-Gomez, Ana M. Diaz Gonzalez, Mel Slater, and Maria V. Sanchez-Vives. 2019. Decreasing Pain Ratings in Chronic Arm Pain Through Changing a Virtual Body: Different Strategies for Different Pain Types. *The Journal of Pain* 20, 6 (June 2019), 685–697. doi:10.1016/j.jpain.2018.12.001
[45] Anjela Mayer, Augustin Rungeard, Jean-Rémy Chardonnet, Polina Häfner, and Jivka Ovtcharova. 2023. Immersive Hand Instructions in AR – Insights for Asynchronous Remote Collaboration on Spatio-Temporal Manual Tasks. In *2023 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. IEEE, Gammarth, Tunisia, 1–6. doi:10.1109/CIVEMSA57781.2023.10231018
[46] George A. Miller. 1956. The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63, 2 (1956), 81–97. doi:10.1037/h0043158
[47] Mark Roman Miller, Hanseul Jun, Fernanda Herrera, Jacob Yu Villa, Greg Welch, and Jeremy N. Bailenson. 2019. Social Interaction in Augmented Reality. *PLOS ONE* 14, 5 (May 2019), e0216290. doi:10.1371/journal.pone.0216290
[48] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. Association for Computing Machinery, New York, NY, USA, 405–415. doi:10.1145/2807442.2807497
[49] Mai Otsuki, Tzu-Yang Wang, and Hideaki Kuzuoka. 2022. Assessment of Instructor's Capacity in One-to-Many AR Remote Instruction Giving. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*. ACM, Tsukuba Japan, 1–5. doi:10.1145/3562939.3565631
[50] Tabitha C. Peck and Mar Gonzalez-Franco. 2021. Avatar Embodiment. A Standardized Questionnaire. *Frontiers in Virtual Reality* 1 (Feb. 2021). doi:10.3389/frvir.2020.575943
[51] Tabitha C. Peck, Jessica J. Good, Austin Erickson, Isaac Bynum, and Gerd Bruder. 2022. Effects of Transparency on Perceived Humanness: Implications for Rendering Skin Tones Using Optical See-Through Displays. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (May 2022), 2179–2189. doi:10.1109/TVCG.2022.3150521
[52] Henning Pohl and Aske Mottelson. 2022. Hafnia Hands: A Multi-Skin Hand Texture Resource for Virtual Reality Research. *Frontiers in Virtual Reality* 3 (May 2022). doi:10.3389/frvir.2022.719506
[53] T. B. Rogers, N. A. Kuiper, and W. S. Kirker. 1977. Self-Reference and the Encoding of Personal Information. *Journal of Personality and Social Psychology* 35, 9 (1977), 677–688. doi:10.1037/0022-3514.35.9.677
[54] Jeff Sauro and Joseph S. Dumas. 2009. Comparison of Three One-Question, Post-Task Usability Questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1599–1608. doi:10.1145/1518701.1518946
[55] Clara Sayffaerth, Annika Köhler, Julian Rasch, Albrecht Schmidt, and Florian Müller. 2025. Through the Expert's Eyes: Exploring Asynchronous Expert Perspectives and Gaze Visualizations in XR. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Computer Society, 1311–1321. doi:10.1109/ISMAR67309.2025.00136
[56] Alexander Schäfer, Gerd Reis, and Didier Stricker. 2022. A Survey on Synchronous Augmented, Virtual, and Mixed Reality Remote Collaboration Systems. *Comput. Surveys* 55, 6 (Dec. 2022), 116:1–116:27. doi:10.1145/3533376
[57] Valentin Schwind, Sven Mayer, Alexandre Comeau-Vermeersch, Robin Schweigert, and Niels Henze. 2018. Up to the Finger Tip: The Effect of Avatars on Mid-Air Pointing Accuracy in Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM, Melbourne VIC Australia, 477–488. doi:10.1145/3242671.3242675
[58] S. R. Searle, F. M. Speed, and G. A. Milliken. 1980. Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. *The American Statistician* 34, 4 (Nov. 1980), 216–221. doi:10.1080/00031305.1980.10483031
[59] Reza Shadmehr and Henry H Holcomb. 1997. Neural Correlates of Motor Memory Consolidation. *Science* 277, 5327 (1997), 821–825. doi:10.1126/science.277.5327.821
[60] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V. Sanchez-Vives. 2008. Towards a Digital Body: The Virtual Arm Illusion. *Frontiers in Human Neuroscience* 2 (Aug. 2008). doi:10.3389/neuro.09.006.2008
[61] Haruto Takita, Kenta Hashiura, Yuji Hatada, Daiki Kodama, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2025. Do We Still Need Human Instructors? Investigating Automated Methods for Motor Skill Learning in Virtual Co-Embodiment. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (May 2025), 2455–2463. doi:10.1109/TVCG.2025.3549540
[62] Haruto Takita, Yuji Hatada, Takuji Narumi, and Michitaka Hirose. 2023. Effects of Virtual Co-embodiment on Declarative Memory-Based Motor Skill Learning. In *ACM Symposium on Applied Perception 2023 (SAP '23)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3605495.3605797
[63] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative Effectiveness of Augmented Reality in Object Assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Ft. Lauderdale Florida USA, 73–80. doi:10.1145/642611.642626
[64] B. Tarr, M. Slater, and E. Cohen. 2018. Synchrony and Social Connection in Immersive Virtual Reality. *Scientific Reports* 8, 1 (Feb. 2018), 3693. doi:10.1038/s41598-018-21765-4
[65] Santawat Thanyadit, Parinya Punpongsanon, Thammathip Piumsomboon, and Ting-Chuen Pong. 2022. XR-LIVE: Enhancing Asynchronous Shared-Space Demonstrations with Spatial-temporal Assistive Toolsets for Effective Learning in Immersive Virtual Laboratories. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (April 2022), 136:1–136:23. doi:10.1145/3512983
[66] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 161–174. doi:10.1145/3332165.3347872
[67] Michiel Van Veldhuizen and Xubo Yang. 2021. The Effect of Semi-Transparent and Interpenetrable Hands on Object Manipulation in Virtual Reality. *IEEE Access* 9 (2021), 17572–17583. doi:10.1109/ACCESS.2021.3050881
[68] Vighnesh Vatsal and Guy Hoffman. 2017. Wearing Your Arm on Your Sleeve: Studying Usage Contexts for a Wearable Robotic Forearm. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 974–980. doi:10.1109/ROMAN.2017.8172421
[69] Ungyeon Yang and Gerard Jounghyun Kim. 2002. Implementation and Evaluation of "Just Follow Me": An Immersive, VR-Based, Motion-Training System. *Presence: Teleoperators and Virtual Environments* 11, 3 (June 2002), 304–323. doi:10.1162/105474602317473240
[70] Seraphina Yong, Leo Cui, Evan Suma Rosenberg, and Svetlana Yarosh. 2024. A Change of Scenery: Transformative Insights from Retrospective VR Embodied Perspective-Taking of Conflict With a Close Other. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. doi:10.1145/3613904.3642146
[71] Jingjing Zhang, Xiyao Jin, Han Tu, Hai-Ning Liang, Zhuying Li, and Xin Tong. 2025. "One Body, But Four Hands": Exploring the Role of Virtual Hands in Virtual Co-embodiment. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (May 2025), 3634–3644. doi:10.1109/TVCG.2025.3549883