Virtual Physical Task Training: Comparing Shared Body, Shared View and Verbal Task Explanation

Jens Reinhardt
reinhardtj@acm.org
Hamburg University of Applied Sciences
Hamburg, Germany

Marco Kurzweg
marco.kurzweg@beuth-hochschule.de
Beuth University of Applied Sciences Berlin
Berlin, Germany

Katrin Wolf
katrin.wolf@beuth-hochschule.de
Beuth University of Applied Sciences Berlin
Berlin, Germany

Figure 1: Screenshots of the different guidance methods: *shared-body, shared-view, verbal* (symbolized by the subtitles)

ABSTRACT

While video lectures are very successful in explaining theoretical aspects, Virtual Reality (VR) also offers the possibility to teach physical skills. While a large body of research focuses on virtual teaching explaining tasks during the task completion, it remains unclear what type of task instructions work best before task execution.

In this paper, we compare three virtual teaching methods: purely verbal explanation, explanation by observing the trainer's activity from exactly their perspective (shared-body), and explanation by observing the task execution in a shared view overlay within one's view.

Our quantitative results show that the purely verbal explanation can result in slower task replication for some tasks, e.g., path drawing in the air. Our qualitative results indicate that shared-body and shared view techniques work well but can be improved through additional verbal explanations. Moreover, task instruction might be hard to remember, which can be avoided through in-situ instructions or splitting the tasks into sub-tasks and letting the trainees replicate shorter physical actions.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; **Computer supported cooperative work**; *Interaction techniques.*

AHs '21, February 22–24, 2021, Rovaniemi, Finland

ACM ISBN 978-1-4503-8428-5/21/02…$15.00
https://doi.org/10.1145/3458709.3458989

KEYWORDS

virtual reality, virtual task demonstration, cooperative distant learning

ACM Reference Format:
Jens Reinhardt, Marco Kurzweg, and Katrin Wolf. 2021. Virtual Physical Task Training: Comparing Shared Body, Shared View and Verbal Task Explanation. In *Augmented Humans International Conference 2021 (AHs '21), February 22–24, 2021, Rovaniemi, Finland*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3458709.3458989

1 INTRODUCTION

Remote training becomes more and more important. It enables trainees to learn from experts without physically traveling, allowing them to learn over a distance without any travel effort. Such possibility is important during the Corona pandemic. Moreover, it can democratize education in the long term.

Virtual reality (VR) enables us to learn remotely, cooperatively, and independently of the trainee's or trainer's physical presence.

While video lectures have been shown to be a very good solution for remotely (or online) learning theoretical skills [4, 29], practical skills better work with a constructing learning "by doing" approach [18]. Thus, practical skill learning often requires a setup more complex than videos, e,g., to train manual actions in a 3D space using simulations. Such 3D simulations are used for medical training [20], dental training [24], training of machine operations [27], learning to weld [12], learning to drive [5, 19], or professional training, such as as a submarine training program [13]. These training systems usually use complex hardware interfaces that provide realistic haptic feedback and imitate real devices. Thus, they are very specialized, only applicable to a single training purpose, expensive, and not available for a broad range of users.

For physical or manual tasks, learning "by doing" is a commonly established learning approach, either guided through demonstration or instructions given before a learner does the task or simultaneously.

Previous research studied such simultaneous guidance presenting the demonstration of a task through the teacher from an egocentric perspective [6, 11, 28], from an exocentric perspective [3, 8, 16, 21], and through view sharing, where the learner sees what the teacher sees [8, 15, 17, 25]. Further research investigated how different kinds of task explanations influence the learner's task's completion time, accuracy, and task load through exploring purely verbal instruction [9, 14], verbal plus projected gesture [14], instruction via video [8, 9, 14, 15, 17], instruction by a real human helper [8], a virtual realistically represented trainer [9], an abstractly represented trainer [9], overlaying the real view of the user with virtual hands [1], or overlaying the virtual view with virtual arms [6].

While most research focused on skill learning with task explanation given during the performance of the task [6, 11, 14, 15, 21, 28], task interruptions [6–8], or learning effects over a longer period [14, 28], only very few research, e.g. on ballet dancing training used task explanation before executing the demonstrated motion [16]. Therefore, it remains unclear how different pre-task explanations influence task completion time, accuracy, and task load.

In this project, we aim to reduce the described research gap through investigating in a user study three common explanation types used to teach a manual task in VR and given by a trainer before a learner examines the learning task. The explanation types we used were purely verbal explanations (*verbal*), observing the task demonstration from the egocentric perspective of the trainer (*shared-body*), and observing the task demonstration from the shared view of the trainer as an image in image projected overlay displayed in the field of view of the learner (*shared-view*). To get generalizable results, we tested each explanation type across three tasks: a *3D pointing task (oil filling)*, a *3D steering law task (painting)*, and a *bi-manual task (wheel control)*.

After presenting the findings of our experiment, we will deliver design recommendations for virtual physical task training.

2 RELATED WORK

Research related to learning physical tasks using VR to be learned "by doing" is based on task demonstration in VR, which can be realized through various task demonstration techniques. Visual task demonstrations can be captured from different viewpoints of the trainer.

2.1 Demonstration Techniques

Research on techniques on how a learner should be supported to learn and perform tasks through different ways of demonstration differ in how communication is made to explain the activity [14], in designs of instructor or trainer [9, 23], in the technology used for instruction [7], or in setups of cameras for sharing the view of one or all of the co-workers [8].

2.1.1 Demonstrations & Explanations. Kirk and Fraser investigated how voice-only compared to voice-plus-gesture, projected as real video on the work surface, affected assembly speed and assembly accuracy measured during instruction and after instruction [14]. They did not find any significant difference between voice-only and voice-plus-gesture during the instruction.

George et al. investigated the effect of instructor design on social presence and performance for virtual training [9]. They compared an avatar, webcam stream, rendered video-capturing of the instructor on a flat surface in VR, and sound-only as instructor representations for two tasks, a memory task, and an object selection task. While they did not find any significant differences between the different tasks, they found that the webcam and avatar representations yielded a high social presence but resulted in the VR user's lower performance. Their results show that users felt that the sound-only representation condition made them work harder. This is in alignment with their findings of lower performance in the visual representations.

Funk et al. compared for assembly tasks the following instruction techniques: HMD instructions, tablet instructions, in-situ projected instructions, and paper instructions as baseline [7]. Study results show that assembling using in-situ projection is significantly faster, and the localization of positions with HMDs is significantly slower. Furthermore, participants made fewer mistakes and had a lower perceived cognitive load during in-situ instructions than HMD instructions.

Rzayev et al. compared a real-world guide with a realistic, abstract, and audio-only representation of a virtual guide in a virtual museum [23]. Their qualitative results and the participants' feedback show that although an audio-only representation of a virtual guide helps to focus on the content, it can reduce the guide's connection.

2.1.2 View Sharing. View sharing is a common way to work and learn together by giving instructions through a live video captured with a standard camera or an HMD.

Fussel et al. compared for pairs of participants performing robot assembly tasks five camera positioning conditions: Side-by-side, head-mounted camera, scene-scene camera, scene-plus head cameras, and audio-only [8]. Study results show that the assembling quality was rated highest when participants worked side by side with the scene camera and lowest under pure audio and head camera conditions, while the self-assessed ability to help the workers and the communication efficiency of the pairs was highest in side-by-side conditions but significantly higher with the scene camera than in audio-only conditions.

Amores et al. developed *ShowMe*, which enables a remote user to be immersed in another user's first-person's point of view and offers demonstration through three-dimensional hand gestures and voice [1]. The remote expert performs hand gestures that are superimposed in real-time onto the novice user's immediate environment. The novice sees their own hands and the hands of the expert in real-time in their immediate physical surroundings. The system has not been evaluated.

View sharing was also done to work and learn together in VR. Valin et al. compared shared viewpoints and telepointers for collaborative tasks in VR. They found that viewpoint sharing enables effective collaboration and is more effective than telepointers for some tasks [26]. Chellali et al. explored changeable view sharing for an object co-manipulation task in VR and found that shared viewpoints were used spontaneously to optimize communication and to reduce collaborative efforts [2]. Pouliquen-Lardy et al. showed for a virtual collaborative object moving task that the manipulator's point of view was preferred when the participants took a perspective [22].

2.2 Demonstration Viewpoints

A large body of research for training physical tasks focuses on demonstrating the required movements by providing visual demonstration through showing the correct movement using different viewpoints.

2.2.1 Egocentric Point of View. Here research on instructional visualizations that represent a continuous demonstration of a movement that a learner should follow using the first-person viewpoint is presented.

Hoang et al. investigated *Onebody*, a VR training system for remote posture demonstration during sports or physical activity training using the first-person perspective and allowing the user to step into the instructor's body [11]. A comparison with existing techniques, including pre-recorded video, video conferencing, and third-person view in VR, indicated that the instructor's first-person perspective offers better posture accuracy in delivering movement instructions.

Yang and Kim proposed the virtual motion training system *Just Follow Me* [28]. A trainee viewed the trainer's motion (tracked as triaxial coordinate structure) from the egocentric perspective with a VR head-mounted display (HMD), traced, and followed a 3-D trajectory of the ghost by using their hands. A user study showed that the *Just Follow Me* could achieve transfer and learning effect as effective as traditional video-based learning media. They found that the VR environment did not significantly negatively affect the performance and its application to the real, video-based situation.

Dürr et al. investigated different egocentric visualizations and demonstration methods for explaining mid-air arm movements for a musical conducting task, including abstract and realistic visualization [6]. For different visualizations, abstract visualization was compared with the realistic visualization of the human arms. The demonstration methods used are the continuously moving guidance visualization, keyframes, and the guidance visualization pauses at each keyframe until the user reaches it. In contrast to the continuous demonstration method, postures within the movement between two keyframes were not visualized with both keyframe methods. Evaluation results show a higher movement accuracy for a realistic than an abstract visualization of the human arm. Moreover, their results suggest that visualizations with an abstract shape and a demonstration technique that visualizes important postures should not pause at essential postures.

2.2.2 Exocentric Point of View. Research investigating the exocentric view on trainer's demonstrated postures and movements, focus primarily on the feedback to support trainees in learning motor tasks.

Nozawa et al. investigated a VR ski training system enabling users to learn by visualizing the trainer's motion and additional graphs to visualize the angle of feet compared to the expert in the third-person view of the trainee [21]. Study results show that reduced visualizations were often more helpful than displaying all available information, like a graph comparing foot angles or the trainer leaving behind transparent avatars that can be passed through. The latter caused distraction among users since focus needed to be split between the transparent avatars and the trainer's main avatar.

Kyan et al. developed a system for training dancing ballet in a VR cave [16]. The egocentric perspective is used to observe an avatar representation of a virtual trainer by the trainee and watch the feedback provided in an immersive 3D environment. To be able to judge their dance movements compared to the trainer, two different exocentric modes of visualization were provided by this system. To visualize the trainer's instructions, a mode in which the trainer and the trainee are displayed side by side, and an overlay mode in which the movements of the trainer and the user are superimposed, were used. The system also allowed to review the own dance performance from different angles to improve the training and detect mistakes in the movement's execution. In a study, the recognition performance of isolated gesture recordings was evaluated but not any viewpoint effects.

Chua et al. compared different visualization techniques to improve Tai Chi learning from the third-person point of view. They included multiple copies of the virtual teacher around the learner, two designs that superimposed the learner's avatar over the virtual teacher's avatar, and a virtual teacher avatar direct in front of the learner [3]. Study results showed no significant differences in the performance imitating the teacher's movement for the different visualizations, but participants found the superimposed design more complicated than the other designs.

2.3 Summary

Demonstration in VR used for learning motor tasks uses different demonstration techniques, including text-based, video, or life-demo instructions. The latter two can vary between egocentric and exocentric view.

While the exocentric vision works well for demonstrating whole-body motions ([3, 16, 21]), it can lead to occlusion by the body of the trainer when demonstrating manual tasks, where the egocentric view give the most natural perspective on the demonstrated task ([6, 11, 28]

Previous studies on task demonstration techniques found no significant difference for verbal compared to verbal plus gesture demonstration techniques [14]. Other works suggest that sound-only representations led to a higher performance than avatar and webcam stream [9]. In contrast, other work suggests that audio-only led to the worst performance compared to other video-based demonstration techniques [8].

For manual tasks, to our best knowledge, different demonstration techniques (e.g., verbal instructions versus video-like view sharing versus immersive body-sharing) have not been compared for egocentric task demonstration.

3 EXPERIMENT

Our study used a machine hall as an exemplary virtual environment as their physical work is commonly done. Here, we compared for

Figure 2: Machine hall used as exemplary virtual environment where physical work is commonly done.

manual/motor task explanation three common egocentric *demonstration techniques* (verbal instructions versus video-like view sharing versus immersive body-sharing). To see how well the techniques help trainees to mimic demonstrated physical tasks, we measured *task completion time (TCT)*, *accuracy*, and *task load* while three typical tasks of varying challenges were mimicked: an *3D pointing task (filling in oil with a can)*, an *3D steering law task (painting)*, and a *bi-manual task (opening a pressure valve through wheel rotation)*.

3.1 Study Design

The study followed a 3x3 mixed design with the independent between-subject variable *demonstration technique* (*verbal*, *shared-view*, and *shared-body*), and the independent within-subject variable *task* (*3D pointing task (oil filling)*, *3D steering law task (painting)*, and *bi-manual task (wheel control)*). The dependent variables were *TCT*, *task accuracy*, and *task load*.

3.2 Participants

The experiment was conducted with 36 participants (13 females, 22 males, 1 prefer not to say the gender) aged between 20 and 50 years and an average age of 27.42 years ($SD = 6.45$), recruited at our university campus and via e-mail mailing list. The participants were distributed evenly among the three groups of the independent between-subject variable *demonstration technique*.

3.3 Tasks

We aim at investigating effects of different *task demonstration techniques* across tasks to get generalizable results. Hence, we have chosen tasks that fundamentally differ in type and complexity (pointing versus steering, one-manual versus bi-manual) and the structure of task load (especially physical and temporal). Thus, our three *tasks* cover a good range of manual tasks.

A *3D pointing task* was represented through filling oil into an engine as here. A virtual can had to be held above an aperture with 3D accuracy so that the oil does now flow besides. A the *3D steering law task* was represented by copying a drawn line as here not only one 3D point has to be met but also a 3D path. Last, a bi-manual task was provided through a wheel that had to be rotated counter-clockwise using both hands.

3.3.1 3D Pointing Task. For the *3D pointing task*, participants had to refill oil into the virtual replica of a steam engine, see Figure 3. For this purpose, a virtual oilcan has to be grasped, located on a virtual table to the participant's right. The participants had to move this oilcan over a steam engine's glass cylinder and fill oil into the cylinder. This was realized by an arm motion that tilted the can, analogous to pouring out liquids with a can in the real world. The cylinder had a red filling line mark. Participants were told to fill as much oil as required so that the oil reached the marked line. The participants could tilt the oilcan as often as they wanted to fill up oil. If the can was held horizontally to the ground, no more oil flowed, and the task was finished, which the participants had been told before they started that task.

3.3.2 3D Steering Law Task. For the *3D steering law task*, the participants saw a white line in 3D space, which they should copy through drawing a green line at the exact as possible same position. The participants were instructed to trace this line by pressing the index trigger[1] on the controller with their right index finger. As a visual aid where the drawn line was positioned in 3D space, a green dot was displayed directly in front of the index finger of the participants' virtual hand. From here, participants could start drawing. The task ended by releasing the right index trigger.

3.3.3 Bi-Manual Task. Releasing pressure from a steam engine had to be done by rotating a wheel of a steam engine's virtual replica, see Figure 3. The rotary wheel had to be grasped with both hands. To clarify the grasping and to simplify the interaction, we could display different outlines on the wheel. We displayed a thin yellow outline around the wheel when one hand was within the wheel's range, a thin green outline when both hands were within the range of the wheel. Participants were told they had to turn this wheel counterclockwise as far as it stops to open the steam engine's valve. The color of the wheel then changed to yellow to show that the valve was entirely open. The participants had to monitor a manometer on the steam engine while holding the wheel in place and release the wheel when the pressure needle pointed at 12 o'clock. When this pressure level was reached, the rotary wheel should be released. The release marked the task to be finished.

According to our experiment design, each participant had to complete the three tasks as fast and as accurately as possible in VR.

3.4 Demonstration Techniques

When participants were in the VR machine hall (see Figure 3), we explored the three task *demonstration techniques*: *verbal* explanation, task demonstration with *shared-view*, and task demonstration with *shared-body* perspective on the demonstrated task (see Figure 3).

3.4.1 Verbal. For the *verbal* instruction, the experimenter described the different tasks orally before the participant was asked to mimic the task using a prepared manuscript. Each sub-task, all units, and measured values necessary for the task were precisely described.

3.4.2 Shared-View. The experimenter demonstrated all activities to the participant in detail using a shared view so that the participant saw in their FoV what the experimenter saw when demonstrating

[1] https://developer.oculus.com/documentation/unity/unity-ovrinput/

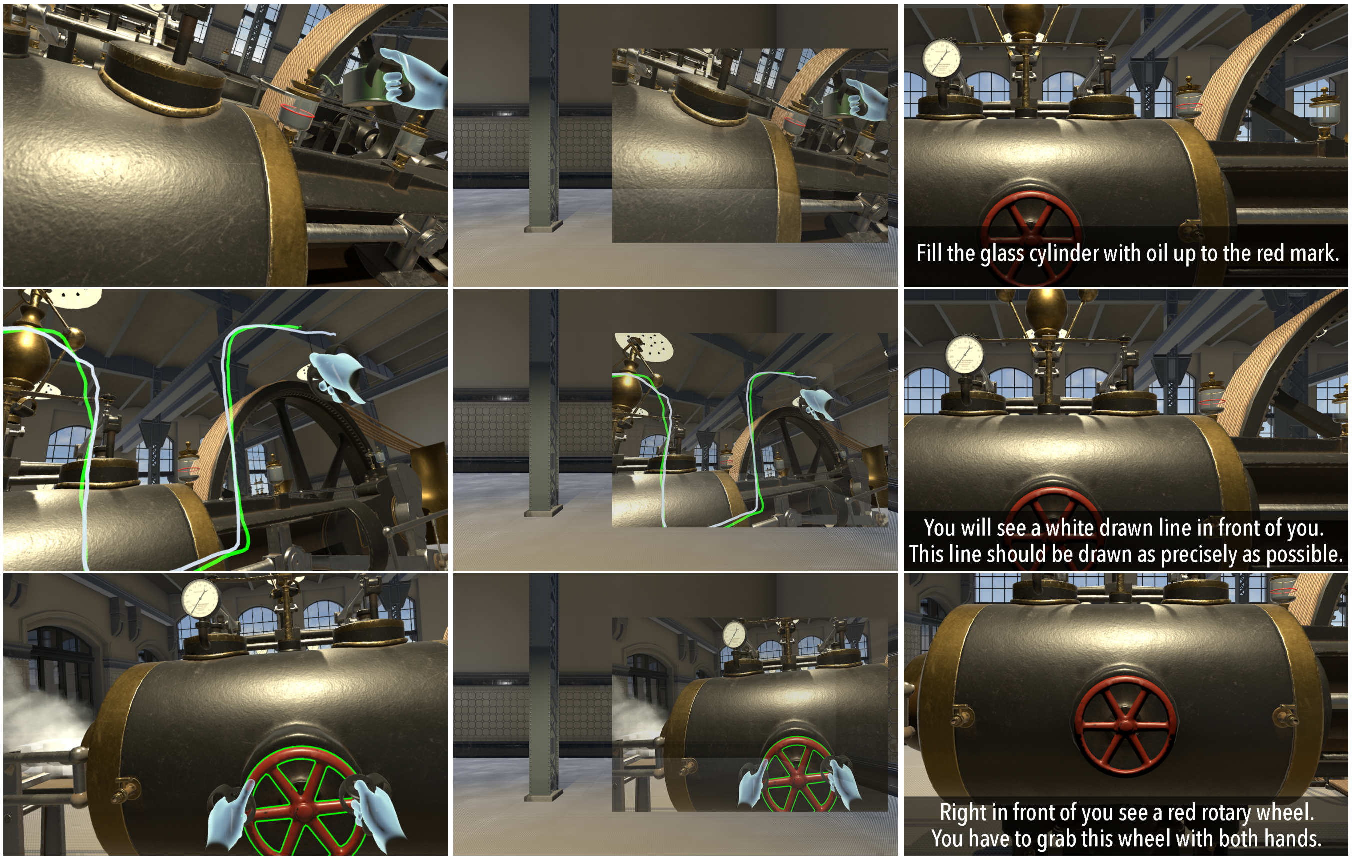

Figure 3: Screenshots of the different tasks for the different demonstration techniques (horizontal: the demonstration techniques: *shared-body, shared-view, verbal* (symbolized by the subtitle); vertical: the tasks: *3D pointing task (oil filling), 3D steering task (painting), bi-manual task (wheel control*)

the task. In contrast to *shared-body* (explained below), in which the 3D view of the experimenter is presented on the participant's HMD in real-time, *shared-view* is a 2D real-time life-video-like view. This view is presented as a overlay in the right FoV above the 3D VR view of the participant. For this purpose, the sequence of activities was captured by a virtual camera from the experimenter's egocentric perspective. To avoid the addition of exocentric views, we turned the participant in the virtual environment in a direction that the experimenter was not seeable. The controllers' handling, which had been explained in training, the experiment started with, was verbally explained through think-aloud to support remembering the UI handling.

3.4.3 Shared-Body. The *shared-body* provided the participant with a 3D view of the experimenter on their HMD in real-time. This means that the participant virtually slipped into the experimenter's avatar while they demonstrated the tasks in detail. Thus, the participant saw all activities of the experiment from the egocentric-perspective, but was not able to influence what happened through interactive actions. Similar to the *shared-view* condition, we advised the participant that learning the task requires observation of the activities shown by the experimenter. Like during the *shared-view* condition, the necessary manipulations of the controllers that the participants could not easily see and learn by observation, were expressed verbally by think-aloud (as memory aid, while all commands have been already introduced during the training phase). To reduce the feeling of motion sickness, the test person was still able to turn their head.

3.5 Measurements

We measured *TCT* for each task to evaluate the effect of our independent variables on time needed to complete each task under the varying conditions.

For the *3D pointing task*, the time measurement ran as long as oil was poured into the glass cylinder. For the *3D steering law task*, the time was measured, when the participant started to paint the line by pressing the right index trigger of the controller. The time measurement was stopped with the release of the right index trigger. If the line drawn with line interruptions, the last release marked the end of the time.

For the *bi-manual task*, we started the time measurement when both hands of the participant grabbed the steam machine's rotary wheel for the first time. The time measurement was stopped when both hands released the rotary wheel. Again, if the grab was interrupted, the last release counted.

Accuracy was also measured differently for each task. For the *3D pointing task*, we measured the fill level of oil in the glass cylinder and calculated the difference to the aimed for filling level indicated by the marker on the glass cylinder.

For the *3D steering law task*, we compared the offset between the line drawn by the participants with the demonstration line. For this purpose, the number of recorded points of the line drawn by the participant was made equal to the given line by interpolation. The distance from each coordinate to the equivalent point of the demonstration line and thus the offset to that line was determined. Finally, the average error to the demonstration line was determined.

For the *bi-manual task*, we measured the virtual pressure level of the steam engine at which the participant released the rotary wheel of the valve. As error and thus as accuracy value, we measured the difference to the pressure value given through the instruction (pointer of the manometer at 12 o'clock means pressure level 50%).

Task load was measured using the NASA TLX questionnaire [10].

We also used semi-structured interviews to better understand the quantitative ratings asking:

- What about this demonstration technique was helpful?
- What about this demonstration technique was not helpful?

3.6 Apparatus & Setup

A virtual machinery hall (see Figure 2) served as exemplary environment, in which two users, the experimenter or trainer and the participant or trainee, could be in the same VR for virtually teaching and training manual tasks. We used two Oculus Rift systems, which consist out of HMDs, controllers, and tracking sensors. The VR scene was rendered on a PC with an Intel i5-7400 CPU, 16 GB RAM, Nvidia GeForce GTX 1060 GPU for the experimenter, and on a PC with an Intel i7-7700 CPU, 16 GB RAM, and Nvidia GeForce GTX 1080 Ti GPU for the participant.

We implemented the VR application in Unity[2]. For implementing the network functionality, we used Photon Unity Networking[3]. The software allowed the free view in the virtual machinery hall and to interact with the steam machine as well as 3D painting in the air (see Figure 3). Participants were able to see their own virtual hands and also the virtual hands of the experimenter for the *shared-view* and *shared-body* condition as transparent blue hands, which is the default setting for Oculus HMDs. Moreover, participants were able to see their own Oculus Touch Controllers. As participants shared the same physical room, they could also hear each other.

During the experiment, participants sat down on a chair to complete the individual tasks. However, they were free to stand up and move around when the corresponding task required this.

3.7 Procedure

After participants signed the consent form and filled in a demographic questionnaire, the *demonstration technique* was randomly assigned to the participants. Each participant was randomly assigned one of the three *demonstration techniques* to complete the three *tasks* guided by the assigned *demonstration technique*. We equipped the participants with the HMD, and both experimenter and participant of the study entered the virtual environment.

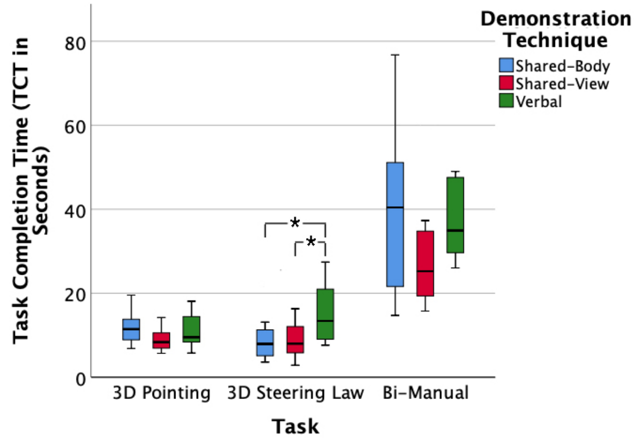

Figure 4: TCT (in seconds) for *demonstration technique*

Participants got a brief training into the controllers' handling and explanations about the *demonstration technique*. During the training, we instructed the participants in the handling of the controllers. Moreover, the participants get the information that all tasks should be performed with the greatest possible precision and as quickly as possible. Furthermore, the participants were told that they could stand up if a task required it.

During the experiment, participants got a demonstration for the first *task* in their randomly assigned *demonstration technique* and performed the task independently and without the help of the supervisor afterwards. After completing the first task, the participant's controller and headset were removed, and participant filled in the questionnaires. Afterward, the second and the third task followed the same procedure.

4 RESULTS

For analyzing the quantitative results, we conducted Kruskal-Wallis H tests to indicate significant differences for the between-subjects variable *demonstration technique*, and post-hoc analysis with Mann-Whitney U tests were conducted with a Bonferroni correction applied, resulting in a significance level set at .017.

Two-way mixed ANOVAs on the continuous dependent variables *task completion time (TCT)* and *accuracy* could not be applied as the assumptions of normality of the distribution of the data as well as of no significant outliers in the data were violated.

4.1 Quantitative Results

4.1.1 Task Completion Time. The *tasks* chosen for our experiment fundamentally differ in type and complexity (pointing versus steering, one-manual versus bi-manual), as well as in structure of task load (especially physical and temporal). Hence, we compare only data for the same *task* between the different conditions of the between-group variable *demonstration technique*.

[2]https://unity.com
[3]https://www.photonengine.com/pun

3D Pointing Task: Descriptive statistics led to the following median values for *TCT* for the *3D pointing task*: $Mdn_{\text{shared-body}} = 11.439$, $Mdn_{\text{shared-view}} = 8.351$, and $Mdn_{\text{verbal}} = 9.539$.

A Kruskal-Wallis H test did not indicate significant differences for *TCT* between the different *demonstration techniques*, $\chi^2(2) = 4.450$, $p = .108$ (see Figure 4).

3D Steering Law Task: Descriptive statistics led to the following median values for *TCT* (in seconds) for the *3D steering law task*: $Mdn_{\text{shared-body}} = 7.936$, $Mdn_{\text{shared-view}} = 7.996$, and $Mdn_{\text{verbal}} = 13.418$.

A Kruskal-Wallis H test indicate significant differences for the *TCT* between the different *demonstration techniques*, $\chi^2(2) = 8.155$, $p = .017$. Post-hoc tests showed that users need significantly less time for the completion of the *task* when the *demonstration technique* was *shared-body* compared to *verbal*, $U = 29.000$, $z = -2,483$, $p = .013$, and a significant shorter time for the completion of the *task* for *shared-view* compared to *verbal*, $U = 30.000$, $z = -2,425$, $p = .015$. No significant difference for the needed time to complete the *task* could be found between *shared-body* and *shared-view* as *demonstration technique*, $U = 69.000$, $z = -.173$, $p = .862$ (see Figure 4).

Bi-Manual Task: Descriptive statistics led to the following median TCT values for the *bi-manual task*: $Mdn_{\text{shared-body}} = 40.422$, $Mdn_{\text{shared-view}} = 25.235$, and $Mdn_{\text{verbal}} = 34.932$.

A Kruskal-Wallis H test did not indicate significant differences for *TCT* between the different *demonstration techniques*, $\chi^2(2) = 3.434$, $p = .180$ (see Figure 4).

4.1.2 *Accuracy.*

Due to the diversity of the individual tasks and the resulting measurement of *accuracy*, only comparisons of the same task between the different conditions of the between-group variable *demonstration technique* were performed.

3D Pointing Task: Descriptive statistics led to the following median values for the calculated difference of the oil fill value in the *3D pointing task*: $Mdn_{\text{shared-body}} = 2.3498$, $Mdn_{\text{shared-view}} = 6.4997$, and $Mdn_{\text{verbal}} = 3.6498$.

A Kruskal-Wallis H test did not indicate significant differences for *accuracy* between the different *demonstration techniques*, $\chi^2(2) = 3.304$, $p = .192$ (see Figure 5).

3D Steering Law Task: Descriptive statistics led to the following median values for the average distance between the given line and the line drawn by the user: $Mdn_{\text{shared-body}} = .1596$, $Mdn_{\text{shared-view}} = .0698$, and $Mdn_{\text{verbal}} = .0820$.

A Kruskal-Wallis H test did not indicate significant differences for the distance to the given line *task accuracy* between the different *demonstration techniques*, $\chi^2(2) = 5.911$, $p = .052$ (see Figure 5).

Bi-Manual Task: Descriptive statistics led to the following median values for the difference between the target pressure value and the value achieved by the user in the *bi-manual task*: $Mdn_{\text{shared-body}} = .0134$, $Mdn_{\text{shared-view}} = .0041$, and $Mdn_{\text{verbal}} = .0082$.

A Kruskal-Wallis H test did not indicate significant differences for *accuracy* between the different *demonstration techniques*, $\chi^2(2) = .258$, $p = .879$ (see Figure 5).

4.1.3 *Task Load.*

We analyzed the *task load* for the independent between-subject variable *demonstration technique*. Comparisons about the different tasks within a demonstration technique were not made, because the tasks vary too widely.

Table 1: ***Mdns*** **of Weighted TLX Ratings**

	Demonstration Technique		
	Shared-Body	Shared-View	Verbal
3D Pointing Task	35.333	58.355	33.000
3D Steering Law Task	32.167	30.834	28.334
Bi-Manual Task	57.333	51.834	51.833

Please see Table 1 for getting an overview of the descriptive statistics and the resulted weighted TLX scores of *Mdns* . A Kruskal-Wallis H test showed no significantly differences in the weighted TLX score for the independent between-subject variable *demonstration techniques*, $\chi^2(2) = .407$, $p = .816$ (see Figure 6).

Independent Kruskal-Wallis H tests were performed to identify significant differences in the weighted TLX score for the same *task* between the different *demonstration techniques*.

3D Pointing Task: A Kruskal-Wallis H test did not indicate significant differences for the *task load* of the *3D pointing task* between the different *demonstration techniques*, $\chi^2(2) = 3.732$, $p = .155$ (see Figure 6).

3D Steering Law Task: A Kruskal-Wallis H test did not indicate significant differences for the weighted TLX score of the *3D steering law task* between the different *demonstration techniques*, $\chi^2(2) = 0.025$, $p = .987$ (see Figure 6).

Bi-Manual Task: A Kruskal-Wallis H test did not indicate significant differences for the weighted TLX score between the different *demonstration techniques* in the *bi-manual task*, $\chi^2(2) = .503$, $p = .778$ (see Figure 6).

4.2 Qualitative Results

We analyzed the qualitative data collected during semi-structured interviews through closed coding. The categories were structured according to our independent variable *demonstration technique* to find explanations for our quantitative analyses' results.

4.2.1 *Shared-Body.*

A majority of participants (12) highlighted the visual presentation of the tasks as beneficial. It was stated for all tasks that it is helpful to see the execution of the task before completing it:

- *"Seeing the steps before doing them, knowing where the objects are."* (P38, *3D pointing task*),
- *"You could see exactly which task you had to fulfill."* (P47, *3D steering law task*), and
- *"Visual guidance made the task simple to replicate."* (P40, *bi-manual task*).

One participant particularly noted that the ability to see the details required to accomplish the task is helpful:

- *"Seeing in which direction to turn the wheel and where the barometer was."* (P38, *bi-manual task*).

Nevertheless, some participants (3) did not recognize or just forget details of the specific tasks. Especially for the *bi-manual task* it was stated that certain details were not recognized, which were important for the completion of the task:

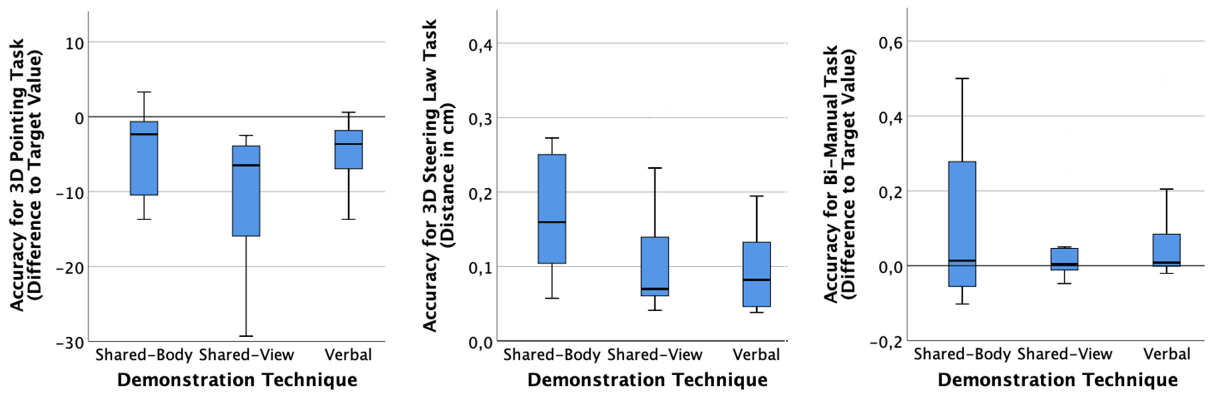

Figure 5: Accuracy measures for the different tasks (*3D pointing task, 3D steering law task*, and *bi-manual task)*

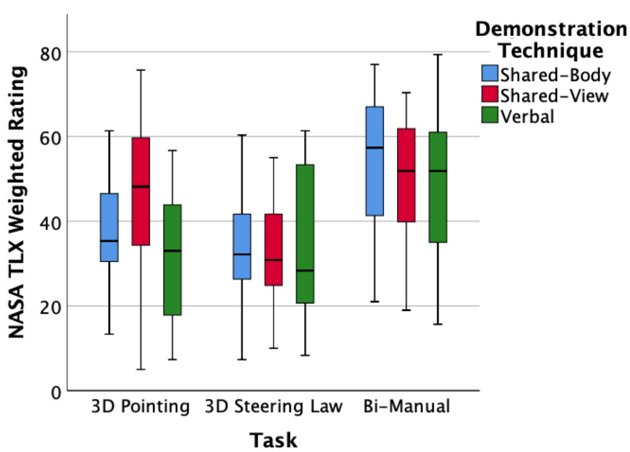

Figure 6: Weighted NASA TLX score

- *"They just said you have to let go of the wheel when the pressure is high enough, but not how to tell."* (P47), and
- *"It was difficult to memorize the direction to steer."* (P45, *shared-body*)

Two participants proposed multi-modal techniques (*shared-body* and *verbal*) in certain sub-tasks, for example, to specify the pressure level in the bi-manual task:

- *"It is hard to being told at which pressure level exactly you have to let go of the wheel, but instead just seeing it."* (P38, *bi-manual task*).

4.2.2 Shared-View. Across all tasks, the majority of the participants (9) in the *shared-view* condition stated that the visual presentation of the task was helpful and useful:

- Beneficial was that the *"whole sequence was shown. Seeing the hands move helps understanding the controls."* (P23, *3D pointing task*),
- *"I can see what I have to do. I get a live commentary to the shown explanation."*(P21, *3D steering law task*), and
- Beneficial was *"seeing the whole process and getting to know the relevant controller keys"* (P22, *bi-manual task*).

Again, two participants missed (additional) verbal explanations:

- *"You were not told all details,so you had to watch closely."* (P27, *bi-manual task*), and
- There was *"no exact explanation of the details of the important objects (which were relevant in this task)."* (P29, *bi-manual task*).

Four participants had problems seeing details, which were:

- *"the small and fixed shared screen."* (P22, *3D pointing task*), and
- *"the pushed button"* (P19, *3D pointing task*), confirmed by
- *"It's hard to see the pressed buttons."* (P21, *3D steering law task*), and
- *"I was not hundred percent sure how much oil I have to fill in."* (P21, *3D pointing task*).

Again, one participant stated that remembering of necessary parts of the tasks shown before may sometimes be hard:

- *"You have to remember some details, like the rotation direction of the wheel"* (P21, *3D steering law task*).

4.2.3 Verbal. The majority of participants (8) indicated that the *verbal* demonstrations technique was helpful. In particular, the comprehensiveness and detail of the instructions was mentioned by the participants (6), such as:

- *"It was intuitive. Identifying the location of the parts involved in the task after being told where they are was simple."* (P8, *3D pointing task*), and
- *"It was clear, concise and on point."* (P5, *3D steering law task*).

Again, two participants found it difficult to remember all details of the task, for example:

- *"After the explanation not all the key points were easy to remember because I wasn't able to reproduce it"* (P8, *bi-manual task*).

Only two participants stated that the demonstration was to fast for the amount of information, such as:

- *"The quick sequence in the explanation when which color appears and with which buttons when what happened was a bit fast"* (P12, *bi-manual task*).

Two participants proposed to show additional visual information, for example:

- *"The 3D aspect was not highlighted, which might confuse some people, otherwise it was great."* (P5), and
- *"A hint about the depth would have been nice."* (P12).

5 DISCUSSION

Comparing quantitatively measurable effects of *demonstrations techniques*, showed that verbal instructions lead to slowest task completion in our 3D steering law task, which was drawing a line in mid-air, while *accuracy* and *task load* were not effected by the *demonstrations technique*.

The longest times needed to draw in 3D after that task was verbally explained, will be discussed in the following paragraphs that are structured according the dependent variable *demonstration technique*. We here further discuss our qualitative results that highlight advantages and disadvantages of the according techniques.

5.1 Shared-Body Technique

By visually demonstrating the tasks, users have oriented themselves on the procedures and handling of the presenter. Learning how to perform these actions through visual demonstration helps performing these actions, which it confirmed through the participants' comments.

A majority of participants stated for all *demonstration techniques* that it was easy to follow and helpful for knowing how to fulfill the task after the demonstration. It was highlighted by our participants that the visual demonstration made the tasks easy to replicate. However, qualitative data also indicated that an exact additional verbal explanation of the details for the tasks would work even better. Especially in the *3D pointing task* and the *bi-manual task*, there was the desire to obtain more information about details like fill level when refilling the oil, or the information about the pressure level, when releasing the wheel. This could compensate for a possible loss of attention during observation of the tasks' actions.

Across all *tasks*, the majority of participants stated that the *shared-body* demonstration was helpful, and the *tasks* were easy to replicate.

5.2 Shared-View Technique

No different quantitative result were found for *shared-body* and *shared-view*. One reason could be that experimenter always performed the demonstration in the same way for *shared-body* as well as for *shared-view*, and that the perspectives in both conditions were almost same.

In summary, sharing the view was also helpful to replicate all tasks. Only the qualitative data indicate that the panel's small size caused visibility problems for details, but this did neither affect accuracy, nor time or task load in any of our *tasks*.

5.3 Verbal Technique

Contrary to the results of Kirk and Fraser, who found no significant difference in *TCT* for the different demonstration methods in an assembling task, we found that the *verbal* led to a significantly slower *TCT* compared to *shared-body* and *shared-view demonstration technique* for the *3D steering law task* of our study [14]. In contrast to the findings of George et al., who found no significant differences for the *TCT* when comparing the instruction methods in any of their tasks [9], we found for our *3D steering law task* that participants needed significantly more time for the *verbal* compared to both visual *demonstration techniques*, which yielded in slower *TCT* for the *verbal demonstration technique*.

Furthermore, we found no significant differences in *TCT* for the *3D pointing* and *bi-manual task*. These results are in line with Kirk and Fraser's findings [14] and the findings of George et al. [9].

We did not find a significant difference for the *accuracy* of any of our *tasks* when comparing the *verbal demonstration technique* with the other demonstration techniques. These results are in line with Kirk and Fraser's findings, who did not find any significant difference between voice-only and voice-plus-gesture for the accuracy in the performed assembly task [14].

Although we have declared for all *demonstration techniques* that the *tasks* should be completed as quickly as possible and with the greatest precision, the missing visual reference during the task's demonstration seems to influence the speed of the execution. When participants see the pace at which the *task* is being executed, they use the speed as a reference for their own execution time to complete the *task*. This time constraint was missing in the *verbal demonstration technique*. Another reason for the significantly slower *TCT* might be that users needs possibly longer to remember the procedure and actions for the task. The user has to transfer the description into an understanding of what to do for the following task.

Interestingly, for the *verbal demonstration technique*, the participants indicated that a reference to the depth in the room, which had to be considered when performing the *3D steering law task*, would have been helpful. However, this did not hurt the *accuracy* of the *3D steering law task* for the *verbal* condition of when comparing to the other *demonstration techniques*.

Still in in line with the lack of significant difference for the *task load* when comparing the *verbal* with the both visual *guidance techniques*, the majority of participants stated that the *verbal technique* helped to easily perform the explained task.

5.4 Multimodal Techniques

Our qualitative data indicates that users have a strong desire for multimodal task demonstration techniques. For both the purely verbal and purely visual instruction methods, it was noted that additional visual cues when verbally instructed and additional verbal cues when visually instructed would be helpful. The simultaneous visual, as well as verbal instruction, can support a task to be demonstrated in learning it since both instruction methods complement each other. Details that cannot be expressed verbally in an exact or very complicated way can often be better demonstrated visually. Also, verbal hints can be given, which must be paid special attention to. Without these additional hints, details in the execution of tasks might not been recognized.

5.5 Releasing the Effort to Remember

The qualitative data show that for some users, too much information was given at once, which led to difficulties in performing the task because details might have been forgotten. Therefore, it is helpful to not rely on too much memory load. For longer and more complex tasks, it may be better to split them into sub-tasks or to give in-situ instructions.

5.6 Design Recommendations

Based on our results and their discussion, we derive the following design recommendations.

Shared-Body Task Demonstration: For pre-task demonstration, we recommend the *shared-body technique*, which has a spatial reference of the objects with which a user has to interact and where a medium number of action sequences have to be communicated. However, due to the loss of control over the execution of the trainee, this instruction method cannot be used for in-situ instructions given at the same time the task is replicated. To release memory load, splitting a complex task into multiple sub-tasks is recommended.

Shared-View Task Demonstration: We also recommend the *shared-view technique*. In contrast to *shared-body*, *shared-view* can also be used for in-situ instructions. Here, users could simultaneously see the instruction visually and perform it directly. A split into sub-tasks can alternatively help to not overload task memory.

Verbal Task Explanation: We only recommend the verbal *demonstration technique* for tasks where not much information needs to be given to communicating the task to be accomplished and where the work processes can be formulated briefly and precisely without visual cues.

Multimodal Instructions: A combination of visually observing the activity of the trainer supported by verbal instructions is preferable to train physical tasks, especially when they are complex. If instructions have to be given simultaneously as the task to be executed, we recommend the combination of *verbal* and *shared-view*. For pre-task execution demonstrations, the combination of *verbal* and *shared-body* is recommended.

6 CONCLUSION AND FUTURE WORK

In this paper, we investigated how different *demonstration techniques* in VR influence the *task completion time (TCT)*, *accuracy*, and *task load* of various tasks. Results of a user study show that the *verbal* instruction resulted in significantly more time for the execution of the *tasks* compared to the *shared-body* and *shared-view demonstrations*.

Through that finding and qualitative feedback of our participants, we derived the following design recommendations. While verbal explanations should only be used for very simple tasks, verbal explanations in addition to visual task demonstrations are recommended. For instructions are given before the tasks are replicated, we propose the *shared-body guidance* with additional verbal hints. For in-situ instructions, we propose the *shared-view guidance* with additional verbal explanation. For the pre-task use of both, *shared-body guidance* and *shared-view guidance*, splitting the task into sub-tasks releases memory load, which is highly recommended for long or complex tasks.

The evaluation of our design guidelines needs to remain future research. Tasks and the extent of the steps required to fulfill and guide them vary widely. Tasks may include several dimensions, such as complexity, length, and type. Future research needs to investigate whether these and other tasks' dimensions are affected differently by the demonstration technique and which demonstration technique should be preferred.

Furthermore, we only looked at demonstration techniques of one type but did not combine them. As we naturally combine demonstration and speech, future research should investigate how *verbal* and *visual demonstration techniques* work together.

With these design recommendations, we hope to help researchers and practitioners design and develop better and usable techniques to train physical tasks in VR.

ACKNOWLEDGMENTS

This work is supported by the German Ministry of Education and Research (BMBF) within the GEVAKUB (01JKD1701B) and the WINK (16SV8425) project.

REFERENCES

[1] Judith Amores, Xavier Benavides, and Pattie Maes. 2015. ShowMe: A Remote Collaboration System That Supports Immersive Gestural Communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI EA '15)*. Association for Computing Machinery, New York, NY, USA, 1343–1348. https://doi.org/10.1145/2702613.2732927

[2] Amine Chellali, Isabelle Milleville-Pennel, and Cédric Dumas. 2013. Influence of Contextual Objects on Spatial Interactions and Viewpoints Sharing in Virtual Environments. *Virtual Real.* 17, 1 (March 2013), 1–15. https://doi.org/10.1007/s10055-012-0214-5

[3] Philo Tan Chua, Rebecca Crivella, Bo Daly, Ning Hu, Russ Schaaf, David Ventura, Todd Camill, Jessica Hodgins, and Randy Pausch. 2003. Training for Physical Tasks in Virtual Environments: Tai Chi. In *Proceedings of the IEEE Virtual Reality 2003 (VR '03)*. IEEE Computer Society, USA, 87.

[4] Jane E Cole and Jeffrey B Kritzer. 2009. Strategies for success: Teaching an online course. *Rural Special Education Quarterly* 28, 4 (2009), 36–40.

[5] J. C. F. de Winter, S. de Groot, J. Dankelman, P. A. Wieringa, M. M. van Paassen, and M. Mulder. 2008. Advancing Simulation-Based Driver Training: Lessons Learned and Future Perspectives. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services* (Amsterdam, The Netherlands) *(MobileHCI '08)*. Association for Computing Machinery, New York, NY, USA, 459–464. https://doi.org/10.1145/1409240.1409314

[6] 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* (Sydney NSW, Australia) *(TEI '20)*. Association for Computing Machinery, New York, NY, USA, 311–322. https://doi.org/10.1145/3374920.3374945

[7] Markus Funk, Thomas Kosch, and Albrecht Schmidt. 2016. Interactive Worker Assistance: Comparing the Effects of in-Situ Projection, Head-Mounted Displays, Tablet, and Paper Instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 934–939. https://doi.org/10.1145/2971648.2971706

[8] Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of Head-Mounted and Scene-Oriented Video Systems on Remote Collaboration on Physical Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 513–520. https://doi.org/10.1145/642611.642701

[9] Ceenu George, Michael Spitzer, and Heinrich Hussmann. 2018. Training in IVR: Investigating the Effect of Instructor Design on Social Presence and Performance of the VR User. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 27, 5 pages. https://doi.org/10.1145/3281505.3281543

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

[11] Thuong N. Hoang, Martin Reinoso, Frank Vetere, and Egemen Tanin. 2016. Onebody: Remote Posture Guidance System Using First Person View in Virtual Environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (Gothenburg, Sweden) *(NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, Article 25, 10 pages. https://doi.org/10.1145/2971485.2971521

[12] Dongsik Jo, Yongwan Kim, Ungyeon Yang, Jinsung Choi, Ki-Hong Kim, Gun A. Lee, Yeong-Do Park, and Young Whan Park. 2011. Welding Representation for Training under VR Environments. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry* (Hong Kong, China) *(VRCAI '11)*. Association for Computing Machinery, New York, NY, USA, 339–342. https://doi.org/10.1145/2087756.2087809

[13] Michael C. Jones. 2008. Simulation across the Spectrum of Submarine Training. In *Proceedings of the 2008 Spring Simulation Multiconference* (Ottawa, Canada) *(SpringSim '08)*. Society for Computer Simulation International, San Diego, CA, USA, 811–815.

[14] David S. Kirk and Danaë Stanton Fraser. 2005. The Effects of Remote Gesturing on Distance Instruction. In *Proceedings of Th 2005 Conference on Computer Support for Collaborative Learning: Learning 2005: The next 10 Years!* (Taipei, Taiwan) *(CSCL '05)*. International Society of the Learning Sciences, 301–310.

[15] Daisuke Kondo, Keitaro Kurosaki, Hiroyuki Iizuka, Hideyuki Ando, and Taro Maeda. 2011. View Sharing System for Motion Transmission. In *Proceedings of the 2nd Augmented Human International Conference* (Tokyo, Japan) *(AH '11)*. Association for Computing Machinery, New York, NY, USA, Article 26, 4 pages. https://doi.org/10.1145/1959826.1959852

[16] Matthew Kyan, Guoyu Sun, Haiyan Li, Ling Zhong, Paisarn Muneesawang, Nan Dong, Bruce Elder, and Ling Guan. 2015. An Approach to Ballet Dance Training through MS Kinect and Visualization in a CAVE Virtual Reality Environment. *ACM Trans. Intell. Syst. Technol.* 6, 2, Article 23 (March 2015), 37 pages. https://doi.org/10.1145/2735951

[17] G. Lee, H. Kang, J. Lee, and J. Han. 2020. A User Study on View-sharing Techniques for One-to-Many Mixed Reality Collaborations. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 343–352.

[18] Richard Light and Rod Fawns. 2001. The thinking body: Constructivist approaches to games teaching in physical education. *Critical Studies in Education* 42, 2 (2001), 69–87.

[19] Maite Lopez-Garate, Alberto Lozano-Rodero, and Luis Matey. 2008. An Adaptive and Customizable Feedback System for VR-Based Training Simulators. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 3* (Estoril, Portugal) *(AAMAS '08)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1635–1638.

[20] Elhassan Makled, Amal Yassien, Passant Elagroudy, Mohamed Magdy, Slim Abdennadher, and Nabila Hamdi. 2019. PathoGenius VR: VR Medical Training. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays* (Palermo, Italy) *(PerDis '19)*. Association for Computing Machinery, New York, NY, USA, Article 31, 2 pages. https://doi.org/10.1145/3321335.3329694

[21] Takayuki Nozawa, Erwin Wu, Florian Perteneder, and Hideki Koike. 2019. Visualizing Expert Motion for Guidance in a VR Ski Simulator. In *ACM SIGGRAPH 2019 Posters* (Los Angeles, California) *(SIGGRAPH '19)*. Association for Computing Machinery, New York, NY, USA, Article 64, 2 pages. https://doi.org/10.1145/3306214.3338561

[22] Lauriane Pouliquen-Lardy, Isabelle Milleville-Pennel, François Guillaume, and Franck Mars. 2016. Remote collaboration in virtual reality: asymmetrical effects of task distribution on spatial processing and mental workload. *Virtual Reality* 20, 4 (01 Nov 2016), 213–220. https://doi.org/10.1007/s10055-016-0294-8

[23] Rufat Rzayev, Gürkan Karaman, Katrin Wolf, Niels Henze, and Valentin Schwind. 2019. The Effect of Presence and Appearance of Guides in Virtual Reality Exhibitions. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/3340764.3340802

[24] Lim Seow San, Djasliana Binte Hussain, Lim Heng Loo, Zhuang Ronglin, Suraj Balasubramanian, Kelvin W. C. Foong, Tok Wee Wah, Muhammad Taufiq Bin Mahmud, Seah her Yuh, Lim Minmin, Emilia Joey Yim Shao'en, Jeremiah K. S. Seah, Sheryl Lim Jia Jin, Low Minghui, and Lu Zhiyin. 2008. Dental VR Application for Dental Training. In *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (Singapore) *(VRCAI '08)*. Association for Computing Machinery, New York, NY, USA, Article 33, 2 pages. https://doi.org/10.1145/1477862.1477906

[25] M. Stefik, D. G. Bobrow, G. Foster, S. Lanning, and D. Tatar. 1987. WYSIWIS Revised: Early Experiences with Multiuser Interfaces. *ACM Trans. Inf. Syst.* 5, 2 (April 1987), 147–167. https://doi.org/10.1145/27636.28056

[26] S. Valin, A. Francu, H. Trefftz, and I. Marsic. 2001. Sharing viewpoints in collaborative virtual environments. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. 12 pp.–.

[27] Wang Xiaoling, Zheng Peng, Wei Zhifang, Sun Yan, Luo Bin, and Li Yangchun. 2004. Development an Interactive VR Training for CNC Machining. In *Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry* (Singapore) *(VRCAI '04)*. Association for Computing Machinery, New York, NY, USA, 131–133. https://doi.org/10.1145/1044588.1044612

[28] 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. https://doi.org/10.1162/105474602317473240

[29] Ahmed Mohamed Fahmy Yousef, Mohamed Amine Chatti, Ulrik Schroeder, and Marold Wosnitza. 2014. What drives a successful MOOC? An empirical examination of criteria to assure design quality of MOOCs. In *2014 IEEE 14th International Conference on Advanced Learning Technologies*. IEEE, 44–48.

A INSTRUCTIONS FOR VERBAL GUIDANCE

A.1 3D Pointing Task

For the *3D pointing task*, the following descriptions were used: *The task must be completed as quickly as possible and with the highest possible accuracy. On the table to your right is an oilcan. Grab the oilcan with your right hand. To grab the oilcan, you make a grab gesture by pressing the trigger on the right controller with your middle finger. Or in other words, your hand closes around the controller like it would in reality. When you have grasped the oilcan, move it over the glass cylinder on your right, which has the red mark. If you cannot reach the glass cylinder from a sitting position, you are welcome to stand up. If the oilcan is above the glass cylinder, you can pour oil into it. Fill the glass cylinder with oil up to the red mark. Then hold the jug horizontally to the ground so that no more oil flows. You can put the oilcan down. The task is now finished.*

A.2 3D Steering Law Task

The *3D steering law task* was explained as follows: *You will see a white drawn line in front of you. This line should be drawn as precisely as possible. By pressing the right pistol trigger with your index finger, painting is activated. By moving your hand, the line can be followed and traced. For the orientation, you will see a green dot in front of you. This green dot shows you the position for drawing. The right pistol trigger must be pressed permanently while drawing. If you let go of the trigger once, you can push it again and continue drawing. When you have reached the end of the line, release the trigger. This ends the task.*

A.3 Bi-Manual Task

The *bi-manual task* was explained through the following verbal explanations: *The task must be completed as quick and as accurate as possible. Right in front of you , you see a red rotary wheel. You have to grab the wheel with both hands. If your hands are in the area of the wheel, it will have a yellow outline. You grab the wheel with the grabbing gesture using both hands. Or in other words, your middle fingers press the triggers on the controller. When the wheel is grabbed, it gets a green contour. Turn the wheel counterclockwise. As long as you can turn the wheel, it is red. If you cannot turn it anymore, it will light up bright yellow, and the valve is open. As long as the valve is open, the wheel must be held. Then the pressure decreases. The pressure must be reduced until the hand on the pressure indicator is at ẗwelve o'clock: Then you must release the wheel. This ends the task.*

B THINK-ALOUD PROTOCOL FOR VISUAL GUIDANCE

The *shared-view* and *shared-body* techniques used the same verbal instructions as think-aloud protocol and actions for the different *tasks.*

B.1 3D Pointing Task

The *shared-view* and *shared-body* techniques used the same verbal instructions as and actions for the *oil filling task*:

(1) **Think aloud**: *I take the oilcan in my hand through the gesture of grasping. The hand encloses the right controller. The middle finger presses the grip trigger.*
(2) **Action**: The experimenter grabbed the oilcan and moved it over the glass cylinder.
(3) **Think aloud**: *I fill a certain amount of oil into the cylinder.*
(4) **Action**: Oil is filled into the glass cylinder up to the mark.
(5) **Think aloud**: *If I hold the oilcan horizontally to the ground and no more oil flows, the task is finished.*
(6) **Action**: The experimenter held the oilcan horizontal and placed it on the table.

B.2 3D Steering Law Task

(1) **Think aloud**: *To start drawing, I press the right index trigger with my index finger and draw the line as accurately and quickly as possible*
(2) **Action**: The experimenter pressed the right pistol trigger with the index finger and started tracing the line
(3) **Think aloud**: *When I reach the end of the line to be copied, I release the right pistol trigger.*
(4) **Action**: The experimenter finished copying the line and released the right index trigger.

B.3 Bi-Manual Task

(1) **Think aloud**: *I grasp the wheel with both hands through the grasping gesture. Both hands close around the controller, both middle fingers press the button.*
(2) **Action**: The experimenter grabs the rotary wheel and turns it counterclockwise until the color of it changes to yellow (this was not expressed orally).
(3) **Think aloud**: *I open the valve.*
(4) **Action**: If the rotary wheel's color changed to yellow, it was held by the experiment leader.
(5) **Think aloud**: *When a certain pressure level is achieved, i release the wheel immediately.*
(6) **Action**: The experimenter looked at the manometer and releases the wheel with both hands exactly when the pointer is at 12 o'clock.