Design and Evaluation of a VR Training Simulation for Pump Maintenance Based on a Use Case at Grundfos

Frederik Winther

Linoj Ravindran

Kasper Paabøl Svendsen

Tiare Feuchtner

Department of Computer Science, Aarhus University, Denmark

Figure 1: Based on a use case at Grundfos, we implemented a virtual training simulation of a structured maintenance task for training of novice technicians. In a user study, we evaluated the potential of this VR Training simulation (C), in comparison to two traditional training methods: (A) Pairwise Training, (B) Video Training.

ABSTRACT

Encouraged by technological advancements, more and more companies consider virtual reality (VR) for training of their workforce - in particular for situations that occur rarely, are dangerous, expensive, or very difficult to recreate in the real world. Thereby the need arises for understanding the potentials and limitations of VR training and establishing best practices. In pursuit of this, we have developed a *VR Training* simulation for a use case at Grundfos, in which apprentices learn a sequential maintenance task. We evaluated this simulation in a user study with 36 participants, comparing it to two traditional forms of training (*Pairwise Training* and *Video Training*). This paper describes the developed virtual training scenario and discusses design considerations for such VR simulations. Further, it presents the results of our evaluation, which support that *VR Training* is effective in teaching the procedure of a maintenance task. However, according to our evidence, traditional approaches with hands-on experience still lead to a significantly better outcome.

Index Terms: Human-centered computing—Virtual reality; Human-centered computing—User studies

1 INTRODUCTION

In industrial research, Virtual Reality (VR) holds the promise of potential low-cost solutions for design and prototyping of new products and simulation of processes for knowledge gain [4]. An obvious advantage of using VR as a training tool, is that it provides a safe environment for making mistakes (e.g., for training surgeries [10]), and makes it possible to practice scenarios that might not occur very often or are difficult to simulate in the real world (e.g., preparing Walmart staff for Black Friday¹). Furthermore, in VR processes can be trained involving machinery and tools that are physically unavailable (e.g., since stopping the production line is not desirable), or because the setup might not be accessible or not even exist yet (e.g.,

training for space missions). Finally, VR (as well as augmented reality) offers the opportunity for supplying information in an engaging and convenient fashion that is not possible in the real world, e.g., highlighted objects and animations for guidance, instead of paper manuals.

While computer-simulated training scenarios have been used for decades (e.g., for training aircraft pilots), there is still a lack of long-term, large-scale evaluations of immersive VR simulations experienced through a head-mounted display that provide evidence of their effectiveness as a training tool. Arguably, not all forms of learning can be equally well supported by such VR simulations (e.g., memorizing a sequence of clearly defined steps by practicing vs. exploring material properties and tools when sculpting a statue). Furthermore, we cannot assume that our lessons learned for the design of traditional computer interfaces will apply equally in this immersive 3-dimensional domain, so there is also a need for new design guidelines and best practices. These must address a whole set of new challenges that immersive VR brings along. For example, the effective simulation of the physical world: in VR nearly all stimuli (visual, haptic and auditive) must be carefully designed and implemented - what is not programmed will not be seen, felt, or heard. A further challenge lies in the limited interactions available to date. A button press, for example, does not allow the user to differentiate between the action of carefully pinching something between the fingertips, or tightly clenching an object in one's fist.

We explore a case at Grundfos, a Danish pump manufacturer, where apprentices are taught sequential maintenance tasks for many different pump models. The company employs various modes and materials to instruct their workers, such as pairwise training and instructional videos. Training of such tasks often poses the following challenge: either a pump needs to be shut down and uninstalled from the system, an additional pump must be acquired and made explicitly available for training purposes, or novices can only train whenever ordinary maintenance tasks are scheduled. Here, VR presents a valuable opportunity for training without interfering with the system in operation, or requiring an instructor to be available, while still providing near-realistic manipulation of the pump and its sub-components in 3D space. The challenge, however, lies in evaluating how effective such VR simulations are in teaching these

¹Walmart Black Friday training: <http://archive.today/9H2qJ>

sequential tasks, in what cases they may be beneficial or detrimental, and whether they are fit to replace or rather complement traditional forms of training.

In this paper, we evaluate the effectiveness of *VR Training* for a maintenance task on a dosing pump, in comparison to traditional training methods (*Video Training*, *Pairwise Training*), based on the results of a user study. The three compared training methods are illustrated in Fig. 1. We expected participants trained in VR to be capable of performing a maintenance task on a real-world dosing pump (H1), that the three training methods would yield comparable performance based on completion time and number of errors (H2 and H3), and that *VR Training* will lead to a higher level of confidence than *Video Training* (H4). In the following section we present relevant related work. We then elaborate on several design considerations for the creation of this VR training simulation, followed by details of our user study. This is followed by a discussion of results and the implications of our findings.

2 RELATED WORK

2.1 Virtual Reality Training

VR and Augmented Reality (AR) are increasingly used for training and reported results from evaluations of such training scenarios vary, depending on what is being trained and how the training is carried out. While some studies report that *VR Training* is inferior to real-world training [5, 12, 16], others report learning outcomes to be equal [20, 26]. In other cases, *VR Training* was shown to even outperform real-world training [30, 31].

Hamblin [16] compared training of a complex manual assembly task in a VE using a Head-Mounted Display (HMD) vs. a desktop computer. The results from the study suggested that VEs can be effective training simulators, but that the transfer of training is better in real-world training. Furthermore, the study showed that virtual training is not suited for all training tasks. Webel et al. [29] state that AR-based training applications should differ from AR-based guiding applications, in the way that training should involve cognitive aspects. Werrlich et al. [30] used this approach while comparing paper-based training with AR-based training when assembling an engine. They implemented four different user modes that ranged from tutorial to expert mode. Results showed that participants trained using paper-based instructions performed the assembly significantly faster but with significantly more errors than those trained in AR.

Carlson et al. [5] investigated retention of a learned task by comparing training in a VE to training in a physical environment. They found that participants who received the physical training learned faster and performed the assembly task more efficiently than the participants who received the virtual training. However, the virtually trained participants did improve their test assembly times from the training to the post-test. Benefits of manipulating physical objects during a cognitive manual task, compared to virtual objects, were also reported by Lok et al. [24].

2.2 Skill Acquisition and Learning Transfer

Fitts and Posner [8] identify three stages of skill acquisition in psychomotor learning tasks: (1) *cognitive*, (2) *associative*, and (3) *autonomous*. In the cognitive stage, the person develops a basic understanding of the task. In the associative stage, performance is improved by drawing associations to the environment. Lastly, in the autonomous stage, the task no longer requires much attention.

One of the main aspects of training users and new potential employees in a specific field of work is to ensure that the learning transfer is high. Learning transfer is the amount of knowledge that is transferred from one context to another. Woodworth and Thorndike [32] originally referred to this as *transfer of practice*. Seidel and Chatelier [27] divide learning transfer into two parts: learning and recall. If these two parts happen in the same environmental context, the subjects involved will have a better memory of

the information. If, however the two environmental contexts differ this leads to *contextual changing*, which decreases the performance of recall significantly. This is also known as *contextual dependent memory*. This concept is very relevant for learning skills in a VE, which should be recalled in the real world.

2.3 Parameters for comparing training methods

To compare different training methods, a common evaluation task and evaluation criteria need to be defined. These criteria or parameters vary among different studies. The most common are described as follows.

Completion time: Completion time is an objective and frequently used evaluation parameter [4, 5, 9, 13, 16, 20, 30]. It is evaluated for tasks that have a defined beginning and end, such as assembly tasks. Completion time is often looked at in conjunction with other parameters such as the number of errors, to evaluate task performance.

Number of errors: Another objective parameter is the number of errors, which is often used in VR training studies [9–11, 13, 16, 20, 27, 30]. This requires prior definition of possible errors, such as picking up a wrong object to interact with [30], or failing to follow the task procedure in the right order [20].

Knowledge retention: Knowledge retention is an objective parameter that measures how much participants remember of the learned task, by comparing completion time and error rates from two tests. Arthur et al. [1] define knowledge decay as "loss of trained [...] knowledge over extended periods of non-use". A series of studies have researched how different training methods influence the participants' retention of knowledge [5, 15, 20, 23]. Knowledge retention can be evaluated at varying intervals after training, ranging from instantaneously to more than 365 days later [1].

Confidence in actions: Confidence in actions is a subjective parameter that evaluates how sure participants feel about their actions. This can provide insights into the effectiveness of training methods from the participants' point of view. Grantcharov et al. [13] measured *confidence of movements* (confidence in actions) in combination with what they refer to as *economy of movements*.

3 TRAINING SIMULATION FOR A MAINTENANCE TASK

3.1 Description of the use case

Companies such as Grundfos mainly apply two different training methods to teach new employees how to handle the machinery: *Pairwise Training* and *VR Training*. In the first, a trainee is personally educated by an instructor. The latter is carried out in a dedicated room, where the trainee is equipped with VR gear and completes the training independently. In this particular case, each *VR Training* session takes about 20 minutes. The task is divided into a series of steps, in each of which narration and visual highlights inform the trainee what object needs to be interacted with and how. Trainees interact with the simulation by selecting and manipulating the highlighted object with their controllers. Apart from these two training techniques, instructional videos are often available as an additional aid when practicing maintenance tasks. Technicians also have access to the respective pump manual, which they may refer to if in doubt about the required steps. However, understanding these manuals inherently requires knowledge of the basic procedures. Furthermore, efficient completion of maintenance tasks is critical, to ensure lowest possible down-time of the system. Thus, it is desirable that technicians are well acquainted with their tasks.

Interestingly, many training experts at companies like Grundfos consider *Pairwise Training* and *VR Training* to be equally effective training methods, without any supporting evidence from a comparative evaluation. To gain such evidence, we developed a simulation of an exemplary maintenance task. This task involves changing the oring of a dosing pump, which requires several steps for disassembly

Table 1: Sequence of steps that need to be carried out in order to replace the o-ring on the dosing pump.

No.	Disassembly of dosing pump	No.	Assembly of dosing pump
1.	Wear gloves and safety glasses	11.	Attach the new O-ring to the drainer
2.	Depressurize dosing pump	12.	Attach the drainer to the dosing pump
3.	Dismantle top and bottom ball valves	13.	Fasten the diaphragm
4.	Detach the back-lid	14.	Close diaphragm digitally
5.	Loosen the screws in the dosing head	15.	Mount the dosing head
6.	Take off the dosing head	16.	Attach each of the four screws
7.	Open diaphragm digitally	17.	Fasten the screws crosswise
8.	Loosen the diaphragm	18.	Attach the back lid
9.	Remove the drainer	19.	Attach bottom and top valve
10.	Remove the old O-ring		

and reassembly (see Table 1). The *VR Training* simulation was designed to closely follow the predefined steps for one specific pump, for which an instructional video was available for comparison.

3.2 Technical Setup and Interaction Design

The *VR Training* simulation was developed in Unity 3D² and experienced through an HTC Vive³ HMD. Matching HTC Vive controllers served as input devices for interaction in the virtual environment (VE). The training simulation was run on a Windows PC with a GeForce GTX 1070 Graphics Card.

The initial design of the VR system was inspired by existing training simulations used at Grundfos and informed by interviews with the company’s training experts and instructors. In particular, these experts assisted in a preliminary informal task analysis, consisting of the definition of the individual steps of the maintenance task and the identification of evaluation criteria. Further, they supplied information on the specific procedure of traditional training methods and the types of skills that were to be acquired thereby. Additional design decisions were made in consideration of central HCI principles, such as supporting direct manipulation (section 3.3) and multimodal guidance (section 3.4). We followed an iterative design process involving a series of pilot studies. These served to test out different aspects of the prototype, such as how hints are presented and how selected objects are visualized in the VE.

3.3 User Interaction

In the resulting prototype the user interacts through two virtual hands (3D-hand-models that appear to be wearing white gloves). These hands follow the movements of the Vive Controllers, thereby reflecting the user’s own hand movements. Further, when the user pulls the trigger on a controller, the fingers of the respective virtual hand curl inward, as if grabbing something (Fig. 2). This suggests to the user that they can reach out and manipulate objects in the VE by simply grabbing them. Such a hand selection pattern generates expectations for the manipulation of interactive objects in the VE, e.g., that objects in the virtual world behave like objects in the real world [21, Chapter 28]. However, for the sake of usability the VE deviates from the real world in some respects. For example, we decided against including a realistic physics simulation to safeguard against components wildly flying through the scene due to wrongly calculated collision forces, thereby ending up beyond the user’s reach. As a result (and to the disappointment of some users) objects cannot be thrown: once a grabbed object is released, it simply drops down while any acceleration in other directions is discounted.

To contextualize the training, the VE resembles an industrial facility (see Fig. 3 A). A special scene was designed, to permit users to get acquainted with VR and freely explore the VE prior to the actual training. In this scene a series of interactive objects were presented on a workbench and included playthings, such as a small rocket that could be launched. By interacting with these

objects, the users playfully acquire the skills for manipulations that are needed in training. The virtual training scenario aims at simulating a real-world maintenance task to decrease contextual changing [22, 27], and includes a detailed 3D model of a dosing pump (supplied by Grundfos). Furthermore, there are two different tools (a torque screwdriver and a needle tool) with which users disassemble components, replace the o-ring, and then reassemble the dosing pump. A Guidance System, described in the following section, informs the user about the required steps for the task.

Apart from the dosing pump, all virtual content was obtained from Sketchup’s 3D warehouse⁴ and modified, or specifically designed in Blender⁵.

3.4 Multimodal Guidance System

During the training a guidance system provides multimodal cues, including voice recordings, textual instructions, highlights and animations, and haptic feedback [23]. To improve knowledge transfer, two levels of difficulty were implemented [30]: *Beginner* and *Intermediate*. When training at beginner level the user is provided with as much assistance as possible. At intermediate level the amount of guidance is decreased by omitting animations and the highlights on objects, thereby requiring increased use of cognitive abilities.

Voice recordings and textual instructions: A female voice guides the user through the sequence of steps, providing instructions and important information regarding safety. A virtual tablet, which is propped up on one side of the workbench, reflects these instructions in written form. The user can refer to these whenever needed (e.g., if they did not pay enough attention to the voice recording).

Highlights and animated hints: When training at beginner level, the regions of immediate interest are highlighted with a yellow glow (Fig. 3 B and C). This makes the objects pop out, leveraging pre-attentive visual processing [18]. Furthermore, the required action is

⁴Sketchup’s 3D warehouse <https://3dwarehouse.sketchup.com/>

⁵Blender (version 2.79b) <https://www.blender.org/>

Figure 2: In the VE virtual hands can pick up interactive components. (A) Relaxed hand with extended fingers, when the user is not grabbing. (B) Pulling the trigger on the controller causes the hand to grab.

²Unity 3D <https://unity.com/>

³HTC Vive HMD <https://www.vive.com/us/>

repeatedly demonstrated by an animated virtual hand (e.g., twisting a component to loosen it, as shown in Fig. 3 B).

Haptic feedback for tool use: A challenge of *VR Training* is to simulate the stimuli that physical objects naturally provide, such as their material properties. Approaches to implement one or more of these stimuli often involve wearable gloves or finger attachments, which allow haptic rendering of textures [17], applying force feedback [3, 6, 7, 14], etc. In our system, important haptic cues were simulated through vibration of the Vive controllers. An example of this is the clicking of a torque screw driver, when the screw was fastened beyond its intended torque. This was accompanied by a recorded clicking sound of the torque screwdriver.

Indication of progress: When manipulation of an object required repeated or prolonged interaction (e.g., fastening a screw, or holding down a button for 2 seconds), a radial progress bar provided the user with feedback.

Snapping: To better facilitate interaction with the small virtual object components, we implemented snapping on two occasions: 1. When disassembling the dosing pump the user must place components in predefined locations on the workbench (see Fig. 3 C). Components snap to these locations when released in proximity. 2. When reassembling the dosing pump, components will snap to their correct position and orientation, when held and released very close to where they should be mounted.

4 EVALUATION

A user study was carried out to compare *VR Training* to traditional training methods for performing maintenance on the dosing pump, and to test the following hypotheses:

- H1** *VR Training* enables participants to successfully perform a structured maintenance task in the real world
- H2** *VR Training* and *Pairwise Training* are comparable with regards to (a) *Task Completion Time* and (b) *Number of Errors*
- H3** *VR Training* and *Video Training* are comparable with regards to (a) *Task Completion Time* and (b) *Number of Errors*
- H4** During the task, participants have higher (c) *Confidence in Actions* after *VR Training* compared to *Video Training*

We expect participants trained in VR to be capable of performing a maintenance task on a real-world dosing pump, since they perform the same sequence of steps as in real-world training (H1), using digital copies of real-world objects. Due to the high fidelity of the VR simulation, we expect the three training methods to yield comparable performance (completion time and number of errors) (H2 and H3). Lastly, we expect that *VR Training* will lead to a higher level of confidence than *Video Training*, since instructions situated in an interactive 3D-environment are easier to understand and therefore cause less uncertainty than pre-recorded instructions presented on a 2D display (H4). We further endeavored to explore (RQ1) which parameters could be used to evaluate the effectiveness of *VR Training* in comparison to traditional training methods, and (RQ2) what design decisions critically impact the learning outcome in a *VR Training* simulation for a maintenance task.

4.1 Participants and Study Design

A total of 36 volunteers (4 female) participated in the study (age range: 20-35, avg. age: 26). These participants were selected from among the experimenters' social network and the campus of Aarhus University. None of the participants had prior experience with dis/assembling dosing pumps, and only 8/36 claimed at least some experience with maintenance of the mechanical and electronic appliances. Their apparent lack of technical expertise was not considered a limitation for this study, since the training scenario was

aimed primarily at novices. In the *VR Training* group 11/12 participants had experienced VR on prior occasions. Participants were quasi-randomly divided into three groups (12 participants per group): (1) *Pairwise Training* group, (2) *Video Training* group, or (3) *VR Training* group (shown in Fig. 1 A, B and C respectively). Care was taken to ensure a balance between male and female participants in each group. A between-subjects design was necessary for this study, since each participant had to learn a sequence of steps to perform the maintenance task on the dosing pump, and we wished to avoid a learning effect across conditions.

In all training conditions, participants were instructed to perform the maintenance task by following the steps presented in Table 1. The *Pairwise Training* group was guided through the steps by a researcher (always the same) who acted as instructor. The *Video Training* group learned the task by watching the instructional video. Both the *Pairwise Training* group and the *Video Training* group performed each of the steps on an actual dosing pump. The *VR Training* group was instructed in VR and completed the task on a virtual dosing pump.

After training, participants performed the task without guidance in a *post-test*. Here we recorded (a) *Task Completion Time*, (b) *Number of Errors*, and (c) *Confidence in Actions*.

Task Completion Time was measured from when participants first started performing maintenance, to when they completed the task - excluding time for donning safety gear, since participants spent a varying amount of time on this, and it was not deemed critical to the task.

Number of Errors was counted based on situations when participants did not perform a step correctly and did not correct the error right away. E.g., if they forgot a step and proceeded to the next one (sequential error), tried to attach a component in wrongly (orientation error), or forgot to use the safety equipment (safety error). These were registered on a checklist of possible errors, which were previously identified by a domain expert in a pilot study.

Confidence in Actions was given as subjective rating on a 5-point Likert scale.

4.2 Apparatus

Both the *Pairwise Training* and *Video Training* required a dosing pump and an extra o-ring that needed to be replaced. The components of a disassembled dosing pump can be seen in Fig. 4. Furthermore, safety glasses, gloves, a torque screwdriver and a needle tool were provided. The needle tool, similar to an awl, was used to remove the o-ring in the dosing pump. For the *VR Training*, participants were seated at a desk and provided with an HMD and two controllers. The VE included a virtual dosing pump, torque screwdriver, needle tool and safety gear, in correspondence with the materials required for the real-world task.

The *post-test* was conducted in a different room on a dosing pump of the same model. However, the setup differed in that the dosing pump was connected to a larger system including two other pumps and several pipes. The necessary tools and safety gear were again provided.

4.3 Procedure

The user study was conducted in a laboratory setting to minimize the risk of extraneous variables. Unless explicitly stated, each participant went through the same evaluation process, independent of their Training Type. Each step of this process is described in detail below, in chronological order.

After giving their informed consent and agreeing to video recording and anonymous data collection, participants provided demographic information and details about prior experience with VR, technical skills etc. Then each participant could inspect the tools and setup that would be used in the *post-test*. Participants were shown the

Figure 3: The VR simulation involves (A) a dosing pump and tools on a workbench. (B) Animations and highlights provide step-by-step guidance (red arrow indicates animation of virtual hand). (C) Yellow target silhouettes show where disassembled components should be placed.

dosing pump, and the components that needed to be dis/assembled in the task were pointed out (Fig. 4).

Independent of the condition, the training always involved the same 19 steps for the maintenance task (Table 1) and took approximately 20 minutes. During this time participants trained the task twice. In the *Pairwise Training* group the experimenter provided instructions from a manuscript. These included explanations for why certain actions were necessary and how different parts should be oriented. On the first run-through, the participant typically relied heavily on the instructor's guidance. On the second run-through they remembered many of the required actions and the instructor provided less guidance, based on how often participants were uncertain or asked for information. However, critical actions, such as using safety equipment, were always mentioned. In the *Video Training* group participants were guided through the maintenance task by a video shown on a PC, while performing the indicated steps on the physical dosing pump. They were allowed to pause the video and replay parts if necessary, but no other aids were given. In the *VR Training* group participants first spent about 5 minutes in a VR "playground" [5], where they could get familiar with the VE. They could also ask questions and request assistance during this introduction phase. Then followed the actual training simulation, first with guidance level set to Beginner, and then to Intermediate.

When the training was completed, each participant spent 6 minutes on a *mental rotation test* [28]. This test intended to distract the participant and diminish the recency knowledge effect between the training session and the post-test [5]. In the *post-test* the participants were asked to replace the o-ring on a physical dosing pump without assistance, to evaluate how well the respective Training Type prepared them for the task. During the design of the post-test multiple factors needed to be considered. Firstly, we needed to find out how much training to provide, to enable participants to successfully learn the task. Secondly, the level of difficulty needed to remain high enough to result in quantifiable errors that allow comparison of the Training Types. Thus, in this scenario the dosing pump was connected to a larger system, so that participants had to apply their learned knowledge in a slightly changed situation. Participants were informed that during the test *Task Completion Time* and *Number of Errors* would be measured. An experimenter was present to observe and assist in case of severe breakdowns. In a concluding semi-structured interview, participants were asked how confident they felt while performing the task in the post-test. This was evaluated through the following questions: "What parts of the task did you feel comfortable with?", "Were there parts you were uncomfortable or uncertain about?" and "Is there something you would have liked to do better at?".

The overall study duration for each participant was approximately one hour.

5 FINDINGS

5.1 Objective and subjective task performance

In regard to RQ1 (*which parameters could be used to evaluate the effectiveness of VR Training in comparison to traditional training methods*) the following parameters were identified as most relevant from among the related work and were chosen as basis for our evaluation to allow for comparison with earlier studies: (1) *Task Completion Time* and (2) *Number of Errors* provided objective measures, whereas (3) *Confidence in Actions* was subjective. These three parameters make it possible to compare the Training Types on two aspects of task performance: how well each participant performed objectively, and how comfortable participants felt subjectively, while performing the task. To test for significant differences between the Training Types, a Kruskal-Wallis test was carried out ($\alpha = 0.05$). This non-parametric test was also used for *Task Completion Time*, since Bartlett's test showed a violation of homogeneity of variances ($\chi^2 = 12.852, p < 0.01$). Pairwise comparisons were made by post-hoc Mann-Whitney test with Bonferroni correction.

Task Completion Time and Number of Errors: The average *Task Completion Times* are visualized in Fig. 5 (A), and given in Table 2. On average participants took about 1.6 times longer for the maintenance task in the post-test after *VR Training*, compared to both other groups. This effect is significant ($\chi^2 = 12.65, p < 0.01$; Posthoc: *VR Training* - *Pairwise Training* $p < 0.01, r = 0.61$ — *VR Training* - *Video Training* $p < 0.01, r = 0.62$).

There is also a significant effect of Training Type on *Number of Errors* ($\chi^2 = 17.44, p < 0.01$), where pairwise comparisons again showed differences between the *VR Training* and both the *Pairwise Training* ($p < 0.01, r = 0.76$) and the *Video Training* ($p < 0.01, r = 0.63$). As is evident from Fig. 5 (B) and the values in Table 2, the *VR Training* group made more than 3 errors on average, while both other groups made about 1.

A Pearson's correlation test shows a positive association between *Task Completion Time* and *Number of Errors* ($r = 0.73, p < 0.01$). We found no connection between the task performance and participants' prior technical expertise.

Table 2: Mean (M) and standard deviation (SD) of the objective task performance measures for each training group: *Task Completion Time* (in minutes and seconds) and *Number of Errors*.

Training Type	Task Completion Time		Number of Errors	
	M	SD	M	SD
Pairwise Training	07:41	01:21	0.67	0.78
Video Training	07:21	02:07	1.3	0.98
VR Training	12:04	04:21	3.75	1.82

Figure 4: Components of the disassembled dosing pump: (a) Top ball valve (including the multi-function valve); (b) Drainer; (c) Dosing head; (d) Back lid; (e) Bottom ball valve; (f) Screws; (g) Diaphragm; (h) o-ring.

Confidence in Actions: As illustrated in Fig. 5 (C), the participants' *Confidence in Actions* may be a reliable indicator of their objective task performance, since *Pairwise Training* and *Video Training* generally led to higher confidence than *VR Training*. This effect was again significant ($\chi^2 = 9.38, p < 0.01$; Posthoc: *VR Training - Pairwise Training* $p < 0.01, r = 0.57$ — *VR Training - Video Training* $p < 0.01, r = 0.49$).

5.2 Common types of Errors

Based on direct observation and analysis of video recordings, the following errors occurred most commonly.

Orientation error: When reassembling the dosing pump some participants had difficulty figuring out the correct orientation of components, such as the drainer and dosing head (Fig. 4 b and c). In the *VR Training* group 5/12 participants required multiple attempts to make components fit, versus 2/12 in the *Video Training* group and 0/12 in the *Pairwise Training* group.

Sequential error: Another common error was that participants forgot a step and proceeded to the next one, e.g., forgetting to remove the top and bottom ball valves (Fig. 4 a and e). This happened most frequently in the *VR Training* group (7/12), compared to 1/12 in the *Pairwise Training* group and 0/12 in the *Video Training* group.

Safety issue: The third, and perhaps most critical error, was neglecting the safety equipment, e.g., not putting on the safety glasses. Here both the *Video Training* group (4/12) and the *VR Training* group (3/12) were more negligent than the *Pairwise Training* group (1/12).

5.3 Observations and interview results

Pairwise Training: As expected, there was a recurring behavioral pattern in the first and second training session in *Pairwise Training*: On the first run-through participants engaged in a conversation with the experimenter and asked questions while they were being instructed. The instructor only pointed towards the parts that needed to be (dis)assembled. On the second run-through participants did the maintenance more independently, requiring less help. During the post-test, most participants struggled with the orientation of the drainer (see Fig. 4 b) when assembling the dosing pump, which one participant mentioned explicitly during the interview as a step that they wanted to be better at. Otherwise reported difficulties were unremarkable.

Video Training: In the *Video Training* group participants exhibited three different strategies: The first strategy was to watch the video, while simultaneously performing maintenance on the dosing pump; the second strategy was to fully focus on the video first, and

only do the maintenance task on the dosing pump while watching the video the second time; the final strategy was to perform maintenance on the dosing pump while watching the video the first time, and trying to perform the maintenance without help of the video in the second round, referring to it only when they got stuck.

Participants in the *Video Training* group seemed to struggle with certain parts of the *Video Training*, specifically when trying to remove the diaphragm (Fig. 4 g) from the dosing pump which one participant also mentioned during the interview. This was mainly due to how the video was structured, and the viewing angle from which the video was recorded, which made it difficult to see certain details, such as that the diaphragm needed to be twisted to be removed. This issue occurred in several parts of the video, as many components of the dosing pump are symmetrical, and was something multiple participants from the *Video Training* group commented on during the interview. Further, 4/12 participants in this group mentioned having struggled with one of the valves.

VR Training: During this training method, multiple challenges were observed, both regarding the interaction techniques and in regard to remembering the sequence of steps.

When interacting with objects in the Sandbox Environment, some participants were confused about how to loosen or fasten screws when holding a torque screwdriver in front of it. Also, some participants were confused about how to release a screw from the torque screwdriver. Both actions were clarified by the observing experimenter. It appears that with the detailed step-by-step guidance given in the first iteration (beginner level) participants failed to reflect on what they were doing.

In respect to memorizing the sequence of steps, some participants struggled in the second run-through of the training, when less guidance was given (intermediate guidance level).

Another observation was that many of the participants did not pay attention to the orientation of individual components, while reassembling the dosing pump (e.g., the drainer component that had to be mounted in a specific orientation). This was probably because all the objects in the virtual training simulation automatically snapped to their correct orientation when placed within the target silhouette. Therefore, participants did not learn the specific orientation of components, but only the specific step in which these needed to be attached to the dosing pump.

During the semi-structured interview, a majority of participants in this group (7/12) mentioned having felt very unsure about the order in which components needed to be dis-/assembled. No specific components or steps were singled out as problematic.

Figure 5: These boxplots show the effectiveness of each training method, based on three evaluation parameters: (A) *Task Completion Time*, (B) *Number of Errors*, and (C) *Confidence in Actions* (rated on a 5-point Likert scale).

6 DISCUSSION OF FINDINGS

We predicted that participants trained in VR would be capable of performing a maintenance task on a real-world dosing pump, since they performed the same sequence of steps in a high-fidelity simulation (H1). This hypothesis was supported, since all participants in the *VR Training* group were capable of completing the maintenance task after just 20 minutes of training. VR training has the benefit of not requiring the availability of specific equipment or tools, apart from the VR setup. Further, it is scalable to service multiple trainees at the same time and can serve to train many tasks on an arbitrary number of machines in various environments and configurations. However, the success of this medium is arguably task-dependent, and it remains to be explored what kind of skill acquisition is best supported by *VR Training*.

Based on the opinion of training experts, we expected the three training methods to result in comparable task performance (H2 and H3). Our results do not support these hypotheses, since in the *post-test* the overall performance (*Task Completion Time*, *Number of Errors*) was significantly better in both *Pairwise Training* and *Video Training* compared to *VR Training*. This is in line with results reported by Werrlich et al. [30], according to which HMD-based training for an assembly process led to significantly longer completion times, compared to paper-based training. However, in contrast to our results, the authors found that HMD-based training led to fewer errors. Overall, our results do not reflect the promising prospect of VR portrayed by other researchers [23]. This contrast may be due to the richness of hands-on experience, which is very difficult to simulate in VR, and may present a great benefit to learning. This is supported by the work of Lok et al. [24], which shows that when performing a cognitive manual task within a VE, manipulating physical objects results in significantly improved performance, compared to manipulating virtual objects. A further contributing factor may be *contextual changing*, which is arguably lower in hands-on training methods, since training environment and equipment vary less between training and post-test scenarios [22]. In this regard, a promising avenue to explore in future may consist of combining *VR Training* with short hands-on sessions, to allow a fairer comparison with other forms of training. It is also notable that *Pairwise Training* allows for guidance to be tailored to the trainees' needs based on the questions asked, which could be considered an explanation for its superiority over *VR Training*. However, this flexibility to tailor instructions is not given in *Video Training*, which led to comparable results as *Pairwise Training*, thus we don't see this line of argumentation supported.

Lastly, we expected that *VR Training* would instill a higher level of confidence than *Video Training*, since multimodal instructions presented in an interactive 3D-environment are richer and provide a better understanding of what to do, compared to an instructional

video seen on a 2D display (H4). This hypothesis was flawed, since it is based solely on how instructions are presented, and does not consider how the trainee follows these instructions: While in *VR Training* participants manipulated a virtual pump, in *Video Training* participants could explore the physical dosing pump and learn through trial and error.

While the comparison of *VR Training* with two Training Types that involve hands-on experience may be unfair, these present the training methods that are currently applied for such tasks and *VR Training* is widely being considered equally effective. Our evaluation may serve to highlight limitations of *VR Training* and demonstrate how hands-on experience impacts the learning process. The presented results may also suggest that supporting training with AR could be superior to a VR simulation, since this allows the combination of hands-on experience on the physical machine with digital assistance through animations and step-by-step instructions (e.g., via HMD [30] or projection [9]). However, AR poses additional requirements, such as reliable tracking of relevant physical components, which may be challenging due to visual and structural properties (e.g., highly reflective materials or near-symmetrical components [2]). Furthermore, it relies on the availability of the physical machine, components and tools, and the ability to stage the scenario that should be trained, which may incur additional costs, difficulties and safety concerns. Hence, VR remains a compelling option, not least due to its purely artificial nature, which enables training of otherwise impossible scenarios, e.g., including products that do not yet exist.

7 LIMITATIONS AND FUTURE WORK

7.1 Simulating hands-on experience in VR

The presented VR training simulation takes a first step towards simulating hands-on experience in VR by requiring tool-use and realistic hand motions (e.g., turning the wrist to tighten a screw), as well as providing haptic feedback through vibrations of the controller and clear visual indications of progress. However, our results show that still richer multimodal feedback needs to be provided - in particular for critical cues, such as components snapping together only when assembled with correct orientation. Future research should explore better communication of such cues, e.g., through force feedback [19]. Further, enabling interaction through hand tracking, instead of controllers, might allow better training of object manipulation steps and correct use of tools in the VE. Hand tracking might be achieved with optical sensors, like the Leap Motion controller⁶, or through hand-tracking gloves (e.g., Cyber Glove⁷ or Manus VR⁸). The

⁶Leap Motion <https://www.leapmotion.com/>

⁷Cyber Glove <http://www.cyberglovesystems.com/>

⁸Manus VR <https://manus-vr.com/>

latter also provide the opportunity for integrating haptic feedback modules (e.g., vibration motors, pressure pads, exoskeleton).

7.2 Balancing the level of guidance

When designing a *VR Training* tool, it is important to ensure effective step-by-step guidance through the given task, while setting an adequate level of challenge to foster learning. The extensive amount of hints given in the first VR training session (*beginner* level), may prevent users from reaching the *cognitive stage* of skill acquisition. Therefore, we reduced the level of guidance in the second session (*intermediate* level). Future work could explore the use of more or dynamic levels of difficulty. It should thereby be considered that not only too much but also too little guidance may be harmful. If the user is left not knowing how to progress in the training, this can lead to frustration and diminish their willingness to learn. Thus, the possibility to request additional information should be provided.

This also raises the question, which type of guidance is most beneficial (e.g., instructions via audio recordings vs. text). Based on our observations, in the first run-through participants consistently listened to the full audio recording and relied on this information to perform the next task. During the second run-through they performed the steps without waiting for the recordings to finish, whenever they remembered enough of the instructions from the first round. At the same time, participants were observed to rely more on the visual text in this second round, which may stem from the diminished amount of guidance they received through highlights and animations. This may suggest that it is beneficial to offer redundant information across multiple sensory channels. Beyond this, the impact of individual types of guidance remains for future work to explore.

Further, we acknowledge that one of the main limitations of our *VR Training* simulation may be not supporting a trial-and-error approach for learning. In contrast to *Pairwise Training* and *Video Training*, the *VR Training* simulation only allowed users to pick up the *correct* component in each step. This removed the possibility of errors by choosing wrong components, which arguably reduces the need to think about what objects to interact with. Also, the snapping of components into their correct orientation may have hindered learning. Arguably, the transfer of learning from VR to the physical world may be increased by conveying more details, such as the structural properties of components. For example, when assembling the pump, components should only lock into their target position if placed there in correct orientation, instead of automatically snapping into place as soon as they are brought into close enough proximity. Future work on training of dis-/assembly tasks could aim to address this, by allowing interaction with all available virtual components at any given time, and more faithfully modelling structural properties that indicate how they may fit together.

It remains to be mentioned that these factors, which we have highlighted as critical for the learning outcome, may not be of equal importance at all levels of expertise. In the presented user study all participants were non-experts, with very limited experience in general maintenance tasks. While we would expect similar results if the study was repeated with (novice) trainees at Grundfos, it may well be that users who already possess general knowledge of the basic processes and characteristics of common dosing pump components, may have less need for exploring the steps of this maintenance task through trial and error.

7.3 Defining the duration of training sessions

In each of the three training methods we compared, participants practiced the full maintenance task exactly twice and we limited the overall training duration to 25 minutes. This was based on guidelines set by Grundfos, where VR training sessions are generally limited to a duration of approximately 20 minutes to diminish the probability of cybersickness. In our experiment, all participants in the *VR Training* group were informed about the risk of cybersickness and were

asked to report any symptoms, so that the study could be interrupted. However, no issues were reported. Further, across all training methods, there was rarely a need to enforce the time limit, since none of the participants felt pressed for time. However, no lower limit for training duration was defined, which means that actual training times could vary between participants, if some completed training steps faster in *VR Training*, paused and replayed the video less often in *Video Training*, or asked less questions in *Pairwise Training*.

Interestingly, in the *Pairwise Training* group most participants commented that they already felt confident of their ability to perform the maintenance task after only a single training session. This may indicate that with some training methods the *associative stage* of skill acquisition may be reached earlier than with others. Required training times and achieved competency may depend on the training method, as is supported by Oren et al. [25] who found that it took participants three times longer to train in VR, but that they were also three times faster in assembly during a post-test, when comparing to real-world training. Future evaluations should therefore consider avoiding setting a time limit or limiting number or training sessions, but instead allow participants to practice to competency. Different training methods could then be compared, based on the time and number of sessions needed to achieve a certain level of expertise, by incorporating training sessions [5] with test phases in between. Such studies might also explore the question of how many VR Training repetitions would be necessary to successfully accomplish a given task. This also raises the question whether our findings might have been different, if participants had been allowed to repeat the VR training more often. While this remains to be confirmed, we deem this unlikely, since the current VR simulation does not reflect some characteristics of the physical task, which the participants struggled with in the post-test (e.g., placing the drainer in the correct orientation and loosening/tightening the diaphragm).

8 CONCLUSION

This case study presents a *VR Training* simulation for a maintenance task on a dosing pump and an evaluation thereof in comparison to traditional training methods applied at Grundfos (*Pairwise Training*, *Video Training*). The three experimental groups were compared based on the participants' performance in terms of (a) *Task Completion Time*, (b) *Number of Errors*, and (c) *Confidence in Actions*. Study results reveal that *VR Training* was effective in teaching the training task. However, both hands-on training methods, *Pairwise Training* and *Video Training*, led to significantly better performance than *VR Training*.

9 ACKNOWLEDGEMENTS

We would like to thank Grundfos for generously supplying the material needed to explore this case. Furthermore, we would like to thank all study participants. This project was partly funded by MADE Digital (Innovation Fund Denmark, grant no. 6151-00006B).

REFERENCES

- [1] W. Arthur Jr, W. Bennett Jr, P. L. Stanush, and T. L. McNelly. Factors that influence skill decay and retention: A quantitative review and analysis. *Human performance*, 11(1):57–101, 1998.
- [2] J. Belo, A. Fender, T. Feuchtnner, and K. Grønbaek. Digital assistance for quality assurance: Augmenting workspaces using deep learning for tracking near-symmetrical objects. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, pp. 275–287, 2019.
- [3] J. Blake and H. B. Gurocak. Haptic glove with mr brakes for virtual reality. *IEEE/ASME Transactions On Mechatronics*, 14(5):606–615, 2009.
- [4] A. C. Boud, D. J. Haniff, C. Baber, and S. Steiner. Virtual reality and augmented reality as a training tool for assembly tasks. pp. 32–36, 1999.

- [5] P. Carlson, A. Peters, S. B. Gilbert, J. M. Vance, and A. Luse. Virtual training: Learning transfer of assembly tasks. *IEEE transactions on visualization and computer graphics*, 21(6), 2015.
- [6] I. Choi, E. W. Hawkes, D. L. Christensen, C. J. Ploch, and S. Follmer. Wolverine: A wearable haptic interface for grasping in virtual reality. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 986–993. IEEE, 2016.
- [7] L. Connelly, Y. Jia, M. L. Toro, M. E. Stoykov, R. V. Kenyon, and D. G. Kamper. A pneumatic glove and immersive virtual reality environment for hand rehabilitative training after stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(5):551–559, 2010.
- [8] P. M. Fitts and M. I. Posner. *Human performance*. Brooks/Cole, 1967.
- [9] M. Funk, T. Kosch, and A. Schmidt. 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*, pp. 934–939, 2016.
- [10] A. G. Gallagher, N. E. Seymour, J.-A. Jordan-Black, B. P. Bunting, K. McGlade, and R. M. Satava. Prospective, randomized assessment of transfer of training (tot) and transfer effectiveness ratio (ter) of virtual reality simulation training for laparoscopic skill acquisition. *Annals of surgery*, 257(6):1025–1031, 2013.
- [11] N. Gavish, T. Gutiérrez, S. Webel, J. Rodríguez, M. Peveri, U. Bockholt, and F. Tecchia. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6):778–798, 2015.
- [12] S. Goerger, R. Darken, M. Boyd, T. Gagnon, S. Liles, J. Sullivan, and J. Lawson. Spatial knowledge acquisition from maps and virtual environments in complex architectural space. 1996.
- [13] T. P. Grantcharov, V. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch-Jensen. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British journal of surgery*, 91(2):146–150, 2004.
- [14] X. Gu, Y. Zhang, W. Sun, Y. Bian, D. Zhou, and P. O. Kristensson. Dexmo: An inexpensive and lightweight mechanical exoskeleton for motion capture and force feedback in vr. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1991–1995. ACM, 2016.
- [15] C. R. Hall, R. J. Stiles, and C. D. Horwitz. Virtual reality for training: Evaluating knowledge retention. pp. 184–189, 1998.
- [16] C. J. Hamblin. Transfer of training from virtual reality environments. 01 2005.
- [17] T. Han, F. Anderson, P. Irani, and T. Grossman. Hydroring: Supporting mixed reality haptics using liquid flow. In *The 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 913–925. ACM, 2018.
- [18] C. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE transactions on visualization and computer graphics*, 18(7):1170–1188, 2011.
- [19] R. Hinchet, V. Vechev, H. Shea, and O. Hilliges. Dextres: Wearable haptic feedback for grasping in vr via a thin form-factor electrostatic brake. In *The 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 901–912. ACM, 2018.
- [20] C. R. Jacquet. A study of the effects of virtual reality on the retention of training. Technical report, University of Central Florida, Orlando, 2002.
- [21] J. Jerald. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, 2015.
- [22] J. Kozak, P. Hancock, E. Arthur, and S. Chrysler. Transfer of training from virtual reality. *Ergonomics*, 36(7):777–784, 1993.
- [23] A. Langley, G. Lawson, S. Hermawati, M. D’Cruz, J. Apold, F. Arlt, and K. Mura. Establishing the usability of a virtual training system for assembly operations within the automotive industry. *Human Factors & Ergonomics in Manufacturing & Service Industries*, 26(6), 2016.
- [24] B. Lok, S. Naik, M. Whitton, and F. P. Brooks. Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(6):615–628, 2003. doi: 10.1162/105474603322955914
- [25] M. Oren, P. Carlson, S. Gilbert, and J. M. Vance. Puzzle assembly training: Real world vs. virtual environment. In *2012 IEEE Virtual Reality Workshops (VRW)*, pp. 27–30. IEEE, 2012.
- [26] F. Rose, E. A. Attree, B. Brooks, D. Parslow, and P. Penn. Training in virtual environments: transfer to real world tasks and equivalence to real task training. *Ergonomics*, 43(4):494–511, 2000.
- [27] R. J. Seidel and P. R. Chatelier. *Virtual reality, training’s future?: perspectives on virtual reality and related emerging technologies*, vol. 6. Springer Science & Business Media, 2013.
- [28] S. G. Vandenberg and A. R. Kuse. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and motor skills*, 47(2), 1978.
- [29] S. Webel, U. Bockholt, T. Engelke, M. Peveri, M. Olbrich, and C. Preusche. Augmented reality training for assembly and maintenance skills. In *BIO web of conferences*, vol. 1, p. 00097. EDP Sciences, 2011.
- [30] S. Werrlich, A. Daniel, A. Ginger, P.-A. Nguyen, and G. Notni. Comparing hmd-based and paper-based training. 10 2018.
- [31] B. G. Witmer, W. J. Sadowski, and N. M. Finkelstein. Training dismounted soldiers in virtual environments: Enhancing configuration learning. Technical report, ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ORLANDO FL ..., 2000.
- [32] R. S. Woodworth and E. L. Thorndike. The influence of improvement in one mental function upon the efficiency of other functions (i). *Psychological review*, pp. 247–261, 1901.