Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (34^{th}cycle)

eXtended Reality for Education and Training

By

Filippo Gabriele Pratticò

******

Supervisor(s):
Prof. Fabrizio Lamberti

Doctoral Examination Committee:
Prof. Sofia Seinfeld Tarafa, Referee, Universitat Politècnica de Catalunya, Spain
Prof. Lucio Tommaso De Paolis, Referee, Università del Salento, Italy
Dr. Giacinto Barresi, Istituto Italiano di Tecnologia (IIT), Italy
Dr. Fabrizio Nunnari, German Research Center for AI (DFKI), Germany
Prof. Andrea Bottino, Politecnico di Torino, Italy

Politecnico di Torino
2022

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Filippo Gabriele Pratticò
2022

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Acknowledgements

I wish to give thanks to all people who in one way or another have supported, collaborated, and inspired me during the last four years.

First of all, I would like to thank my academic supervisor, Prof. Fabrizio Lamberti for the mentorship he provided me during the Ph.D. period, by making constructive criticisms and invaluable suggestions that often were way beyond his bare duties.

I would also bring recognition to the VR@POLITO initiative which provided me with the equipment needed for conducting my research, and this fact was of exceptional value during the pandemic period.

Also, I would like to gratefully thank all my lab colleagues which shared the joys and tears of this journey.

A special place is surely dedicated to my friends for their patience and the comprehensive support they gave to me during these years. I especially want to thank, among the others: *Gialex*, whose path is entangled so much to mine that life brings us together whenever we need the most; *Simone*, that had the love and the commitment to teach me priceless lessons by example from which I drew extensively during the Ph.D. period and aided me to reach this result way easier compared to what could have been if I never met him; *Giorgio*, which I love as a brother, and makes me every time feel like having additional family, and I thank for all the deep chats, and I'm grateful for having learned my preferred joy and expression tool from him; *Francesca*, as the person that, even if our life paths diverged in recent times, more than anyone else has motivated and supported me in manifold and unique ways every day of my life for the all Ph.D. period, and to which I acknowledge that this result is in a relevant part thanks to.

A final, but most emotional, thought goes to the family and in particular to my parents that nurtured me towards this goal with blind faith and pure love.

Abstract

The last decade has witnessed unprecedented breakthroughs in the field of eXtended Reality enabling technologies, accompanied by a growing variety of cost-effective devices hitting the market, also at the consumer level. The increased accessibility and the disruptive potential of this family of immersive media are catalyzing the interest of both the academia and the industry, which are putting many efforts into helping them attain maturity and become commonplace in a wide range of application fields, encompassing arts, design, engineering, architecture, medicine, and so forth.

Undoubtedly, training and education were advocated as two of the most promising applications which can benefit from such immersive media, thus being the subject of studies since the early days of Virtual and Augmented Reality. Many advantages have been recognized to eXtended Reality training systems, such as the possibility to experience a given scenario under repeatable and controlled conditions even under circumstances that could be potentially hazardous, impractical, or very resource-intensive if recreated for real. With the advancements in the field, novel challenges to address and limitations to overcome emerged towards the seamless adoption of such training system at a mass scale.

The work that the author carried out during the Ph.D. period, partly presented in this thesis, was aimed at expanding the boundaries of eXtended Reality-based tools used in the education and training contexts. Specifically, the attention was focused onto three research directions: firstly, with the aim of supporting their deployment at scale, the employment of such systems with a self-learning approach, i.e., without the need for a human trainer to partake was explored; secondly novel ways to employ pedagogical agents in eXtended Reality training systems were investigated, both to ameliorate the social-related aspect of such experiences and to enable unconventional pedagogical models; lastly, the exploitation of eXtended Reality-based tools from the often-overlooked training provisioner perspective was considered.

It is the author's hope that the work presented in this document can offer interesting insights and pave the way for future research in the considered domain.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and Objectives

Since the first steps moved by I. Sutherland back in the 60's in the computer graphics domain [1], there have been many speculations about ways in which a user could experience a simulated virtual environment (VE).

The perspective of having a human interacting within a perfect VR was undoubtedly fascinating. More formally, in the 90's, Milgram and Kishino introduced the so-called "reality-virtuality continuum" [2] to classify such experiences and respective media. In the light of developments in the field that followed, this seminal idea was later extended and revised [3]. The classification is based on two factors: the extent of world knowledge, and the degree of immersion. The first factor considers how much the experience traces the real world and how much the system is aware and capable of integrating the user's body, objects, and surroundings into the simulation. The second factor refers to the capability of the system to deceive the human senses by superseding the stimuli coming from the reality with those of the simulated VE.

Today, this family of media is referred to under the eXtended Reality (XR) umbrella term. Formerly sharing the meaning with Mixed Reality (MR), recently the term started to be used more often to frame a kind of experiences that are not fully immersive, like the VR ones are, but are featured with an extent of world knowledge greater than that of Augmented Reality (AR) experiences (which are instead similar in terms of the degree of immersion [4]); in other words, the key difference between AR and MR is that, in the former, the digital content is just overlaid on top of the real world within the user's field-of-view (FOV), whereas, in the latter, it is fully blended and contextualized with the real environment [4].

In the last decade, there has been an unprecedented breakthrough in the field of XR enabling technologies, together with a growing variety of cost-effective devices hitting the market, also at the consumer level. This hardware usually comes in the form of Head-Worn Devices (HWDs), which provide the user with the visual and aural feedback coming from the simulation by tracking the head-movements in six degrees of freedom (DOFs) [5]. Furthermore, interaction plays a key role in XR interactive experiences. Considering, e.g., the education domain, the authors of [6] demonstrated that if a trainee is not allowed to interact with the VE, the resulting training effectiveness is lower than that of a bare slide-supported classroom lecture. Hence, the HWD is often complemented with some pieces of technology capable to track at least the user's hands and allowing him or her to interact with the VE. This is implemented via built-in hand tracking features of the HWD based on cameras and computer vision techniques or, more frequently, with tracked hand-held controllers [5]; the latter is indeed the most affordable, usable, and robust choice to deploy at scale, considering the current limitations of the former option [7].

The increased accessibility and the disruptive potential of these media are catalyzing the interest of both the academia and the industry, which are putting many efforts into helping them attain maturity and become commonplace in a wide range of application fields encompassing arts, design, engineering, architecture, medicine, and so forth [8].

Undoubtedly, from the early days of XR, training and education were advocated as two of the most promising applications, which attracted a lot of the interest on the various media in the family (VR, AR, and MR). Indeed, many advantages have been recognized to XR training systems (XRTSs). With the aid of a simulated VE, trainees can experience a given scenario under repeatable and controlled conditions, enabling training circumstances that could be potentially hazardous, impractical, or very resource-intensive if recreated for real [9, 10]. This is for instance the case of first responders (FRs) or pilots training [11–13]. As a matter of fact, the aviation industry adopts simulation-based training as a well-established practice, since it was demonstrated that pilots previously trained in simulators need less in-flight training to meet a satisfactory level of proficiency [14]. Thanks to the induced sense of presence and embodiment, XR can also represent an effective way to overcome distance barriers, by preventing to relocate people to training facilities and providing a shared VE in which multiple people can act, thus enabling social experiences at distance.

Under these premises, it is not surprising that XRTSs are stepping outside from the academic laboratories and military-grade facilities, getting more and more frequently integrated into the training programs of companies and institutions [15]. Nonetheless, there are still challenges and open points towards the seamless adoption of XRTSs at a mass scale, and even greater potential to unlock for what it concerns XR-based tools usages in educational contexts besides the conventional, simulation-based pedagogical approaches [9, 10, 16, 17].

1.2 Thesis Organization

Thesis Goal: This document presents part of the work that the author carried out during the Ph.D. period with the aim to expand the boundaries of XR-based tools usage in the education and training contexts.

Thesis Structure: Chapter 2 focuses on the scalability of VR training systems (VRTSs), by exploring the employment of such systems with a self-learning approach, hence without the need for a human trainer to intervene; specifically, a state-of-the-art VRTS is evaluated against a real-world training, by outlining the strengths and limitations, as well as possible countermeasures that could ameliorate some of the VRTS distressing factors that emerged from the comparison. Stepping onto the concept of guidance systems (GSs) used in self-learning-oriented VRTSs, Chapter 3 studies how to leverage the so-called pedagogical agents (PAs) in novel ways; in particular, it explores the use of a PA to enable unconventional pedagogical models in a MR-based training system, and by proposing the use of a PA to improve the learning experience in an immersive distance learning setting. Finally, Chapter 4 considers the (often overlooked) training provisioner perspective focusing, in particular, on the research direction of authoring tools, and presenting an approach that leverages an XRTS to support training design.

Chapter 2

XRTSs for Self-Learning

The work described in this chapter has been formerly published in [18, 19]

Despite the rising interest and integration in the companies training programs, XRTS are still mainly used as a complement to established training practices rather than as a cogent replacement [9, 20]. Surprisingly, this fact still stands as true also in the case of XRTSs that supports the self-learning/-assessment of trainees. These are systems that inherit the concepts and features from simulation-based XRTSs, and evolve them by adding the capability to instruct and provide step-by-step guidance to the trainee by using multi-modal (video clips, voice-overs, visual hints, instructional animations, etc.) scaffolding/guidance systems [10]. As a matter of fact, the prospect to use an XRTS to transfer the intended knowledge and skills to a trainee without the need for a human trainer intervention sounds to be a remarkably valuable opportunity, especially for learning subjects that could benefit from *hands-on* and *learning-by-doing* training approaches [21]. Having at disposal an intelligent tutoring system (ITS) that automatically trains the given audience is indeed a key factor for deploying training at scale. Certainly, this envisaged scenario is subordinate to proving that such automatic XRTSs are at least as effective as traditional approaches in achieving the learning goals [9].

One of the domains that is currently giving great momentum to XRTSs development is the industrial one, being XR recognized as a key element in the Industry 4.0 vision [16]. This domain offers plenty of use cases for training, which are indeed challenging from a didactic standpoint as they fall in the *mixed-tasks* category [9]. Those are tasks in which coalescence of both cognitive and physical requirements shall be reached in order to successfully achieve the goal [21]. Within such domain, a representative class of tasks that fits these attributes are the so-called machine tasks (MTs). MTs are those in which a human operator, in a factory or workshop setting, is mandated to execute a mixed series of different interactive steps on a machine or a piece of equipment [22]. Thus, it should not be surprising that industrial assembly and maintenance (IMA) tasks are the most studied tasks in this domain, being them a perfect exemplification of MTs (and consequently of mixed-tasks): in fact, in such a context a trainee is asked to spend the gained skills while simultaneously recalling applicable procedural information, and be aware of possible safety risks [23, 21, 9]. Common actions encompassed by an IMA task are, for instance, fixing connectors, manipulating and mounting objects either directly by hand or through tools or machine-specific remote controllers, machinery programming, and so forth.

In this context, MR training systems (MRTSs) have already proved their capability to be as effective as (if not more effective than) traditional methods when it comes to transferring the desired knowledge to trainees. Despite the still existing technological limitations, paper-based instructions are easily outperformed by MR solutions capable of providing timely step-by-step instructions directly superimposed to the piece of equipment or tool being used [24], without impeding physical interaction.

The above qualities, coupled with the somehow less complicated process of arranging a MR experience [25] w.r.t. a VR experience (considering, e.g., the lower simulation complexity and the fewer 3D assets required), set MRTSs as a compelling solution. Nevertheless, the necessity to count on physical equipment to deploy the MR experience represents at the same time the primary drawback of MRTSs, which hampers their flexibility and scalability compared to VRTSs [9].

Unfortunately, according to the literature, the status of self-learning VRTSs is yet less defined [9, 10, 21]. The body of research is scattered, and results are oftentimes controversial or hardly generalizable to real scenarios (because, e.g., of the hardware employed, of over-controlled experimental conditions, etc.). Furthermore, this situation is heightened by the arguably task-dependent efficacy of VRTSs [26].

This chapter first evaluates the effectiveness of self-learning-oriented VRTSs in the context of mixed-tasks training by reporting the findings in Section 2.1. By building on the aftermath of this first study, Section 2.2 proposes a potential solution to mitigate one of the identified issues.

2.1 Effectiveness of Self-Learning VRTSs

Even though it has been demonstrated that the intrinsic qualities of VRTSs are enough to ensure a more effective knowledge transfer w.r.t. self-studying printed, static material (like in the case of an aircraft's safety card [27]), this is far from the scenario of MTs training. As a matter of fact, learning systems arranged by companies to train operators on MTs are more articulated than bare manuals reading (paper-based instructions): for instance, they often rely on human tutors for face-to-face or pairwise teaching [26], and promote the trainees to accumulate hands-on experience on the physical machinery [21].

Motivated by the paucity of studies that examined the learning transfer performance of automatic VRTSs outside laboratory settings, and in particular, against the well-established strategies adopted "on-field" by companies training departments [9, 10], this section presents a user study that was aimed at assessing the training effectiveness of a VRTS for the self-learning of an IMA task. Specifically, it was decided to focus on a real maintenance procedure that is performed on industrial robots (IRs). More explicitly, this work is driven by the following two research questions:

- RQ.1 Can an automatic VRTS guarantee a transfer of knowledge and skills comparable to that of traditional training? Specifically, is an individual trained with such a system able to perform a self-learnt procedure on a real IR with a level of performance comparable to that of an individual trained in a traditional way?
- RQ.2 From a trainee perspective, is the training experience with a self-learning VRTS less satisfactory compared to a traditional learning session with a human instructor?

2.1.1 Background

Authors of [26] have recently experimented with a VRTS for IMA by using just consumer-grade hardware and evaluated it in a concrete training scenario. The scenario was about pump maintenance and the VRTS performance was compared against the two training approaches adopted on daily-basis by a company involved in the study, i.e., video-based training and pairwise tutoring. According to the reported evidence, it was observed that the effectiveness of VR-based training was significantly inferior to that of the usual approaches, despite the fact that trainees who used the VRTS acted adequately when asked to repeat the procedure on a physical pump. Even though the VR experience was designed with the appropriate care to facilitate the understanding of tackled content (i.e., the correct steps to perform were exemplified using 3D animations), the VRTS failed to convey the feedback that the trainees actually needed to manipulate tools and small parts involved in the task.

In training situations using IRs, however, the operating conditions may differ from those considered in [26]. In the above study, apart from the dissimilarities in the interactions involved, only small-scale movements were considered. The task could have been performed by the trainees in a seated-only position or on a table, which might not be possible when operating with medium to large-sized equipment. When accomplishing a maintenance task on an IR, an operator may be obliged to walk about the surroundings, which necessitates that he or she uses proprioception and spatial awareness skills in order to be always aware of the hazards and perhaps of the dangerous behaviors of a moving machine. Even tough, in VR, enabling interaction is key to foster proprioception [28], unfortunately this is not enough to also uphold the transfer of spatial knowledge, since other influencing factors are involved like the simulation fidelity [9], the time spent in the VE [29, 30], etc. Studies like, e.g., [31] indicated that immersive VR is superior w.r.t. non-immersive experiences (desktop-based VR) when it comes to foster spatial knowledge transfer. An investigation that specifically focuses on spatial awareness is reported in [32]. In particular, a simulated manufacturing task with a cooperative IR was chosen to challenge an operator on space sharing management, and an immersive VRTS was used to train him or her. The qualitative findings outlined that the fist encounter with an IR could be substantially ameliorated by employing the VRTS; nonetheless, it must be noted that the effectiveness of the training was not further assessed on a real IR.

Authors of [20] instead arranged a user study using also a real robot to evaluate if a VRTS could be an effective tool for training robot programming abilities. It was observed that users who underwent the standard training performed worse than those trained with the VRTS. However, diversely from the previous study, the VRTS was employed just in a simulation-based fashion, i.e., complementing the instructional material provided in a traditional classroom setting with a human instructor, rather than for an automatic self-tuition. Furthermore, the study devised to evaluate the VRTS entailed as control condition users who has been just trained with theoretical notions. Even though this condition was indeed appropriate for the goals of the investigation, this education approach may not be entirely representative of real-world circumstances in which theoretical training is frequently accompanied with hands-on practice.

In [33], another study that utilized a VRTS to teach robot programming and display trajectories execution is reported. The presented system just delivers an immersive view of a conventional desktop-based programming software, rather than allowing the users to interact with the VE using VR interfaces (hand-held controllers were not considered since it is stated that the aim was to work with an inexpensive setup). Therefore, if there is a need to train IR operators on the execution of tasks involving manual operations, it could be argued that this system would represent a far from the ideal solution.

Robot programming was also the target of the work described in [34]. Differently than in the previous work, the authors enabled the interaction through hand-held controllers. By considering the current limitations of the technology, they decided to simplify the traditional robot programming interface by also adapting it to better harness the opportunities of VR. Albeit the user study participants deemed as promising the newly designed interface, due to the fact that, like in [32], no assessment was arranged on a real robot it is unknown how they would have performed in a real-world programming task. As already demonstrated in the literature [35], the "encoding specificity" principle indicates that performance tends to decline when there is a (not negligible) discrepancy between the environment in which the learning effectiveness is assessed and the learning environment that was used for the training. Therefore, it is assumed that should such a system be used to train an operator, he or she would need to experience an additional, real-robot programming interface to master the considered task.

A viable solution to deal with the aforementioned issue may be to allow the trainee to experience the same task in several situations by changing for instance configuration, environment, and so forth. In this regard, it is relevant to mention the method proposed in [36], which procedurally generates different VEs that may be later used for IR operators training. Unfortunately, since the aim was to solely determine the quality of the generated VEs, no analysis was carried out to evaluate the impact on training effectiveness.

The few works reviewed above are, to the best of the author's knowledge, the only published papers exploiting a VRTS in a training context that consider IRs. Aside from decisions made in those work that may hamper applicability/replicability or representativeness of the experiments (hardware used, ways in which the robot's interfaces were reproduced, etc.), the major limitation of the reported literature is that no study was performed to validate if the VRTS was actually able to guarantee a satisfactory transfer of knowledge and skills with a self-learning approach, by asking the trainees to perform, on a physical robot, the particular procedures that were taught to them by the system on a virtual replica.

2.1.2 Materials and Methods

This section depicts the steps undertaken to devise an immersive VRTS for IRs self-training and to arrange a user study with the aim to validate its effectiveness by employing a real IR.

2.1.2.1 Case Study

In order to select a relevant use case, several experienced instructors from a world-wide manufacturer of IRs were involved in the process; in particular, instructors were affiliated with the KUKA College [1]. The study was aimed at evaluating whether such a VRTS could guarantee a training experience as well as a transfer of knowledge and skills comparable to the company's regular training program used globally.

An industrial manipulator's *mastering procedure (MP)* was selected as the subject for the study. This decision was made since, despite the fact that the MP is taught in introductory robot programming classes, mastering a robot is basically a

[1] The training department of KUKA.

light maintenance task (a mixed-task). Furthermore, it is an occasional duty, that operators (typically not belonging to the maintenance staff) undertake directly on the production line. Thus, a VRTS might allow operators to access training information autonomously when required without overwhelming the IR manufacturer's technical support (dwindling production line delays and increasing knowledge retention), as well as enable asynchronous remote training approaches (reducing the number of persons that need to be moved to training locations).

2.1.2.2 Mastering Procedure

Calibrating an IR is essential to ensure that it performs precise, repeatable, and accurate movements. For every axis of the robot, the purpose is to align the mechanical zero-point to the internal references of electrical/software one [37]. It is a recommended practice to calibrate a robot in case of load distribution or payload changes, before the first usage (commissioning), during/after maintenance, and in the event of failures or collisions. The key steps of the MP that are required to calibrate a KUKA IR are described below.

1. By using the KUKA SmartPad (SP), depicted in Figure 2.1, the operator needs to manually move all the axes so to have them attain a so-called *pre-calibration position*, i.e., a peculiar axis angle which is distinctive of each robot class, and somewhat dissimilar within the class (for each robot). This must be accomplished by visually aligning explicit references located in the vicinity of axis joints (an example is shown in Figure 2.2). Neither the robot nor the SP provide the operator with further feedback. The order is irrelevant.

2. Considering a given axis (and starting from the #1), a supplementary external sensor, the *Electronic Mastering Device (EMD)* [37], must be screwed to a definite pawl (on the axis) and connected electrically to the robot.

3. For the given axis, the operator must configure and launch a semi-automatic mastering program by means of the SP's graphical user interface. As a consequence, the *mastering motion* of that axis will be controlled by the program; for safety reasons, the operator has to keep pressing a physical button on the back of the SP (the *enabling switch*) all along with the axis motion.

Fig. 2.1 KUKA Smart Pad (SP): the device that the operators needs to use to manually control and program the robot is depicted. It consists of a touch-based graphics interface, as well as of some physical buttons and switches for safety-critical operations.

4. The EMD must be disconnect and unscrewed (these operations need to be performed in this exact order to prevent damages).
5. Steps 2–4 must be performed again for each of the robot axes by following the order that they are numbered.

The steps just summarized are those to follow in case no errors occur. Yet, the operator must also be familiar with multiple error recovery procedures (ERPs) and potentially apply them grounding on the acquired procedure-dependent decision-making and analytical skills. In reality, the robot typically does not provide explicit feedback on errors occurrence; the main reason for that is intrinsic to the kind of the errors that may happen during the MP, which can be primarily detected solely when the operations are completed or as a consequence of equipment damage. In addition, unlike other IMA tasks, a simple rollback or step undoing is not enough to correct the considered errors. These peculiarities translate into challenging didactic features that any training method shall address in order to make the operators competent in autonomously carrying out the procedure.

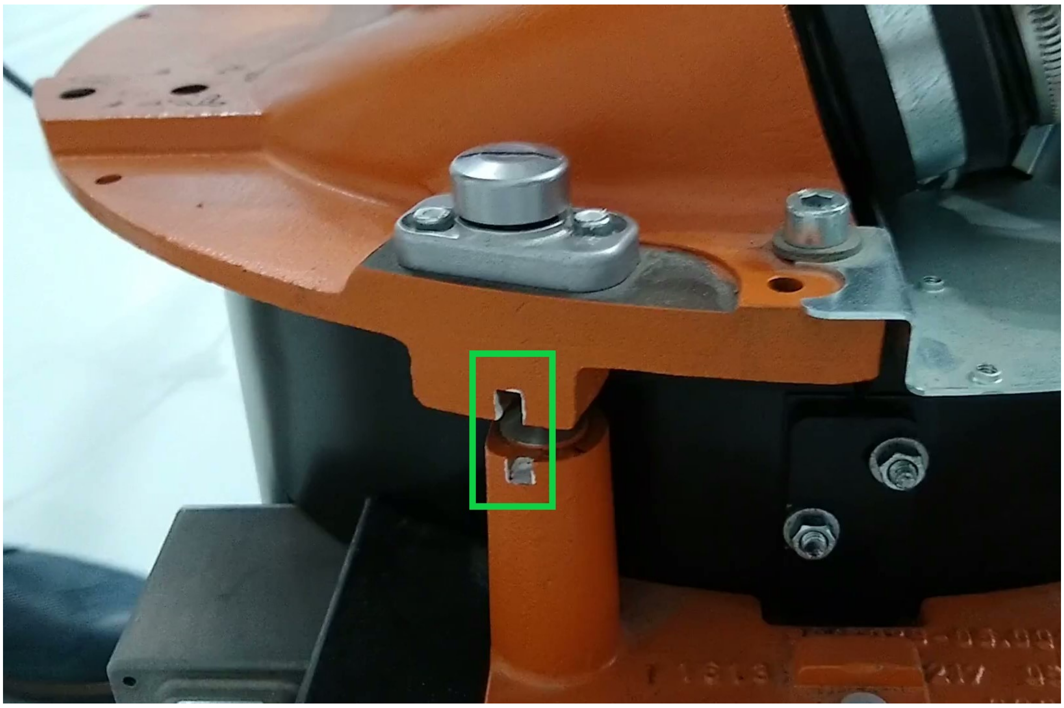


Fig. 2.2 Axis #1 of a KUKA KR-16 robot in pre-calibration position. The green box highlights the visual references to be aligned by the operator in the first step of the MP.

Consequently, what follows is a summary of the most noteworthy aspects of the intended training from the trainee's perspective.

- develop procedural knowledge of the MP;
- being capable of recognizing variations in the system status by relying solely on typically subtle visual and/or aural cues (e.g., extremely slow axis motions);
- exploiting proprioception skills and spatial awareness so to become able in safely moving in the robot proximity;
- manage the equipment appropriately in order to prevent damages;
- manage errors by running appropriate ERPs depending on the current state of the system and the type of error.

(a) (b) (c)

Fig. 2.3 The three stages of the standard training (ST) at KUKA College: a) In-class learning; b) hands-on demonstration learning; c) peer-tutoring with supervision.

2.1.2.3 Standard Training

KUKA College's master instructors devised a training process (deployed in every company's training facility) that, according to internal audits, certifies trainees' acquisition of required learning outcomes.

As depicted in Figure 2.3, in a typical course a three-stage learning process is followed by the trainees for each of its salient parts. A classroom lecture approach is used for the first stage, by having small classes of no more than 12 individuals. During this stage, for backing his or her teaching, the instructor might utilize a variety of projected supports such as slides, drawings, videos, or SP emulators. This stage is aimed at teaching both theoretical and practical concepts (like procedures), and to prepare the trainees to conduct activities on the robot in the following two stages. A demonstration learning approach is used for the second stage in which the instructor performs the practical exercises and the MP hands-on. This demonstration is executed using a dedicated teaching environment, named *didactic cell (DC)*, which incorporates an IR along with multiple safety devices and props (depending on the subject being taught). For the last stage, trainees are grouped in teams of three and are allowed to practice directly on the IR (each team is assigned to a separate DC) supervised by an instructor who can step in if necessary. This supervised, peer-tutoring approach was chosen over alternate layouts (e.g., having a trainee for each DC) to optimize the allotment of limited physical resources (the DC) and to take advantage of the benefits of cooperative learning dynamics [38].

The just described approach, which alternates in-class and hands-on stages, has been used to devise a so-called *standard training (ST)* on the MP. The ST arranged for the user study was made up of two modules extracted *as is* from the KUKA basic course on robot programming. The first module was about the essential function of a KUKA IR (how to identify and recognize the robot primary components, how to move its axes by using the SP, etc.), and the second one on the MP. The rationale behind introducing the first module was the author's intention to include study participants lacking familiarity with IRs (as further explained in Subsection 2.1.3). The ST was estimated to last about 90 minutes, according to KUKA College's internal statistics.

2.1.2.4 VR Training

To train the operators on the same modules of the ST, it was devised a self-learning tool in the form of a state-of-the-art VRTS. For the sake of brevity, the design process, which comprised numerous iterations with constant feedback from company's instructors, and the early validation stages will not be described in detail.

Technologies and Implementation: The VRTS was developed considering the HTC Vive Pro kit [39] as a target immersive VR HWD with hand-held controllers. This HWD is endowed with a display resolution of 1400×1600 pixels per eye, covering a horizontal FOV of $110°$ at a $90Hz$ refresh rate. The tracking supports a 6DOFs of each tracked item and is provided by Valve's Lighthouse technology that relies on infrared laser emitters. The tracked hand-held controllers and their built-in physical buttons are used to interact with the VE. The implementation of the VRTS application leveraged the Unity (v2018.4) game engine [40] together with the SteamVR framework (v2.7.2).

To create the basic VE, free of charge 3D assets and custom-created ones modeled with Blender (v2.91) [41] were employed. To complete the VE, 3D assets were extracted from KUKA SimPro (v3.0.5) and polished with Blender [41] (specifically, to model a high-fidelity replica of the KUKA KR-16 robot and of the rest of the DC). Appropriate care was taken to reproduce auditory cues from the actual DC in VR; in particular, sound recordings of interaction with props (such as, for example, safety doors) and of physical robot movements were harvested and integrated into the VE by means of spatialized 3D audio simulation.

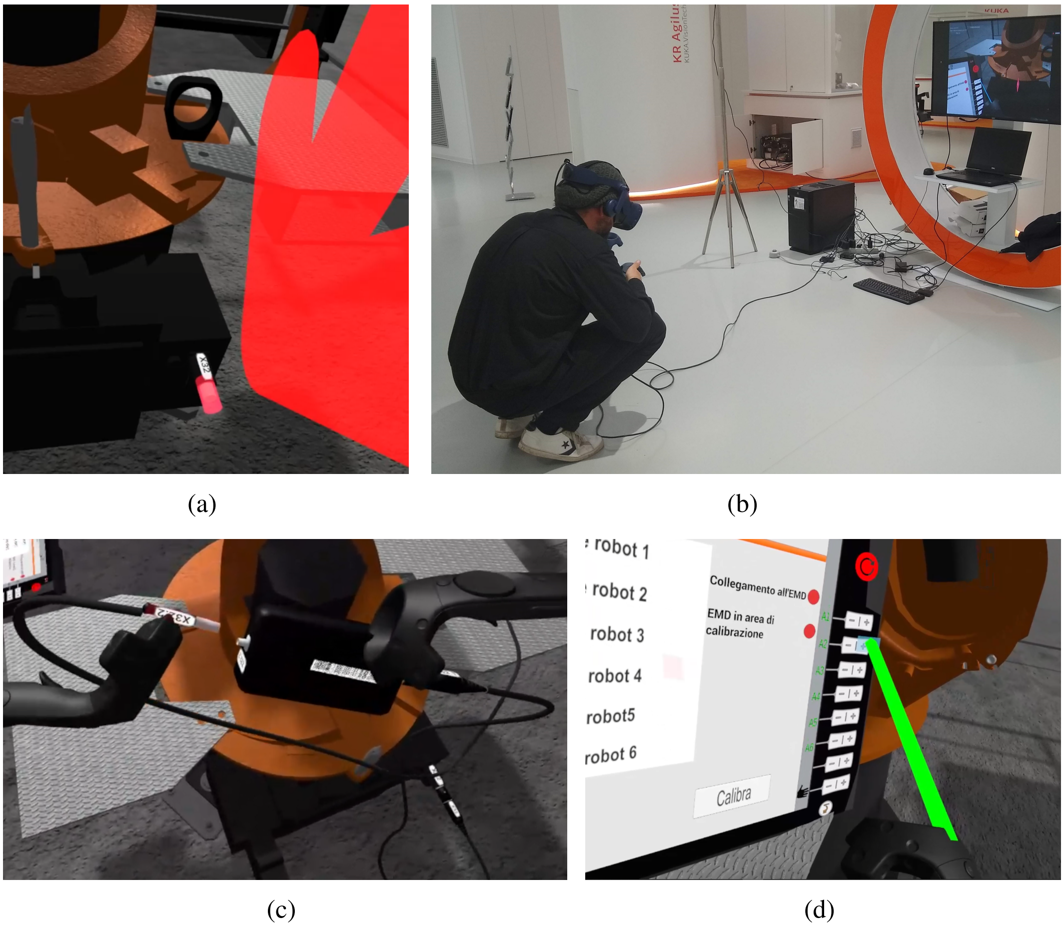


Fig. 2.4 Collection of moments from the VRTS experience: a) Example of MR-like hints; b) trainee intent on aligning the robot's axis #1; c) interaction with virtual tools; d) interaction with the virtual SP.

User Experience: According to [42], trainees' performance in VR starts to degrade after 55 minutes. Thus, staying below this parameter the VRTS was designed with an estimated fruition time of around 40 minutes in mind. To define the user experience for the self-learning (summarized in Figure 2.4) it was followed the well-known scaffolding approach, largely adopted in VRTSs [43, 44], which consists of a GS to provide step-by-step instructions.

Differently than in the ST, in the VRTS the three stages are condensed as a single, interactive information flow. The GS delivers chunks of instructions through a voice-over implemented with pre-defined audio tracks of a synthesized female voice (generated using a state-of-the-art text-to-speech tool). In order to make the user pay attention at the instructions delivered, they are not delivered also in a different, duplicated manner, e.g., through text (like in [26]). Anyhow, by clicking a button on the virtual SP, the user is made able to replay the last chunk at discretion.

In order to direct the user's attention towards the so-called *hot-spots*, as in [24] were used simulated MR-like hints (Figure 2.4a) in order to evoke the pre-attentive visual processing [45]; specifically, these hints were implemented in the form of arrows, objects highlighting with blinking outlines, and other graphics signs. Some examples of hot-spots include a visual element to gaze at, an object to pick up, and so forth.

There are steps in the MP in which the trainee is requested to learn from which perspective a certain hot-spot must be looked at. The GS stresses this aspect by observing the user's gaze direction and head position. This functionality was introduced after monitoring the users' behavior during the multiple design iterations, and noticing that they tended to overlook the voice instructions when there was an alternative way to complete the assigned task in a "quick & dirty" manner. As a matter of example, when the operator needs to move the robot's axis #1 in its pre-calibration position, he or she shall close look at visual elements located at the robot basement, and this is achievable only by kneeling down in the nearby of that hot-spot (Figure 2.4b); it was also noticed that the trainees had the tendency to neglect this action, especially in the case in which the GS unwittingly provided supplementary feedback (not matched with the one of the real scenario).

An established way to design a step-by-step GS is to have it block the instruction provisioning on each step. Applying this modality to the previous example reflects an implementation that invites the trainee to proceed in aligning the successive axis as soon as the pre-calibration position of the axis #1 has been successfully reached. Nevertheless, this would produce a very different experience w.r.t. that with the real robot in which, on the contrary, the trainee would not be provided with any feedback about the successful task completion. The GS must account for this circumstance accurately, deferring the check (e.g., until after the next user action) to ensure that the related skills are correctly transferred to the trainee.

As mentioned in Subsubsection 2.1.2.2, further skills to develop pertain to the ability to manage the equipment without damaging it. Specifically, how to execute the required micro-manipulations with provided tools (Figure 2.4c) and cable management need both to be mastered by the trainee; for instance, a certain connector may have to be either fixed using a bayonet, taking care of the orientations, or screwed, depending on the case. Considering the limitations of the current technology for hand tracking and force feedback (at least, outside laboratory settings), it was decided to demonstrate these interactions and the relative required actions through 3D animations, by illustrating them on the given components (for instance, when two connectors are brought close enough with proper orientation).

Moreover, the locomotion that was allowed in the VE was mostly restricted to *real walking* [46]; this choice was deemed to be capable of preserving a high sense of presence, which was essential considering that, in order to safely move in the proximity of the robot, conspicuous proprioception and spatial awareness are required. A *teleportation* mechanism was nevertheless included, but the user was allowed to move in that way just to reach carefully predefined locations that were chosen far enough from hot-spots; hence, the trainee was anyway forced to reach such locations using real walking.

Finally, since a core element of the real robot MP is to gain dexterity with the SP, particular care was taken in implementing its simulated version. Specifically, the safety-critical SP's enabling switch was mapped onto the gripper button of the hand-held controller. A ray-casting selection using the free hand (Figure 2.4d) was used to handle the other interactions. The interactions with the touch-screen were emulated utilizing the trigger button of the respective controller.

A video showcasing the VRTS experience is available at http://tiny.cc/s2p6tz.

2.1.3 Experiment

This section presents the design of the user study that was run to evaluate the devised VRTS against the two research questions in Section 2.1 by comparing it to the company's ST. The training was conducted in a situation modeling the worst-case scenario, in which the majority of learners are domain-agnostic.

2.1.3.1 Experiment Design and Metrics

The sample of the study consisted of 18 participants aged between 23 and 35 years, recruited as volunteers among engineering students enrolled at Politecnico di Torino and the available networks of contacts.

The study was arranged following a between-subjects design ($n = 9$) so to avoid learning effects biases, randomly assigning the participants to two equal-sized groups. One group, named STG ($\mu = 24.6$, $\sigma = 1.3$, made of 1 female and 8 males), experienced the ST; as already described in Section 2.1.2.3, the participants were grouped into teams of three people for undergoing the second and third stage of the ST. Participants of the other group were trained via the VRTS, referred to as VRG ($\mu = 26.1$, $\sigma = 3.3$, made of 1 female and 8 males); prior to being exposed to the VRTS, the trainees were allowed to practice locomotion in the VE and interaction with objects using a "sandbox" VR scenario.

After having completed the training, independent of group they had been assigned to the participants underwent an evaluation phase. In particular, they were tested using a quiz and by a company's instructor who was asked to assess the acquisition of the expected learning outcomes by analyzing their execution of the MP on a real robot (further details will be provided in the following subsection).

To complement the evaluation of the training effectiveness, aspects pertaining to the training experience were also analyzed by collecting trainees' feedback using a questionnaire. The questionnaire included 109 statements to be scored on a 1-to-5 Likert scale (from strongly disagree to strongly agree). In order to minimize possible interview fatigue effects, the questionnaire was arranged into three sections, and the administration intervalled by activities that allowed for mental rotation as described below. Futhermore, for the longest section (80 items), the participants were allowed to take a few minutes break at three different administration checkpoints (one every 25 items, approximately).

Before starting the training, a first section of the questionnaire (BTQ) was administered to collect information pertaining to prior knowledge and level of proficiency with technology pertinent to the experiment, demographics, and the perceived *self-efficacy* (i.e., attitudes towards/expectations from the training to be experienced) [47]. A second section was administered after the training was completed (ATQ), consisting of items concerning post-experience self-efficacy, and statements adapted from the Instructional Materials Motivation Survey (IMMS) [48]; this section was completed by the NASA-TLX tool [49] for measuring the trainees' cognitive load. The participants belonging to the VRG were also asked to fill in a dedicated section (ATQ-VR) with statements on the usability of the VR application based on the VRUSE [50] and the System Usability Scale Questionnaire (SUS) [51] tools. Lastly,

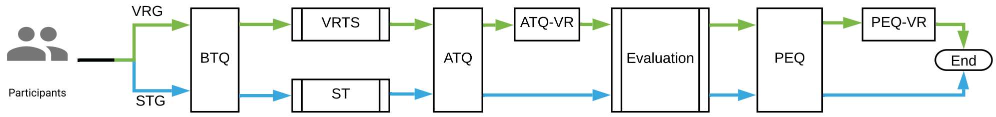


Fig. 2.5 Experiment design and analysis tools.

a post-evaluation (PEQ) section was administered with the aim of complementing the assessment coming from the instructors by collecting feedback about trainees' overall satisfaction and self-evaluation (about the perceived proficiency to perform the task on the real robot). As for the ATQ, a part of this section was dedicated to VRG participants (PEQ-VR), with the aim of investigating their level of satisfaction with the VRTS and its simulation fidelity. It is worth mentioning that, to avert conditioning, the PEQ was administered before making the trainees aware of the instructor's evaluation outcome. Finally, open feedback about the experience was additionally collected.

Figure 2.5 summarizes the experiment design, whereas the full questionnaire is available at http://tiny.cc/l2p6tz.

2.1.3.2 Evaluation Procedure

As anticipated, an evaluation procedure was specifically devised in order to compare the effectiveness of the VRTS against the ST in terms of knowledge and skills transfer capabilities, arranged as follows.

Right after completing the training, the trainees were asked to answer a multiple-choice quiz. The quiz was made of seven questions (each with five options, only one correct) with a maximum allowed time for filling it of 5 minutes. Questions were either borrowed from the KUKA's robot programming certification exam, or constructed ad-hoc in collaboration with the company's instructors. The aim of the quiz was to preliminarily verify the acquisition of procedural knowledge, in conjunction with examining theoretical aspects of the MP.

The trainees were then required to accomplish autonomously the whole MP by operating on an real robot. It is relevant to recall that this was the first time that the participants belonging to the VRG performed hands-on operations in the physical DC. The DC was organized as follows:

- all the tools necessary for the procedure were placed outside it in their carry-on case;
- the pawls' protective cups on the robot axes were removed;
- the safety door was securely closed;
- the robot was setup by giving all the axes a -10° offset w.r.t. their pre-calibration position.

The trainees were scrutinized and evaluated by a KUKA's instructor throughout their performance, but the communication between the two was forbidden. An exception was made in the case of safety reasons (i.e., to avert damages to equipment or individuals); in these situations, the instructor could discretionally intervene by stopping the activity either temporarily or permanently.

The instructor was blind w.r.t. to the experimental condition of the given trainee, and a structured evaluation sheet was devised to support the instructor during the assessment, letting him or her record both subjective scores and objective measures. For the subjective aspects, the instructor scored the trainee's performance by assigning a 1-to-10 grade and considering the following operations: equipment management (tools, cables, and connectors), SP management, safety aspects, ability to move nearby the robot, and overall performance. Regarding the objective scores, the following aspects were considered: a set of time intervals on some milestones identified in the MP numbered from I to V (safety door unlocked, axis #1 in pre-calibration position, all the axes in pre-calibration position, EMD correctly connected the first time, axis #1 mastered, all the axes mastered), the overall task completion time, plus the number of errors made. Specifically, the instructor was asked to note down the possible errors made by using a checklist included in the evaluation sheet that was constructed organizing the most common errors in two classes (i.e., minor and major, based on their severity).

The evaluation sheet is available for download at http://tiny.cc/i2p6tz.

2.1.4 Results and Discussion

The two-tailed Mann-Whitney U-tests (p-value ≤ 0.05) were applied to the experimental data to look for significant differences between the two groups. After a

Table 2.1 Per group demographic features according to the data collected thorugh the BTQ.

Group	Not Familiar with IRs	Low Familiarity with VR
STG	22%	56%
VRG	22%	56%
p-value	.364	.863

description of the population features, in the following the most relevant findings are discussed (including all the statistically significant results); the discussion articulates along the two major dimensions set by the posed research questions, i.e., the effectiveness of the ST against the VR-based training (primarily based on the evaluation step results), and the trainees' experience with the two approaches (based on the questionnaire results). Thereafter, overall considerations and remarks are provided.

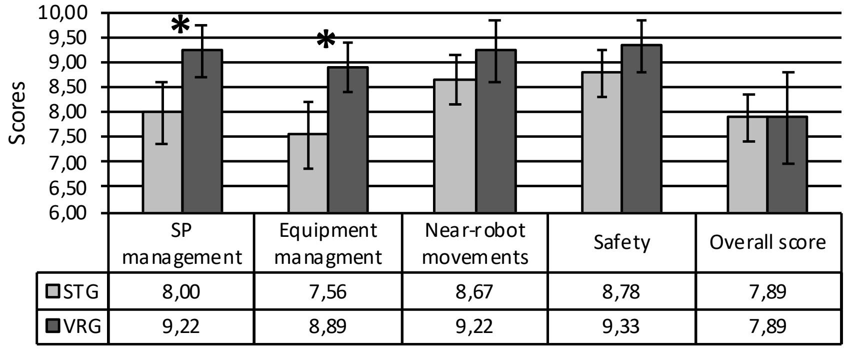
2.1.4.1 Sample Features (BTQ)

By looking at the data acquired via the BTQ (Table 2.1), 22% of the sample had low familiarity with IRs, whereas 78% had no familiarity with them. About immersive VR, 44% of the participants reported using HWD quite often, whereas 56% had never or seldom used immersive VR technology. No statistical differences between the two groups were found for any of the collected demographic features, hence the random designation of participants did not seem to have introduced any bias in this regard.

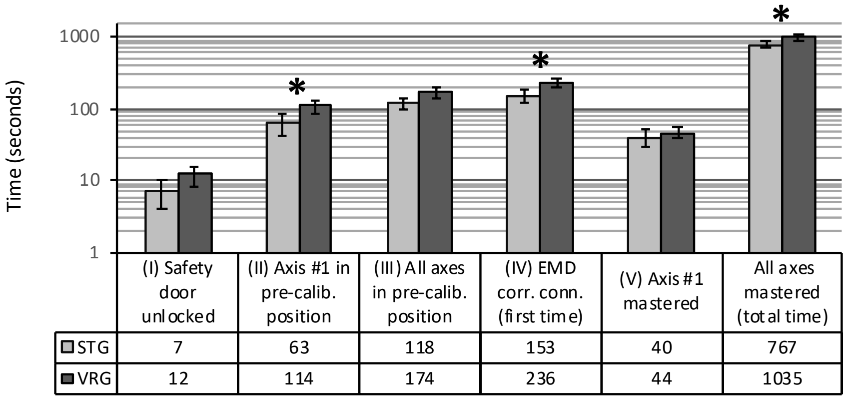
2.1.4.2 Training Effectiveness

Hereafter, training effectiveness is analyzed by comparing the duration of the two experiences in conjunction with the data collected in the evaluation phase (also summarized in Figure 2.6), i.e., the time spent to achieve the various milestones of the MP, the grades given by the instructor, the errors made, and the acquired knowledge (quiz scores).

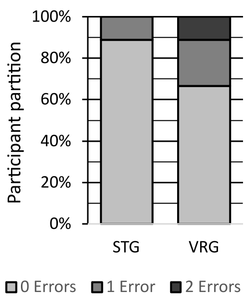
Training Time: By looking merely at the training time (i.e., regardless of the evaluation outcome), in terms of efficiency the ST was outperformed by the VRTS,



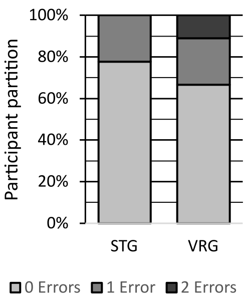
(a) Instructor's evaluation grades



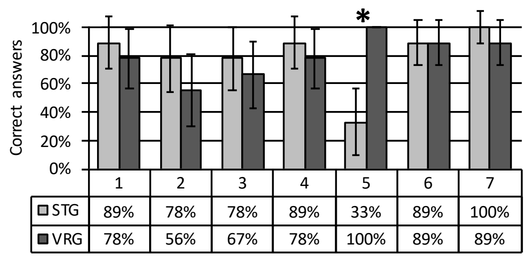
(b) MP milestones (intervals)



(c) No. of major errors



(d) No. of minor errors



(e) Quiz (percent. of correct answers to questions)

Fig. 2.6 Data collected in the evaluation. Significant differences are marked with *.

on average, since training in the STG was roughly 2.5 times slower than in the VRG (90.7 ± 4.8 min vs 36.3 ± 8.2 min).

Instructor's Evaluation Grades: As reported in Figure 2.6a, the performance of the VRG was quite good and on par with the STG. Regardless of the group, all the trainees were able to conclude the MP on the real robot in autonomy, and for the *overall grade* given by the instructor, no significant differences were found. Hence, the requirements and challenges of the MP (Subsubsection 2.1.2.2) were effectively handled by the designed VRTS.

More into details, about *near-robot movements* and *safety*, the VRG was graded comparably to the STG. Notably, the learning effectiveness about the SP and the other equipment management was superior for the VRTS w.r.t. the ST. For what it concerns the *SP management*, it is possible to explain the better performance of the VRG by considering the fact that the evaluation also comprised aspects of the interaction with the touch-based graphics user interface of the SP that were more straightforward to teach interactively by the GS than during the first and second stage of the ST (whereas in the third stage, the trainees were generally less willing to dive into the details of the interface). Even though parallel deductions about the *equipment management* might be made based on the respective grades, a deeper examination reveals that the situation is somewhat different. The VRG obtained indeed a high enough average grade that confirms the VRTS effectiveness also in this regard; however, this outcome is likely attributable to an overly cautious behavior of the trainees, since it should be recalled that the evaluation stage was for the VRG subjects the first time in which they got in touch with the physical equipment.

MP Milestones Completion Time: The latter findings are also substantiated by the measured completion times of the MP milestones (Figure 2.6b). For the milestones II and IV, i.e., those in which the greater number of critical actions pertaining to equipment management occur (milestone II encompasses the time to gain familiarity with the physical DC, whereas milestone IV includes the connection of many small props and the management of delicate connectors), the VRG was found significantly slower than the STG. Conversely, for milestones in which the trainees are mainly asked to manage the SP (III and V), no significant differences were spotted.

The previous consideration is important because it relates to another training aspect outlined in Subsubsection 2.1.2.2 regarding how well and accurately the VRTS is able to provide the aural and/or visual cues needed to recognize changes in system (mostly the IR) status, since the respective procedural knowledge is crucial to fulfilling milestones III and V.

By looking at the *overall completion time*, the VRG was slower than the VRG (17.24 ± 2.20 min vs. to 12.78 ± 1.37 min); nevertheless, in the context in which the MP is executed, this difference can be deemed as acceptable. Furthermore, it is important to note that this slight difference in additional time spent on average by the trainees in the VRG, of less than five minutes, must be contrasted to the about 50 minutes longer duration of the ST w.r.t. to the VR-based training.

Errors and ERPs: The higher number of ERPs executed by the VRG w.r.t. the STG was also a contributing factor to the differences in completion times. As illustrated in Figure 2.6c and Figure 2.6d, the instructor noted more errors for the VRG than the STG, on average; nevertheless, this difference was not found as statistically significant. For both the groups, most of the errors were associated to equipment management, although even in case of major errors, the trainees were able to recover on their own without the instructor intervention.

Quiz: Regarding the quiz (Figure 2.6e), significant differences were spot only for *question #5*. Specifically, procedural knowledge was tested in this question since trainees were asked to put in the right order actions to be performed when calibrating a given axis. Interestingly, the VRG obtained better scores than the STG. Nonetheless, it should be noted that 67% of STG participants who gave an incorrect answer picked the second-best option (the two best options differ in which was the right time for pressing the enabling switch). Manifold interpretations could be given for this phenomenon. On the one hand, it could have been due to an over-learning effect experienced by the STG; in other words, it is possible that this group of trainees were able to carry out the actions without actually recognizing them as a result of having learned them so well, therefore substantiating the interpretation that the skills transfer was greater with the ST. On the other hand, it could be that the bad habit of continuously pressing the enabling switch even when not needed was developed by the participants of the ST during the third stage, in which the instructor ward is less conspicuous; thus, the safety property of this device could have been erroneously transferred to other SP functionalities. The latter interpretation would further support the already examined superiority of VRTSs in backing the knowledge transfer for cyber-physical devices that are characterized by hybrid haptic and digital interfaces.

2.1.4.3 Training Experience

In the following it is reported an analysis of the subjective feedback based on the multiple sections of the administered questionnaire.

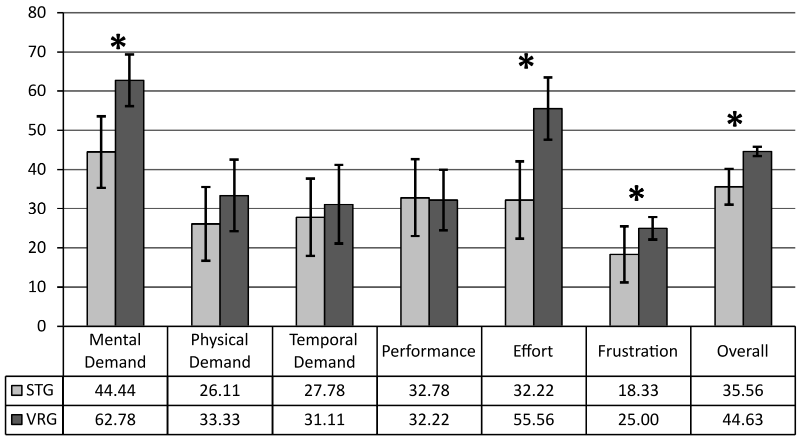


Fig. 2.7 Cognitive load (based on NASA-TLX). Statistically significant differences are marked with *.

Table 2.2 Scores collected through the ATQ. Significantly better options (based on statement phrasing) are highlighted. Mean (SD) values for both the groups are reported, with p-values.

#	Statement/Item	STG	VRG	p-value
1	There was something interesting since from the beginning of this training experience that attracted my attention	3.89(0.31)	4.56(0.25)	.050
2	The approach used by the teacher was captivating	4.00(0.44)	4.00(0.41)	.931
3	The quality of information/instructional material helped to hold my attention	4.67(0.40)	3.56(0.25)	.004
4	The way information was provided helped keep my attention	4.33(0.33)	3.67(0.24)	.063
5	The amount of repetition in this training caused me to get bored sometimes	1.33(0.67)	3.44(0.53)	.001
6	The variety of information helped keep my attention on the training	4.22(0.43)	3.33(0.41)	.050
7	The information provided by the teacher was boring	1.56(0.46)	2.56(0.42)	.031
8	The amount of information provided was so large that the training was irritating	1.22(0.46)	2.33(0.47)	.019
9	The teacher showed the relevance of taught content	4.33(0.31)	3.56(0.25)	.024
10	It is evident how this information should/could be used after the training	3.89(0.42)	4.22(0.31)	.605
11	The amount of information provided was so large that it was difficult to identify the most relevant to remember	1.44(0.33)	1.89(0.37)	.258
12	During the training, I was confident that I would have been able to learn the taught content	4.44(0.37)	4.22(0.39)	.605
13	The training activities planned in the experience were too difficult	1.00(0.29)	1.67(0.43)	.067
14	During the training, I was confident that I would have been able to pass a test on it	4.11(0.31)	3.44(0.25)	.050
15	The training topics are relevant to my interests	2.78(0.72)	2.78(0.52)	.796
16	Completing this training successfully was important to me	3.00(0.60)	4.22(0.46)	.050
17	I am overall satisfied with the training experience	4.06(0.34)	4.00(0.28)	.605
18	I feel confident that I will correctly perform the MP on the real robot on the first try and without further help	3.78(0.37)	3.56(0.42)	.667

ATQ, ATQ-VR: By looking at the NASA-TLX (Figure 2.7), significantly lower scores were registered for the STG w.r.t the VRG for what it pertains the *effort*, *frustration*, *mental demand*, and *overall score* components, whereas comparable scores were reported for *temporal demand*, *physical demand*, and *performance*.

Regarding the ATQ investigated factors (Table 2.2), significant differences were spot in favor of the ST for what it concerns boredom (items #5 and #7), frustration (item #8), attention difficulties (items #3 and #6), ability to underline the taught content relevance (item #9), and perceived confidence to pass an examination (item #14). Contrariwise, w.r.t. to the ST, the VRTS was better at motivating the trainees (item #16) and found as more captivating (item #1).

Table 2.3 Section of the questionnaire on simulation perception dedicated only to subjects in the VRG. Statements with negative phrasing are marked with °.

#	Statement/Item	Mean (SD)
	ATQ-VR	
1	*SUS* Score	81.39(3.64)
2	I can see a real benefit in this kind of didactic experience	4.56(0.25)
3	The didactic experience was too simplistic to be useful°	1.78(0.52)
4	I wasn't aware of making mistakes°	1.78(0.31)
5	It was difficult to learn how to use the VR system°	1.44(0.25)
6	It was not so important to feel present to successfully complete the assigned tasks°	2.11(0.60)
7	In the VE I had a sense of "being there" (presence)	4.56(0.34)
8	The information was clearly presented	4.33(0.33)
9	The voice-over explanations (audio) were clear enough	4.33(0.33)
10	The MR-like visual cues helped understanding	4.78(0.21)
	PEQ-VR	
1	Overall, I think that simulation fidelity was very high	4.44(0.25)
2	I wished I had more time to spend free-practicing in the VRTS	3.00(0.62)

Apart from the distinctions just discussed, for all the other components investigated via the ATQ, the VRTS and the ST were deemed as comparable. This is the case of critical aspects such as perceived training effectiveness (item #18), information clarity (item #10), and overall satisfaction with the training experience (item #17).

Considering the ATQ-VR (Table 2.3), results worth to mentions are about the VRTS usability (SUS tool scores, item #1), which was evaluated as remarkably high ("excellent", according to [52]), together with the fact that it was judged as easy to learn (item #5) and able to induce a high sense of presence (item #7).

PEQ, PEQ-VR: Also considering the trainee's evaluation on the real robot similar trends were observed. This is testified by the fact that the paired items in the ATQ and the PEQ (Table 2.4), i.e., *perceived training effectiveness* (PEQ.#1, ATQ.#18) and the *overall satisfaction* (PEQ.#12, ATQ.#17), scored comparably.

Also most of the remaining PEQ components, i.e., the perceived capability of the given training method to support the acquisition of required proprioception skills and transfer the necessary spatial knowledge (items #3, #4, #7, and #9) were considered as comparable. Furthermore, the ST and the VRTS were considered as equally effective in transferring the equipment management abilities and procedural knowledge (items #2, #5, #6, and #10), despite the fact that the VRG felt more

Table 2.4 Scores collected through the PEQ. Significantly better options (based on statement phrasing) are highlighted. Mean (SD) values for both the groups are reported, with p-values.

#	Statement/Item	STG	VRG	p-value
1	I felt confident while performing the MP on the real robot	4.33(1.00)	4.00(0.33)	.258
2	I felt the need to frequently stop throughout the MP to remember what should have been done next	1.67(0.87)	2.00(0.41)	.436
3	It was easy to safely move around the robot	4.44(0.73)	4.44(0.34)	.931
4	It was easy to locate the pawls on the robot's axes	4.00(1.12)	4.22(0.31)	.863
5	I struggled to use the SP	1.44(1.01)	1.11(0.16)	.666
6	The cable management and the connectors coupling/fastening was as expected	4.44(0.73)	4.11(0.28)	.297
7	Working on a real robot made me feel anxious	1.11(0.33)	1.33(0.24)	.436
8	I believe that I managed the equipment with the appropriate care	4.78(0.44)	4.22(0.21)	.050
9	It was easy to locate the visual references close to axis joints	4.44(0.73)	4.11(0.16)	.222
10	It was difficult to move all the axes in their pre-calibration position	1.78(0.83)	1.78(0.39)	.931
11	It was difficult to couple/fasten the connectors without risking to damage them	1.56(0.73)	2.33(0.33)	.050
12	I am still overall satisfied with the training experience	4.56(0.53)	4.44(0.25)	.730

anxious when handling the provided equipment (item #8), and was a slightly more challenged by operations involving small components (item #11).

Finally, regarding the PEQ-VR (Table 2.3), the simulation fidelity was judged as remarkably good.

2.1.4.4 Discussion and Remarks

Based on the findings reported in Subsubsection 2.1.4.2, for what it concerns training effectiveness, and in particular regarding the ability to support the transfer of knowledge (spatial knowledge included) and the learning of relevant skills (also related to proprioception), the VRTS emerged as a convincing option to the ST. More in depth, scores obtained by the VRG were on par or even superior w.r.t. those of the STG, and the individuals in the former group were indeed able to successfully conduct the procedure on the real IR. Although, on the one hand, the obtained results indicate the effectiveness of VR applied to training experiences that include the handling of cyber-physical devices (the SP, in this case), on the other hand they may be misleading for what it concerns the management of small, physical equipment, since the VRG was observed behaving way too careful with this kind of objects, consequently taking more time to complete the MR w.r.t. the STG. Nevertheless, the extra time needed by the VR trainees is counterbalanced by considerably quicker training times.

For what it pertains to the training experience, the VRTS and the ST were deemed as comparable for the majority of the investigated aspects, including critical ones like the perceived training effectiveness and overall satisfaction.

For some aspects, dissimilarities were found too. One of them relates to the fact that the cognitive load of the VRG was found to be greater than that of the STG. This outcome also accords with previous works in the literature, since it is commonly acknowledged that immersive VR experiences, particularly those including rich interaction and enticing visual-auditory stimuli, are more likely to elicit a high extraneous cognitive load [53]. Considering the explored case study, the VRG higher *effort* and *mental demand* scores were probably influenced also by the VRTS's higher pace and density of instructional content if compared to the ST, whereas the usability of the VR-based system did not appear to play a major effect.

A factor that should be ameliorated in the future is the VRTS workload, albeit considering that anyhow for both the VRTS and the ST the *overall score* was more than adequate and not overwhelming, being below the 50^{th} for the category of interest (*Video-games* and *Robot operation*) reported in [54].

The greater levels of frustration observed for the VRG depict another area that shall be considered for improvement in the future. Albeit contributed by side factors (e.g., the reported VRTS higher attention difficulties and boredom w.r.t. the ST), the open feedback collected from the VRG revealed that the rating of these factors was principally ascribable to the fact that the GS was felt as not adapting to the trainees' needs and to the voice-over experience (rather than the synthetic text-to-speech, a real human recording would have been preferred). Specifically, participants were particularly frustrated and annoyed with the modality the scaffolding system was adopting to instruct the pre-calibration pose. It should be recalled that, in such a step, the trainee is enforced to assume a specific head pose, and in case he or she is not matching the correct observation point, in addition to the MR hints the voice-over provides an error feedback repeating it till the trainees performs correctly.

One promising aspect that should be addressed in the future since it has a high potential to level off the VRTS with the ST regards the social elements of the ST experience. This hindsight is substantiated by the fact that the STG trainees gained more confidence that they would have been able to pass an examination, probably due to encouragements and reassurances that were perhaps given during the training by the instructor; an additional corroborating element could be the fact that the VRTS was deemed as less effective in emphasizing the relevance of the various content being taught (it is speculated that this came as a consequence of the fact that, in the ST, the human instructor integrated the lecture with anecdotes based on past-experience, eliciting rich Q&A interactive sessions).

It is worth noting that, although there were no differences between the two groups in terms of the intrinsic motivation of the participants about the topic taught, the VRTS was able to motivate the participants to complete the training more effectively than the ST. Moreover, the trainees in the VRG were more captivated by the training experience than those in the STG, which is quite remarkable considering that also the ST trainees were expected to be stupefied by their first encounter with an IR.

2.2 Scaffolding on Parallax-Dependent Tasks

As previously stated, one of the aspects in need of improvement for a VRTS pertains to the frustration levels induced by the scaffolding system, with the chance to possibly boost both the pleasantness and the efficacy of the training. From the experiment described in the previous section, it emerged that trainees were particularly sensitive to this issue during the teaching of parallax-dependent tasks (PDTs). By definition, PDTs are a subgroup of MTs in which an operator shall not only *look at* a certain point-of-interest (POI), but also do so *from* a specific observation point in order to accomplish the task successfully (and/or more readily). In other words, a training provisioner demands a preferential *head-pose* that the trainee shall match to complete the PDT. This requirement presents a unique training challenge in terms of expressing to the trainees the importance of simultaneously looking at the POI and doing so from a preferred point-of-view (POV), a skill that the VRTS (automated or not) has to transfer to them without negatively impacting the experience. Moreover, this challenge is even intensified by the fact that trainees may behave way too lazy and relaxed during the training experience, especially in VR as a side effect of the *sandbox* quality of the medium. Although a trainee being sluggish when taught on a MT by an automatic VRTS does not necessarily imply a deteriorated learning outcome [22], this could be a major stumbling block in the PDT scenario.

So far, a great number of works studied the problem of leading the user's attention towards a deliberate element [55–64]; numerous studies also presented methods for persuading the user to reach a desired target location in the work area [65–67], or inducing him or her to copy a given full-body pose [68–71]. Nevertheless, since a compound of these three aspects is demanded while teaching a PDT, it is still unknown to what extent the proposed strategies are still viable and able to fulfill the aforementioned requirements.

In order to address this issue, this section reports on a study in which three metaphors inspired by the literature are proposed and compared, with the aim to adopt them in VRTSs that encompass PDTs training.

2.2.1 Background

As anticipated, there has been a varied and growing body of literature that faced the problem of leading the user's gaze with the aim to have him or her look at a deliberate POI. Unsurprisingly, the interest of academics in the research area of immersive cinematography has been captivated by this aspect [55]. As a matter of example, the authors of [56] explored strategies to entice the viewer to focus on a director's selected POI by either applying a black vignetting to a non-salient portion of the scene or constantly narrowing the in-focus region of the FOV. Regardless of the positive findings achieved, these strategies presuppose that the user's observation point in the space is planned or settled. Also in studies exploring collaboration distance with immersive media, like [64], analogous limitations can be seen. A configuration is described in the latter piece of research in which a remote expert that is wearing a VR HWD is allowed to see a live 360° video stream of a user that uses a MR HWD and provides assistance to him or her. Notably, in this configuration it is the MR user who determines the VR user's observation point. Apart from this drawback, there are anyhow relevant insights. The VR user's gaze is constantly signaled to the MR user by means of a rectangular frame. This metaphor allows both the users to achieve view independence that, in conjunction with ray-pointing and deictic gestures, was proven to have a beneficial influence on cooperation and on the efficacy of the users' interaction. Alongside, also studies that explored the shoulder-to-shoulder cooperation paradigm employed these and additional view sharing strategies so that the users are allowed to independently and freely move inside the collaborative environment (regardless of the medium used, VR or MR)[57].

Overall, plenty of metaphors have been explored by the literature such as the use of non-immersive video sharing from ego-views, of a 3D avatar (or part of it), or of gaze-rays or field of view/frustum visualization methods [58, 59].

Authors of [61] suggest that two-dimensional video sharing should be forsaken in favor of avatars. Moreover, as indicated by the findings reported in [72], avatars that feature eye-gaze animated harvesting eye-tracking data from the user are inferior, from the suitability and ease of use perspectives, to the simpler addition of a head-gaze visualization to the same avatar. This latter visualization strategy has been effectively substituted by a view-frustum in conjunction with both a full-body 3D avatar [73] and a simplified head representation [62]. Despite the encouraging findings, it is uncertain if these metaphors may also be used to effectively deliver training on a PDT since in the aforementioned studies, the user is not required to match the head pose (both position and orientation) of the reference metaphors.

A few works have attempted to address the challenge of leading the user to a desired location in the VE. For instance, in [67] was explored the employment of metaphors allowing the user to relocate in the VE: it was observed that all the evaluated approaches were well received by users in terms of sickness and sense of embodiment. The analogous transition style (infinite velocity) of the strategies proposed in [67] was further analyzed by the authors of [66], who confirmed that it is a better choice w.r.t. alternative types of continuously animated position interpolation that are not under the control of the user. In [65], four approaches aimed at pushing the user to assume a common perspective (which is a scenario close enough to the PDT one) were compared. Partially in contrast with the previous works, it was observed that, by considering the discomfort and disorientation component, the continuously animated position interpolation (fly) should be preferred to the infinite velocity method. Nevertheless, this contradictory result (w.r.t. [66]) might had been due to the used VE (a white void), since the users had solely the object of interest as a reference frame. This is indeed quite a different scenario from the one of MT training typical of industrial settings. Another takeaway is that in case the users are restricted to only one locomotion style, they will go for a manual controllable relocation method instead of flying. Lastly, the authors of [22] evaluated multiple metaphors to direct the user's focus in the context of MTs training using a self-learning XRTS. It was observed that certain users had an instinctive desire to stay into the instructor avatar and copy its movement even when not explicitly asked to behave so by the system.

It should be remarked that, even though all the aforementioned results may theoretically be considered for adoption into an XRTS, none of the reviewed works examined the effectiveness of the presented strategies from the perspective of skills transfer to learners. Moreover, by looking at the PDT training context, it is of particular interest not only whether the trainees will assume the desired head-pose during the scaffolded training, but also whether they will be able to recall that pose when asked to perform the taught task on their own.

There are also works that, by using XRTSs, focused on teaching the user to learn a full-body pose by exploiting the so-called avatar follower effect [71]. This approach can be easily applied in the context of teaching sports or dance movements [68–70]. Notwithstanding, for the sake of PDTs, it is irrelevant whatever full-body posture the operator uses to match the desired head-pose; hence, these approaches might be deemed a little excessive for the scenario being considered.

The study described in the following was grounded on these insights, with the aim to identify a suitable metaphor to be used in a XRTS for the training of PDTs.

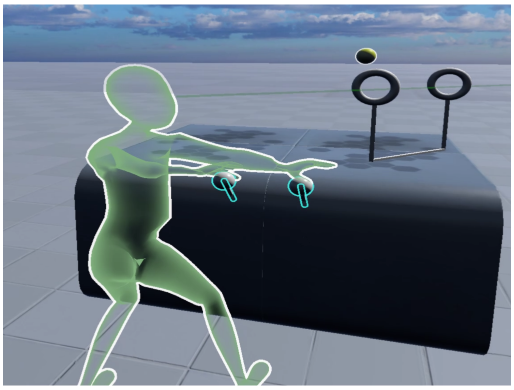
2.2.2 Materials and Methods

In the following, three metaphors are proposed and described, intended to be used to promote trainees to copy a desired head-pose when trained via an automatic VRTS on the execution of a PDT. A testbed was arranged in order to conduct experimental evaluations under repeatable settings and to incorporate a whole range of PDT categories. Specifically, six PDTs were considered into a procedure that is taught using a devised self-learning VRTS allowing two operating modes: one for training the user with the aid of a scaffolding system, and another for evaluation. Using such a testbed, a user study was run to evaluate and compare the proposed metaphors.

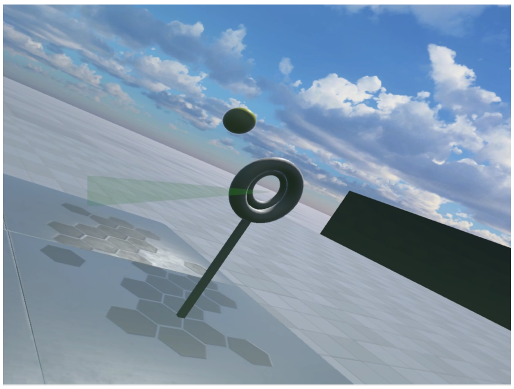
2.2.2.1 Technologies

The testbed used in the study was deployed considering the HTC Vive Pro kit [39] as a target immersive VR HWD with hand controllers (device features are already reported in Section 2.1.2.4). The tracked hand-controllers and their built-in physical buttons are used to interact with the VE.

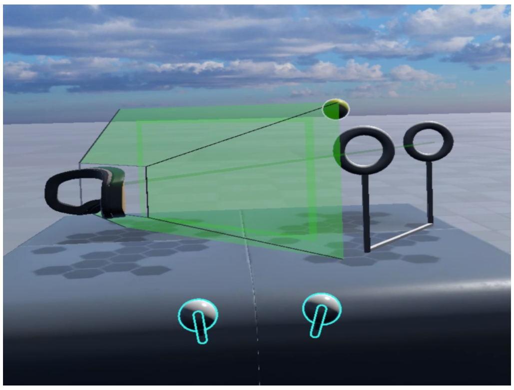
The implementation of the VRTS application leveraged the Unity (v2020.2.2) game engine [40] together with the SteamVR framework (v2.7.2). To create the VE, free of charge 3D assets and custom-created ones modeled with Blender (v2.91) [41] were employed.



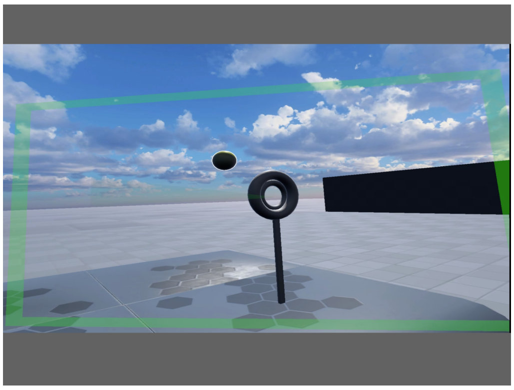
(a) Avatar outside



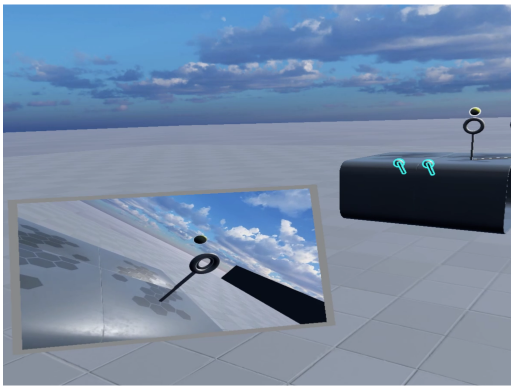
(b) Avatar in pose



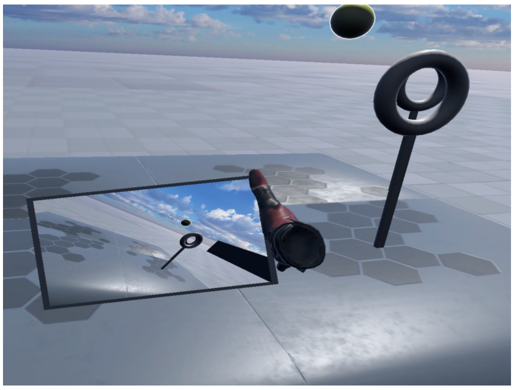
(c) Frustum outside



(d) Frustum in pose



(e) PiP



(f) Relocating PiP

Fig. 2.8 Proposed metaphors.

2.2.2.2 Metaphors

Here below it is detailed the implementation of the three metaphors (Figure 2.8) intended to be used when teaching a PDT to an operator by adding them to a traditional scaffolding system.

It is worth recalling that a suitable metaphor needs to clearly inform the trainee about the head-pose to match (given by the training provisioner) or, in other words, to jointly instruct *from* which observation point he or she should look at a designated target POI (like an object), and *where* is such POI to gaze at.

- *Avatar (A):* As already pinpointed in Subsection 2.2.1, the usage of 3D avatars seems to be one of the most promising metaphors for teaching PDTs. Specifically, it was devised an avatar consisting of a simplified human model devoid of sexual connotations; its stature was automatically modified to match that of the actual user. Similarly to what done in [68], the model was shaded with a Fresnel effect, so to stimulate the trainee stepping inside the avatar. Also, a head-gaze line was included (Figure 2.8a) as proposed by [72]. In addition, to offer feedback to the user that he or she is successfully matching the intended head-pose, it was chosen to modulate the shading transparency such that the user is encouraged to fit the head-pose by minimizing occlusions. The fading behavior of the head-gaze line has been implement to be proportional to the angular disparity between the user's current head-gaze direction and the desired gaze direction (Figure 2.8b), and by enabling this modulation only when the user's head was contained in the avatar's head.

- *Frustum (F):* Since only the goal head-pose matters in PDTs and not how the user gets to match it, a frustum visualization metaphor was devised in an analogous form to what done in [62] and [72]. Differently than in those works, though, the simplified head model was changed with one resembling a facemask (Figure 2.8c) so as to better distinguish the Avatar metaphor from the Frustum one and to deliver a greater affordance. The frustum is shown in two distinct ways, depending on the actual head position of the user: if the user's head is far enough from the deliberated pose, the frustum is shown in the form of a pyramid trunk (like in [62]); in the other case (trainee wearing the mask), a rectangular frame is used to show the frustum, like in [64] (Figure 2.8d). Lastly, to complete the metaphor, a gaze-line with the same fading behavior as in the Avatar metaphor was included.

- *Picture in Picture (PiP or simply P):* It has been decided to include also this metaphor in the study since quite common both in real-world applications and in the literature [22]. The metaphor was implemented as a floating board showing in a loop a video prerecorded from an ego-view perspective (Figure 2.8e). Notably, to avoid introducing an unwanted advantage w.r.t. Frustum and Avatar, the video shows just one of the multiple ways to assume the target head-pose and to gaze at the deliberated POI, but does not provide any explanation on how to complete the assigned task. The board was designed to automatically follow the user by locating itself in his or her FOV at a distance needed to have the video fill 15% of the FOV. Also, by grabbing the board (Figure 2.8f), the user may move it anywhere he or she wants.

2.2.2.3 Scaffolding

When the VRTS is used in training mode, a step-by-step GS is used to scaffold the user throughout the procedure to learn (illustrated in Subsubsection 2.2.2.4). A voice-over (a recorded human voice) informs the user of the steps to perform in order to complete the current task. In order to ensure that the user pays attention to the spoken instructions, until the voice-over is complete he or she is prevented from interacting with the VE or from moving. The last voice-over piece may be replayed at the user's discretion. To complement the instructions, mild MR cues (blinking object outlines) are used in conjunction with the given metaphor, which the GS triggers simultaneously with the voice-over. When the learner completes the task satisfactorily, the GS provides acoustic feedback and moves to the next task. When the VRTS is configured in evaluation mode, the scaffolding system will not be activated but the last described auditory feedback on task completion is anyhow provided.

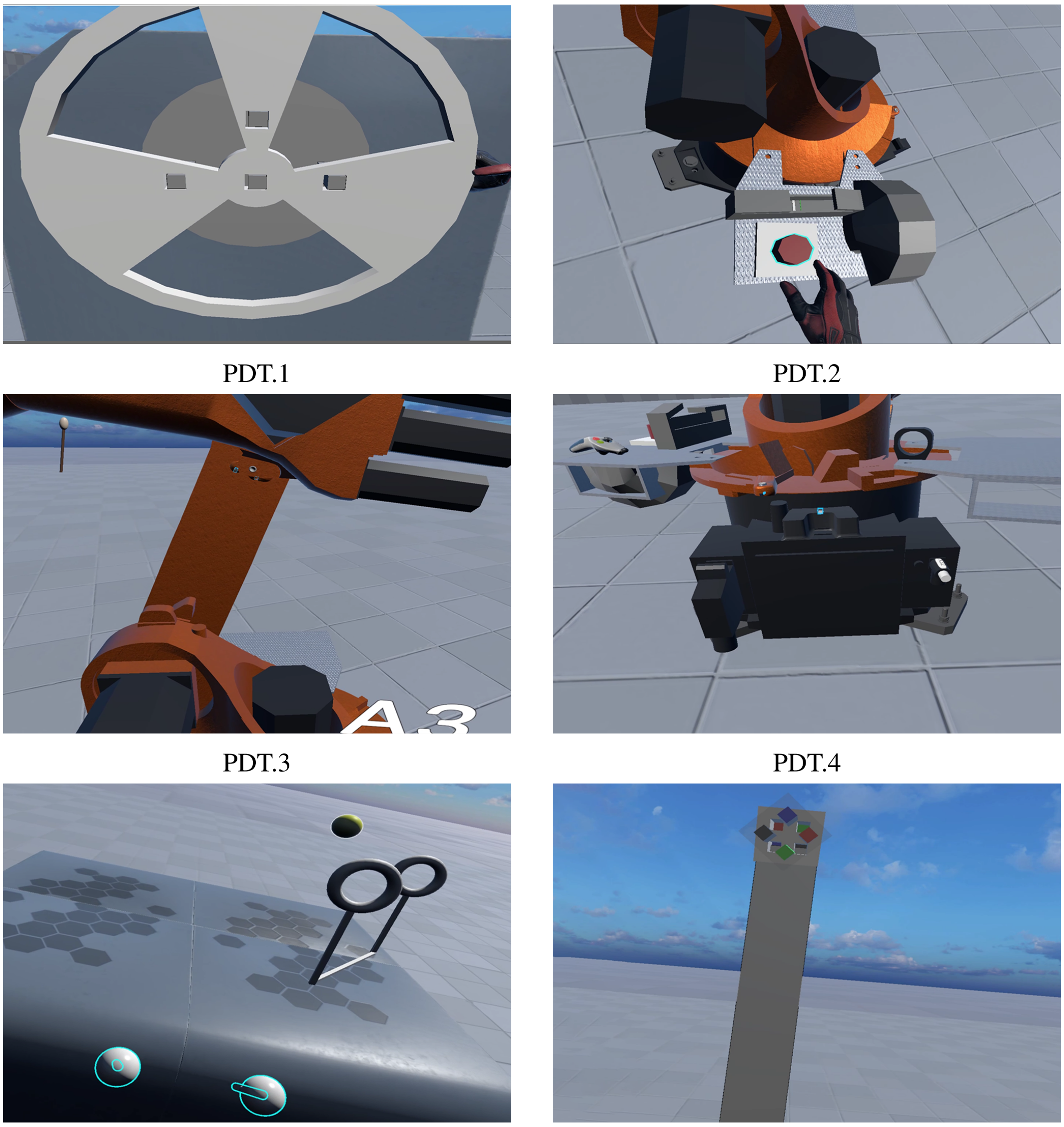
2.2.2.4 Procedure and Tasks

In the following it is reported a description of the procedure devised for the experimental evaluation, which includes the six PDTs in the testbed. It was chosen to construct a fictitious, though realistic, procedure as opposed to copying a real one. Motivations that led to this decision are manifold. On the one hand, it has been possible to arrange a scenario with the specific aim of lining up the various PDTs in order to stress the metaphors. On the other hand, this choice allowed to make the assumption that participants had no prior knowledge of the procedure they would be instructed onto. This was of the utmost relevance, since it is speculated that previous knowledge can have a significant effect when, in a PDT, someone is asked to match a certain head-pose. Accordingly, when designing the PDTs, particular care was taken to avoid trivial privileged observation points, since a task with high intrinsic affordance would have resulted in an easy-to-decipher target head-pose for the participants. Thus, the encompassed tasks were conceived as possible to complete even in case the user is not adopting the deliberated head-pose, even though this may have adverse effects on the learner's comprehension of task execution and on the score assigned by the training provisioner.

When defining the PDTs to include in the testbed scenario, two distinct aspects were considered. Firstly, the categorization of machine-tasks presented in [22] was considered, i.e., the so-called *local* (L), *spatial* (S), and *body-coordinated (C)* tasks: a local task is one that can be completed with one hand from the current operator's location (i.e., within arm's reach); a spatial task requires the operator to move to another location before performing the required operations; finally, to complete the interaction of a body-coordinated task, an operator needs to coordinate his or her body (e.g., by simultaneously using both hands).

Secondly, each of the three classifications above was further split depending on the type of machine control, considering that a machine may be either *directly* (D) or *remotely* (R) controlled (it is speculated that the D category implicitly conveys cues to the operator about the desired head-pose to match). Notably, the distinction between the two categories is not linked to the way the operator executes the task, so if he or she uses a tool (e.g., screwdriver, leverage, and so forth) or directly his or her hands, but to the fact that the operator is confined to a certain position while operating the machine rather than he or she is able to freely move, since in principle it is possible in both the cases to stay far or close to the controlled machine. By combining these two factors (3×2), the PDTs were designed as follows:

PDT.1 <D,L>: The task of installing a wheel on a hub was taken as inspiration for defining this PDT. In order to complete the task, the operator is asked to maintain the wheel as much parallel as possible w.r.t. the wall-mounted hub plane, and align the wheel holes to the spines of the hub (Figure 2.9–PDT.1). The operator moves the wheel by directly grabbing it (D). Ideally, the task shall be executed matching a head-pose that is aligned with the center spine at about 50cm distance from the hub plane. This L kind task was deliberately selected as the first one so the operator will be already in the vicinity of the



PDT.1 PDT.2 PDT.3 PDT.4 PDT.5 PDT.6

Fig. 2.9 PDTs devised for the VRTS testbed.

target head-pose when the metaphor is triggered (first activation). This is critical for priming the avatar follower effect for the A metaphor [71].

PDT.2 <D,S>: To get to a remote controller needed for the subsequent task, the operator has to unlock a container located few meters away from the hub (S). A button (in the VE) must be clicked to open the container. Near the button there is box-shaped compartment in which a green mark moves back and forth passing over a white tick sign. The remote will be disclosed just in case the button click (D) happens when the green and white ticks collimate (Figure 2.9–PDT.2). To correctly check the alignment, the white mark shall be gazed with an angle of 20° from the plane normal.

PDT.3 <R,L>: The MP of IRs [18] (Subsubsection 2.1.2.2) was taken as inspiration for this task and the following one. To control the robot, the operator uses the remote controller (R) gathered from the previous task and he or she is asked to operate the robot to have it match a target pose. This pose is identified by aligning two ticks on the robot moving parts (Figure 2.9–PDT.3). In this task, the ideal head-pose is obtained by looking at one of the two ticks (the remote control affects just one of the two ticks, the other remains still) at about 1m from this POI and 1.50m above the floor (L).

PDT.4 <R,S>: The same as before but on a diverse robot axis. The ideal head-pose was located a few meters away from the previous one (S); to reach it appropriately, the operator needs to assume an annoying and uncozy pose (Figure 2.9–PDT.4).

PDT.5 <D,C>: This task uses a machinery mock-up placed a few meters away from the robot. The machinery is made of two handles that can be operated horizontally (D) and should be moved simultaneously (C) to control a floating ball-shaped object. The operator is asked to place the ball in the middle of two reference rings so to align the ball with the holes. The optimal head-pose is a bit uncomfortable to match, since the rings should be gazed by keeping the hand on the handles (Figure 2.9–PDT.5).

PDT.6 <R,C>: A connector fixing activity inspired this last task. A pillar is hosting a socket located approximately 2.5m above the ground. For this task, another (belt-mounted) remote controller (R) made of two knobs is used to move and rotate the connector. The operator needs to employ both hands (C) with the aim to match the socket color code to the one of the connector colored spines (Figure 2.9–PDT.6). The deliberated head-pose may induce neck-strain in the operators (wearing an HWD) caused by the high positioning of the socket that could be a little bit demanding to maintain.

To exemplify, videos showing the pose matching process for PDT.5 for each of the three metaphors are available at http://tiny.cc/phd_th_ch1-2vids.

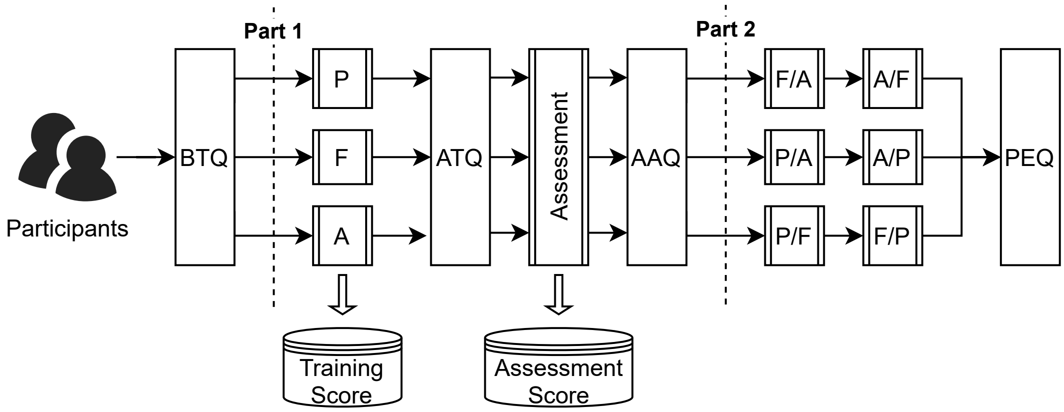


Fig. 2.10 Experiment design and analysis tools.

2.2.3 Experiment

In this section, the user study that was arranged to evaluate the three proposed metaphors is presented.

2.2.3.1 Experiment Design

The sample was made of 15 volunteers (2 females, 13 males) of height between 160cm and 189cm ($\mu = 175.2cm$, $\sigma = 6.9cm$) and aged between 24 and 30 ($\mu = 26.9$, $\sigma = 2.5$), recruited among students enrolled in computer engineering courses at Politecnico di Torino and the available networks of contacts. None of the participants suffered from color blindness, and no further exclusion criteria were considered.

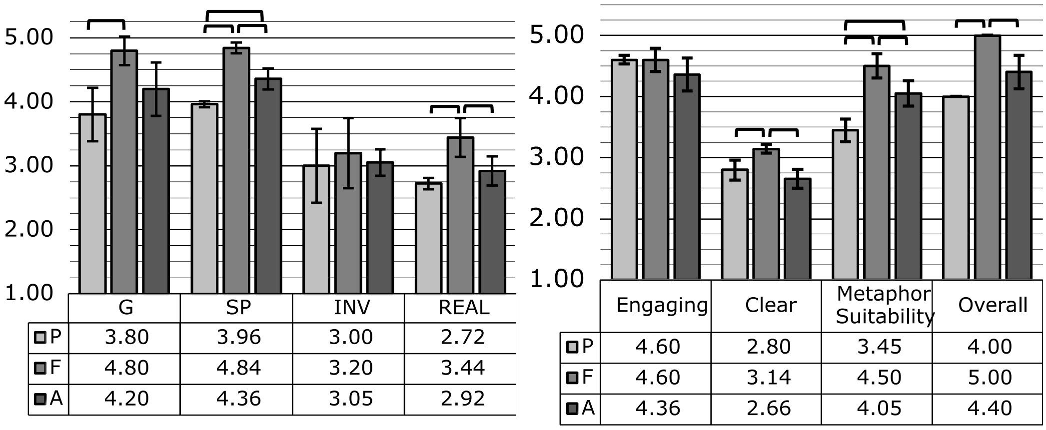
The experiment follows a mixed-design by having it arranged into two phases as depicted in Figure 2.10. Initially, the participants were assigned to three different groups of equal-size by following a between-subject design ($n = 5$). Prior to entering the VRTS, the participants used a "sandbox" VR environment for practicing locomotion in the VE and interaction with objects. After that, each group was assigned a different metaphor and was allowed to experience the VRTS in the training configuration. When the training was over, the participants were asked to execute the procedure autonomously via the evaluation mode of the VRTS. To be able to grade the participant by assigning him or her a score for each PDTs performance, video recordings of the procedure (for both the training and the evaluation runs) were harvested and then annotated. To minimize potential evaluation biases, the video annotation was performed by one confederate who was excluded from the user study activities. It is worth to mention that the recordings were collected within the VRTS software to replace the current metaphor with a dummy representation equal for all the participants; hence, at the evaluation time the annotator was unaware of which metaphor was actually used by the participant. The score was assigned considering both the exactitude of task execution (MARK) and the correctness of matching the target head-pose (POV) along with whether the right POI was looked at or not.

In addition to the scores derived from the video annotation, a multi-part questionnaire was also used to collect subjective feedback from the participants. Before initiating the training, a first part was administered (BTQ) to collect information pertaining prior knowledge and level of proficiency with technology pertinent to the experiment, demographics, and Simulator Sickness Questionnaire (SSQ) data [74]. At the training completion, the second part (ATQ) was administered, consisting of items aimed at evaluating the suitability of the metaphor and the perceived training quality; to complete this part the i–group Presence Questionnaire (IPQ) [75] was used as well. Lastly, after the assessment (AAQ) the part aimed at collecting feedback about self-evaluation and confidence while executing the learned procedure was administered, together with a post SSQ.

Thereafter, the participants joined the second phase of the user study aimed at analyzing the human-computer interaction (HCI) aspects of the three metaphors; this phase, a within-subject design was adopted. In particular, the participants were asked to perform the training again but to experience the two metaphors remaining in Latin square order, and to finally fill in a post-experience questionnaire (PEQ). The PEQ was made of the SUS tool [51], and an ad-hoc part in which, for each examined HCI component, a rank of the metaphors must be provided (without ties) by the participants. Finally, open feedback about the experience was additionally collected. All the standard questionnaires mentioned above (SSQ, IPQ, and SUS) were adopted in their original form.

2.2.3.2 Results and Discussion

By looking at the data acquired via the BTQ, 46.7% of the sample reported to use HWD quite often, whereas 53.3% had never or seldom used immersive VR technology. There were no statistical differences between the three groups, so the random designation of the participants did not seem to have introduced any bias in this regard.



(a) IPQ

(b) VRTS perception

Fig. 2.11 ATQ Results. Significant differences are marked with baffles.

In the following, a discussion of the most relevant findings is reported. Phase I data were analyzed using the Kruskal-Wallis test and Conover post-hoc, whereas Phase II data were examined with the Friedman test and Conover post-hoc. The Mann-Whitney U test was used to investigate the pre/post exposure effects.

Phase I: None of the Pre/Post SSQ indicators highlighted any significant difference, so there was no metaphor that stimulated a greater level of sickness compared to the other. By considering the IPQ (Figure 2.11a) results, the F metaphor was judged to be fairly superior in the respect of the other two, i.e., delivered a higher level of realism and spatial presence w.r.t to both A and P, and fostered a greater sense of presence w.r.t. P. In addition, it was found that the perception of the VRTS (Figure 2.11b) was significantly influenced by the metaphors. Specifically, w.r.t both A and P, the F group perceived the quality of the training as higher overall and the clarity of the information delivered by the VRTS as superior. Also in the case of the metaphor suitability, it was spotted a significant difference in favor of F, which was judged to be superior to A and P (in this order).

The scores obtained from the video annotation throughout the training and the evaluation lead to somewhat different considerations. It should be noted that no statistical differences were found from the pre/post analysis (training vs evaluation scores) for any of the tasks (so, for brevity, in Figure 2.12 just the evaluation scores are reported). Basically, when evaluated, the trainees were able to faithfully recall

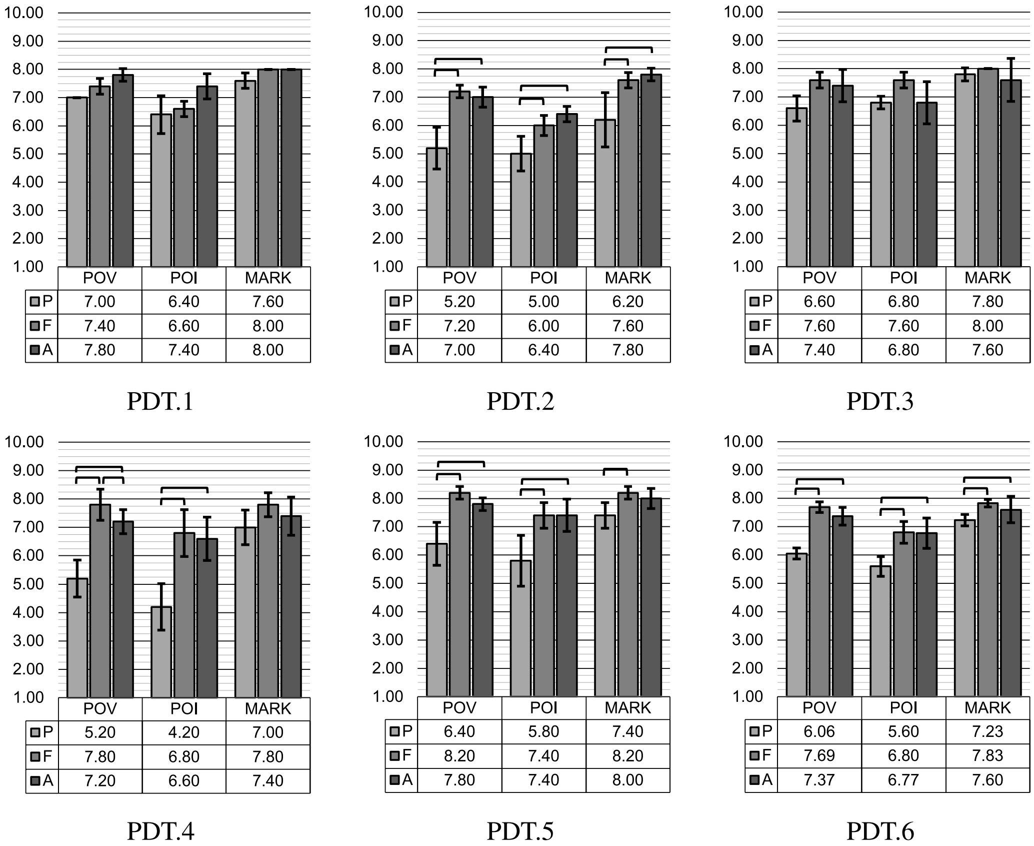


Fig. 2.12 Evaluation scores: Correctness in matching the target head-pose (POV), looking at the correct point-of-interest (POI), task execution exactitude (MARK). Significant differences are marked with baffles.

the procedure by copying what learned during the training, regardless of whether they had correctly executed the procedure or any erroneous head-pose had been memorized due to an ineffective metaphor. As guessed, the D tasks (PDT.1, PDT.3) were the less sensitive to the kind of metaphor employed, and the limited impact on trainees performance is underlined by lacking of any significant differences for these tasks. In the other tasks, a markedly better score was obtained by A and F w.r.t. P; no significant differences were instead found between F and A. There is an exception for PDT.4 (R, S), in which F was able to more effectively encourage the trainees matching the required head-pose. Observing the AAQ respective items, statistical differences were not found among groups, implying that trainees in P had no idea of their not-so-optimal performance.

Table 2.5 Rankings and p-values for Friedman and Conover post-hoc pairwise comparisons. Rankings are obtained directly from PEQ answers. Inverted items are marked with *.

	P	F	A	P-F-A	P-F	P-A	F-A
Overall	3	1	2	**<.001**	**.003**	**.012**	.751
Assertive	3	2	1	**<.001**	**.009**	**.002**	.643
Learnability	3	2	1	**.003**	**.044**	**.032**	.930
Affordance	3	1	2	**<.001**	**.003**	**.012**	.751
Mental workload*	3	1	2	**.010**	**.044**	.113	.850
Efficiency	3	1	2	**<.001**	**.003**	**.012**	.751
Ambiguity*	3	1	2	**<.001**	**.009**	**.012**	.982
Ease of use	3	2	1	**<.001**	**.003**	**.012**	.751
Latency*	3	2	1	**<.001**	**.009**	**.005**	.930
Frustration*	3	2	1	.642	.982	.643	.751
Boredom*	3	2	1	**.010**	.113	**.044**	.850
Visibility	1	3	2	.778	.751	.930	.930
SUS score (SD)	74.8(17.4)	86.3(7.5)	86.0(7.3)	**.031**	**.030**	**.021**	.599

Phase II: After having experienced all the metaphors, the participants corroborated to some degree the preference trend just described. As also indicated in Table 2.5, a significantly better SUS score was reported for F and A w.r.t. P, whereas it was not possible to identify a clear winner between these two. An analogous pattern can also be found in the rankings provided for the several HCI factors investigated (Table 2.5).

2.2.4 Limitations

The representativeness of these findings is subject to certain limitations. Specifically, the power of the study was likely insufficient to discern any potential dissimilarities between A and F. Furthermore, even though the sample features were balanced among the groups, the population of interest may be not fully represented by the considered sample. Despite these limitations, it is believed that this exploratory study could pave the way in understanding better the potential subtleties of using an XRTS for the training of PDTs.

2.3 Considerations and Remarks

The work in this chapter was aimed at framing the effectiveness of VRTSs for the self-tuition of procedures entailing *mixed-tasks*.

The first study that was reported challenged a state-of-the-art VRTS supporting the self-learning of a MT involving an IR against a consolidated training system based on in-class and hands-on sessions, by performing an on-field evaluation. After learning with the two distinct training approaches, the study participants were requested to complete the learned task autonomously on a real IR. By observing both the subjective and objective results, it was concluded that the effectiveness of the VRTS in terms of enabling the trainees to successfully complete the task was overall comparable to that of the ST. Furthermore, the development of proprioception skills and spatial awareness was well backed by both the VRTS and the ST, being the two approaches on pair regarding this aspect and about the capability to foster the transfer of the required procedural knowledge. This finding about the VRTS ability to develop the transfer of spatial awareness is in line with was previously found in [32], and further completes it. However, some of the design decisions drove the VRTS to be more mentally taxing and frustrating for the trainees. Anyhow, the participants trained with the VRTS were satisfactorily instructed on how making decisions based on the system (IR) status and on executing ERPs if demanded; notably, they achieved better proficiency in handling hybrid digital-physical devices. Thus, although redesining IRs control interface specifically suited for a VR setting is indeed valuable [32], it appears that there is not an actual need to do that from a training perspective. The training with the VRTS lasted significantly less than the ST, but more time was needed on average by the VRTS trainees to complete the learned task on the real IR. Yet, the VR-based training was deemed as a pleasant and time-efficient learning method, albeit it fell behind the ST in aspects that are typically strengthened by trainer-trainee social interactions. The findings of the study should represent a litmus of the status of VRTSs for the training of MTs and pave the way for future experimentation in the field by considering some aspects requesting further research. For instance, emphasis should be placed on enhancing the skills transfer efficiency of VRTSs in crucial areas involving interaction, especially with small and delicate items. Moreover, steps should be taken towards making the VRTSs able to ably adjust to the trainees' learning pace (e.g., by adapting the instruction flow to their needs) and to encourage and motivate them during the training (e.g., by giving congruent feedback to both achievements and mistakes). Moreover, one of the aspects needing further improvement appeared to be related to the frustration levels induced by the VRTSs scaffolding system. Specifically, it emerged that the trainees were particularly sensitive to this issue for PDTs learning.

To cope with that specific aspect, a second study was devised, with the prospect of conceivably boost both the pleasantness and the efficacy of the training. In particular, three metaphors inspired by the literature are proposed and compared, with the aim to adopt them in VRTSs that encompass PDTs training, by nudging the trainees to match a head-pose deliberated by the training provisioner. A testbed scenario in the form of an automatic VRTS for training a fictional procedure that included six classes of PDTs was devised to evaluate the metaphors. The findings reported that the most popular methapor (P) was actually outperformed by both the A and F metaphors, which is in line with what has been stated in [61] and further substantiate the value of the A and F metaphors already used for related task in the literature [62, 73] also in the context of PDTs training, albeit it was not evident which was the better between A and F. Discerning this difference should be the subject of further studies, possibly extending the evaluation to MRTSs. Furthermore, an aspect worth investigating is to what extent the metaphors would impact the social presence in a scenario in which they are triggered by a human trainer in shared XRTSs experiences instead of by a scaffolding system.

Chapter 3

Pedagogical Agents in XRTSs

The work described in this chapter has been formerly published in [76, 77]

An aspect that is left as a bequest from the work reported in the previous chapter and could be tackled to potentially ameliorate the effectiveness of a (possibly) automatic XRTSs pertains to the social elements of the training experience.

When putting a scaffolding system side-by-side with a real-world training delivered by a human instructor, it could seem trivial to ascribe the differences between the two approaches to the lack of a human instructor in the XRTS, and to expect that adding such an element in the form of an avatar could mitigate this fact. The core idea of using computer-based characters in an ITS can be traced back to the 70's. These characters, typically referred as to PAs, have been extensively studied in the context of traditional, non-immersive media [78, 79]. PAs are virtual life-like characters, often but not necessarily coming in the form of anthropomorphic avatars, that are used in multimedia educational contexts to support the individuals' learning. It should be noted that a PA can act in a variety of roles, not limited to that of a teacher/instructor; it could also serve as a self-tutor, motivator, mentor, learning companion, and so forth [78, 80].

The aim of including a PA is to mimic the social processes that are typical of real-life teaching. In fact, the PA concept is backed by the *social agency theory* [81, 82]. According to this theory, introducing social cues in a multimedia learning context can prime a sense of social presence in learners that can lead to deeper cognitive processing and, eventually, to an increased learning performance [83, 78].

Although PAs have been proved to be valuable tools in non-immersive multimedia settings [83, 82, 79], their effectiveness when incorporated in XRTSs is still controversial [79, 82]. Recently, the authors of [82] found that including a teacher PA in a VRTS could be a two-edged sword, since it was observed that it helped learners to gain knowledge on conceptual elements, but it actually had a detrimental effect on the learning performance when considering the learning of factual elements.

In this chapter, two studies are presented and discussed with the aim of shedding some light on the use of PAs in XRTSs, as well as of proposing novel ways in which PAs may be exploited in such tools. The first study focuses on the use of a PA in an automatic XRTS by leveraging it to move from a traditional pedagogical approach to a learning-by-teaching (LBT) paradigm. In the second study, the PA is used in a VRTS featuring a real teacher with the purpose of fostering students' participation while managing possible social pressure phenomena.

3.1 Learning-by-Teaching with Robotic PA

Despite the opportunity offered by XRTSs in terms of flexibility, at present both applications and studies focused mostly on the adaptation of the foremost pedagogical model, in the following referred to as traditional learning (TL) approach. In TL, a pedagogue teaches a given content to one or more learners, perhaps taking advantage of supplemental materials such as blackboards, books, or slides. As already stated, in an automatic XRTS, the teacher's role is usually superseded by a scaffolding system. Even though these systems have been proved to reach learning performance on par with the real-world training (Subsection 2.1.4), there is much more that could be done.

Since the 50's, pedagogues have invested a lot of time and effort in designing didactical models to enable pupils to climb the learning pyramids [84] more and more effectively. Pivoting on the didactic model spectrum, on the opposite end of the TL there is the so-called LBT model. LBT roots in the naïve practice of peer-tutoring, in which pupils tutor other peers by teaching each other self-learned domain knowledge from traditional (or non-traditional) sources. Even though the classic

LBT method (humans teach humans) has been demonstrated to be significantly more effective than TL [85, 86], particularly for long-term retention of gained knowledge, it also has some disadvantages. Leaving aside the fact of being less efficient (more time-consuming) compared with TL, the learning boost strictly depends on the role taken by the learner in a given moment, being the benefits mostly evident for the teacher role rather than for the tutee role, and considering the fact that the two roles have to rely on diverse stimuli and feedback [87, 88].

With the aim to have all the students experience the LBT in the most profitable role (teacher) and avert the immolated tutee, researchers have attempted to replace the tutored learner with so-called teachable agents. These agents are PAs that can be taught by the learners about a given subject; by doing so, learners will achieve a more in-depth understanding of that subject [89]. In other words, the actual objective is not to eventually program the teachable agent, but to leverage it in order to stimulate the cognitive processes implicated in the LBT model, enabling the learner to obtain a deeper understanding of the subject through the activity of lecturing someone else.

3.1.1 Background

Given the importance of the social components in LBT [86], empathy included, one of the most promising implementations of teachable agents makes use of service robots [90]. Robotic Teachable Agents (RTAs) have been the subject of several investigations and confirmed to be equally or even more effective w.r.t. their usage in a TL fashion [91, 76], as well as of being capable of activating the cognitive processes necessary for an effective LBT experience [91]. Nevertheless, an ITS employing a bare RTA is quite limited in terms of functionalities, since the principal (and, often, only) form of Human-Robot Interaction (HRI) available is the oral explanation from the learner [92]. Hence, with the aim to expand the possibilities of an ITSs leveraging a RTA-based LBT approach, a handful of researchers proposed to combine the RTA with a MR environment.

Authors of [93] studied the use of a mobile robot in a spatial MR system to teach a topic related to geometry. They observed a different reaction of the learners to the variation of the robot social attribution feedback (diverse pronouns and negative or positive connotations), indicating that the MR environment does not affect significantly the social interaction. Yet, no direct comparison with a TL version was conducted. The same MR robotic training system (MRRTS) was used in a second study [94] to investigate whether the physical RTA offers a significant learning advantage compared to a desktop-like application (no MR, no robot) and to a digital replica of it (just MR). This concern exemplifies a key challenge when it comes to the design of MR robotic experiences, in which the augmented content may supplant the need for a physical robot resulting in a valueless additional complexity [76]. Considering that no significant differences were noted among the three versions about the learning performance, and being this result in contrast to what found in the literature for classic setups, it is conceivable that the devised MRRTS suffered from the said issue.

In response to a call to action from the research community [95, 90] and with the aim to explore the LBT pedagogical model in an XRTS (and, specifically, to clarify if the MR might be detrimental to the RTA features enabling the LBT approach), this section presents a preliminary user study analyzing the learning effectiveness of a MRRTS implementing the LBT paradigm against a TL approach.

3.1.2 Materials and Methods

The implementation of the MRRTS leveraged a commercial off-the-shelf programmable toy robot together with a table-top projected spatial MR setup.

3.1.2.1 Technologies

In particular, among the many options, the robot selected was the *Anki Cosmo* [96], since quite popular and embedded with multiple anthropomorphic features that enhance its compliant social behaviors and emotional connotation (Figure 3.1). An official Python programmable SDK is provided by the manufacturer [97].

Cozmo is a non-holonomic robot sized $6 \times 7 \times 11$cm (at rest) that incorporates two movable elements (apart from the wheels). The first movable element, i.e. the "head", can rotate by 45° upward and 20° downward with one rotational DOF. A "face" implemented with a LED matrix display of 2×2cm completes the Cozmo's head. The display is used to show a stylized anthropomorphic facial expression by means of eye-like animations (selectable from a pre-defined list included in the SDK). Located underneath the display there is a 640×480 pixels RGB camera with a field

Fig. 3.1 Anki Cozmo [96].

of view of 60° (however, the SDK allows to read an image limited to a 320×240 in grayscale). The camera can be used, by leveraging the respective SDK built-in feature, to have the robot simulate a *look-at* behavior by orienting the head and the whole robot towards the user (Cozmo can automatically track the user's face). The second movable element is a front lifter (controllable via the SDK, with one positional DOF), that although mainly designed to interface with some interactive cubes (tangible objects) provided in bundle with the robot, can also be programmed for custom needs like being retargeted to have the robot mimicking interactions with the projected environment [98] (i.e., tap-like animation). Cozmo features WiFi connectivity, and the SDK also contains a Text-To-Speech (TTS) module that can be used to output sounds on the robot built-in speaker. The SDK architecture is based on an event-driven approach with a plethora of features available (for brevity, in the following only the subset of features that were actually exploited in the work are mentioned).

As illustrated in Figure 3.2, the high-level architecture of the MRRTS arranged for the study is composed of Cozmo, a projector, an RGB-D camera, a PC, and an Android smartphone. As previously stated, a spatial MR setup was selected for the devised system with a table-top projection of the augmented digital contents. In the following, just a concise description of the implementation is given, since the

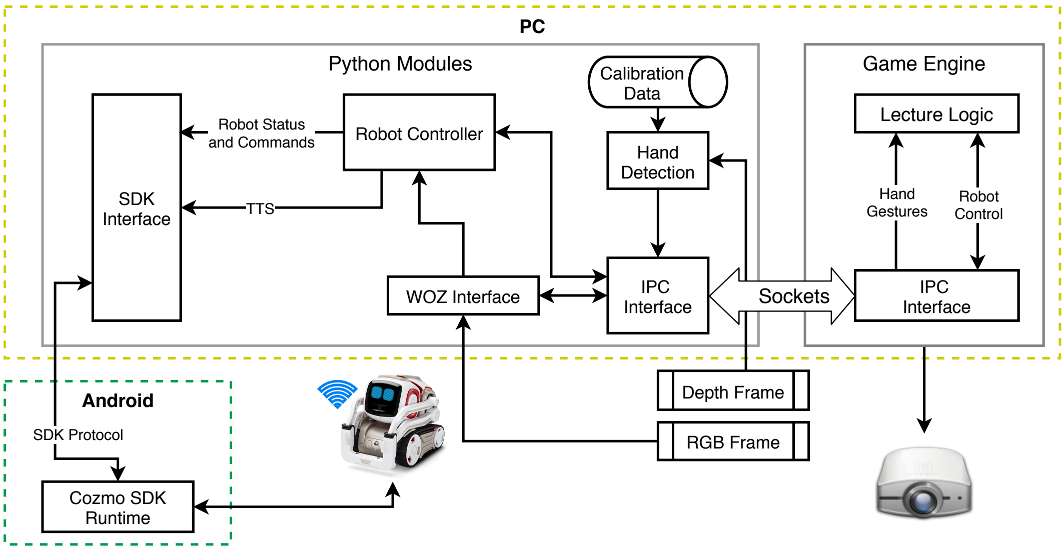


Fig. 3.2 MRRTS Architecture.

selected setup (Figure 3.3) was one of the most popular in previous works [98] and was used also in [93, 94] when studying the subject of LBT with RTA. To deliver the projected environment onto the table, the projector was mounted close to the ceiling. The table was also blanketed with a black cardboard sized 85×65cm (the projected surface has coincident size) to ameliorate the projected image clarity. In addition, to enable natural gesture interaction with the MR environment [98], a Microsoft Kinect v2 was affixed in close proximity to the projector. In the arranged configuration, both the depth-sensing 512×424 pixels camera and the 1920×1080 pixels RGB camera were used. The first camera enables touch-based interaction with the projected surface by performing hand gesture recognition. Specifically, well-known computer vision methods (background subtraction, opening, depth-level thresholding, contour detection) were used to process the depth image. These functionalities, implemented using the OpenCV (v3.2) library [99], were used to identify the hand position and its configuration (i.e., closed or open), as well as to detect three touch gestures, i.e., tap, drag, and slide. The second camera was instead used by the Wizard-of-Oz (WOZ) interface described later (Section 3.1.2.2).

Due to the fact that the estimated odometry provided by the SDK does not provide an accuracy matching the minimal necessity of the scenario of interest (mostly due to issues related to error drift), in order to track the robot position a depth image processing parallel to that performed for the hand gesture detection was adopted, by resorting to an *outside-in* approach. On average, this tracking algorithm is capable of

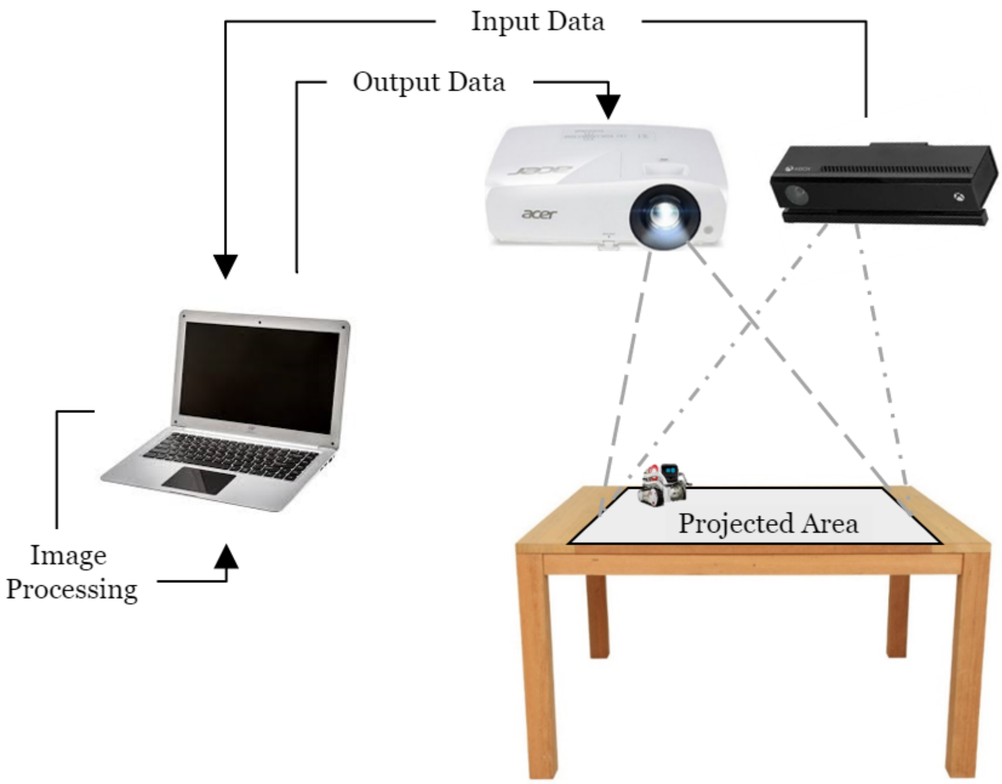


Fig. 3.3 MRRTS Setup.

$\overline{Err} = 0.81cm \pm 0.62cm$. For lining up the coordinate system used by the projection and the external tracking with the robot internal coordinate system, a calibration step is performed to compute the needed transformation matrices.

The TTS functionality of the SDK was used to provide voice feedback when needed (in English). The implementation of the lecture graphics and logic leveraged the Unity (v2018.3) game engine [40], to be deployed on a Windows PC. An additional Python application was scripted for the robot control logic and the gesture detection so to exploit the robot SDK functionalities. The WOZ interface was instead delivered through a webpage developed with Flask and programmed in HTML5 and JavaScript. The various modules were then allowed to interface with inter-process communication (IPC) by exploiting ZeroMQ sockets [100]. Since to execute the SDK runtime an Android smartphone is needed (which uses a WiFi network hosted by Cozmo itself), this device was connected via USB cable to the PC hosting the applications.

3.1.2.2 Experience Design and Implemented Variants

As already stated, this study was aimed at evaluating the learning effectiveness of the TL didactic model against the LBT in a MRRTS. With that in mind, two variants of a new training experience, named *MireLab*, were designed and implemented.

Topic: The *Thèvenin Theorem* from the electronic engineering domain was selected as training topic for MireLab. Since the population of undergraduate students from electronic engineering was selected as a target for the study, the topic was chosen as a tradeoff between not being overly complex so that the students should not feel overwhelmed having they a fitting level of prior knowledge about the domain and not being too basic so to keep the learners challenged and avoid boredom. Specifically, at least the Kirchhoff's Circuit Laws and the Ohm's laws were assumed as background knowledge. A hypothetical lecture that could be given in an electronic laboratory was taken as inspiration to design *MireLab*. In such a context, students would have been provided with additional lecture material, such as paper sheets for notes/calculation, slides, and definitely a test bench with components to build a circuit and check the acquiring knowledge.

Common Foundation: With the aim to reduce the discrepancies between the two variants, both of them were built upon a common foundation to what concerns the robot features exploited and the projected environment (interface). State-of-the-art guidelines for MR-based robotic experiences [98] were considered in the design of MireLab. The primary interface is made of three zones (Figure 3.4a). The first zone (top-left), which occupies the greatest portion of the projection, is made of a whiteboard area in which is possible to construct the circuits and introduce other information (e.g., pictures, equations, etc.). On the right, it was devised a *components area*, in which either the student or the robot are able to select the desired electrical components by performing a coherent gesture: tap-animation (moving the lever) for the robot, and finger-tap for the student. As soon as a component is selected, it is possible to configure it in a buffer space (bottom-right) by picking the desired value from a dropdown list. In the bottom-left panel is instead possible to access few auxiliary buttons to orient the component (while in the buffer) or erase the whiteboard. From the buffer, a drag & drop gesture (coherent for both the robot and the student) can be used to position the component into the whiteboard area.

Afterwards, it is possible to connect (wire) all the desired components by pressing the *cable button* and subsequently choosing the appropriate component terminals. Finally, a pop-up tool is provided that serves as a calculator or as an input tool for the LBT variant.

As said, Cozmo is allowed to move freely across the projected area and interact with it by mimicking the respective student actions. Moreover, to promote the illusion of a living being, micro-choreographies are continually played by Cozmo. The robot is also allowed to interact with the student by communicating via the speaker (TTS feature), or by displaying elements in the shared projection.

Traditional Learning: In this variant, the robot acts as the teacher whereas the user assumes the role of the tutee. For such configuration, well-known practices were used for the implementation. A software controls the robot by relying on a a finite-state machine logic. The robot adopted as a teaching style the *I-do, We-check, You-practice* delegation pattern, essentially regulating the lecture pace through feedbacks to the tutee and milestones reaching. In other words, the robot explains (the concept), portrays and solve (breakdown exercises) while communicating to clarify content to the tutee. Furthermore, the robot also occasionally involves the tutee during the lecture, for instance asking him or her to remove a given component (indicating it) or to choose the component value. The purpose of these collaborative interactions is to maintain the tutee's focus throughout the explanation, which should lead to an active learning and to a more engaging experience.

Learning-By-Teaching: As anticipated, for the LBT variant the learner (user) acts as a teacher lecturing the RTA. In the design of this variant, the four critical steps that have been demonstrated in the literature to effectively maximize the learning gains [101] for an LBT approach were considered. Specifically, the learner has to:

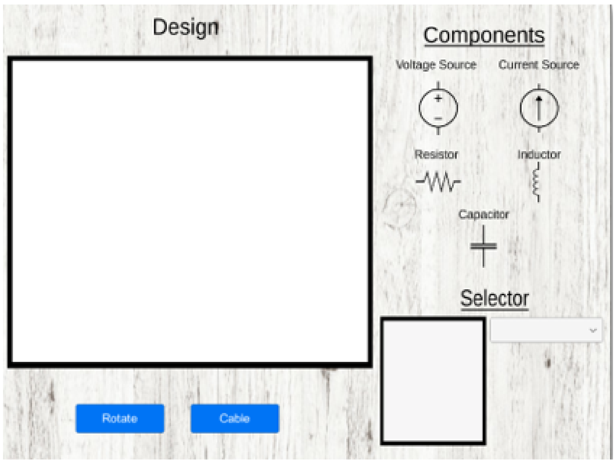
1. prepare to teach (expectation to teach);
2. explain to others/RTA (teaching);
3. interact with others/RTA (Q&As, feedback to the RTA);
4. observe the RTA spending the acquired knowledge (recursive feedback) [102].

For the first step, the learner is asked to prepare the lecture on his or her own by studying a provided one-sheet long document [93, 94]. The cheat sheet, that is available for download at http://tiny.cc/s8utsz, encompasses a terse dissertation of the topic matching the information that the robot provides in the TL variant. Although it is organized in such a manner that it alludes to a particular order to be followed by the learner when later teaching, some points of the document provide the learner considerable directions about how he or she could conduct the lecture. All the included circuits, equations, and images are referenced with a numerical code. The learner can use this code during the lecture, via the input tool feature, to add on the fly these snippet elements to the whiteboard.

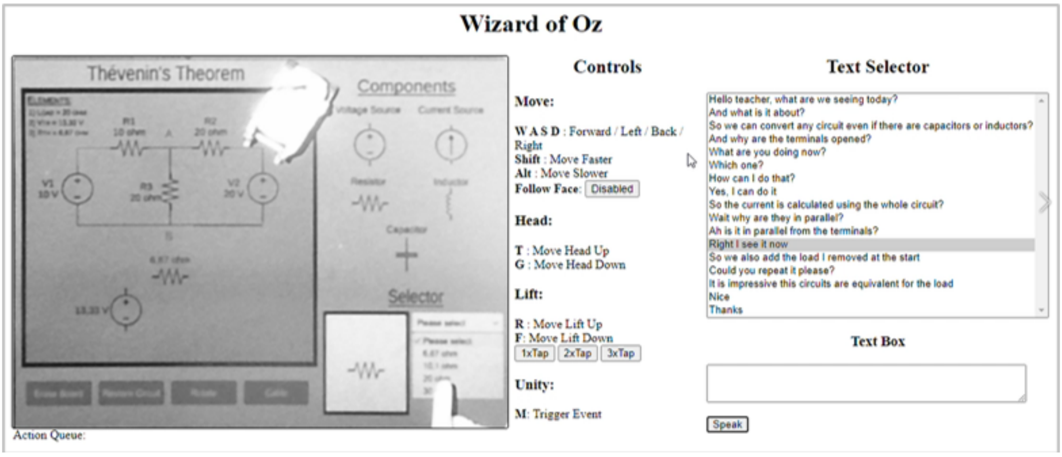
Afterwards, the learner carries out the lecture using the MireLab interface and, meanwhile, interacts with the RTA using the voice and via the MR environment (steps 2 and 3). The robot participates in the lecture by asking for clarifications and possibly executing tasks as per learner's request, so as to rise up the involvement level of the robot and avoiding the possibility for it to go unnoticed [98].

Lastly, the learner sees the robot solving an exercise about the lecture topic while it interacts with the MR environment (step 4). In order to minimize possible biases, it was decided to implement this by having the participants watch a prerecorded video (identical for all of them) of the robot performing such activity. It is worth mentioning that, in the LBT variant, the robot behavior is controlled with a WOZ approach, differently from what it happens in the TL variant where it acts autonomously. This choice was due to the RTA need for intricated interaction, in light of the fact that there are no or very immature Artificial Intelligences (AIs) available for that particular purpose (and that creating such AIs was outside the scope of the study).

Wizard of Oz: As shown in Figure 3.4b, the *"wizard"* (an user that controls the robot by mimicking an implemented AI logic, for the experiment this role was taken by a confederate) is provided with a WOZ interface that allows him or her to have the robot behave comparably to how it acts in the TL with the AI control. The robot is not controlled only manually, but the wizard is provided with assisted controls to minimize the interaction discrepancies w.r.t. the TL robot behavior and simplify operation. The wizard uses the Kinect RGB camera feed to observe remotely the MR environment and teleoperate the robot. To this aim he or she can use keyboard and mouse input to either control the robot directly or by clicking on a point of the



(a) The devised projected interface of MireLab



(b) The WOZ interface as seen from the wizard's viewpoint

Fig. 3.4 Mirelab interface.

camera feed so to have the robot automatically move to the respective location with the shortest path motion. Particular care was taken to systematize recurring Q&As that the RTA could have to provide to the handle. The wizard may select one from a list and possibly edit the text or combine multiple items together to subsequently have that spoken by the robot (via the TTS engine). In addition, the wizard may also activate a variety of preloaded animations encoding various reactions and emotions. Finally, in order to trigger particular application events, dedicated buttons were included. This feature is critical for emulating a robot interaction with the MR environment.

3.1.3 Experiment

In this section, the preliminary user study arranged to evaluate the training effectiveness of the LBT against the TL approach in a MRRTS scenario is presented, and the results discussed.

3.1.3.1 Experiment Design

As anticipated, the target population for the study was set to the undergraduate students from an electronic engineering degree that satisfied the criteria of having enough background knowledge but scant regarding the chosen training topic, i.e., the Thévenin's Theorem. To select accordingly the participants, each volunteer was tested using a multiple-choice screening quiz, which encompassed both practical exercises and theoretical questions about the two areas of interest (five questions about the topic and 10 questions on background knowledge). The study sample was then made of volunteers that obtained a score lower than 6/10 about the topic and greater than or equal to 6/10 on background knowledge. The resulting study population consisted of 6 male participants aged between 22 and 25 ($\mu = 23.83, \sigma = 1.07$). To minimize the learning effect bias, the experiment was arranged with a between-subjects design, hence forming two equal-sized groups (TL and LBT) by randomly allotting the participants ($n = 3$).

Before undergoing the training, a before-training questionnaire (BTQ) was administered to the participants so to collect information pertaining prior knowledge and level of proficiency with technology pertinent to the experiment, familiarity in teaching other individuals, study habits, and behavior while attending classes. Afterwards, a tutorial on the system features and interface was given to the participants of both the groups with minor differences (mainly regarding the functionalities of the snippets input tools in the LBT variant). After that, the participants experienced the training. For the LBT, it was permitted to annotate the cheat sheet while studying it to prepare the lecture. The participants were also allowed to check the notes while teaching the robot, nevertheless, to avert the possible negative impact on the HRI that is caused by this communication barrier (will the learner constantly keep the paper notes in their hands), they were instructed to consult their notes sporadically and leaving them on the table (outside the projected environment) for the rest of the time. Furthermore, in order to thwart the participants from preparing to lecture/study in a shallow way, the maximum allowed time for consulting the notes was set to five minutes cumulative. After that, the participants were brought to another room where they watched the video recording of the robot spending the knowledge just taught. For the training stage of the other group (TL), no particular expedients were adopted. Regardless of the group, for each participant the time elapsed to complete each step was recorded. At the training completion, a post-experience questionnaire (PEQ) was administered, consisting of items aimed at analyzing the participants' robot perception (via the Godspeed questionnaire [103]), together with few self-efficacy items to examine the perceived learning gains, and ad-hoc additional items regarding the peculiarities of the experiment. The SUS tool [51] completed the PEQ. To assess objective learning gains, a post-training test (PTT) was also administered. This quiz is an extended version of the screening quiz (with eight additional questions to the five questions of the screening quiz). Lastly, a final questionnaire (FQ) investigating the satisfaction with the training experience and its perceived quality was administered. Given the fact that, according to the literature, the major benefit of the LBT w.r.t. TL is the superior long-term retention of the learned content, for this study a retention test (RT) was also performed. Specifically, one week after having being exposed to content associated with the Thevenin's Theorem, the participants filled in again the PTT quiz. All the devised questionnaires and quizzes can be accessed at http://tiny.cc/p9utsz.

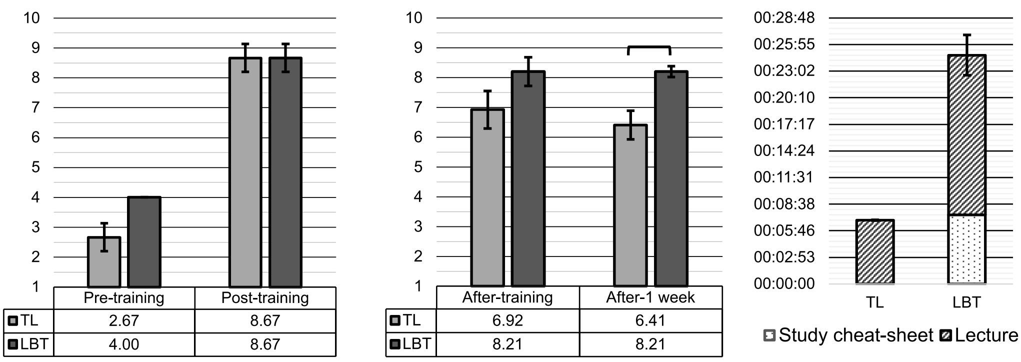
3.1.3.2 Results and Discussion

The two-tailed Mann-Whitney U-test ($p \leq 0.05$) was applied to the experimental data to look for significant differences between the two groups. No significant differences between the groups were found for any of the investigated BTQ elements. On average, the participants were extremely familiar with touch-screen interfaces and were occasional videogames players. Conversely, they were little to none familiar with either toy and service robots. Additionally, five participants reported teaching individuals at least once a month, whereas one seldom or never (belongs to the LBT group).

Concerning the PEQ, the five elements of the Godspeed questionnaire did not show any significant difference (likeability p-value $= 1.00$, anthropomorphism p-value $= .70$, animacy p-value $= 1.00$, perceived safety p-value $= .40$, perceived intelligence p-value $= .70$), indicating that in both the groups the robot was perceived to behave similarly. This finding also seems to corroborate the assertion that any bias was introduced in the comparison by the LBT implementation of the WOZ. For what it concerns the overall usability, both the variants were judged as barely acceptable according to SUS results ($\bar{x}_{TL} = 68.3$, $SD_{TL} = 14.6$ vs. $\bar{x}_{LBT} = 61.7$, $SD_{LBT} = 7.2$), although not differing significantly (p-value $= .70$). By looking at the open feedback collected from the participants, the somewhat low ratings were mostly attributable to the touch surface sluggishness.

Notably, a significantly higher self-efficacy was observed for the LBT ($\bar{x}_{LBT} = 4.0$, $SD_{LBT} = 0.00$) compared to the TL ($\bar{x}_{TL} = 3.11$, $SD_{TL} = 0.38$), along with the participants' confidence about "*successfully pass a test on a thévenin's theorem without further training*", ($\bar{x}_{LBT} = 4.0$, $SD_{LBT} = 0.00$ vs. $\bar{x}_{TL} = 2.33$, $SD_{TL} = 0.58$). Contrarily, no significant differences were found for the FQ items, indicating comparable perceived quality of the training and satisfaction levels.

Objective Learning Gains: The objective outcomes regarding learning gains (pre/post learning scores) are illustrated in Figure 3.5. Regardless of the group, all the trained participants passed the evaluation successfully. Specifically, significantly higher learning gains were spot in both the groups for the shared items of the PTT (Figure 3.5a), suggesting that both the variants were effective. It is worth noting that, even though the LBT scored higher than the TL in the complete PTT, this difference was not significant (Figure 3.5b). Hence, it could be speculated that, differently than with other media, the intrinsically interactive nature of MRRTSs and



(a) Learning Gains (b) Retention (c) Training Time

Fig. 3.5 Objective results of the study. All scores are normalized to 10 and significant comparisons are marked with baffles.

a good implementation of the related best practices tend to thin the efficacy disparity between the two learning approaches. Yet, the limited sample size probably affected this outcome. Nonetheless, a significant difference in the retention test (full PTT scores, immediately after training vs. after one week from the exposure) was found. As it can be noted in Figure 3.5b, the LBT participants reported a lower knowledge decay w.r.t. the TL ones; they were also able to satisfactorily pass the after-1-week test with a minimum score of 7.7, whereas this was the top score obtained for the TL (and one of the participants was not able to achieve a sufficient grade). This finding corroborates the LBT superiority when it comes to granting long-term retention. Since these outcomes are in line with previous ones in the LBT literature, it may be inferred that the MRRTS correctly activated the required mental processes, as well as that the MR addition had no overly adverse impact on the LBT efficacy.

More in depth, by applying the Scheirer–Ray–Hare Test to the retention test outcomes it is possible to observe that the training approach was the main differential factor. Specifically, no significant interaction effects were observed between the exposure time and the learning approach (p-value $= .80$), and this difference is unlikely affected by just the exposure time (p-value $= .46$), whereas a remarkable significance was spotted for the training approach (p-value $= .003$).

Finally, regarding the training efficiency, the reported findings are compatible with those of previous studies (Figure 3.5c, being the TL significantly brisker than the LBT (about four times). This is attributable to the time spent studying the cheat sheet, but it is also largely ascribable to the larger time spent while interacting with the robot (teaching).

3.2 Stimulating Student Participation with a Peer PA

Indeed, it is possible also to envisage XRTSs that renounce to replace the human tutor but allows him or her to interact with the learners in a shared XR environment. Diversely from the pure simulation-based approach, these tools take advantage of networking to allow multiple users to act in the same VE despite potentially being physically located far from each other. There are plenty of commercial platforms available that are frequently used also with education purposes, like, e.g., *AltspaceVR* [104] and *Horizon Worlds* [105].

Broadly speaking, online education has become one of the most popular and indispensable means for accessing educational content. In addition to its intrinsic value, the COVID-19 pandemic has prompted institutions and individuals to welcome remote learning (RL) as a fallback choice due to the temporary infeasibility of on-site education.

RL can be delivered primarily in two ways, either by following a synchronous or an asynchronous strategy. The asynchronous strategy heavily confides in the learner's ability to self-study prepared didactic material, such as prerecorded audiovisual lectures, books, and so forth. Because of their proneness to scalability, most massive open online courses (MOOCs) have adopted this approach of RL in the last decade. Nevertheless, the process of organizing such a type of didactic material and, even more demanding, implementing substantial modifications to the pedagogical strategy may be regarded as a waste of time if the RL is adopted just as a transitory option. Consequently, the most practical style of RL used globally is the synchronous strategy, being it, between the two, the most similar to traditional classroom-based learning because it leverages real-time communication tools such as video-conferencing platforms.

When it comes to backing teachers and students, albeit these platforms typically offer a set of digital tools that are often superior to what is generally achievable in conventional classroom environments, there are still some deficiencies in the social and emotional factors of teaching and learning [106, 107]. The use of platforms based on VR environments has been considered a cogent and valuable option to cope with this issue [108, 18].

The higher level of co-presence and embodiment promoted by VR, if compared to video-conferencing systems, had been reported to well support the social features of learning without compromising knowledge transfer [107, 109]. Still, numerous open points remain. One of the most frequently noted social-related concerns with synchronous RL is that students are less inclined to communicate with the teacher even if polled to interact with him or her [106]. Within this context, it is rather common to witness a teacher opening the floor for questions or even asking something to the classroom and receiving no answers in response, fostering a displeasing and awkward silence (hence nourishing a vicious spiral). To complete this picture, it is worth mentioning the higher social pressure that may be experienced in RL scenarios that may contribute to worsening this induced bashfulness, with teachers often ineffectual to remedy such situation. The depicted circumstances may distract students and can potentially result in a less enjoyable and engaging learning experience w.r.t. to an on-site one [106, 107].

With the aim to mitigate that problem, in the following is proposed a strategy that, by leveraging a PA in an unconventional way, it is expected to foster interaction between students and the teacher in a synchronous, VR-based RL context. The core idea behind it is that the addition of a PA in the form of an “undercover" CA acting as a participating peer-student could lessen the social pressure by ice-breaking the interaction between the students and the teacher, lay down an excuse to demonstrate to students which dynamics occurs when the teacher answers one of their questions, and potentially reassure them in rising questions themselves and interact more confidently. In short, in terms of lecture participation and code-of-conduct, the CA presents real students with behavioral examples to follow.

3.2.1 Background

CAs are software capable of holding a meaningful conversation with a user by processing his or her speech and/or written textual information [110]. In recent years, CAs have increasingly become widespread in everyday life, thanks to the ground-breaking advancements in Natural Language Processing (NLP) and AI domains, as well as to the ease of access to commercial software services. Definitely, in the form of PAs, CAs have been used in education for a while [110].

As already stated, having a PA acting in the teacher role has been indeed primarily investigated [81]. For instance, in [111], it was found that a PA is able to foster enhanced learning gains w.r.t. multimedia instructional material, by also stimulating the areas of the brain associated with social behaviors and relations. Unfortunately, it is axiomatic that having a PA in the role of a teacher is purposeful solely when the human teacher is absent, i.e., only in asynchronous remote learning layouts. Consequently, previous works that studied the use of CAs in the context of RL proposed new ways of using them, either in their declination of chatbots (textual only) or as fully-fledged animated avatars. For example, the use of CAs to ameliorate collaborative and peer learning has been the subject of multiple studies [110, 112]. The authors of [113] devised a chatbot with the capability to scaffold online collaborative learning discussions by encouraging productive practices (i.e., stating disagreement and agreement, explaining, reading and reusing other students' arguments, etc.) leveraging the so-called *academically productive talk* strategy. The proposed strategy was confirmed to be valuable and the CA was able to facilitate interaction and promote reasoning in such an asynchronous remote learning scenario.

In [114], it was demonstrated that chatbots may be used in a group chat of strangers to stimulate interaction among them and relieve social pressure; unfortunately, this study was not conducted with learning objectives in mind. Analogously, the authors of [115] found that group conversations can be effectively encouraged by a chatbot agent by governing the discussion time, organizing members' opinions, and stimulating members to participate equitably. Interestingly, the findings of the study reported in [116], were the chatbot is still acting in the facilitator role in a group conversation but with a learning perspective in mind, indicate that a CA may be utilized effectively to promote interaction among students participating in a synchronous study group of a MOOC (without a teacher present), by also improving explicit reasoning and individual domain knowledge acquisition.

Differently, in [117], the CA is not behaving as a scaffolding PA, but instead bring to life virtual classmates in a VR RL classroom. In particular, previous learners' time-anchored messages were condensed to implement the CAs behavior. It was found that both social interactivity and learning outcomes were positively affected by the CAs. However, the devised system was arranged in asynchronous remote learning, but without a human teacher to interact with and with predefined instructional material. Therefore, it could be argued that the system is actually based more on a stigmergy implementation (to deliver FAQs to students) rather than truly pretending to use the CA-based classmates as genuine student impostors.

To summarize, it appears that the majority of previous works were interested in chatbots designed without the aim to exploit them in VR-based RL, with a paucity of studies that analyzed in synchronous learning scenarios the function of CAs in non-teacher roles. Above all, to the best of the author's knowledge, the study reported herewith is the first one investigating the CAs as an interaction catalyst in a synchronous learning context.

3.2.2 Materials and Methods

In order to evaluate the undercover CA concept by means of a user study, a VR RL system including a CA was arranged whose details are described in this section.

3.2.2.1 Classroom and Lecture

By using the outcomes of [107] as a foundation, to deliver the virtual classroom experience the open-source Mozilla Hubs [118] project was selected. Mozilla Hubs is a social platform that enables multiple users to gather in a VE while using customizable avatars. It is possible to access this platform via a web browser from different devices, including HWD thanks to the WebXR supports. In this study, the built-in *science class* environment was used, experienced with the Meta Quest 2 HWD [119]. The environment adopts an auditorium style layout, with seven columns of three chairs rows available for the student avatars. There is also a whiteboard serviceable to display lecture material, and a table with a chair dedicated to the teacher avatar. The undercover CA was arbitrarily named *Aria* and seated in the first row. For the experiment, the participant avatars (real students) were seated in the same spot, both to ensure that he or she can simultaneously view the teacher and *Aria*, and to prevent possible biases. In the classroom, as shown in Figure 3.6, aside from the user, the undercover CA, and the teacher, four auxiliary dummy students (i.e., inactive avatars) were also sparsely deployed. To facilitate the user in effortlessly understanding the roles, a green avatar was used for the teacher, whereas an identical blue avatar was assigned to all the students. In order to restrict the focus of the study on the student perspective, the teacher role was played by a confederate of the author's research group. To minimize possible discrepancies for the study participants, and since implementing a CA that can address any topic under any circumstance is out of the scope of this exploratory study, the teacher lectured by following as strictly as possible a common script prepared in advance about the *solar system* topic, with an intended lecture duration of 15 minutes. A set of slides was supplied to the teacher for use during the class. Before jumping into the main lecture content, it was decided to have a *warm-up* phase in which the teacher starts greeting the students and attempts to engage in small talk with them (e.g., asking for their names). This was to give the user an opportunity to become acquainted with the

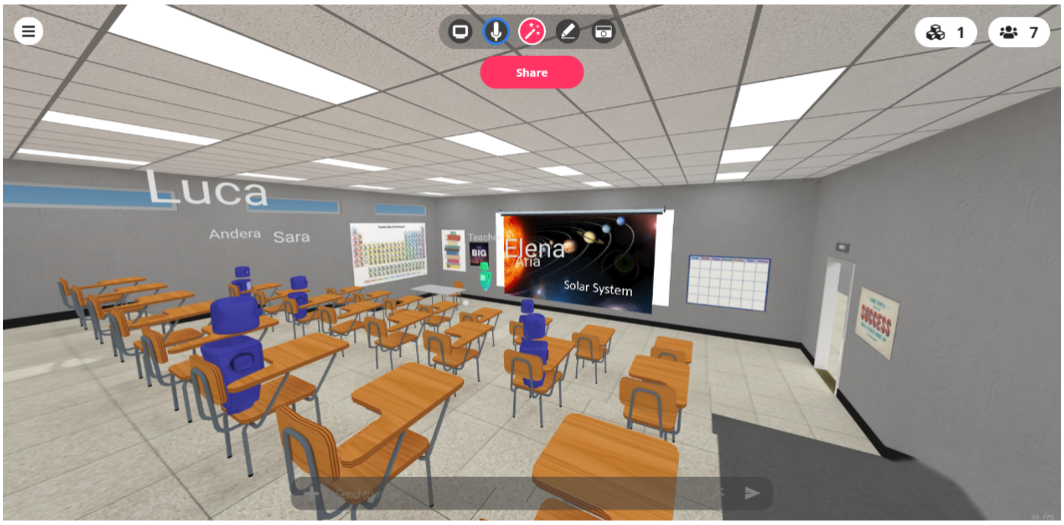


Fig. 3.6 Mozilla Hubs science classroom used for the study; the CA is seated in the front row.

teacher and to encourage social interaction. For similar reasons, a brief *farewell* phase was also conceived towards the conclusion of the class.

3.2.2.2 Undercover CA

Design: By scrutinizing an ideal participative student's behavior, a minimal set of features that should be supported by the undercover CA was derived. Specifically, the following interaction patterns occurring between the teacher and the students were pinpointed.

- *Student clarification*: the student requests permission to speak by activating the Mozilla Hubs raise-hand emoji; then, the student asks the teacher a question, after getting permission.
- *Teacher question*: the teacher poses a question to the class and waits for a response; if no response is received after about 10-15 seconds, the lecturer offers the answer and then resumes the lesson; a *student clarification* that the teacher redirects to the class also falls in this case.
- *Student answer*: the student responds to a question from the teacher as a follow-up to the preceding exchange.

- *Teacher lecture*: the teacher carries on the lecture by illustrating the content to the students; a response to a *student clarification* also falls in this case.

Particular care was taken to implement the *student answer* interactions for *Aria*, in order to not only demonstrate by example a participative behavior, but also allow the user to effectively be the participative student. Specifically, the CA delays its response by nine seconds. This threshold was chosen as a result of preparatory trials made during the development steps, as it was determined to provide a good balance among the considered aspects. If the user does not begin a *student answer* interaction during this time interval, *Aria* will intervene at the timeout, otherwise will not. Additionally, if the user steps in while the CA is uttering a (*student clarification* or *student answer*), *Aria* will instantly cease its activity. Also, a set of predefined questions (*student clarification* interactions), six in total, one for each of the lecture's key topics, were preloaded into the CA.

Furthermore, with the aim to avert *Aria* acting over-participatory (thus being regarded as a sycophant) when paired with a particularly reticent user, it was defined an activity rate that is used by the CA when it comes to deciding to initiate a *student answer/student clarification* interaction. The activity rate is defined as the ratio between the number of words uttered by an agent (CA/student) and the teacher. This statistic is continuously updated, and until the threshold (set at 20%) is not exceeded, the CA will stay inactive.

Implementation: To implement the CA, a custom client was developed in JavaScript by extending the Mozilla Hubs codebase [120], and deploying it on a self-hosted NodeJS server. In Figure 3.7, a high-level architecture of the system is illustrated. The CA was endowed with the ability to communicate with the classroom through speech, and Google's DialogFlow was used to implement its logic. The Speech-to-Text (STT) module of Mozilla Hubs (WebSpeechAPI) is responsible to process both the user's and the teacher's utterances. The former are parsed by the STT engine, but the resulting text is actually ignored by the DialogFlow instance and utilized just for detecting the user's activity together with computing the corresponding activity rate needed to properly adjust the *Aria*'s behavior. Contrariwise, DialogFlow continuously analyzes the teacher's transcription to look for matches with the deliberated intents that define the *Aria*'s behavior within the given lecture script. With the aim to have the lecture flowing as much freely and naturally as possible, the following

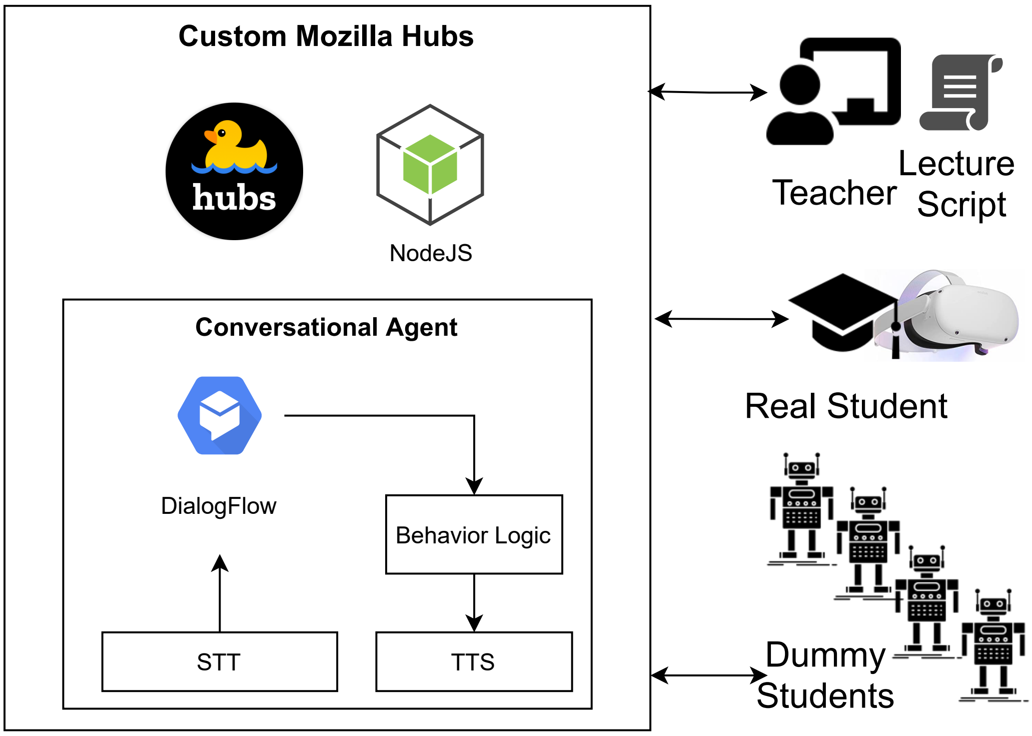
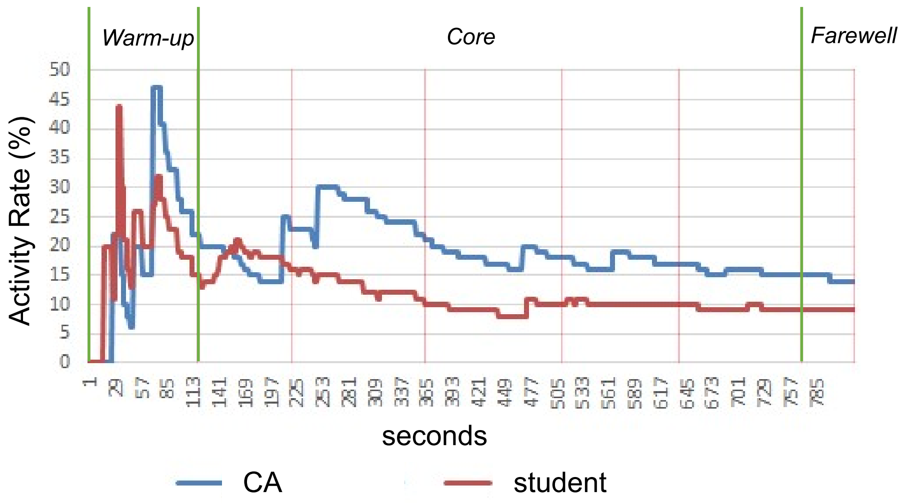


Fig. 3.7 High-level architecture of the devised setup and CA.

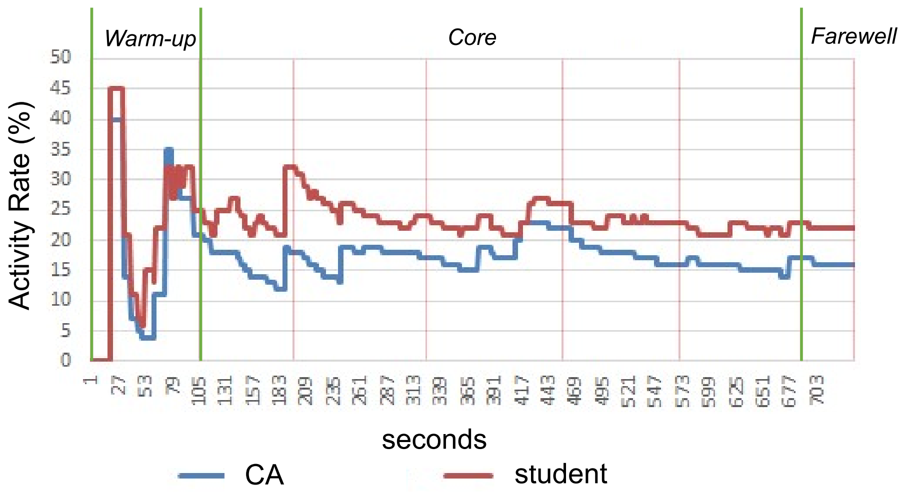
expedients were taken. To better allow the CA to orient in the ongoing lecture, the script was segmented into chunks, and for each of them multiple intents were defined. Moreover, the CA was conceived as sufficiently flexible to identify interaction patterns associated with various verbal constructs. The DialogFlow output is processed and interpreted so as to have the CA behave as intended (e.g., generating a hand emoji to start a *student clarification* interaction). Finally, the CA was given the ability to communicate using the voice by exploiting the TTS module of the WebSpeechAPI. Considering the fact that the synthetic CA's voice could have been detrimental to the suspension of disbelief (making the CA less convincing than a real human-controlled avatar), this factor was specifically analyzed with structured interviews, as it will be described after.

3.2.2.3 System Quality

The system was qualitatively assessed with preliminary testing to evaluate whether the CA could adapt its behavior based on the interactivity level of the real student. Figure 3.8 depicts the activity rate evolution of both the real user and the CA under



(a) Bashful Student



(b) Participative Student

Fig. 3.8 Example of two runs of the system during preliminary qualitative evaluation. Plots of the activity rate metric for the CA and the student. In one of the runs, the student was behaving in a participative way, in the other he/she was acting bashfully.

two separate sample runs; each run emulates a distinct user's behavior: that of a user behaving in participative way, and that of a bashful user. As it can be observed, the CA's activity rate differs in the two cases, and ultimately plateaus at a lower value in the case of a participative student.

3.2.3 Experiment

This section presents the exploratory user study conducted to analyze the impact of the undercover CA in the training experience.

3.2.3.1 Experiment Design

The sample of the study was made of 12 participants aged 16 to 52 ($\mu = 23.6$, $\sigma = 5.0$), invited to volunteer through the available network of contacts. None of them was suffering from color blindness.

The study was arranged following a between-subjects design by randomly assigning the participants to two equal-sized groups ($n = 6$). One group underwent the lecture with the CA as an active agent (CAG); the CA was instead deactivated for the second group, by replacing it with a dummy student (NCAG). Before beginning the lecture, all the participants were briefed on the fundamental features of the system (i.e., how to raise the hand for questions, how to use the chat and speak). With the aim to increase social pressure, the participants were also informed that all the other avatars in the classroom were controlled by genuine users attending the lecture. As previously stated, the participants did not have any prior relation with the teacher (a professor member of the author's research team). The teacher gave the lecture by seeking as strictly as possible the provided script, though dynamically adapting it to the audience's behavior (e.g., answering questions).

A multi-part questionnaire was devised to collect subjective feedback. Before the lecture experience, the first part of the questionnaire (BEQ) made of items about prior knowledge and familiarity with technologies associated with the study, demographics, and mindset of attending synchronous lectures wad administered. In addition, this part included a knowledge test consisting of eight multiple-choice questions about the lecture content (Pre-QUIZ).

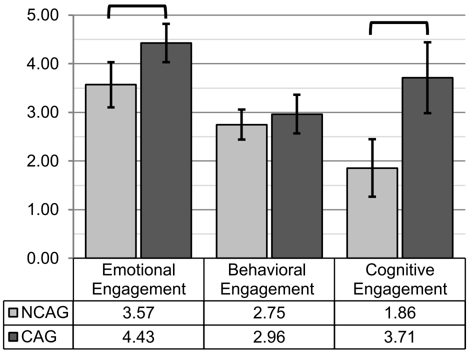
The second part of the questionnaire was administered after the lecture (AEQ), and was aimed at evaluating several components of the experience. Specifically, the participant's engagement along the emotional, behavioral, and cognitive dimensions was measured by adapting [121] (BEC-ENG). To investigate the potential interference effects caused by the CA on the social presence perceived by the user relative to the teacher, a Networked Minds Measure Scale Questionnaire (NMMQ) was adopted [122]. The Godspeed questionnaire [103] was instead used to examine the perceived quality of the CA. Furthermore, in order to evaluate factors such as perceived value of the CA, relief in social pressure, CA influence on the learning experience, and overall satisfaction, an Ad-Hoc part was arranged. Finally, open feedback about the experience was additionally collected. Completing that part is a post-test (Post-QUIZ), made of the same questions of the Pre-QUIZ which was employed to assess the learning gains.

3.2.3.2 Results and Discussion

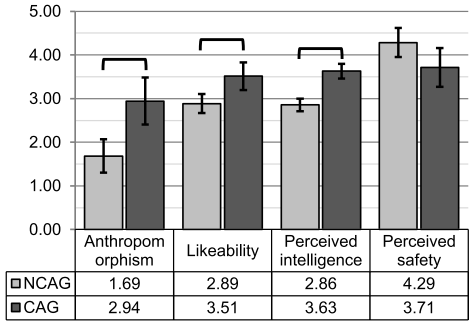
The two-tailed Mann-Whitney U-tests (p-value ≤ 0.05) were applied to the experimental data to look for significant differences between the two groups. The BEQ reported that 17% of the sample was unfamiliar with videoconferencing tools and interactive applications; 50% were used to immersive VR technology, whereas the other half stated had never or just seldom used an HWD; finally, 40% stated to frequently use a CA (in the form of a personal assistant). No statistical differences were found between the two groups for any of the BEQ elements, hence the random designation of participants did not seem to have introduced any bias in this regard.

Interesting findings can be observed looking at the AEQ results. About the BEC-ENG (Figure 3.9a), the CAG was deemed as significantly superior regarding the cognitive and emotional engagement, whereas for the behavioral component, no differences were spotted. Hence, the CA presence led to a much more pleasurable class experience, as well as an increased propensity to stop the lecture to ask for explanations on difficult topics based on the Figure 3.10 (item #15).

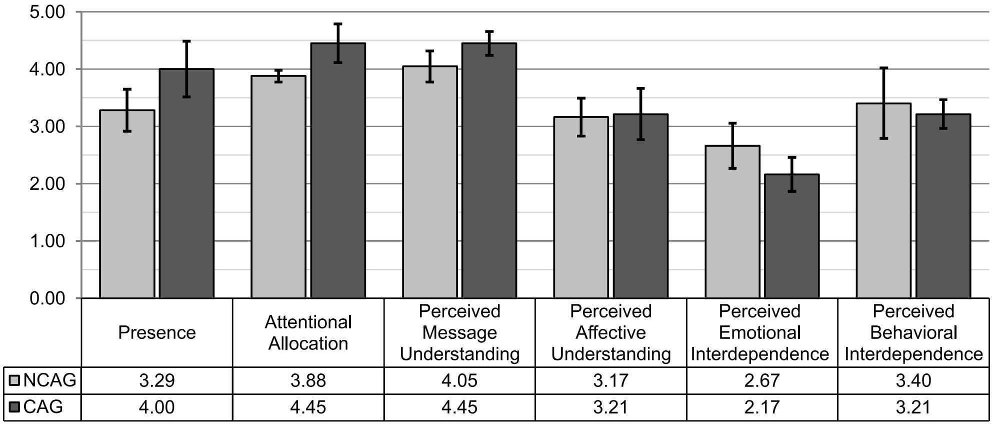
None of the aggregated NMMQ dimensions (attention allocation, presence, perceived behavioral and emotional interdependence, and perceived affective and message understanding) showed significant differences. This outcome suggests that the reciprocal social presence between student and teacher was neither improved not hampered by the addition of the CA. For the sake of completeness, two significant differences were spotted for the single elements, highlighting that the CAG found the teacher less negatively influenced by the students' attitude ($\bar{x}_{NCAG} = 3.00$, $\bar{x}_{CAG} = 1.71$, $CI_{NCAG} = 0.56$, $CI_{CAG} = 0.73$, p-value $= .011$) and at the same time it was perceived as easier for the teacher to understand them ($\bar{x}_{NCAG} = 3.57$, $\bar{x}_{CAG} = 4.57$, $CI_{NCAG} = 0.52$, $CI_{CAG} = 0.76$, p-value $= .026$).



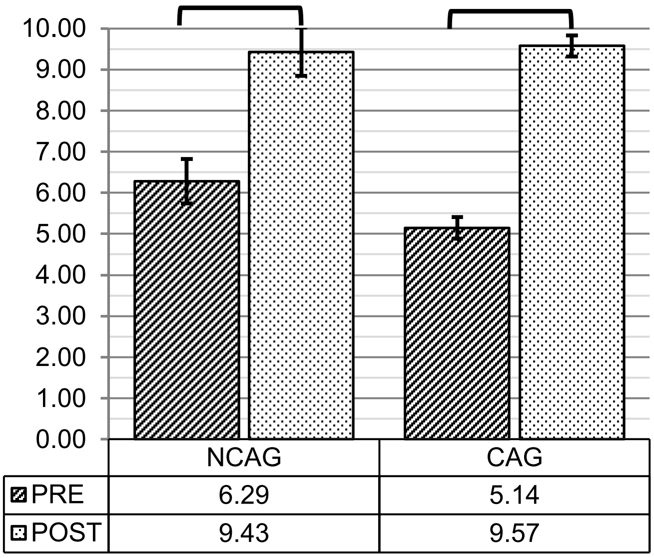
(a) BEC-ENG



(b) Godspeed on CA

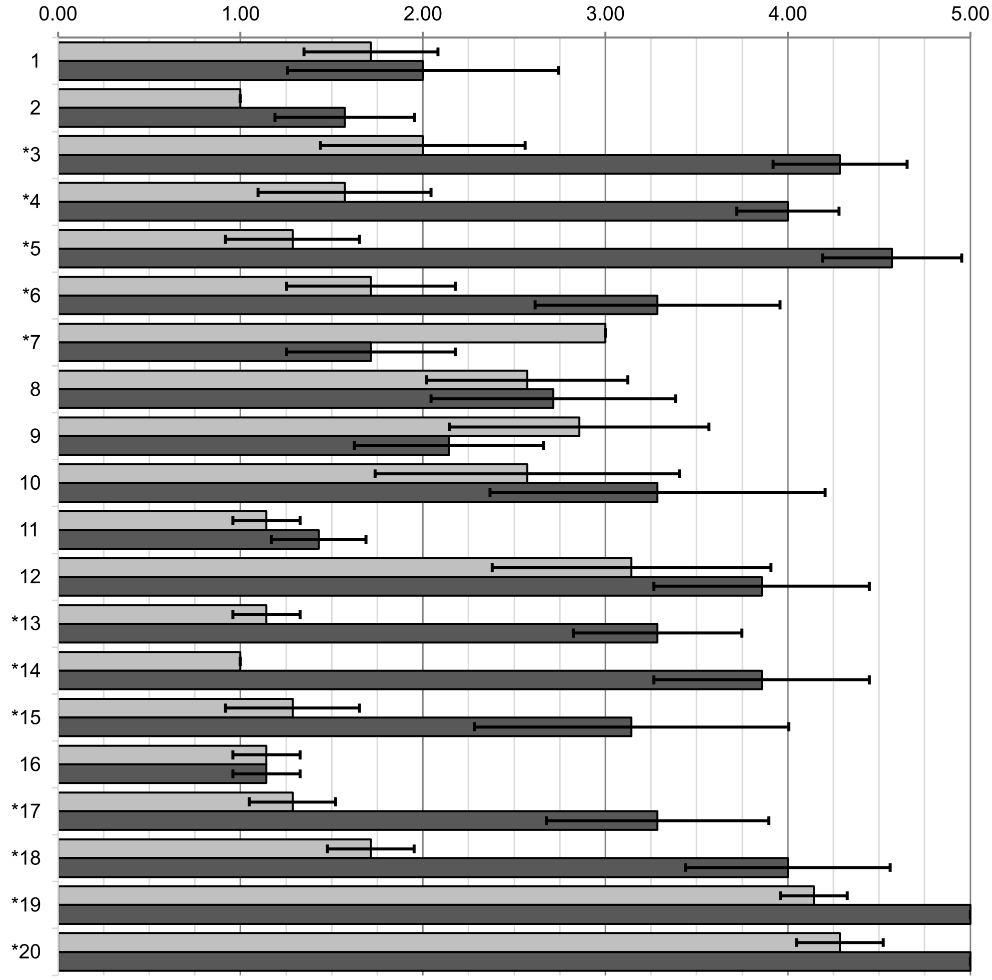


(c) NMMQ Student-Teacher



(d) Learning Gains. Scheirer-Ray-Hare Test:
NCAG–CAG (p-value $=$ .673); PRE-POST (p-value $<$ .001);
Interaction effect (p-value $=$ .452).

Fig. 3.9 AEQ Results (Part 1). Significant differences are marked with baffles.



	*20	*19	*18	*17	16	*15	*14	*13	12	11	10	9	8	*7	*6	*5	*4	*3	2	1
NCAG	4.29	4.14	1.71	1.29	1.14	1.29	1.00	1.14	3.14	1.14	2.57	2.86	2.57	3.00	1.71	1.29	1.57	2.00	1.00	1.71
CAG	5.00	5.00	4.00	3.29	1.14	3.14	3.86	3.29	3.86	1.43	3.29	2.14	2.71	1.71	3.29	4.57	4.00	4.29	1.57	2.00

Fig. 3.10 AEQ Results (Part 2) – Ad-Hoc. Significant differences are marked with *. #01. I think the way Aria was acting distracted me from the main goal of the experience; #02. I felt like Aria was intentionally reacting to my actions; #03. I felt like Aria was intentionally reacting to the teacher's actions; #04. I clearly understood the suggestion provided by Aria when needed; #05. I clearly understood what Aria was saying; #06. I think Aria was controlled by a Human; #07. I think other classmates were involved; #08. I think other classmates were controlled by a human; #09. I didn't notice other classmates except Aria; #10. I preferred to use the chat to interact with the teacher instead of the voice; #11. I would have preferred being the only student in the classroom; #12. I feel more confident when I see there is a classmate in the class; #13. If Aria was not in the class, I could not learn as much as I did with her present; #14. I felt legitimated to interact with the teacher since Aria did it first ("icebreaker"); #15. If Aria would have not been there, I think I wouldn't have interacted with the teacher in the same manner; #16. I think Aria was cahooting with the teacher; #17. I understood better the teacher questions thanks to Aria; #18. Aria was an example to me (e.g., how to interact with the teacher); #19. I think I could successfully pass a test on the topic of "solar system" without further training; #20. Overall I am satisfied with the learning experience. [77]

Regarding the perception of the CA, with the exception of the perceived safety, the CAG was found significantly better for all the aggregated elements of the Godspeed questionnaire (Figure 3.9b). This finding indicates that the CA was judged as quite likable and alive w.r.t. a dummy student (from the participant's viewpoint, a shy real student peer), and not more dangerous or frightening compared to it. It is worth nothing that, as it can be seen in Figure 3.10 (items #03 and #16), the CA was also perceived as an intelligent friendly peer, not cahooting with the teacher.

Other worth discussing outcomes from the Ad-Hoc part of the questionnaire are that the CA was not judged as distracting Figure 3.10 (item #01 and #06). Remarkably, according to the the CAG, the CA aided to reduce social pressure in the early phases of the lecture (Figure 3.10, items #14 and #17), enhanced the learning performance (Figure 3.10, items #13 and #19), in general positively affected the student-teacher interaction (Figure 3.10, items #15 and #18), ultimately leading to a more satisfactory learning experience, overall (Figure 3.10, item #20).

Conversely, looking at the objective learning performance data (Figure 3.9d) and running a pre/post analysis with the Scheirer-Ray-Hare test, no significant differences were found between the two groups (p-value $= .673$), despite the learning gains being indisputable for both the NCAG and the CAG (p-value $< .001$).

To conclude, the interaction analytics (number of interactions of each participant) were significantly lower in the NCAG w.r.t. the CAG ($\bar{x}_{NCAG} = 2.14$, $\bar{x}_{CAG} = 5.71$, $CI_{NCAG} = 2.66$, $CI_{CAG} = 2.00$, p-value $= .038$), hence substantiating the formerly discussed subjective results and supporting the speculation about the fact that the undercover CA can actually promote interaction.

3.3 Considerations and Remarks

The work presented in this chapter aimed at exploring the use of CA in XRTSs by particularly focusing on having it acting the role of a student peer and, by doing so, unlocking potential use cases and exploitation opportunities.

The first study explored the use of a RTA in a MRRTS to arrange a LBT pedagogical experience. Specifically, the LBT effectiveness against the consolidated TL approach in a MRRTS was investigated. The study targeted a population of electronic engineering students, and the selected learning subject was the *Thévenin Theorem*.

The obtained results indicated that both approaches were capable of sufficiently supporting the knowledge transfer to the student. Despite the limited sample size of the exploratory investigation, it was noticed that the students who received LBT training were able to retain the gained knowledge better than those who experienced the TL, with the compromise of a more protracted time to complete the process. This finding is in line with previous works in the classical pedagogical domain [101, 102] as well as with those employing a RTA (though not leveraging MR) [91], and designates the LBT as a profitable approach also for MRRTS settings that deserve the community's attention. Moreover, this result backs the speculation about the fact that the robot replaceability issue which probably affected [94] could be effectively addressed by following emerging design guidelines for MRRTSs [98] also by exploiting the LBT model. In light of this outcome, future research should focus on validating the preliminary findings with a larger sample size, by also widening the scope to multiple target populations (K-12, High School, etc.); moreover, the focus should also be shifted to consider autonomous RTA control with emphatic and believable behavior by developing appropriate tools and AIs, and cultivating the natural HRI which appears to be an essential factor that could lead to the success of this particular kind of MRRTSs in terms of efficiency and effectiveness.

The second study considered the scenario of a synchronous VR-based multi-user RL environment, and proposed the use of an undercover CA mimicking the behavior of a participative student with the aim of relieving social pressure and stimulating teacher-student interaction. An exploratory user study was run to validate the devised CA and the concept behind it. Despite the limited sample size, findings suggest that the undercover CA was indeed an interaction catalyzer, able to lessen the social pressure (similarly to what has been found in [114]), and eventually lead to a more engaging and satisfactory learning experience overall. It was found that the group of participants who experienced the CA perceived greater learning gains w.r.t. the TL group, albeit this was not corroborated by the quantitative measure of learning performance, which was indeed high in both the groups but anyhow not significantly different. In the future, it is recommendable to further substantiate the findings of this exploratory study by improving its statistical power. This would also permit to analyze possible correlations between the students' personality traits and attitudes against the CA's activity rate impact on their conditional behavior. Furthermore, it could be of interest to deploy more CAs simultaneously in the same classroom, by devising an orchestration system for them, so to investigate if and how the CAs-real students proportion influences the learning experience and social interaction in RL. Lastly, considering the teacher's perspective, it might be advisable to increase the system flexibility by allowing the CA to automatically synthesize the answers starting from the lecture material, and let the lecturer feed on-demand information at lecture time so as to better adapt to the ongoing classroom circumstances.

Chapter 4

Training Provisioners’ Perspective

The work described in this chapter has been formerly published in [123]

The majority of efforts carried out so far in the research field of XRTSs [9], and correspondingly the work presented in the previous chapters of this dissertation, focused on training as experienced from a learner’s perspective. In some ways, since in the classic simulation-based VRTS the role of the trainer still represents a significant part in the equation, it may be conjectured that this element was already studied in the past and it is therefore no more current or of interest for the community. In fact, on the contrary, it is just since a few years that efforts have been devoted also to the training provisioners’ perspective. For instance, with the ever-increasing diffusion of XRTSs it has been acknowledged the need to provide trainers and domain experts with tools to create XRTSs experiences, particularly through authoring tools that do not require programming expertise to achieve the goal [124]. Indeed, the study reported in Section 3.2 to some extent took into account the teacher’s perspective by providing a tool that helps him or her to carry out the lecture as s/he considers more appropriate pedagogically-wise. Nonetheless, this is still far from an XR-based tool oriented to a training provisioner as the primary target user.

This chapter presents and discusses an approach that leverages VR from a training provisioners’ perspective, with the specific aim to support them during the training design phase.

4.1 Leveraging VRTSs as a Mock-up Tool

Arranging a lecture or a training experience is never a trivial activity, but there are indeed contexts more challenging than others for the TP. FRs training, for instance, has been acknowledged as one of the most stressful context [125, 126] for many reasons. First, FRs have to be prepared to face low-frequency, high-severity events, and often operate with time-pressure constraints and life-threatening responsibility. So, in order to devise a meaningful training experience for such a category of trainees, particular care must be taken to include relevant experiences in the design phase [127]. Second, the training procedures are subject to continuous changes and evolutions as a consequence of integrating takeaways from the aftermath of on-the-field experiences or simply due to the introduction of new or updated equipment [128]. Hence, given the many aspects involved, such as the intricacy of the operational procedures or the necessity of having multiple operators to coordinate and perform as a team, drafting, updating or even re-designing a training experience can be troublesome for TPs.

With that in mind, it is imaginable to use a bare simulation-based VRTS depicting a possible training scenario that includes all the equipment and contexts and let some users experience it as a sandbox; in other words, having the TP using it as a mock-up tool supporting the ongoing design process by iterating and validating on aspects such as time and spatial constraints, task order execution, and so forth.

The consolidated result of such a pipeline can be either foreseeable as a pure VRTS to be used in a common way to train operators, or as a step to deploy such training experiences in real-life, on-site exercise.

In this section, the aforementioned idea is explored and evaluated by employing a VRMT to aid TPs in a training experience design and validation process. The activity was carried out in collaboration with Italian Civil Protection bodies (Piedmont Region Civil Protection [129]) and the Piedmont Region Coordination body of Civil Protection Volunteering [130]. In practice, a case study in which Civil Protection operators are trained on the operation of a High Capacity Pumping (HCP) module was put in place for evaluating a VRMT-based approach against a design approach that is commonly adopted for the definition of experiential courses, based on dramaturgical techniques [131]. Experts TPs from the involved FR bodies were asked to participate in the study and providing their feedback.

4.1.1 Background

The exploitation of VR-based tools in the design and review pipelines is definitely not a novelty for application domains such as BIM modeling [132] and CAD-based product development. As a matter of example, in [133] it has been proved that VR might be used profitably to double-check the threats to which construction staffers could be exposed and to optimize building design in order to avert these hazards. Nevertheless, using VR as a support tool for designing training experiences is still an under-explored research area.

Efforts in this direction have been put in place by the authors of [134], who presented an empirical method to synthesize training requirements leveraging a collection of VR tools with the final aim to arrange a VRTS. Specifically, two training scenarios were designed and implemented by having multiple application developers and stakeholders (i.e., physicians and navy officials) to jointly collaborate in the process. It was observed that, with respect to classic requirement elicitation techniques, that VR facilitated the developers in spotting the contextual simulation issues and improved the stakeholders communication towards the developer by clarifying the objects placement and scale in the VE. A more formal approach was presented in [135], were a conceptual framework was proposed for the production of simulation-based VRTSs. The fundamental concept of this framework is the introduction of core rules for arranging the simulator architecture by relying on two layers: an object-closed design layer to implement the training scripts, and a user-centered design layer, intended to model the subjects' orientations, guaranteeing that the operations will be executed in the deliberated manner as well as the relevance of the 3D assets used to populate the VE.

4.1.2 Materials and Methods

4.1.2.1 Case Study

The case considered in this work study was selected in collaboration with the hydrological risk divisions of the involved FRs bodies, by applying the VRMT-based approach to the arrangement of a simulation-based VRTS of one of their operational modules (the Civil Protection HCP module).

The HCP module is aimed at securing deluged areas so as to hamper possible additional causalities and damages by draining water from flooded buildings or waterholes. Specifically, when the module is operating in river proximity, the major objective is to prevent the river bank from shattering by draining critical puddles.

For this procedure a standard set of guidelines to train the operators were already available [136]; however, these guidelines were strictly entangled with real-world training and, more importantly, its constraints. For example, reproducing a deliberate catastrophic event (weather conditions, location, etc.) is often impractical or unfeasible in a real-world exercise context, whereas it is reasonably practical in a VE. Hence, applying straightforwardly this set of guidelines would be quite limiting for designing a training experience to be delivered in VR.

A hands-on practice that took place in the city of Ivrea, Italy during an exercise day was considered as a reference for the training experience. Specifically, during that practice, 20 operators from the *Coordination body of Civil Protection Volunteering* and *Piedmont Region Civil Protection Unit* rehearsed an intervention on a riverbank by operating actual equipment.

A group of four researchers harvested information and documented with videos the exercise over the span of five hours, with the aim to identify and outline the fundamental structure of the HCP procedure. In particular, several operators were scrutinized during the action, and later interviewed about the intervention and conceivable variants that could have occurred (and the relatively different actions to be performed). As a result, four main stages that outline the training structure were delineated. Since a given operator is not enforced to specialize in a specific role, the four stages were not summarized based on the roles, but on functional areas and the possible operation dependencies. For each stage, a set of guidelines to pursue and the equipment utilized were also pinpointed. By working as a team, the operators underwent the stages identified as follows.

1. *Setting-up the operational field*: The zone of the flooding must be inspected by the team in order to define which areas are in need of draining (aspiration), where to position the pump and where to flow out such water (delivery). Factors contributing to the decision of where to place the pump (Figure 4.1.iv) are: avoiding raised terrain so as not to overload the pump, and locating it close to the target aspiration zone. After placing the pump, poles and signaling tape shall be used to delimit the operational field. In the case of having decided

(i) (ii) (iii) (iv) (v) (vi)

Fig. 4.1. Equipment used by the Piedmont Region Civil Protection HCP module.

another river as the delivery zone, in order to prevent possible wear of the respective riverbank a protective sheet (Figure 4.1.iii) must be positioned, thus avoiding additional damages to the bank. Regardless of the stage, Personal Protective Equipment (PPE), e.g., helmet, gloves, life vest, and safety boots, must be worn by every operator.

2. *Assembly of aspiration and delivery chains*: The aspiration and delivery chain are assembled by the team. Regardless of the chain, to act safely and transport the tubes the operators shall work in couples. Semi-rigid tubes are used to form the aspiration chain from the pump to the draining zone, and a filter (Figure 4.1.i) needs to be attached to the extremity that shall be dunked in the puddle; in order to facilitate its retrieval, the filter will be tether fastened. In addition, the operator in charge of sinking the filter shall be fastened with another rope which is controlled by its buddy. The delivery chain, in turn, can be made of either semi-rigid or stiff-rigid tubes as well as foldable hoses (Figure 4.1.ii). The far end of the chain, which starts from the pump and reaches the delivery zone, shall be completed by a rigid element to be placed on the protective sheet and secured to the terrain with multiple pegs and strings. All the hoses and tubes are made with both a female and a male type on the two endpoints. During the assembly process, the various elements are coupled by matching a male with a female endpoint and fixing the connection with hooks and rings available on the endpoints through the help of lever tools (Figure 4.1.v).

3. *Pump start-up and supervising*: An operator is responsible of preparing the pump by closing both the impeller and the filter valves. When the chains are correctly set up, he or she can use a key switch on the control panel to turn the pump on. While the pump is running, all the doors shall remain closed to limit the noise. The aspiration and delivery chains (Figure 4.1.vi) are monitored by the other operators, who look for potential leakages. In case of such an unfavorable event, the operation must be immediately stopped by turning the pump off and fixing (or replacing) the flawed chain element. At the end, when there is no more water to be drained, the pump shall be turned off.

4. *Disassembly*: By keeping the pump inactive, both the impeller and filter valves are opened by an operator in order to drain the water still stagnating in the pump. Afterwards, the filter is retrieved from the aspiration zone by using the respective rope. The chains are dismantled starting from the delivery one, by prioritizing the foldable hoses.

4.1.2.2 Dramaturgy-Based Prototyping

As documented in [137], to satisfy the design objective of prototyping a training that encompasses procedural tasks it is possible to resort to the approach defined as dramaturgy-based. It is a holistic approach, whose benefits have been extensively studied [131, 138, 137], which enables the design of experiential courses by combining a spectrum of activities such as cultural, social, physical, emotional, creative, and reflective ones [131].

The application of DP is not restricted to a limited phase of the design process, but can be applied during pristine ideation, testing, and maintenance. Usually, the FRs bodies involved in this work do not rely on DP for their training design, since the exercise is usually arranged directly on-the-field, without a preliminary design iteration back in the office.

In order to fairly evaluate the VRMT approach against a traditional one, the process used by the involved FRs bodies for training design was completed with a DP testing step. Hence translating into a methodology capable to both meet the case study learning requirements and allowing to perform the training design session in an indoor location.

The key steps that the DP prescribes are as follows:

1. a large enough indoor location has to be prepared with the material potentially needed by a team of experienced trainers for the training prototyping (blackboards, stationery, cardboard props, etc.);

2. after explicitly defining the objective of the DP activity, the team of trainers engage in a first activity consisting of a brainstorming, hence, gathering everyone's ideas, recommendations, and concerns; notes, sketches, and reference material (e.g., videos, photographs, manuals, etc.) are the primary tools used during this activity;

3. a written and visual storyboard is peer-drafted by reorganizing and converting the discussed brainstorming outcome; consensus shall be granted among the team members;

4. this is the step in which dramaturgy is actually exploited; the trainers physically act to mimic the novel designed training, with the aim to assess and potentially validate the produced prototype prior to deploying it in the real world;
5. the team of trainers returns to the second step and resumes the design if the training prototype does not fulfill the initial requirements.

In the fourth step, improvised props are exploited while performing, with some of the actions reproduced when practicable, whereas most of them are mimicked. Props are needed since the majority of the HCP equipment cannot be readily placed or used in an indoor location. In addition, it should be recalled that the procedure itself is tailored to a deliberate disaster-oriented context (e.g., urban flooding, river overflow). Even though these circumstances can (and shall) be simulated for the actual training, during the design phases it is unfeasible or unpractical to emulate them; as a result, the trainers have to resort to their mental fiction. Video recording of past exercises or adverse events can be used as a reference to support the mimicking; for example, to review the procedure timings and validity. Actions way too impractical to mimic (such as ladder climbing) are simply pretended.

Despite the relative good affordability and effortless deployability of DP, this approach is severely dependent on the trainers' ability to predict real life effects of just imagined or surrogated actions. Undoubtedly, the training so produced needs further real-world validation, which can result in a quite cost- and time- consuming activity if it is born in mind the intricacy of settings and the equipment involved. This situation can be further worsened if eventually the training prototype appears to be flawed and there is the need to further iterate on the design by putting in place a new DP session.

4.1.2.3 VRMT Implementation

The VRMT was developed considering the HTC Vive Pro kit [39] as a target immersive VR HWD (device features are already reported in Section 2.1.2.4) and the Valve Index [139] as hand-held controllers. The latter were preferred to the bundled HTC wands controllers since able to provide higher fidelity interactions with, and finer control on, the virtual objects. The tracked hand-controllers and their built-in physical buttons are used to interact with the VE.

The implementation of the VRMT application leveraged the Unity (v2018.4) game engine [40] together with the SteamVR framework (v2.7.2). To populate the VE, 3D assets modeled with Blender [41] (specifically, to reproduce a high-fidelity virtual counterpart of the pump) were used. Particular care was taken also to replicate in VR the auditory cues of the real equipment (pump included).

In order to provide a faithful replica of the real HCP procedure, all its stages were included in the VRMT. In addition, the VRMT was featured with the option to let the trainees experience the whole training (Figure 4.2.) or just some chunks. The chunks consider the placement of the pump, the placement of the protective sheet on the riverbank, the delimitation of the operational field using poles and tape, the operation and control of the pump, and the operation of the safety and retrieval rope.

The trainees experienced the training by acting in just one role (i.e., in the aspiration chain team or in the delivery one), and the remaining roles were automatically managed by the simulation. Tubes and hoses were simulated carefully by implementing all the actions to be performed for building the chain, with particular care for the elements coupling.

Since the majority of the considered tasks requires the cooperation of operators who need to work in couples, the VRMT included Non-Player Characters (NPCs) to act as trainee's buddies; for instance, transporting tubes (Figure 4.2.v), alerting the companion in the case of a leakage, or aiding during the mounting tasks (Figure 4.2.vi, Figure 4.2.viii). As anticipated, the chain that must be assembled by the application (pertaining to the team the trainee is not assigned to) is actually carried out by NPCs. All the NPCs logic was implement with a standard event-driven strategy via finite-state machines. The scaffolding was implemented with visual cues (e.g., highlighting) and diegetic audio instructions in the form of walkie-talkie voice communications.

Two officers of the Piedmont Region Civil Protection were asked to validate the implementation. Overall, it was judged positively and even exceeding the expectations in terms of engagement and fidelity.

4.1.2.4 Experiment Design

The value of devised VRMT as a support tool for the design of training experiences was assessed from a TP's perspective by means of a user study described in the

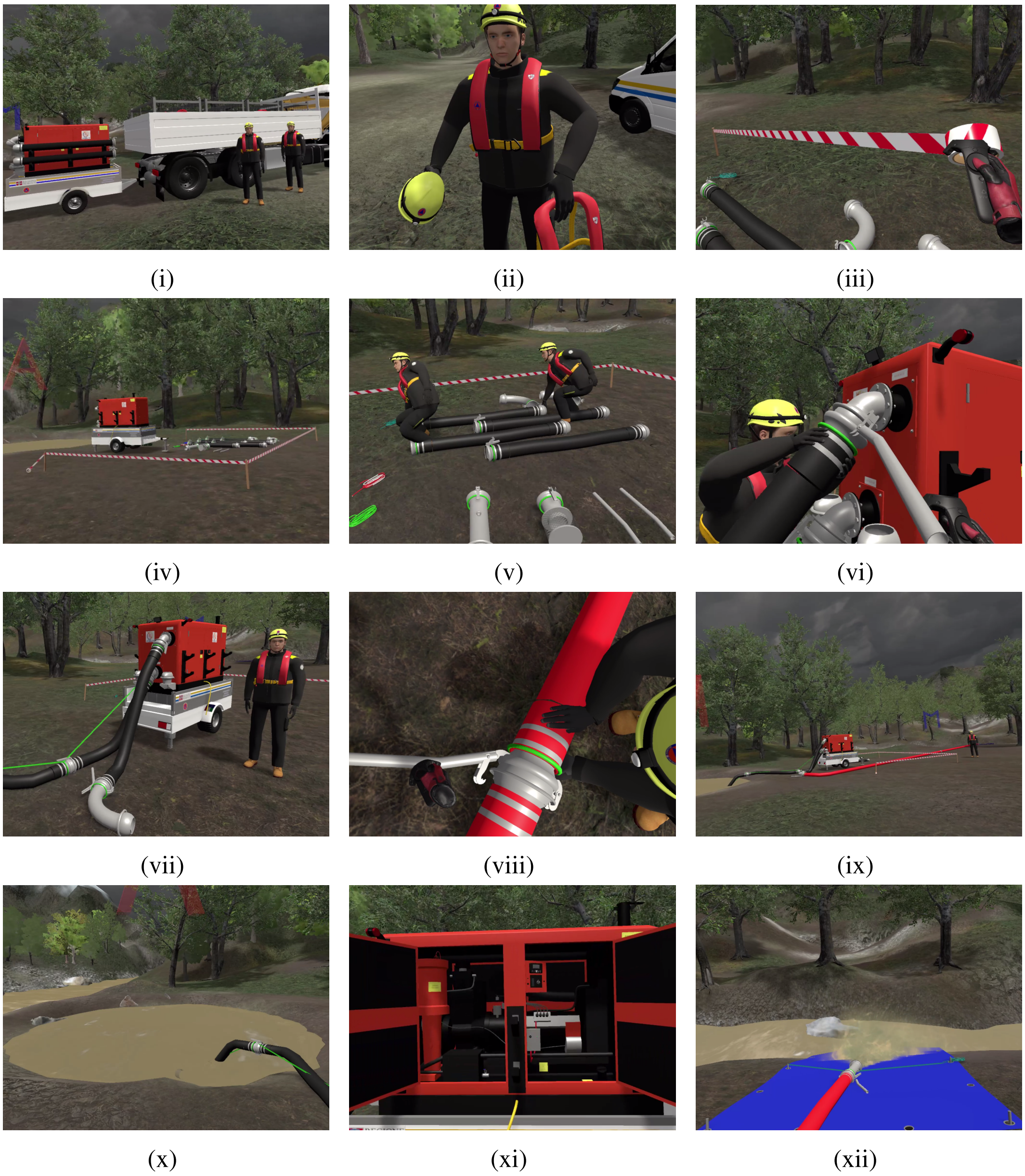


Fig. 4.2. Excerpt of few key moments of the arranged training procedure in the VRMT.

following. The roles of TPs were played by the two most experienced trainers of the Piedmont Region Coordination body of Civil Protection Volunteering, having both of them an established background in training preparation and real-world exercise on the HCP procedure.

Firstly, they were requested to produce a revised training procedure by undergoing a DP session as a team (observing the phases described in Subsubsection 4.1.2.2).

Secondly, the TPs were asked to analyze the behavior of three trainees experiencing the VRMT (one at a time); the trainees, belonging to the participating Civil Protection bodies, were selected with diverse degrees of proficiency about the HCP procedure. The TPs watched the trainee's performance on a projected screen (6m distant, 110" diagonal) streaming the HWD first-person perspective. In addition, the TPs were allowed to take notes and communicate both between them and with the trainee. In the following, the two TPs will be quoted to as TP 1 and TP 2.

A three-part questionnaire was devised to collect structured feedback from the TPs. The first part was made of items adapted from the IMMS [48] with the aim to assess the learning performance of the trainees from the TPs' perspective w.r.t. the planned learning goals. In the second part, broad opinions about the appropriateness of the VRMT as a mock-up tool were collected. The components analyzed were about the potential to effortlessly identify possible risk-exposure and equipment misuse, the capacity to recognize flaws in the arranged training, the envisioned real-world deployability, and so forth. In the third and last part, the TPs rated the preferred approach by scoring seven factors from 1 (DP) to 5 (VRMT), whereas the items of the first two parts were scored from "totally disagree" (1) to "totally agree" (5).

4.1.3 Results and Discussion

This section documents and discusses the results collected with the questionnaire. For each part of the questionnaire, the respective Cronbach's α is reported.

No participants suffered from cybersickness after experiencing the VRMT and all of them were able to complete the training.

4.1.3.1 Observed Trainees' Performance

The trainees were deemed to performing decently ($\bar{x} = 4.17$, $SD = 0.37$, $\alpha = .86$), indicating that, according to the TPs, the learning effectiveness of the VRMT training was convincing and, above all, that the TPs were able to detect and dissect factors pertaining to the proper execution of the taught procedure. More precisely, they were capable to observe that the trainees made few severe errors and eventually became aware of them (Table 4.1, items *#1* and *#2*). As foreseeable, the TPs were uncertain

Table 4.1 Results of the first section of the TP questionnaire on trainees' performance ($\alpha = 0.86$). Statements with negative phrasing are marked with $^\circ$.

#	Statement/Item	TP 1	TP 2
1	Trainees made severe errors$^\circ$	2	1
2	Trainees were unaware of the errors made$^\circ$	2	1
3	I think the trainees acquired the knowledge and skills required to autonomously operate in a real situation	4	3
4	Overall, I am satisfied with the arranged training experience	5	4
5	I am satisfied with the competences acquired by the trainees	4	4
6	I am satisfied with the way the trainees executed the HCP procedure	4	4

about the expected learning transfer, that is, whether the trainees would have been able to act independently and effectively in the field without additional real-world training (Table 4.1, item *#3*). In spite of this outcome, they rated both the arranged training and the trainees' performance as satisfactory (Table 4.1, items *#4–6*).

4.1.3.2 Suitability of VR as a Mock-Up Tool

Remarkably positive feedback was collected from the TPs about the VRMT suitability in terms of supporting the training design. This includes critical factors like the potential to pinpoint didactic flaws in the designed training (Table 4.2, item *#2*) and the capacity to derive take-home lessons to further develop and improve specific part of it (Table 4.2, items *#5* and *#6*). Broadly speaking, the VRMT was deemed as excellent for prototyping real-world training (Table 4.2, item *#7*), and the "as-is" deployability (from VR to a real-world setting) was judged as remarkably good (Table 4.2, item *#3*). An aspect that should be subject to improvement in the future pertains to the ease of observing the trainee's actions, since the selected first-person view was found suboptimal in this perspective (Table 4.2, item *#1*).

4.1.3.3 VRMT vs. DP

From the direct comparison between the two approaches less unequivocal findings were obtained. Overall, the VRMT was preferred to the DP, though by a little margin ($\bar{x} = 3.15$, $SD = 0.91$, $\alpha = .84$). Multifaceted preferences are visible in Figure 4.3. The VRMT was definitely preferred when it comes to validating in advance the training structure and organization (items *#1* and *#2*), and was deemed as somewhat

Table 4.2 Results of the second section of the TP questionnaire on the suitability of VR as a mock-up tool ($\alpha = 0.71$).

#	Statement/Item	TP 1	TP 2
1	From the provided point of view, it was easy to get all the operations performed by the trainees (execution order, equipment interaction, coordination, etc.)	5	3
2	Observing the trainees operate in VR allowed me to analyze potential issues of the arranged training	4	4
3	I think it is foreseeable to deploy the arranged training in a real-world exercise	5	5
4	I think observing the trainees executing the arranged training in VR made me a better trainer	4	4
5	Observing trainees operate in VR enabled me to meditate on how to arrange and explain specific parts of the training	5	4
6	I could identify which parts of the arranged training should be reworked and/or revised	4	4
7	I think that VR is a great tool to prototype a FR training exercise	5	4

valuable for identifying potential setbacks and flaws (item *#3*). In turn, the DP was judged as able to enable faster iteration on training design, as it could have been expected (item *#6*). Results for items *#4*, *#5* and *#7* were instead a bit surprising. The two ties and the small preference for the DP (as being better for detecting potential equipment mishandling) can be possibly explained by considering the fact that the DP involves more active participation of the TPs w.r.t. to watching a trainee's performance on the screen. This interpretation also fits with Table 4.2–*#1*. Hence, this factor should be ameliorated in the future, e.g., by letting the TPs join the VE together with the trainee so as to enable the former to freely view the latter and interact with the simulated environment.

4.1.3.4 TPs' Remarks

Interesting considerations emerged from the TPs' open feedback.

About the level the expertise, TP 1 self-declared as extremely proficient in the considered HCP procedure, whereas TP 2 self-declared as more skilled on other procedures. This fact also justifies the relatively lower number of errors detected by TP 2 w.r.t. TP 1 when supervising the trainees in the VRMT.

According to TP 1, the VRMT and DP could be considered as complementary. The TP elaborated on this remark by noting that, despite the VRMT being superior on many aspects, it suffers from a shortcoming w.r.t. the DP; precisely, in the TP's opinion, identifying mistakes and incorrect behaviors while watching the trainees in

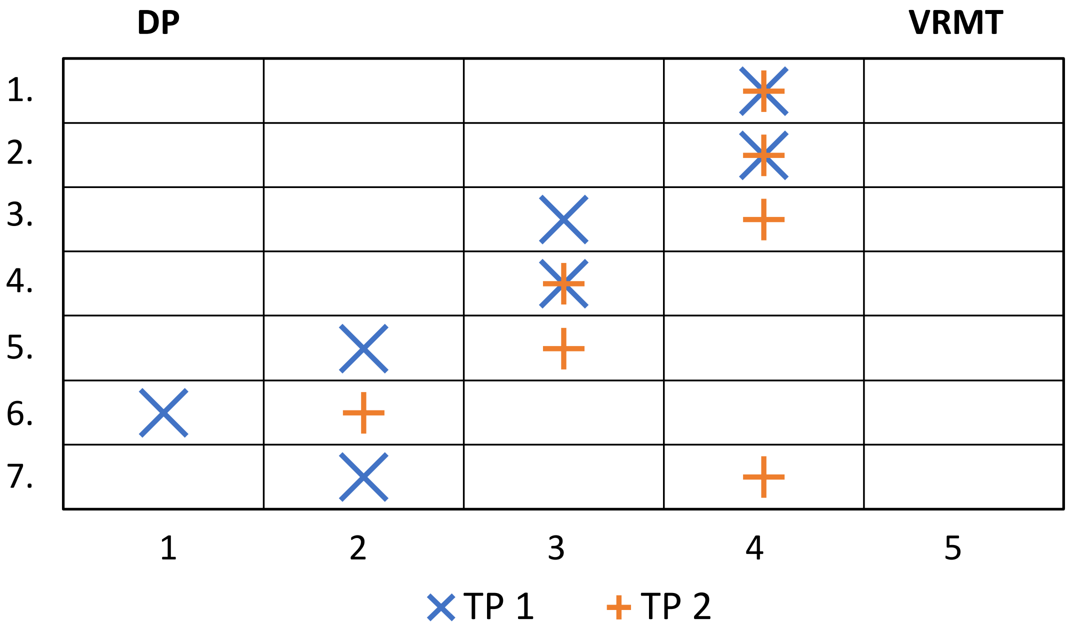


Fig. 4.3 Results of the third section of the TP questionnaire on the direct comparison between VRMT and DP ($\alpha = 0.84$): #1 Which tool allows me to better focus on how the trainees will experience the arranged training; #2 Which tool allows to iterate faster on the prototype; #3 Which tool allows me to better foresee possible issues of equipment managment training; #4 Which tool allow me to better prevent/limit risk for trainees during the training; #5 Which tool allows me to better identify possible training flaws; #6 Which mock-up tool I think is preferable for prototyping a training that is ready to be deployed in the real world; #7 Which mock-up tool allows me to better arrange/organize the training contents.

the VRMT is rather tricky. The TP suggested as a possible countermeasure to adopt different points of view rather then the only one used (first-person). These remarks are consistent with the results of the direct comparison part of the questionnaire, in which the TP 1 favored DP more than TP 2.

Also the TP 2 spotted the same drawback regarding the VRMT; however, the TP 2 also heralded the excellent VRMT usability and opinionated that it could be of great support to the training organization. Specifically, the TP 2 underlined the problems connected to arrange on-field validation, which could be even exacerbated by anomalous external factors, such as the experienced pandemic conditions. As a consequence, the VRMT appears as an even more compelling alternative to DP, since it could reduce the necessity for in-person, physical interactions by just utilizing video-conference tools for streaming or by expanding it with networked multi-user capabilities.

4.1.3.5 Trainees Survey

For the sake of completeness, the feedback from the three trainees (in the following as, TR 1, TR 2 and TR 3) collected via a separate questionnaire to evaluate the user experience with the VRMT is also reported. The questionnaire followed the structure of [140], including the Hedonic Quality Stimulation (HQ-S) and the Attractiveness (ATT) components of the *AttrakDiff* questionnaire [141] . According to the results illustrated in Figure 4.4, the user experience was rated from neutral to positive.

In addition, factors such as the trainees' satisfaction with the devised training, the communication with the TPs, and the NPCs coordination were instead analyzed with an ad-hoc set of items.

Having the TPs observing them was not deemed as detrimental to their experience (Table 4.3–*#1*), which was overall judged as satisfactory (Table 4.3–*#3*). On the contrary, the trainees were not fully convinced about the simulated equipment interaction (Table 4.3–*#2*); according to TR 2 and TR 3, this was mainly due to the lack of force feedback (tube weight) which is reflected in odd and slightly incoherent handling.

The NPCs' behavior was overall judged as satisfactory. Nonetheless, TR 1 lamented the fact that there have been situations in which the NPCs were waiting for the advancement of the procedure instead of acting more productively. It was

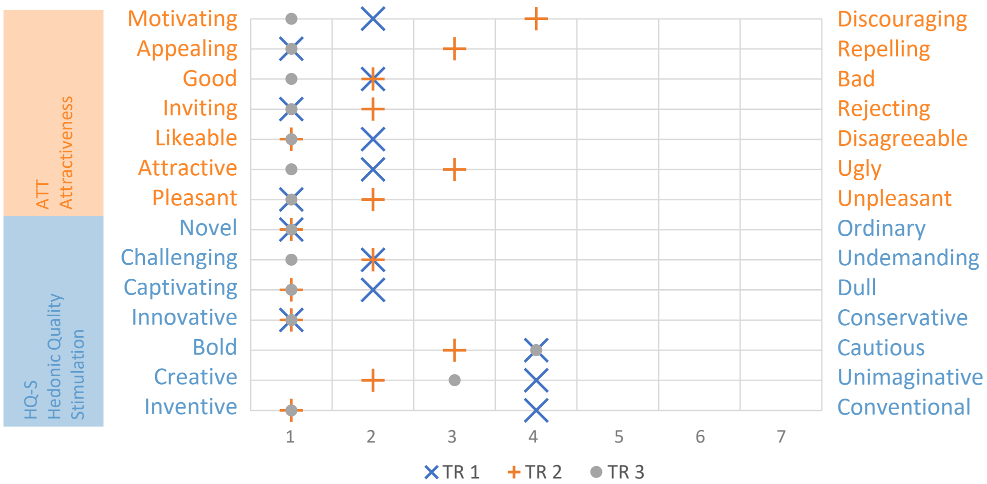


Fig. 4.4 Results of the first section of the trainee questionnaire regarding Attractiveness (ATT) and HQ-S [141].

Table 4.3 Results of the custom section of the trainee questionnaire ($\alpha = 0.84$). Statements with negative phrasing are marked with $^\circ$.

#	Statement/Item	TR 1	TR 2	TR 3
1	Being observed by a trainer negatively influenced my performance$^\circ$	2	1	1
2	I think I handled the (simulated) equipment appropriately	4	3	3
3	Overall, I am satisfied with this experience	4	4	5
4	I found that the Non-Player Characters (NPCs) simulation was well designed	3	4	5
5	I would have preferred that the other operators were human-controlled (instead of NPCs)	5	5	3

also suggested to have the remaining agents controllable by other users to evolve the experience into a multi-user one; this would reflect also in an enhanced, multi-user simulation fidelity which is worth considering for future studies.

4.2 Considerations and Remarks

This chapter focused on the TPs' perspective by proposing the use of VRTSs as mock-up tools for backing the design of experiential courses, regardless of their future deployment (on-field or as VRTSs). The proposed design approach was evaluated against a commonplace training design approach (DP). A real training procedure of the partnering FRs bodies was used as a reference for implementing the VRMT, and two experienced TPs were asked to output a revamped and extended version of it by following the two different approaches and, then, to provide their feedback.

Based on the obtained results, from a TP's perspective the VRMT was judged as a valuable approach to back the training design. This appraisal includes relevant aspects, such as the ability to detect weaknesses and gather insights about improving specific training parts, the ability to examine the trainees' behavior, as well as the predicted deployability in a real-world exercise. With respect to the DP approach, the VRMT was deemed as capable of simplifying the validation of the training structure and organization. The main limitation that was found for the VRMT was its more laborious design iteration process w.r.t. DP. Therefore, in the future efforts should be devoted to ameliorating this element by exploring, e.g., *no-code programming* [142] approaches to enable unskilled TPs to modify the mock-up by themselves. Lastly, the TPs judged as more demanding to spot possible equipment mishandling in the VRMT than in DP, albeit surprising this is probably ascribable to the fact that they were able to observe the trainees operating in the real world or with the

VR first-person perspective. Future research should consider letting the TPs enter a shared VRMT and allow them willingly interact with the simulation and observe the trainee within the VE.

Chapter 5

Conclusions

In this chapter, a summary of the performed studies and of achieved outcomes is first reported. Afterwards, the main limitations and possible future works are discussed.

5.1 Summary of Research Outcomes

The previous chapters summarized the research that was carried out by the author within the Ph.D. period with the objective to contribute to the development of XR-based tools used in education and training contexts.

The activities focused on three different aspects of XRTSs. In Chapter 2, in order to facilitate the conceivable deployability at scale of these tools, it was investigated their self-learning variant, i.e., their usage without the need of a human pedagogue. Even though studies regarding the use of XRTSs for the self-tuition of a trainee are quite present in the literature, the body of research is scattered, and results obtained by previous works are oftentimes controversial or hardly generalizable to different scenarios. This is due to the fact that the efficacy of the medium appears to be task-dependent, and that most of the studies focused on benchmarking such tools in over-controlled experimental conditions, hence giving limited information, e.g., about how they would perform w.r.t. certified training procedures arranged by real companies, especially in the context of teaching MTs.

With that in mind, a user study aimed at assessing the training effectiveness of VR on an IMA task (in the domain of industrial robotics) was arranged, by challenging a state-of-the-art VRTS supporting self-learning against a consolidated training system based on in-class and hands-on sessions, and subsequently performing an on-the-field evaluation of the trainees' performance. The main finding was that the effectiveness of the VRTS in terms of letting the trainees successfully complete the task was overall comparable to that of the ST. More in-depth, the development of proprioception skills and spatial awareness was well backed by both the VRTS and the ST, being them on par regarding this aspect as well as regarding the capability to back the transfer of the required procedural knowledge. Also the skills pertaining to decision-making after having observed the system (an IR) status and to executing ERPs, if needed, were satisfactorily transferred; notably, the VRTS trainees achieved better proficiency in handling hybrid digital-physical devices. Even though the VRTS fell behind the ST in factors that are typically strengthened by trainer-trainee social interactions, the former was judged as a pleasant and indisputably time-efficient learning method. However, the more packed and dense experience of the VRTS led to a higher cognitive workload for the trainess w.r.t. the ST, and some aspects induced, by design, higher annoyance to them. Specifically, one of these aspects in need of improvement appeared to be related to the frustration levels induced by the scaffolding system when teaching a PDT.

To cope with this specific issue, a second study was devised with the purpose of boosting both the pleasantness and the efficacy of the training in the context of self-learning PDTs via a VRTS. In particular, three metaphors inspired by the literature were proposed and compared, with the aim to find which one(s) can better encourage a trainee to match the head-pose deliberated by the TP. A testbed scenario in the form of an automatic VRTS for training a fictional procedure that included six classes of PDTs was arranged for the evaluation. The obtained results indicate that the conventional way of instructing, based on a video that exemplifies the task execution, is outperformed by both the proposed metaphors and, in particular, by the one that was designed with superior affordance in mind.

By considering the other notable shortcoming found for the self-learning XRTS regarding the limited expression of social elements, in Chapter 3, the employment of PAs was investigated. In particular, the attention was focused on the potential use cases and exploitation opportunities enabled by having a PA acting in the unconventional role of a student peer.

The first of the two studies carried out along this research direction explored the use of a RTA in a self-learning MRRTS to arrange a LBT pedagogical approach. Specifically, the LBT effectiveness was evaluated against the TL approach (conventional scaffolding). Results indicated that both the approaches were capable of sufficiently supporting the knowledge transfer to the learners, but those who underwent the LBT training reported higher self-efficacy and confidence levels compared with those who experienced the TL; they also exhibited a superior knowledge retention when tested after a one-week period, at the cost of a longer time spent in the learning process. Thus, the LBT appeared to be a profitable model also for MRRTS settings that deserve the community's attention.

Moving to VR scenarios in which it is desired to maintain a human pedagogue involved, a second study was carried out by considering a synchronous, multi-user RL environment. Within this context, it is rather common to witness a teacher opening the floor for questions or even asking something to the classroom and receiving no answers in response, fostering a displeasing and awkward silence; these situations can nourish a vicious spiral, corroborated by the higher social pressure that may be experienced. In such scenarios, the learning experience can be possibly less enjoyable than an on-site one. With the aim to mitigate this issue, a PA was exploited in the form of an undercover CA that mimicks the behavior of a participative student so as to possibly relieve social pressure and stimulate teacher-students interaction. The findings of an exploratory user study indicated that the undercover CA was an interaction catalyzer, able to lessen the social pressure, and eventually leading to a more engaging and satisfactory learning experience, overall. It was found that the group of learners who experienced the CA perceived greater learning gains w.r.t. the TL ones, albeit this outcome was not supported by the quantitative measure of the learning performance, which was indeed high in both groups but anyhow not significantly different.

Finally, in the last part of this thesis, the focus was shifted from the trainee's to the TP's perspective. Specifically, Chapter 4 proposed and discussed an approach that leverages a VRTS as a mock-up tool for backing the design of experiential courses, regardless of their future deployment (in the real world or as VRTSs). The proposed design approach was evaluated against a commonplace training design approach (DP). A real training procedure of the partnering FRs bodies was used as a reference for implementing the VRMT, and two experienced TPs were asked to produce a revamped and extended version of it by following the two different approaches and, then, to provide their feedback. The outcome of a user study highlighted that, from a TP's perspective, the VRMT was judged as a valuable approach to back the training design. This appraisal includes relevant aspects, such as the ability to detect weaknesses and gather insights about improving specific training parts, the ability to examine the trainees' behavior, as well as the predicted deployability in a real-world exercise. With respect to the DP approach, the VRMT was judged as capable of simplifying the validation of the training structure and organization, albeit the design iteration process of the VRMT was found to be more laborious compared to that of DP.

In conclusion, summarizing the research presented in this document it appears that XRTSs, even in their automatic form, are now sufficiently mature to be used for effectively delivering training on procedures encompassing mixed-tasks, being on par with on-site learning provided by a human instructor at least for what it concerns their performance for pure procedural knowledge transfer. Yet, to meet this goal, it should be considered the caveat that, for some peculiar tasks, ad-hoc solutions may have to be adopted in addition to the vanilla GSs (like in the case of PDTs). There are also pedagogical models that could be integrated into XRTSs which are capable to boost long-term retention w.r.t. to a basic GS; specifically, the LBT approach was found to be particularly interesting from this viewpoint. This approach was enabled by exploiting a PA in the unconventional role of a peer student. The same role was considered also in the diverse scenario of immersive RL, where the introduction of the PA was able to successfully stimulate students-teacher interaction. Furthermore, considering the TPs' perspective, it was found that VR-based mock-up tools may represent a cogent option to support them during the design of training experiences.

5.2 Limitations and Future Work

The main limitation of the reported studies, especially for those following a between-subject design, is about statistical power, which might have not been reached. The limited sample size of the performed experiments was mainly due to the difficulties in rercuiting and running user studies involving an HWD as a consequence of the restrictions imposed by the COVID-19 pandemic situation, which was overlapped in time with the research activities.

Despite this limitation, and considering the explorative nature of the studies and solutions that have been proposed, it is the author's hope that the work presented in this document offered interesting insights and paved the way for future research in the considered domain. In particular, with regards to automatic XRTSs, it is advisable in the future to improve aspects pertaining their ability to adapt the delivered training experience, making such systems capable of adjusting to the trainees' learning pace (e.g., by tailoring the instruction flow to their needs), as well as of focusing on social-related aspects by encouraging and motivating them during the training (e.g., by giving congruent feedback on both achievements and mistakes). With respect to the integration of PAs into XRTSs, in the future efforts should be devoted to improve the AIs in order to have such agents acting with emphatic and natural behaviors, and allowing a pedagogue (if present) to feed the algorithm on-demand/in real-time to make it better adapt to the ongoing lecture circumstances. Finally, by considering the perspective of TPs, there is the need to devote further efforts in order to make them able to create and edit XRTSs experiences also if lacking computer programming skills, so to iterate faster and on their own on the training design.

Acronyms

AI Artificial Intelligence. 55, 62, 74, 97

AR Augmented Reality. 1, 2

CA Conversational Agent. viii, 62–71, 73–75, 95

DC didactic cell. 13, 14, 19, 23

DOF degree of freedom. 2, 14, 49, 50

DP dramaturgy-based prototyping. viii, 82, 83, 85–89, 91, 95, 96

EMD Electronic Mastering Device. 10, 11, 20

ERP error recovery procedure. 11, 12, 24, 44, 94

FOV field-of-view. 2, 14, 30, 35

FR first responder. 2, 77, 78, 82, 91, 95

GS guidance system. 3, 15, 16, 23, 28, 35, 96

HCI human-computer interaction. 40, 43

HCP High Capacity Pumping. 77–80, 83–86, 88

HQ-S Hedonic Quality Stimulation. 90

HRI Human-Robot Interaction. 48, 57, 74

HWD Head-Worn Device. 2, 14, 21, 30, 32, 38, 40, 64, 70, 83, 86, 96

IMA industrial assembly and maintenance. 5–7, 11, 93

IMMS Instructional Materials Motivation Survey. 18, 86

IPC inter-process communication. 52

IPQ i–group Presence Questionnaire. 40

IR industrial robot. 6–10, 13, 14, 21, 23, 27, 29, 38, 44, 94

ITS intelligent tutoring system. 4, 46, 48

LBT learning-by-teaching. 47–49, 51, 53–55, 57–60, 73, 74, 95, 96

MOOC massive open online course. 61, 63

MP mastering procedure. 9–14, 16–23, 38

MR Mixed Reality. 1–3, 5, 15, 16, 27, 28, 30, 35, 48–51, 53, 55, 56, 60, 74

MRRTS MR robotic training system. 49–53, 57, 59, 60, 73, 74, 95

MRTS MR training system. 5, 45

MT machine task. 5, 6, 29, 31, 44, 93

NLP Natural Language Processing. 62

NMMQ Networked Minds Measure Scale Questionnaire. 69

NPC Non-Player Character. 84, 90

PA pedagogical agent. 3, 46–48, 62, 63, 94–97

PDT parallax-dependent task. 29–37, 39, 42, 43, 45, 94, 96

POI point-of-interest. 29, 30, 34, 35, 38, 40, 42

POV point-of-view. 29, 40, 42

PPE Personal Protective Equipment. 80

RL remote learning. 61–64, 74, 75, 95, 96

RTA Robotic Teachable Agent. 48, 49, 51, 54–56, 73, 74, 95

SP KUKA SmartPad. 10, 13–17, 20, 23, 24, 27

SSQ Simulator Sickness Questionnaire. 40, 41

ST standard training. 13–15, 17–19, 21, 23–29, 44, 94

STT Speech-to-Text. 66

SUS System Usability Scale Questionnaire. 18, 26, 40, 43, 58, 59

TL traditional learning. 47–49, 53, 55, 57–60, 73, 74, 95

TP training provisioner. viii, 77, 84–92, 94–97

TTS Text-To-Speech. 50, 52, 54, 56, 67

VE virtual environment. 1–3, 7–9, 14, 17, 18, 31, 32, 35, 37, 39, 60, 64, 78, 79, 83, 84, 88, 92

VR Virtual Reality. vi, viii, 1, 2, 5, 7, 8, 14, 15, 18, 21, 24, 27–30, 32, 39, 40, 44, 61–64, 70, 74, 76, 78, 79, 83, 84, 87, 88, 92, 93, 95, 96

VRMT VR-based mock-up tool. viii, 77, 78, 82–92, 95, 96

VRTS VR training system. 3, 5–10, 14, 15, 17–19, 21, 23–30, 32, 35, 37, 39–41, 44, 45, 47, 76–78, 91, 94, 95

WOZ Wizard-of-Oz. 51, 52, 55, 56, 59

XR eXtended Reality. 1–4, 60, 76, 93

XRTS XR training system. 2–4, 31, 32, 43, 45–47, 49, 60, 73, 76, 93, 94, 96, 97

References

[1] Ivan E Sutherland. Sketchpad: A man-machine graphical communication system. *Simulation*, 2(5):R–3, 1964.

[2] Paul Milgram and Fumio Kishino. A taxonomy of Mixed Reality visual displays. *IEICE Transactions on Information and Systems*, 77(12):1321–1329, 1994.

[3] Richard Skarbez, Missie Smith, and Mary C. Whitton. Revisiting Milgram and Kishino's Reality-Virtuality continuum. *Frontiers in Virtual Reality*, 2, 2021.

[4] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. What is Mixed Reality? In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, page 1–15. ACM, 2019.

[5] Becky Spittle, Maite Frutos-Pascual, Chris Creed, and Ian Williams. A review of interaction techniques for immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2022.

[6] Johannes Leder, Tina Horlitz, Patrick Puschmann, Volker Wittstock, and Astrid Schütz. Comparing immersive Virtual Reality and powerpoint as methods for delivering safety training: Impacts on risk perception, learning, and decision making. *Safety Science*, 111:271–286, 2019.

[7] Nicholas Ho, Pooi-Mun Wong, Matthew Chua, and Chee-Kong Chui. Virtual Reality training for assembly of hybrid medical devices. *Multimedia Tools and Applications*, 77(23):30651–30682, 2018.

[8] Stephanie Hui-Wen Chuah. Wearable XR-technology: Literature review, conceptual framework and future research directions. *International Journal of Technology Marketing*, 13(3-4):205–259, 2019.

[9] Alexandra D Kaplan, Jessica Cruit, Mica Endsley, Suzanne M Beers, Ben D Sawyer, and Peter A Hancock. The effects of Virtual Reality, Augmented Reality, and Mixed Reality as training enhancement methods: A meta-analysis. *Human Factors*, 63(4):706–726, 2021.

[10] D Hamilton, J McKechnie, E Edgerton, and C Wilson. Immersive Virtual Reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education*, pages 1–32, 2020.

[11] Janne Heirman, Shivam Selleri, Tom De Vleeschauwer, Charles Hamesse, Michel Bellemans, Evarest Schoofs, and Rob Haelterman. Exploring the possibilities of extended reality in the world of firefighting. In *IEEE International Conference on Artificial Intelligence and Virtual Reality*, pages 266–273. IEEE, 2020.

[12] Annette Mossel, Mario Froeschl, Christian Schoenauer, Andreas Peer, Johannes Goellner, and Hannes Kaufmann. VROnSite: Towards immersive training of first responder squad leaders in untethered Virtual Reality. In *IEEE Virtual Reality*, pages 357–358. IEEE, 2017.

[13] Aurélie Conges, Alexis Evain, Frederick Benaben, Olivier Chabiron, and Sebastien Rebiere. Crisis management exercises in Virtual Reality. In *IEEE Conference on Virtual Reality and 3D User Interfaces – Abstracts and Workshops*, pages 87–92. IEEE, 2020.

[14] Esa M Rantanen and Donald A Talleur. Incremental transfer and cost effectiveness of groundbased flight trainers in university aviation programs. In *Proceedings Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 764–768. HFES, 2005.

[15] Berten Hurkmans and Kamakshi Rajagopal. The use of extended reality technologies for learning in industry. In *EdMedia+ Innovate Learning*, pages 900–903. Association for the Advancement of Computing in Education (AACE), 2020.

[16] Tariq Masood and Johannes Egger. Augmented Reality in support of Industry 4.0 – Implementation challenges and success factors. *Robotics and Computer-Integrated Manufacturing*, 58:181–195, 2019.

[17] Johannes Egger and Tariq Masood. Augmented Reality in support of intelligent manufacturing–a systematic literature review. *Computers & Industrial Engineering*, 140:106195, 2020.

[18] Filippo Gabriele Pratticò and Fabrizio Lamberti. Towards the adoption of Virtual Reality training systems for the self-tuition of industrial robot operators: A case study at KUKA. *Computers in Industry*, 129:103446, 2021.

[19] Filippo Gabriele Pratticò, Federico De Lorenzis, and Fabrizio Lamberti. Look at it this way: A comparison of metaphors for directing the user's gaze in eXtended Reality training systems. In *7th International Conference of the Immersive Learning Research Network*, pages 1–8. IEEE, 2021.

[20] Dimitris Nathanael, Stergios Mosialos, and George-C Vosniakos. Development and evaluation of a virtual training environment for on-line robot programming. *International Journal of Industrial Ergonomics*, 53:274–283, 2016.

[21] Nirit Gavish, Teresa Gutiérrez, Sabine Webel, Jorge Rodríguez, Matteo Peveri, Uli Bockholt, and Franco Tecchia. Evaluating Virtual Reality and Augmented Reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6):778–798, 2015.

[22] Yuanzhi Cao, Xun Qian, Tianyi Wang, Rachel Lee, Ke Huo, and Karthik Ramani. An exploratory study of Augmented Reality presence for tutoring machine tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM, 2020.

[23] Ziyue Guo, Dong Zhou, Qidi Zhou, Xin Zhang, Jie Geng, Shengkui Zeng, Chuan Lv, and Aimin Hao. Applications of Virtual Reality in maintenance during the industrial product lifecycle: A systematic review. *Journal of Manufacturing Systems*, 56:525–538, 2020.

[24] Stefan Werrlich, Austino Daniel, Alexandra Ginger, Phuc-Anh Nguyen, and Gunther Notni. Comparing HMD-based and paper-based training. In *International Symposium on Mixed and Augmented Reality*, pages 134–142. IEEE, 2018.

[25] Microsoft. Dynamics 365. https://dynamics.microsoft.com/it-it/mixed-reality/guides/, [accessed: 10-10-2020].

[26] Frederik Winther, Linoj Ravindran, Kasper Paabøl Svendsen, and Tiare Feuchtner. Design and evaluation of a VR training simulation for pump maintenance. In *Extended Abstracts Conference on Human Factors in Computing Systems*, pages 1–8. ACM, 2020.

[27] Luca Chittaro and Fabio Buttussi. Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):529–538, 2015.

[28] Mark R Mine, Frederick P Brooks Jr, and Carlo H Sequin. Moving objects in space: Exploiting proprioception in virtual-environment interaction. In *Proceedings 24th Annual Conference on Computer Graphics and Interactive Techniques*, pages 19–26. ACM, 1997.

[29] David Waller, Earl Hunt, and David Knapp. The transfer of spatial knowledge in virtual environment training. *Presence*, 7(2):129–143, 1998.

[30] Salman Nazir, Linda Johnstone Sorensen, Kjell Ivar Øvergård, and Davide Manca. Impact of training methods on distributed situation awareness of industrial operators. *Safety Science*, 73:136–145, 2015.

[31] Rafael Molina-Carmona, María Luisa Pertegal-Felices, Antonio Jimeno-Morenilla, and Higinio Mora-Mora. Virtual Reality learning activities for multimedia students to enhance spatial ability. *Sustainability*, 10(4):1074, 2018.

[32] Elias Matsas and George-Christopher Vosniakos. Design of a Virtual Reality training system for human–robot collaboration in manufacturing tasks. *International Journal on Interactive Design and Manufacturing*, 11(2):139–153, 2017.

[33] Vicente Román-Ibáñez, Francisco A Pujol-López, Higinio Mora-Mora, Maria Luisa Pertegal-Felices, and Antonio Jimeno-Morenilla. A low-cost immersive Virtual Reality system for teaching robotic manipulators programming. *Sustainability*, 10(4):1102, 2018.

[34] Luis Pérez, Eduardo Diez, Rubén Usamentiaga, and Daniel F García. Industrial robot control and operator training using Virtual Reality interfaces. *Computers in Industry*, 109:114–120, 2019.

[35] Endel Tulving and Donald M Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5):352, 1973.

[36] Christian E López, James Cunningham, Omar Ashour, and Conrad S Tucker. Deep reinforcement learning for procedural content generation of 3D virtual environments. *Journal of Computing and Information Science in Engineering*, 20(5), 06 2020. 051005.

[37] Peter Mako. Principles of mastering at KUKA robots. *International Scientific Journal about Machatronics*, 2(2):7–12, 2017.

[38] Angela M O'Donnell. The role of peers and group learning. *Handbook of Educational Psychology*, pages 781–802, 2006.

[39] HTC VIVE Pro kit. https://www.vive.com/us/product/vive-pro-full-kit/, [accessed: 12-04-2021].

[40] Unity. https://unity.com, [accessed: 12-04-2021].

[41] Blender. https://www.blender.org, [accessed: 12-04-2021].

[42] Panagiotis Kourtesis, Simona Collina, Leonidas AA Doumas, and Sarah E MacPherson. Validation of the Virtual Reality neuroscience questionnaire: maximum duration of immersive Virtual Reality sessions without the presence of pertinent adverse symptomatology. *Frontiers in Human Neuroscience*, 13:417, 2019.

[43] Krantiraditya Dhalmahapatra, Souvik Das, and J. Maiti. On accident causation models, safety training and Virtual Reality. *International Journal of Occupational Safety and Ergonomics*, 28(1):28–44, 2022. PMID: 32374205.

[44] Zhenan Feng, Vicente A González, Robert Amor, Ruggiero Lovreglio, and Guillermo Cabrera-Guerrero. Immersive Virtual Reality serious games for evacuation training and research: A systematic literature review. *Computers & Education*, 127:252–266, 2018.

[45] Christopher Healey and James Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, 2011.

[46] Alberto Cannavò, Davide Calandra, F Gabriele Pratticò, Valentina Gatteschi, and Fabrizio Lamberti. An evaluation testbed for locomotion in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, pages 1871–1889, 2020.

[47] David Heyne, Neville King, Bruce Tonge, Stephanie Rollings, Melinda Pritchard, Dawn Young, and Nicole Myerson. The self-efficacy questionnaire for school situations: Development and psychometric evaluation. *Behaviour Change*, 15(1):31–40, 1998.

[48] John M Keller. Development and use of the arcs model of instructional design. *Journal of Instructional Development*, 10(3):2, 1987.

[49] Sandra G Hart and Lowell E Staveland. Development of NASA–TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52:139–183, 1988.

[50] Roy S Kalawsky. VRUSE – A computerised diagnostic tool: For usability evaluation of virtual/synthetic environment systems. *Applied Ergonomics*, 30(1):11–25, 1999.

[51] John Brooke. SUS: A ‘quick and dirty’ usability scale. *Usability Evaluation in Industry*, page 189, 1996.

[52] Bangor Aaron, Kortum Philip, and Miller James. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3):114–123, 2009.

[53] Jocelyn Parong and Richard E Mayer. Cognitive and affective processes for learning science in immersive Virtual Reality. *Journal of Computer Assisted Learning*, 37(1):226–241, 2021.

[54] Rebecca A Grier. How high is high? A meta-analysis of NASA–TLX global workload scores. In *Proceedings Human Factors and Ergonomics Society Annual Meeting*, volume 59, pages 1727–1731. HFES, 2015.

[55] Sylvia Rothe, Daniel Buschek, and Heinrich Hußmann. Guidance in cinematic Virtual Reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction*, 3(1):19, 2019.

[56] Fabien Danieau, Antoine Guillo, and Renaud Doré. Attention guidance for immersive video content in head-mounted displays. In *Proceedings of IEEE Virtual Reality*, pages 205–206. IEEE, 2017.

[57] Minghao Cai and Jiro Tanaka. Remote shoulder-to-shoulder communication enhancing co-located sensation. In *Proceedings of 12th International Conference on Advances in Computer-Human Interactions*, pages 80–85. IARIA, 2019.

[58] Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. Look together: Using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing*, 21(1):173–186, 2017.

[59] Austin Erickson, Nahal Norouzi, Kangsoo Kim, Ryan Schubert, Jonathan Jules, Joseph J LaViola, Gerd Bruder, and Gregory F Welch. Sharing gaze rays for visual target identification tasks in collaborative Augmented Reality. *Journal on Multimodal User Interfaces*, 14(4):353–371, 2020.

[60] Thammathip Piumsomboon, Youngho Lee, Gun A Lee, Arindam Dey, and Mark Billinghurst. Empathic Mixed Reality: Sharing what you feel and interacting with what you see. In *International Symposium on Ubiquitous Virtual Reality*, pages 38–41. IEEE, 2017.

[61] Geonsun Lee, HyeongYeop Kang, JongMin Lee, and JungHyun Han. A user study on view-sharing techniques for one-to-many Mixed Reality collaborations. In *Conference on Virtual Reality and 3D User Interfaces*, pages 343–352. IEEE, 2020.

[62] Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. A user study on Mixed Reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM, 2020.

[63] Patrick Renner and Thies Pfeiffer. Attention guiding using Augmented Reality in complex environments. In *Conference on Virtual Reality and 3D User Interfaces*, pages 771–772. IEEE, 2018.

[64] Theophilus Teo, Gun A Lee, Mark Billinghurst, and Matt Adcock. Investigating the use of different visual cues to improve social presence within a 360 Mixed Reality remote collaboration. In *17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pages 1–9. ACM, 2019.

[65] Joshua Coburn, John Salmon, and Ian Freeman. The effects of transition style for collaborative view sharing in immersive Virtual Reality. *Computers & Graphics*, 92:44–54, 2020.

[66] Daniel Medeiros, Eduardo Cordeiro, Daniel Mendes, Maurício Sousa, Alberto Raposo, Alfredo Ferreira, and Joaquim Jorge. Effects of speed and transitions on target-based travel techniques. In *Proceedings of the 22nd Conference on Virtual Reality Software and Technology*, pages 327–328. ACM, 2016.

[67] Sebastian Oberdörfer, Martin Fischbach, and Marc Erich Latoschik. Effects of virtual environment transition techniques on presence, illusion of virtual body ownership, efficiency, and naturalness. In *Proceedings of the Symposium on Spatial User Interaction*, pages 89–99. ACM, 2018.

[68] Ping-Hsuan Han, Kuan-Wen Chen, Chen-Hsin Hsieh, Yu-Jie Huang, and Yi-Ping Hung. AR-arm: Augmented visualization for guiding arm movement in the first-person perspective. In *Proceedings of the 7th Augmented Human International Conference*, pages 1–4. ACM, 2016.

[69] Thuong N Hoang, Martin Reinoso, Frank Vetere, and Egemen Tanin. One-body: Remote posture guidance system using first person view in virtual environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, pages 1–10. ACM, 2016.

[70] Alberto Cannavò, Filippo Gabriele Pratticò, Giuseppe Ministeri, and Fabrizio Lamberti. A movement analysis system based on immersive Virtual Reality and wearable technology for sport training. In *Proceedings of the 4th International Conference on Virtual Reality*, pages 26–31, 2018.

[71] Mar Gonzalez-Franco, Brian Cohn, Eyal Ofek, Dalila Burin, and Antonella Maselli. The self-avatar follower effect in Virtual Reality. In *Conference on Virtual Reality and 3D User Interfaces*, pages 18–25. IEEE, 2020.

[72] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. The effects of sharing awareness cues in collaborative Mixed Reality. *Frontiers in Robotics and AI*, 6:5–23, 2019.

[73] Thammathip Piumsomboon, Gun A Lee, Andrew Irlitti, Barrett Ens, Bruce H Thomas, and Mark Billinghurst. On the shoulder of the giant: A multi-scale Mixed Reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–17. ACM, 2019.

[74] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993.

[75] Thomas W. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Z. für Medienpsychologie*, 15:69–71, 2003.

[76] Filippo Gabriele Pratticò, Francisco Navarro Merino, and Fabrizio Lamberti. Is learning by teaching an effective approach in Mixed-Reality robotic training systems? In *Intelligent Technologies for Interactive Entertainment*, pages 177–190. EAI, Springer International Publishing, 2021.

[77] Filippo Gabriele Pratticò, Javad Alizadeh Shabkhoslati, Navid Shaghaghi, and Fabrizio Lamberti. Bot undercover: On the use of conversational agents to stimulate teacher-students interaction in remote learning. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, pages 277–282. ACM, 2022.

[78] George Veletsianos and Gregory S Russell. Pedagogical agents. In *Handbook of research on educational communications and technology*, pages 759–769. Springer, 2014.

[79] Tiffany D. Do. Designing virtual pedagogical agents and mentors for extended reality. In *IEEE International Symposium on Mixed and Augmented Reality – Adjunct*, pages 476–479, 2021.

[80] Noah L. Schroeder, Olusola O. Adesope, and Rachel Barouch Gilbert. How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research*, 49(1):1–39, 2013.

[81] Steffi Heidig and Geraldine Clarebout. Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, 6(1):27–54, 2011.

[82] Gustav Bøg Petersen, Aske Mottelson, and Guido Makransky. Pedagogical agents in educational VR: An in the wild study. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–12. ACM, 2021.

[83] Luis Alfaro, Claudia Rivera, Jorge Luna-Urquizo, Elisa Castañeda, Jesús Zuñiga-Cueva, and María Rivera-Chavez. New trends in pedagogical agents in education. In *International Conference on Computational Science and Computational Intelligence*, pages 923–928, 2020.

[84] Perry Sprawls. Evolving models for medical physics education and training: a global perspective. *Biomedical Imaging and Intervention Journal*, 4(1), 2008.

[85] Rod D Roscoe and Michelene TH Chi. Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of educational research*, 77(4):534–574, 2007.

[86] David Duran. Learning-by-teaching. evidence and implications as a pedagogical mechanism. *Innovations in Education and Teaching International*, 54(5):476–484, 2017.

[87] Keith J Topping, Rayenne Dehkinet, Silvia Blanch, M Corcelles, and David Duran. Paradoxical effects of feedback in international online reciprocal peer tutoring. *Computers & Education*, 61:225–231, 2013.

[88] Roxana Moreno. Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32(1-2):99–113, 2004.

[89] Gautam Biswas, Krittaya Leelawong, Daniel Schwartz, Nancy Vye, and The Teachable Agents Group at Vanderbilt. Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19(3-4):363–392, 2005.

[90] Justin Werfel. Embodied teachable agents : Learning by teaching robots. Harvard, 2014.

[91] Fumihide Tanaka and Shizuko Matsuzoe. Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, 1(1):78–95, 2012.

[92] Fabrizio Lamberti, Filippo G Pratticò, Davide Calandra, Giovanni Piumatti, Federica Bazzano, and Thiago RK Villani. Robotic gaming and user interaction: Impact of autonomous behaviors and emotional features. In *IEEE Games, Entertainment, Media Conference*, pages 1–9. IEEE, 2018.

[93] Kasia Muldner, Victor Girotto, Cecil Lozano, Winslow Burleson, and Erin A Walker. The impact of a social robot's attributions for success and failure in a teachable agent framework. Boulder, CO: International Society of the Learning Sciences, 2014.

[94] Erin Walker, Victor Girotto, Younsu Kim, and Kasia Muldner. The effects of physical form and embodied action in a teachable robot for geometry learning. In *16th International Conference on Advanced Learning Technologies*, pages 381–385. IEEE, 2016.

[95] Amy L Baylor. Three research directions for affective learning technologies. International Society of the Learning Sciences, Inc., 2018.

[96] Anki Cozmo. https://anki.com/en-us/cozmo.html, [accessed: 10-06-2022].

[97] Cozmo SDK. http://cozmosdk.anki.com/docs/index.html, [accessed: 10-06-2022].

[98] Filippo Gabriele Pratticò and Fabrizio Lamberti. Mixed-Reality robotic games: Design guidelines for effective entertainment with consumer robots. *IEEE Consumer Electronics Magazine*, 2020.

[99] OpenCV v3.2. https://opencv.org/opencv-3-2/, [accessed: 10-06-2022].

[100] ZeroMQ. https://zeromq.org/, [accessed: 10-06-2022].

[101] Logan Fiorella and Richard E Mayer. Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology*, 39(2):75–85, 2014.

[102] Sandra Y Okita and Daniel L Schwartz. Learning by teaching human pupils and teachable agents: The importance of recursive feedback. *Journal of the Learning Sciences*, 22(3):375–412, 2013.

[103] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81, 2009.

[104] AltspaceVR. https://altvr.com/, [accessed: 10-06-2022].

[105] Horizon Worlds. https://www.oculus.com/horizon-worlds/, [accessed: 10-06-2022].

[106] Mitchell L. Thom, Blair A. Kimble, Kelli Qua, and Susanne Wish-Baratz. Is remote near-peer anatomy teaching an effective teaching strategy? Lessons learned from the transition to online learning during the Covid-19 pandemic. *Anatomical Sciences Education*, 14(5):552–561, 2021.

[107] Andrew Yoshimura and Christoph Walter Borst. Remote instruction in Virtual Reality: A study of students attending class remotely from home with VR headsets. In *Mensch und Computer - Workshopband*. GI, Gesellschaft für Informatik e.V., 2020.

[108] Fabrizio Lamberti, Alberto Cannavò, and Paolo Montuschi. Is immersive Virtual Reality the ultimate interface for 3D animators? *Computer*, 53(4):36–45, 2020.

[109] Alberto Cannavò and Fabrizio Lamberti. A Virtual Character Posing System based on Reconfigurable Tangible User Interfaces and Immersive Virtual Reality. In Marco Livesu, Gianni Pintore, and Alberto Signoroni, editors, *Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference*, pages 2617–4855. The Eurographics Association, 2018.

[110] Santi Caballé and Jordi Conesa. Conversational agents in support for collaborative learning in MOOCs: An analytical review. In *International Conference on Intelligent Networking and Collaborative Systems*, pages 384–394. Springer, 2018.

[111] Wenjing Li, Fuxing Wang, Richard E Mayer, and Tao Liu. Animated pedagogical agents enhance learning outcomes and brain activity during learning. *Journal of Computer Assisted Learning*, pages 1–17, 2021.

[112] Stergios Tegos and Stavros Demetriadis. Conversational agents improve peer learning through building on prior knowledge. *Journal of Educational Technology & Society*, 20(1):99–111, 2017.

[113] Gregory Dyke, Iris Howley, David Adamson, Rohit Kumar, and Carolyn Penstein Rosé. Towards academically productive talk supported by conversational agents. In *Productive Multivocality in the Analysis of Group Interactions*, pages 459–476. Springer, 2013.

[114] Donghoon Shin, Sangwon Yoon, Soomin Kim, and Joonhwan Lee. Blahblahbot: Facilitating conversation between strangers using a chatbot with ml-infused personalized topic suggestion. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2021.

[115] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. Bot in the bunch: Facilitating group chat discussion by improving efficiency and participation with a chatbot. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM, 2020.

[116] Stergios Tegos, Georgios Psathas, Thrasyvoulos Tsiatsos, Christos Katsanos, Anastasios Karakostas, Costas Tsibanis, and Stavros Demetriadis. Enriching synchronous collaboration in online courses with configurable conversational agents. In *International Conference on Intelligent Tutoring Systems*, pages 284–294. Springer, 2020.

[117] Meng-Yun Liao, Ching-Ying Sung, Hao-Chuan Wang, Wen-Chieh Lin, and Fu-Yin Cherng. Embodying historical learners' messages as learning companions in a VR classroom. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–6. ACM, 2019.

[118] Mozilla Hubs. https://hubs.mozilla.com/, [accessed: 10-06-2022].

[119] Meta Quest 2. https://store.facebook.com/it/quest/products/quest-2/tech-specs/, [accessed: 10-06-2022].

[120] Mozilla Hubs codebase. https://github.com/mozilla/hubs, [accessed: 10-06-2022].

[121] Mohamed Ali Nagy Elmaadaway. The effects of a flipped classroom approach on class engagement and skill performance in a blackboard course. *British Journal of Educational Technology*, 49(3):479–491, 2018.

[122] Chad Harms and Frank Biocca. Internal consistency and reliability of the networked minds measure of social presence. In *7th annual international workshop: Presence*. Universidad Politecnica de Valencia Valencia, Spain, 2004.

[123] Filippo Gabriele Pratticò, Federico De Lorenzis, Davide Calandra, Alberto Cannavò, and Fabrizio Lamberti. Exploring simulation-based Virtual Reality as a mock-up tool to support the design of first responders training. *Applied Sciences*, 11(16), 2021.

[124] Hugo Coelho, Pedro Monteiro, Guilherme Gonçalves, Miguel Melo, and Maximino Bessa. Authoring tools for Virtual Reality experiences: A systematic review. *Multimedia Tools and Applications*, pages 1–24, 2022.

[125] William Wilkerson, Dan Avstreih, Larry Gruppen, Klaus-Peter Beier, and James Woolliscroft. Using immersive simulation for training first responders for mass casualty incidents. *Academic Emergency Medicine*, 15(11):1152–1159, 2008.

[126] Danny M Peterson and Ronald W Perry. The impacts of disaster exercises on participants. *Disaster Prevention and Management: An International Journal*, 1999.

[127] Helen Sinclair, Emma E. Doyle, David M. Johnston, and Douglas Paton. Assessing emergency management training and exercises. *Disaster Prevention and Management: An International Journal*, 21(4):507–521, Jan 2012.

[128] Robert Hays and Michael J. Singer. *Simulation Fidelity in Training System Design: Bridging the Gap Between Reality and Training*. Springer-Verlag New York, 1988.

[129] Piedmont Region Civil Protection. https://www.regione.piemonte.it/web/temi/protezione-civile-difesa-suolo-opere-pubbliche/protezione-civile, [accessed: 30-07-2021].

[130] Piedmont Region Coordination body of Civil Protection Volunteering. https://www.coordinamentoregionaleprotezionecivilepiemonte.it/, [accessed: 30-07-2021].

[131] Sarah I. Leberman and Andrew J. Martin. Applying dramaturgy to management course design. *Journal of Management Education*, 29(2):319–332, 2005.

[132] Reza Zaker and Eloi Coloma. Virtual Reality-integrated workflow in bim-enabled projects collaboration and design review: A case study. *Visualization in Engineering*, 6(1):1–15, 2018.

[133] Ryszard Klempous, Konrad Kluwak, Radosław Idzikowski, Tomasz Nowobilski, and Tomasz Zamojski. Possibility analysis of danger factors visualization in the construction environment based on Virtual Reality model. In *8th IEEE International Conference on Cognitive Infocommunications*, pages 363–368. IEEE, 2017.

[134] Vivian Gómez, Kelly Peñaranda, and Pablo Figueroa. Requirements gathering for VR simulators for training: Lessons learned for globally dispersed teams. In *IEEE Conference on Virtual Reality and 3D User Interfaces – Abstracts and Workshops*, pages 452–453. IEEE, 2021.

[135] Igor Petukhov, Liudmila Steshina, Andrey Glazyrin, and Dimiter Velev. Design model of a training simulator in Virtual Reality. In *Proceedings of the 5th International Conference on Fundamentals and Advances in Software Systems Integration*, pages 1–7. IARIA, 2019.

[136] 2008/73/ec,euratom: Commission decision of 20 december 2007 amending decision 2004/277/ec, euratom as regards rules for the implementation of council decision 2007/779/ec, euratom establishing a community civil protection mechanism. *OJ*, L 20:23–34, 2008-01-24.

[137] Andy Martin, Dan Franc, and Daniela Zounková. *Outdoor and Experiential Learning: An Holistic and Creative Approach to Programme Design (1st ed.)*. 2004.

[138] Andrew Martin, Sarah Leberman, and James Neill. Dramaturgy as a method for experiential program design. *Journal of Experiential Education*, 25(1):196–206, 2002.

[139] Valve Index. https://www.valvesoftware.com/en/index, [accessed: 30-07-2021].

[140] Patrick Jost, Sue Cobb, and Isabella Hämmerle. Reality-based interaction affecting mental workload in Virtual Reality mental arithmetic training. *Behaviour & Information Technology*, 39(10):1062–1078, 2020.

[141] Marc Hassenzahl, Franz Koller, and Michael Burmester. Der user experience (ux) auf der spur: Zum einsatz von www.attrakdiff.de. In Henning Brau, Sarah Diefenbach, Marc Hassenzahl, Franz Koller, Matthias Peissner, and Kerstin Röse, editors, *Tagungsband UP08*, pages 78–82. Fraunhofer Verlag, 2008.

[142] Lei Zhang and Steve Oney. Flowmatic: An immersive authoring tool for creating interactive scenes in Virtual Reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 342–353. ACM, 2020.