Assessing the Usability of Different Virtual Reality Systems for Firefighter Training

Fabrizio Corelli, Edoardo Battegazzorre, Francesco Strada, Andrea Bottino, Gian Paolo Cimellaro
Department of Control and Computer Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
fabrizio.corelli@studenti.polito.it, {edoardo.battgazzorre, francesco.strada, andrea.bottino, gianpaolo.cimellaro}@polito.it

Keywords: VR Training, Serious Games, Locomotion, Usability, User Study.

Abstract: The use of Virtual Reality (VR) based learning environments for training firefighters is becoming more and more common. The key advantages of these approaches is that they allow the development of experiential learning environments, where trainees can be involved into and interact with complex emergency scenarios, including those that cannot rely for the training on real world systems and environments due to costs or security concerns. Despite that, current VR training systems are still affected by a number of weaknesses, mainly related to usability and to the (limited) sense of presence conveyed by the virtual environment (VE), which can negatively affect the expected learning outcomes. To this end, in order to gain further insight into this problem, this work aims at assessing the usability of a firefighter training application deployed in three VR systems and exploiting serious games in the educational approach. The VR systems under analysis provide different levels of immersion and offer different approaches to manage interaction and locomotion inside the VE. Experimental results, obtained through a user study, show differences among the three systems. In particular, the devices and metaphors used to manage locomotion in VR seem to be the most critical parameters with respect to usability and learners' achievements.

1 INTRODUCTION

Current VR technologies are providing fire services with new and compelling opportunities for transforming the way firefighters are trained. According to (Engelbrecht et al., 2019), the key advantages of VR-based firefighter training can be summarized as follows. First, firefighters need to learn how to address critical situations where they can risk their life, and VR provides a safe training environment where emergencies can be simulated without putting trainees in any real danger. Second, a VR application allows trainees to experience different scenarios and emergency procedures and the same application can be used to train different categories of personnel. Third, VR training can be also delivered remotely, thus helping reduce costs and overall training times, and it can even support collaborative activities, which are beneficial for learning (Rojas-Drummond and Mercer, 2003). Then, VR offers compelling experiential learning environments that improve the learner's engagement and, ultimately, the learning outcomes. Another advantage is that VR promotes repetition (i.e., users can repeat the training session at their own pace), which, in turns, improves content retention. Finally, training sessions can be logged and reviewed in a debriefing sessions, where learners can critically analyze what they did, get insights from their experience and analyze mistakes and explore alternative solutions. Thus, debriefing sessions enhance the transfer of knowledge and skills from virtual to real world, and promote active learning (Garris et al., 2002).

The idea of using VR for firefighter training is not a new concept. Several works have been presented in the literature. The VR systems used to support the training sessions can vary from desktop VR (Lee et al., 2010) to immersive systems, such as those based on Head Mounted Displays (HMDs), (Argasinski et al., 2018), (Tate et al., 1997) or on CAVE environments (Backlund et al., 2007). Simulations can also exploit dynamic and physics based systems for recreating the correct behaviour of fire and smoke (Cha et al., 2012). This is a crucial element in this context, especially when training targets emergency procedures inside buildings or tunnels. Some attempts have been even made to increase the sensory spectrum of these simulations by implementing odor (Cater, 1994) and heat radiance generators (Lee et al., 2010) to improve the sense of immersion into a fiery environment. Given the relevance of the decision-making aspect in the emergency management, VR environment have been used as well to analyze the relationships among firefighter experience and the decision-making processes (Bayouth et al., 2013).

Despite the many advantages, there are a number of weaknesses that still affect current VR systems in this specific context (Engelbrecht et al., 2019). One relevant issue is that, despite the many efforts, actually there are no devices or interaction metaphors that allow natural movements in large VR spaces, such as those that are typical in firefighter training scenarios. Moreover, some of these navigation approaches might even induce cybersickness (Rebenitsch and Owen, 2016), with a negative effect on learning. A second weakness is that technology still needs a leap forward to maximize the sense of presence, which ultimately is one of the main factors that help transfer the learned skills from virtual space to real life. Finally, since trainees' acceptance of the technology is essential to achieve the desired learning outcomes (Heldal et al., 2016), the usability of the system and the overall user experience should be maximized as well. However, achieving this result is not trivial and involves a careful selection of the interaction devices/metaphors and an ad-hoc design of the application.

The goal of this work is to assess the usability of three different VR systems (using different hardware configurations) in a firefighter training context. In details, these systems are a desktop VR, and two immersive environments, both leveraging an HMD for visualization and hand-held controllers for interaction but exploiting two different locomotion techniques: a gaze-directed-steering metaphor (Bowman et al., 1997) and an active re-positioning technique (Nilsson et al., 2018) that allows trainees to naturally walk through large VEs leveraging an omni-directional treadmill. The training scenario used in the experiments reproduces a fire-following-earthquake event in a school. The educational path of the application comprises a learning mode (where trainees are instructed step-by-step on the sequence of actions they have to perform in the emergency procedure), and an evaluation mode (which leverages a serious game to assess the learned skills).

The three VR system have been compared by means of an user-evaluation study that involved 45 volunteers divided in three separate groups (each using a different system). Experimental results showed differences in the usability of the three system under analysis, highlighting that the way locomotion is managed is the most critical parameter that affects both usability and users' achievements in terms of learning outcomes.

The rest of the paper is organized as follows. Section 2 details the design of the training application and the characteristics of the different VR systems under analysis. Section 3 introduces the experimental protocol and Section 4 presents and discusses the results. Finally, Section 5 draws the conclusions of this work.

2 APPLICATION AND SYSTEM DESIGN

The firefighter training programs, irrespective of the actual simulation scenario addressed, require to carry out certain actions in a specific order. The completion of an action involves navigation in the VE and interaction with virtual places, objects and avatars inside the environment. That said, the application design should be flexible enough to support different scenarios and procedures. To this end, we modeled action dependencies as a directed graph, where nodes represent individual actions and edges correspond to dependency requirements. The control of the node execution flow leverage context awareness (i.e., environment state, user's interaction, internal and external events). Composite nodes can be used to orchestrate various sub-nodes (according to different algorithms, such as sequential or parallel execution and loop management), thus allowing the definition of complex action inter-dependencies.

As another constraint, the design should support the application deployment on different hardware configurations. This has been done by leveraging design patterns specific for multi-platform applications and exploiting the cross-platform features offered by the Unity engine for the development.

2.1 Learning path design

The learning path of the application is organized in two different parts: a *learning* and an *evaluation* session. In the learning session, users are guided step-by-step through the correct sequence of actions they have to perform. For each individual action, visual and audio clues instruct trainees on what they have to do (and why) and on the sequence of steps required to complete the activity. During task accomplishment, users are supported by prompt and clear feedback that inform them when an interaction is available (such as adding a glow to interactable objects) and notify the success/failure in performing an action.

The evaluation session is envisioned as a serious game, where users can freely perform any action involved in the procedures learned in the previous phase, but they cannot benefit from any of the cognitive aids available in the learning mode. The game design leverages two main elements. A timer enforces a time limit for completing a given task. A score value translates the player's success in the game into a numerical representation, which is then used as an automatic assessment of the learned skills. The current score and timer values are displayed in the VE and audio and visual cues highlight specific events associated to them (e.g., timer expiration, new points obtained). At the end of the game, players receive a summary report of their achievements and can see their placement in the overall ranking. This feature aims at fostering competitive behaviors (i.e., by challenging players to beat their colleagues' scores), which ultimately are beneficial to learning (Cagiltay et al., 2015).

2.2 VR environments

For the evaluation, the application was deployed in the three types of VR environments depicted in Figure 1, whose main differences consist in the interaction techniques and the level of immersion offered.

The first is a desktop VR system (referred to as DVR in the following), where the VE is displayed on a large monitor. The mouse movements control camera orientation, while the player's spatial movements are handled with the keyboard arrow keys. The behaviour associated to an interactable object can be triggered by first selecting the object (pointing the camera at it) and then pressing a mouse button.

On the contrary, the immersive VR system (IC) uses an HMD for the visualization and hand-held controllers for managing interaction and locomotion. The head movements (captured by the HMD tracking system) are used to control camera position and orientation. As for the navigation, the active area of the HMD trackers is not large enough to let user explore the VR by physical walking. Therefore, locomotion is managed by using the joystick included within the controllers to translate the user in the current gaze yaw direction. Interaction with objects leverages the virtual hand metaphor (Poupyrev and Ichikawa, 1999), where the controller movements are mapped into that of two virtual hands that players can use to grab or activate interactable objects in the VE, thus providing a natural and immediate interaction tool.

The last system (KAT) is similar to IC, since it leverages HMD as display system and hand-held controllers for interaction with objects, but uses a KAT-Walk treadmill to manage locomotion. Treadmills are "body-centric" re-positioning systems (Nilsson et al., 2018) that translate physical gestures (e.g. walking or running) into virtual movements. With the KAT-Walk, the user is strapped into a harness, attached to a supporting structure and slightly lifted over a concave platform (Figure 1). The platform has a low friction surface that prevents the forces generated during each step from physically moving forward the user. The walking gestures are captured by two inertial sensors that are placed on special overshoes, while a third sensor, placed on the back of the harness, tracks the movement direction and triggers the locomotion input when the user tilts forward or backward.

Figure 1: VR environments considered in this study: desktop VR (DVR, left), immersive VR (IC, center) and treadmill based immersive VR (KAT, right).

3 EXPERIMENTS

The scenario selected for our experiments is a fire-following-earthquake event in a school. In order to make it as realistic as possible, the emergency scenario is set into a real building (the Mascagni middle school of Melzo, Italy), which was recreated in the VE importing its Building Information Model (BIM). In the emergency scenario under analysis, a short circuit caused by an earthquake generates a fire inside a classroom, located in the first floor of the school. Some students in the classroom are wounded or trapped by fallen debris and, thus, they need to be assisted and rescued, while the remaining students escape from the classroom and evacuate through the external emergency stairs. The activities of the emergency procedure trainees are supposed to learn, as well as their completion times, have been defined with the help of expert firefighter trainers according to the characteristics of the building where the procedure takes place. These activities are the following. The firefighter should enter the main building, identify the fire location by analyzing the smoke propagation and the screams coming from the classroom. Since the operator is likely to enter the school for the first time, he/she must observe an evacuation plan chart hanging on the wall in order to orient himself/herself into the unfamiliar environment. Then, he/she must

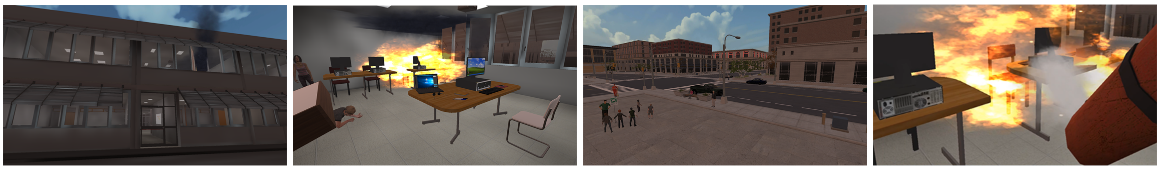

Figure 2: Snapshot of the VR training scenario (from left to right): the main entrance of the school, the computer room where the fire started, the evacuation meeting point seen from the emergency stairs and the fire-extinguishing step.

find his/her way to the first floor (navigating corridors and using internal stairs). When the operator has reached the classroom, he/she should first indicate to the unharmed students the closest emergency exit so that they can promptly evacuate the building. Then, he/she must provide first aid to the injured students and help them evacuate to a safe area. Subsequently, the learner has to go back in the classroom and operate a fire extinguisher (available in the building). Once the fire is extinguished, the operator should return to the safe area where the evacuated students are waiting for him. Some snapshot of the procedure actions can be seen in Figure 2. We underline that the scenario taken in consideration allows us to stress the interaction in general and the locomotion interfaces in particular, since users had to (virtually) walk for long distances and move inside narrow spaces (thus requiring a fine-grained and detailed control of their movements) in order to complete the assigned tasks.

In order to compare the different setups described in Section 2.2, we performed a user study that involved 45 volunteers (35 men and 10 women, aged between 29 and 30) selected among students and personnel of our University. Six of them are frequent users of VR, 24 have had a previous experience with VR and the remaining 15 had never experienced VR before. Users were divided in three groups of 15 people, each experiencing the training application through a different system. The experimental protocol applied was the following. First, since users are likely to have low familiarity with the VR devices to be used, the learning experience started with an interaction training session in a test environment where users could get acquainted with the interaction and locomotion modalities available. This preliminary step aimed at reducing barriers related to VR technologies (especially for novices). When users felt confident with the system, they were invited to repeat twice the learning session and, finally, to perform twice the evaluation session.

The assessment of the different systems encompassed the analysis of both the learning outcomes and the usability of the system. In the human-computer interaction (HCI) field, *usability* is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (Iso, 1998). Following this definition, we assessed usability by measuring *effectiveness* in terms of correctness of the performed procedures, *efficiency* from time on task and paths navigated inside the environment (where both effectiveness and efficiency metrics using objective in-game analytics), and *satisfaction* through standardized questionnaires (i.e., leveraging subjective measurements) that users were required to fill after they completed the experience. The proposed questionnaire is organized in different sections aimed to collect (i) information about usability in broad terms (by considering the System Usability Scale, SUS (Brooke, 2013), and the five attributes defined by Nielsen (Nielsen, 1993), i.e., *learnability*, *efficiency*, *memorability*, *possibility to recover from errors* and *satisfaction*), (ii) ergonomic aspects of the proposed system (through statements derived from the ISO 9241-400 standard (iso, 2007)) and (iii) detailed usability information on the VR system (on the basis of the VRUSE questionnaire (Kalawsky, 1999)). All questionnaire items had to be scored by users expressing their agreement on a five point Lickert scale (1, totally disagree; 5, totally agree).

As for the expected learning outcomes, since trainees should learn how to execute the various steps of the procedure in the correct order and within a predefined time interval, their assessment was based on the analysis of users' behaviour in terms of completion time, correctness of the performed procedures and learning progresses.

4 RESULTS AND DISCUSSION

In this section, we will first discuss the results concerning usability of the VR systems under analysis (Section 4.1. Then, we will assess the learning outcomes achieved with these systems and with the help of the serious game used during the evaluation phase.

In the following, both subjective and objective results obtained from the three volunteers groups were analyzed with a One-Way ANOVA test with post-hoc Tukey's test to eventually identify groups showing statistically significant differences.

4.1 Usability assessment

A first indication of the usability of the three environments comes from the SUS normalized results in the 0–100 range (with odd items reverted). According to (Brooke, 2013), a score above 68 shall be considered as above average. As shown in Fig. 4, the KAT SUS score is significantly lower than that of the other two systems (DVR 86.0, IC 81.7, KAT 68.3, $p = 0.00002$) and, thus, its usability can be considered as barely acceptable. Similar conclusions can be drawn by considering Nielsen's attributes (second section of the questionnaire). The three systems did not show any significant difference in the learnability, memorability and satisfaction attributes. However (Figure 3), the KAT had significantly lower values than DVR and IC (DVR 4.7, IC 4.5, KAT 4.1, $p = 0.001$) regarding efficiency and was also perceived as having lower possibilities to recover from errors than IC (DVR 3.9, IC 4.4, KAT 3.7, $p = 0.03$).

These efficiency results are confirmed by the objective results collected. As a matter of fact, if we consider the average distances travelled with the three systems in the first and second run of the evaluation session (Figure 6), we can observe a large difference between the KAT and the other systems (DVR 262.2, IC 241.5, KAT 284.4, $p = 0.01$ for the first run; DVR 254.8, IC 213.6, KAT 277.2, $p = 0.001$ for the second run, where all measurements are expressed in meters).

Similar differences between KAT and DVR/IC are obtained from the analysis of the percentage of actions completed in time (first run: DVR 95.0%, IC 90.08%, KAT 75.8%, $p = 0.02$; second run: DVR 97.5%, IC 96.4%, KAT 83.0%, $p = 0.01$; see Figure 6). This result highlights as well the negative effect of the lower KAT usability on the learning outcomes. Trainees should not only learn to perform all the actions required (and in the correct order), but they should also learn to complete them in a suitable time, since (for some actions) time is a critical parameter (e.g., in our case, the room where fire started should be reached as soon as possible in order to provide immediate first-aid to the people in that place).

A possible explanation of this last result is provided by both direct observations and analysis of the path traveled. Observations highlighted that KAT users were less capable than others to exert a fine-grained control of their movements. As a result, they were not always in the condition of following an "optimal" path to reach their destination. These difficulties in controlling small and detailed movements in narrow spaces[1] are probably the main reason that led to increase the completion times of individual actions and, thus, of the full procedure.

Another element that contributes to highlight the navigation issues with KAT is the analysis of the travelled paths, which are summarized in the heatmaps shown in Figure 5. In each heatmap, colors on the floor correspond to different traffic intensities (red areas are the ones lots of people walked in, while green regions had lower traffic). It can be seen that user tracks for DVR and IC are much more compact than the KAT's one (despite few outliers that took "bizarre" paths to reach their targets).

As for ergonomics (third section of the questionnaire), learners were asked to evaluate the interaction with the VR systems according to the four statements summarized in Figure 3. It can be seen that three out of these four statements underline the low KAT ergonomics. In details, the locomotion device is more cumbersome than the ones provided by DVR and IC (DVR 1.4, IC 1.6, KAT 2.4, $p = 0.005$) and, to be operated, it requires a higher mental effort than DVR (DVR 1.7, IC 1.8, KAT 2.5, $p = 0.03$) and higher physical efforts than both DVR and IC (DVR 1.1, IC 1.5, KAT 3.8, $p = 5.53e-13$).

The fourth questionnaire section (VRUSE) underlines again the lower user appreciation received by KAT. The VRUSE is divided into different subsections (related to the analysis of *functionality, locomotion, interaction with objects, flexibility, error correction/handling, simulation fidelity*, and *sense of immersion and presence*). Each subsection is then concluded by a closing summary question aimed at capturing from users an overall evaluation on the aspects investigated by the subsection. If we start analyzing these closing questions, it can be observed that the answers are similar for all the systems[2], exception made for the question corresponding to the locomotion section, where the KAT had significant lower values than DVR and IC (DVR 4.1, IC 3.9, KAT 3.2, $p = 0.04$).

Thus, in order to gain further understanding about the issues encountered by users, we deemed interesting to discuss in details the questions of this subsection and their answers (which are summarized, again, in Figure 3). A first negative comment is that the KAT

[1] In our simulation, this is, for instance, the situation trainees are faced with when they have to operate in the classroom where the fire started, or when they have to climb staircases.

[2] As a note, the fact that the level of immersion and presence delivered by DVR was similar to the one offered by IC and KAT was quite a surprising finding. A possible explanation is that each user group experienced a single VR system (thus, volunteers had no possibilities to make comparisons among them) and DVR group was experiencing the VE through large monitors, which probably helped mitigate the "through the window" effect inherent in desktop VR systems.

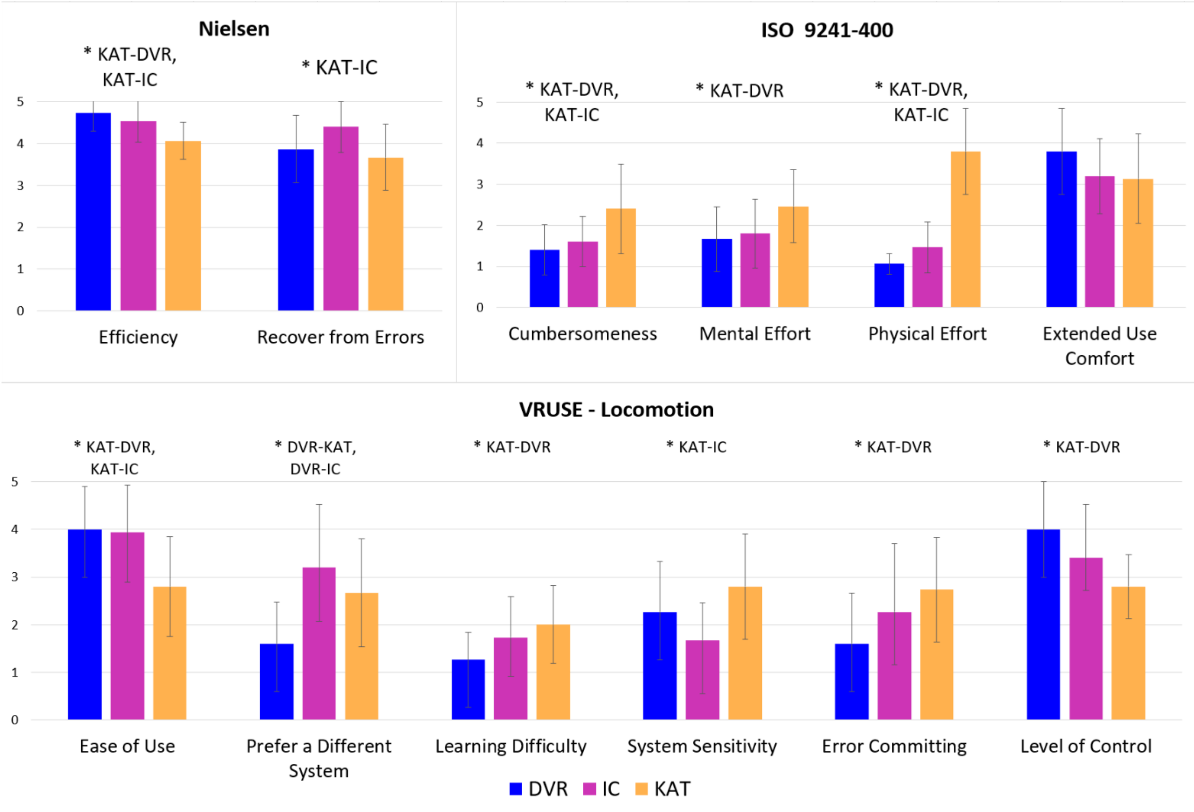

Figure 3: Excerpt of the subjective evaluation results. Top row: overall scores for the most significant Nielsen attributes (Left) and the ISO 9241-400 properties (Right). Bottom row: usability factors regarding the locomotion system for which we found a significant difference among systems (questions adopted from the VRUSE questionnaire). For all graphs, the "*" symbol indicates a statistically significant difference and standard deviations are expressed through error bars.

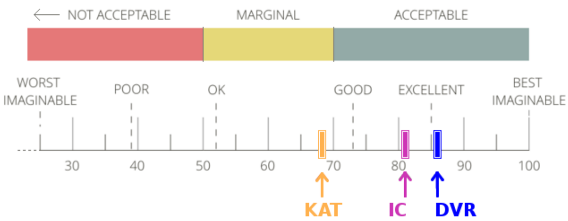

Figure 4: SUS summary results. The scales displayed are taken from (Brooke, 2013) and show that while IC and DVR score as good/excellent, KAT can be only considered as marginally acceptable.

ease of use is sensibly lower than that of the other two locomotion approaches (DVR 4.0, IC 3.9, KAT 2.8, $p = 0.003$). Both locomotion systems in immersive environments (IC and KAT) are not highly appreciated by their users, since they expressed their preference to use a different locomotion system with a sensibly higher strength than that used by DVR users (DVR 1.6, IC 3.2, KAT 2.7, $p = 0.002$). Then, learning how to use the KATWalk appears to be more difficult than the mouse and keyboard controls offered by DVR (DVR 1.3, IC 1.7, KAT 2, $p = 0.04$) and the KAT sensitivity is higher than that of IC (DVR 2.3, IC 1.7, KAT 2.8, $p = 0.02$). Finally, the probability of committing errors with KAT is higher than with DVR (DVR 1.6, IC 2.3, KAT 2.7, $p = 0.04$) and the level of control is lower with KAT than with DVR (DVR 4.0, IC 3.4, KAT 2.8, $p = 0.005$). These level of control values confirm the difficulties experienced by KAT users in the (fine) control of the navigation.

4.2 Learning outcomes

As we already observed, the VR application was effective in producing the expected learning outcomes. The percentage of actions completed in the correct order in the second evaluation run was 100% for all users except one volunteer of the DVR group who made a single mistake. The percentage of actions completed in the correct order and in a timely fashion (Figure 6, second run) was close to 100% for all groups, except for KAT where this value was only 83% (and significantly different from that of the other groups). This result, as we discussed before, was mainly due to the issues with the locomotion device.

Figure 6 allows to appreciate as well that, despite

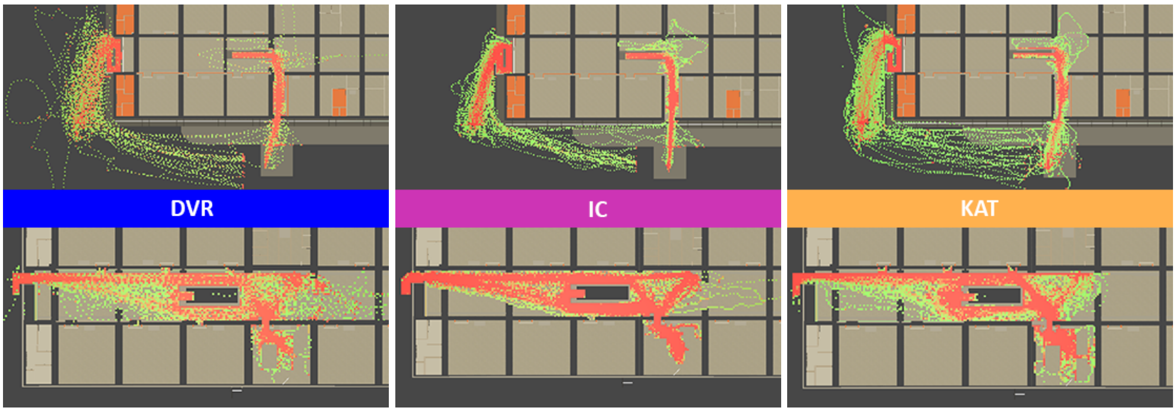

Figure 5: Heatmaps indicating users' movements in the virtual environment. Images are color coded (green: less frequently travelled paths, red: more frequent ones) and are divided for group (from left to right: DVR, IC and KAT) and floor (for each group, top image shows the ground floor and bottom image the first floor).

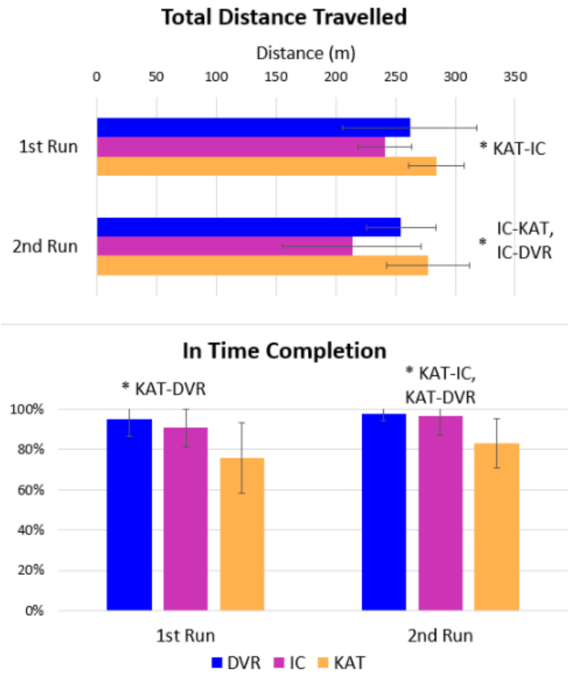

Figure 6: Total distances traveled (up) and percentage of activities completed in time (down) for the two runs of the evaluation session.

Statement	Score (SD)
I had fun using a game to learn intended content	4.56 (0.58)
I would have preferred an instructor-based learning modality	2.22 (0.94)
I would have preferred a learning modality, based on books, notes, slides, etc.	1.38 (0.74)
The designed game is a valid learning tool	4.42 (0.61)
The possibility to compare my score with other learners' scores made me try to improve my results	4.40 (1.04)
I found the tool more a game than a system suitable for training	2.78 (1.09)
Trying to improve my results in the game let me learn intended content better	4.02 (1.02)
The presence of a timer stimulated me to quickly carry out required actions	4.07 (1.06)
I found the timer a stressful element	1.78 (0.84)

Table 1: Statements used to assess users' appreciation of the serious game (mean values and standard deviations).

small inter-session differences, all volunteer groups benefits from repetition learning. The percentage of correct action completion increased between the two runs by 1.7% for both DVR and IC and 8.3% for KAT, and that of correct and timely completion of 2.5% for DVR, 5.6% for IC and 7.2% for KAT. Despite that, the lack of follow-up studies does not allow to infer the effects of repetition on users' knowledge retention.

Finally, we can analyze the contribution provided to the learning outcomes by game mechanics introduced in the evaluation session. This analysis is based on the concluding section of the questionnaire, in which we asked volunteers to express their level of agreement with the 9 statements reported in Table 1. We found that the results have no significant differences among the three groups and, thus, we reported them as the cumulative averages among all users of the three groups. These results can be summarized as follows. The game features made the learning process more fun and the approach proposed to deliver the educational contents was appreciated by users. The game elements introduced to solicit the competitiveness (the score and the leaderboard) were able to improve the learners' engagement and foster repetition, and the timer was considered as an effective element in encouraging users to complete the procedure actions in time. Finally, volunteers found that there was a good balance between educational and entertainment elements.

5 CONCLUSIONS

In this work we presented and discussed a user study aimed at assessing the usability, ergonomics and effectiveness (in terms of learning outcomes) of three different VR systems used to deploy a firefighter training application.

The major takeaways of this work are the following. First, the VR training environment was capable of achieving the expected learning outcomes, in terms of both precision and timely completion of the emergency procedures. Second, the introduction of the serious game in the evaluation phase of the application was appreciated by learners and was contributing to support the educational path designed. Third, we found different levels of usability (as well as different levels of users' appreciation) among the different VR systems analyzed.

In particular, one of the most critical parameters influencing the evaluations was the quality of the locomotion management provided by the system. As a matter of fact, locomotion is a relevant task in our scenario. In order to complete the emergency procedures, users have both to travel long distances and to exert a fine control on their movements. With respect to these requirements, when the interaction devices (or the interaction metaphors) are not capable of supporting the users, the results is a negative effect on the trainee performances, in terms of both accuracy and timing of the execution.

Future works will address the evaluation of alternative locomotion interfaces, simple to use and capable of guaranteeing a high level of immersion and an adequate naturalness in the locomotion control. Then, given the relevance of locomotion in other areas of application, we are planning to extend the breadth of this study beyond the firefighter training domain.

REFERENCES

(2007). *ISO 9241-400 Ergonomics of human-system interaction – Part 400: Principles and requirements for physical input devices*.

Argasinski, J. K., Wegrzyn, P., and Strojny, P. (2018). Affective vr serious game for firefighter training. In *Workshop on Affective Computing and Context Awareness in Ambient Intelligence*, volume 41, page 43.

Backlund, P., Engstrom, H., Hammar, C., Johannesson, M., and Lebram, M. (2007). Sidh–a game based firefighter training simulation. In *2007 11th International Conference Information Visualization (IV'07)*, pages 899–907. IEEE.

Bayouth, S. T., Keren, N., Franke, W. D., and Godby, K. (2013). Examining firefighter decision-making: how experience influences speed in process and choice. *International Fire Service Journal of Leadership and Management*, 7:51.

Bowman, D. A., Koller, D., and Hodges, L. F. (1997). Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques. In *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*, pages 45–52.

Brooke, J. (2013). Sus: A retrospective. *Journal of Usability Studies*, 8(2):29–40.

Cagiltay, N. E., Ozcelik, E., and Ozcelik, N. S. (2015). The effect of competition on learning in games. *Computers & Education*, 87:35 – 41.

Cater, J. P. (1994). Smell/taste: odors in reality. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, volume 2, pages 1781–vol. IEEE.

Cha, M., Han, S., Lee, J., and Choi, B. (2012). A virtual reality based fire training simulator integrated with fire dynamics data. *Fire Safety Journal*, 50:12–24.

Engelbrecht, H., Lindeman, R., and Hoermann, S. (2019). A swot analysis of the field of virtual reality for firefighter training. *Frontiers in Robotics and AI*, 6:101.

Garris, R., Ahlers, R., and Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4):441–467.

Heldal, I., Wijkmark, H., and Pareto, L. (2016). Simulation and serious games for firefighter training: Challenges for effective use. *NOKOBIT*, 24:12.

Iso, W. (1998). 9241-11. ergonomic requirements for office work with visual display terminals (vdts). *The international organization for standardization*, 45.

Kalawsky, R. S. (1999). Vruse – A computerised diagnostic tool for usability evaluation of virtual/synthetic environment systems. *Applied Ergonomics*, 30(1):11–25.

Lee, J., Cha, M., Choi, B., and Kim, T. (2010). A team-based firefighter training platform using the virtual environment. In *Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry*, pages 299–302. ACM.

Nielsen, J. (1993). *Usability Engineering*. Academic Press.

Nilsson, N. C., Serafin, S., Steinicke, F., and Nordahl, R. (2018). Natural walking in virtual reality: A review. *Computers in Entertainment (CIE)*, 16(2):8.

Poupyrev, I. and Ichikawa, T. (1999). Manipulating objects in virtual worlds: Categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages & Computing*, 10(1):19–35.

Rebenitsch, L. and Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality*, 20(2):101–125.

Rojas-Drummond, S. and Mercer, N. (2003). Scaffolding the development of effective collaboration and learning. *International journal of educational research*, 39(1-2):99–111.

Tate, D. L., Sibert, L., and King, T. (1997). Using virtual environments to train firefighters. *IEEE Computer Graphics and Applications*, 17(6):23–29.