在虚拟现实培训中提供关键刺激用于决策制定:一个消防员培训场景的评估研究

Delivering Critical Stimuli for Decision Making in VR Training: Evaluation Study of a Firefighter Training Scenario

Pedro Monteiro, Miguel Melo, António Valente, José Vasconcelos-Raposo, and Maximino Bessa

***Abstract*—The goal for a virtual reality (VR) training system is to enable trainees to acquire all the knowledge they need to perform effectively in a real environment. Such a system should provide an experience so authentic that no further real-world training is necessary, meaning that it is sufficient to train in VR. We evaluate the impact of a haptic thermal stimulus, which is of paramount importance to decision making, on trainees performance and knowledge acquisition. A thermal device was created to deliver the stimulus. As a proof of concept, a procedure from firefighter training is selected, in which sensing the temperature of a door with one's hand is essential. The sample consisted of 48 subjects divided among three experimental scenarios: one in which a virtual thermometer is used (visual stimulus), another in which the temperature is felt with the hand (thermal stimulus) and a third in which both methods are used (visual + thermal stimuli). For the performance evaluation, we measured the total time taken, the numbers of correctly executed procedures and identified neutral planes, the deviation from the target height, and the responses to a knowledge transfer questionnaire. Presence, cybersickness, and usability are measured to evaluate the impact of the haptic thermal stimulus. Considering the thermal stimulus condition as the baseline, we conclude that the significantly different results in the performance among the conditions indicate that the better performance in the visual-only condition is not representative of the real-life performance. Consequently, VR training applications need to deliver the correct stimuli for decision making.**

***Index Terms*—Firefighters, haptic cues, multisensory displays, thermal, training, virtual reality (VR).**

I. Introduction

EVERY day, we rely on the capabilities of highly trained professionals, especially first responders, who need to act promptly in an emergency. These professionals are required to go through many hours of training in real life, such as scenarios prepared to simulate, ideally, all the conditions they might encounter in their jobs. Time, cost, and safety [1] as well as ethics [2] are the greatest challenges faced when creating training scenarios; for example, realistic and complex scenarios sometimes require expensive disposable props, they may take many hours to create and they may expose trainees to threats to their lives. From a trainee point of view, traditional learning mechanisms [3], [4] create a gap between real-life situations and training examples, such that $\approx 35\%$ of industry accidents result from poor training [5]. Additionally, rich training scenarios have proven to be beneficial [4], [5]. Thus, the quality, consistency, and efficiency of training have a significant impact on how well trainees learn procedures [1].

The use of virtual reality (VR) for training is not a recent concept [6], although its use for this purpose continues to be discussed [7]. VR refers to an immersive system that offers the ability to surround the user with a highly realistic and interactive virtual environment (VE) [8], [9]. Although the creation of a specific real-life scenario for VR [5] can be expensive in terms of both time and money [3], this process is a one-time task and offers advantages such as reusability of the VE and full control of the VE and its variables; thus, for repeated tasks, the costs are lower [10]. Because VR scenarios pose no physical risks to either trainees or training instruments [11], trainees can learn and attempt procedures as much as needed and have the freedom to experiment with different scenario outcomes [4] and to fail and be corrected without suffering any harm [10]. Moreover, VR can induce the same stress levels required for training as real-world scenarios [12].

A VR experience is usually classified and characterized based on the concepts of immersion, presence, and cybersickness. Presence is a mental state that portrays how immersed a user is in an application [8], and it is closely related to the amount of output generated by the VR application, i.e., the number of stimuli that it provides to the user in order to mentally involve him/her in the experience so that he/she feels present in the VE [9], [13]. In this regard, presence is a state of mind, while immersion is a technological factor [13]. As such, it is desirable to provide an experience that is both interactive and coherent between the VE and the stimuli that the user feels in reality [14]. Cybersickness, on the other hand, is related to how pleasant a VR experience is to its user [15]. The amount of cybersickness experienced by

Manuscript received December 16, 2019; revised April 16, 2020, June 19, 2020, and August 28, 2020; accepted October 4, 2020. Date of publication November 3, 2020; date of current version March 12, 2021. This work was supported by the ERDF—European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation—COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, in part by the FCT—Fundação para a Ciência e a Tecnologia within project POCI-01-0145-FEDER-028618 entitled PERFECT—Perceptual Equivalence in virtual Reality For authEntiC Training. This article was recommended by Associate Editor M.-L. Shyu. *(Corresponding author: Pedro Monteiro.)*

Pedro Monteiro and Miguel Melo are with the INESC TEC, 4200-465 Porto, Portugal (e-mail: monteiro.p@outlook.pt; mcmelo@inectec.pt).

António Valente, José Vasconcelos-Raposo, and Maximino Bessa are with the INESC TEC, 4200-465 Porto, Portugal, and also with the Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal (e-mail: avalente@utad.pt; jvraposo@utad.pt; maxbessa@utad.pt).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2020.3030746

the user is heavily influenced not only by factors intrinsic to the user but also, more importantly, by the coherence of the stimuli the application provides. While some people tend to naturally be more susceptible to cybersickness symptoms [16], it is the sensory conflict between what they see, hear and feel that is the main culprit of symptoms such as nausea or disorientation [17]. Because these concepts are important for describing what a user feels during a VR experience, they should be studied when evaluating any VR application.

It is known that an increase of coherent sensory stimuli leads to an increase in immersion and, consequently, in the sense of presence [13]. Thus, multisensory stimuli are of paramount importance for establishing presence [14]. However, merely increasing the number of stimuli is not enough. If the stimuli that are added are inconsistent with the virtual experience, this inconsistency can increase the cognitive load for users and, thus, lead to a negative effect on the sense of presence [18], [19]. Therefore, it is also important to maintain a balance among the stimuli offered to the user, application interactivity, and control [14]. Some examples of multisensory stimuli in VR beyond visual stimuli [20] include proprioceptive feedback from room-scale user tracking, spatial sound [21], and haptic stimuli [22], [23] such as wind, warmth, and vibration.

In conjunction with the Cruz Verde firefighters brigade, who supervized the experimental implementation and which can widely benefit from such training methodologies [24], the research team chose a real-world training scenario focusing on neutral plane identification for virtual implementation and study. In this scenario, the trainee is required to use his/her ability to detect a significant increase in the temperature of a closed door. The height at which this temperature difference exists indicates that only below that height is it safe to enter the closed space beyond, as above that height, the temperature will compromise the firefighter's physical integrity (please refer to Section III-B for a detailed description of the procedure). Due to the nature of the procedure, thermal cues are important. A previous pilot study by the authors [25] identified key aspects of this scenario, such as the technological options for properly tracking the hands of the user, the navigation methodology, and the threshold to mark a door as correctly signaled. The VR simulator used in this study improved on those aspects, providing not only a more realistic and complete training environment, but also, more importantly, an updated interaction method with the use of room-scale tracking for user navigation and more precise hand tracking. The main goal of this study was to understand the importance of delivering critical stimuli for decision making in VR training. Thus, the performance of the subjects when deciding whether it is safe to enter a room using only a haptic thermal stimulus (baseline) was compared against the performance when using only a visual stimulus of the kind that is usually used in VR scenarios, as thermal devices are not available. As a secondary goal, we aimed to study the performance of the subjects when using both stimuli, haptic and visual, to perform the task. Additionally, we evaluated user performance during the experiment to better understand the applicability of the thermal stimulus delivery device and this VR simulator for training. We further evaluated our simulator considering standard VR metrics such as presence, cybersickness, and user satisfaction. To conduct the study, a thermal prototype was developed, and the details of its development are also presented.

II. Related Work

A. Disaster Training in VR

There are many examples and studies of VR as a training tool. For first responder training, in [26], a VR application for firefighters that allows users to navigate using their bodies, thus requiring physical effort, was presented. The application required users to scan various locations and evacuate victims. The results showed that despite the learning curve for use of the simulator, the performance goals were met, and high levels of enjoyment were reported, indicating that the simulator showed great potential for knowledge transfer. Other studies [27], [28] have proposed VR applications with realistic parameters for fire simulation, allowing trainees to learn and perform standard operation procedures. However, from a firefighter and paramedic perspective [1], the visual fidelity of the VE should be improved to ensure that these simulators can be used for effective training. The authors also stated these professionals have a limited number of times to practice their skills and that VR could be the solution.

Some studies [4], [5] have discussed the use of VR for the disaster training of field operators based on a VE that uses real data from a plant to generate different outcomes. The results showed that trainees' performance increased significantly compared to those who learned using only traditional methods (i.e., classroom and images) since VR can provide a large amount of contextualized information, giving trainees the opportunity to explore different procedures and see their outcomes. VR for the training of construction workers on risky procedures was also studied [3] and found to yield better results than traditional methods regarding trainees' ability to correctly assess risks. Because the VE allowed the trainees to perceive the risks first-hand based on their own decisions, they retained the information for longer than those taught conventionally. Additionally, the immersive nature of the VR experience naturally made the trainees more interested in and receptive to the information provided by the simulator.

In a comparison of various teaching techniques (i.e., videos, a written manual, and VR) for earthquake safety [29], the results similarly showed that VR was most effective. Although the application was presented as realistic, this was true only in relation to the visual and auditory stimuli; no haptic feedback was given. Overall, the results of these previous studies show the advantages of VR technology, with a focus on how realistic a VE can be and the need to faithfully represent the situations trainees might face. In [10], the authors stated that VR training can produce the same results as training in real life. Although most studies have featured VEs with high realism, only audio-visual systems have been used.

B. Thermal Stimuli in VR

The use of thermal stimuli in VR is not a frequent topic of research, especially in training applications, most of which use an audio-visual approach. Nevertheless, some studies [30], [31] have proposed thermal models for VR based on how human skin reacts to a thermal stimulus, as such stimuli are useful for the identification of different kinds of objects based on their thermal properties [32].

In a previous study [23], [33], a multisensory CAVE capable of delivering heat using infrared lights was proposed and developed. It was found that users perceived the existence or nonexistence of warmth after 2.5s and 2.0s following the lamps being switched ON and OFF, respectively. From presence and immersion questionnaires, it was found that the quality of heat intensity and direction was good and that the sense of presence increased with this stimulus, proving that it can serve as an effective addition to visual stimuli.

In the work described in [34] and [35], a ventilator and infrared lamps were used to deliver warm and cold stimuli to large areas, and a Peltier device was used to deliver warm and cold stimuli to the hand. Early results suggested that these devices could serve as the basis of a complete thermal feedback system capable of delivering sensations of comfort, warning of proximity to a thermal source, and thermal sensory substitution, among others. Additionally, the Peltier device could quickly change the user's skin temperature.

To assess the effect of temperature on user performance in a maze task with the presentation of a warm stimulus when the user was on the correct track and a cold stimulus otherwise, a study was conducted using Peltier devices [36]. Due to the device actuation time, a narrow temperature range of 29 to 35 °C was used. The results showed that the temperature stimuli were effective in increasing performance. The authors also noted how important it is to keep the skin temperature neutral when the presentation of a stimulus is not intended and that it is essential to consider the actuation times of these devices, which could lead to users confusing the intended temperature signals.

准实验设计，以定量为重点的横断面研究。采用的抽样技术是非概率方便抽样方法

III. METHODS

The adopted methodology consisted of a quasi-experimental design, cross-sectional study with a quantitative focus. The sampling technique used was the nonprobabilistic convenience sampling procedure.

A. Sample

The experiment was performed by 48 participants (24 men and 24 women) aged 18 to 28 ($M = 21.38,\ \mathrm{SD} = 2.34$) recruited at the university where the experiments took place. The sample was composed of students eligible to enroll in firefighter training and become certified for the job. The participants were divided into the following three groups: the visual group ($N = 16$) was presented with a visual thermometer next to the virtual hand, on which the temperature was shown (see Fig. 1); the haptic group ($N = 16$) only felt the temperature on the hand, without any visual guidance; and the visual + haptic (V and H) group ($N = 16$) both saw the thermometer and felt the temperature on the hand. Each group contained an equal number of participants of each gender.

Fig. 1. Beginning of the tutorial with the virtual instructor (left). Participant performing the procedure on a door during the tutorial in the visual condition (right). The red bar represents the marked height of the neutral plane. The green panel at the bottom of the door instructs the participant regarding where on the door to start checking the temperature (shown during the tutorial). The visual thermometer consists of a bar above the hand that shows the temperature (the area of the colored portion increases with the temperature and shows a larger stepwise increase when reaching the height of the neutral plane).

B. Materials

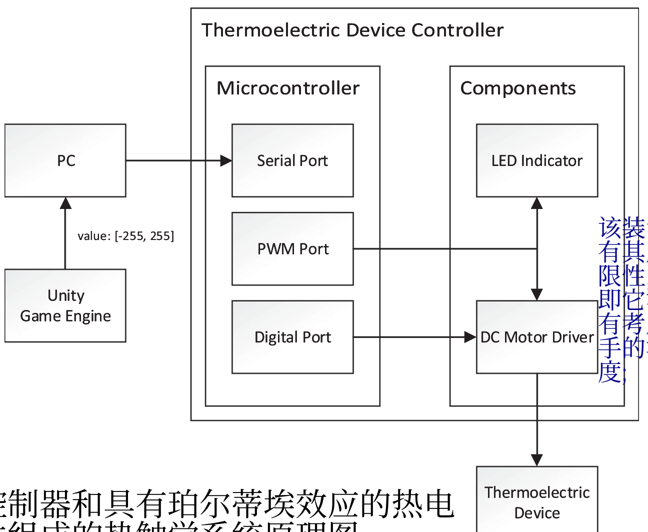
A computer with an Intel i7-6700K CPU and NVIDIA GeForce GTX 1080 GPU was responsible for running the application. The VIVE HMD headset was used to provide the visual stimulus, and its controllers were used to interact with the VE and track the users' hands. This HMD has a 110° viewing angle and a per-eye resolution of 1080×1200 pixels. The audio was delivered through Bose QuietComfort 25 headphones with active noise cancellation.

1) Training Scenario: The real-world training scenario used was the procedure for neutral plane identification in an indoor environment. When approaching a door, the firefighter is required to check whether there is fire inside the room and at which height it is safe to enter the room. In this scenario, fire particles and smoke tend to rise with hot air inside the room, beyond the door, heating the upper part of the door. The firefighter must place him/her next to the door and, with the back of his/her nondominant hand, sense at which height an accentuated increase in temperature occurs. That height is the neutral plane, and it is safe to enter the room below that height. This is a risky procedure on which a trainee needs to be trained multiple times.

The VR simulator was developed under the supervision of the Cruz Verde firefighters brigade and their commander, and as such, the simulator was required not only to teach the neutral plane concept but also show participants the steps they would need to perform to be successful in the training task. This task required participants to learn the neutral plane identification procedure, in which sensing the temperature of a door with one's hand is critical to identify whether it is safe to enter a room.

A realistic VE was created for the required conditions and procedure. This VE consists of two scenarios: a warehouse (see Fig. 1), in which a tutorial is presented to teach the user the training task, and a two-store building (see Fig. 2), in which the procedure is required to be executed multiple times. Both scenarios are filled with props that typically characterize such environments, and the avatars wear the uniform of Portuguese fire brigades.

Fig. 2. View of the VE during the experiment (left). Participant performing the procedure on a door during the experiment in the haptic condition (right). The red bar represents the marked height of the neutral plane.



该装置有其局限性，即它没有考虑手的温度；

由控制器和具有珀尔帖埃效应的热电器件组成的热触觉系统原理图

Fig. 3. Schematic of the thermal haptic system composed of a controller and a thermoelectric device with a Peltier effect.

Movement through the VE is accomplished by means of teleportation between doors, with a teleport target placed in front of each door but not immediately adjacent to it. Once near a door, the user can walk in the real space to position him/her on the correct side of the door and then perform the required door verification procedure.

2) Thermal Haptic Device: A device was created to deliver the thermal haptic stimulus, with the following requirements:

1) can warm up and cool down as needed;
2) has a short response time;
3) does not compromise the physical integrity of the users;
4) is as nonintrusive as possible.

The developed device system is composed of a controller and a thermoelectric device with a Peltier effect. Fig. 3 shows a schematic of the system. As the controller, an Arduino Uno is used. This device receives a value ranging from -255 (maximum cold) to 255 (maximum heat), where a value of 0 is used to turn OFF the device. Based on the sign of the received value, the controller sends a digital value to an L298N DC motor driver to define the polarity of the motor output. The absolute value defines the electric potential to be applied to the thermoelectric device, with a higher value resulting in more voltage applied to the device and, consequently, a stronger heat or cold stimulus, and is sent through a pulsewidth modulation port of the Arduino. The controller also has an LED indicator that shows the intensity of the value applied to the thermoelectric device.

The thermoelectric device is a 12V Peltier cell, which produces heat and cold simultaneously on opposite sides. Changing the polarity of the voltage controls which side becomes warm or cold. During calibration tests, we observed that the device responds with less than 1s between when the intensity value is sent and when it is applied.

An informal evaluation was undertaken to assess the effectiveness of the developed thermal device by asking 16 independent participants to wear it and identify whether they could detect temperature changes from a neutral state (ambient temperature) to a hot state. This change was manually controlled by the researchers and was not presented in the training environment. Each participant was asked to identify whether he/she felt the change in temperature, which was applied after a random amount of time, in each of six trials. It was verified that in 81.3% (78 of 96) of the attempts, users were able to feel the temperature changes on the device and to perceive the change in temperature from a neutral state to a hot state.

It must be noted that the device has the limitation that it does not take the hand temperature into account; therefore, instead of ensuring that there was a significant increase in the temperature on the hand, only a significant increase was applied to the device. Because temperature perception can change with the temperature of the hand and habituation to heat can occur, this may explain the $\approx 20\%$ failure rate; thus, during the training exercise, after checking each door, we apply a small amount of cold to the hand to counter the effects of warming the device and hand.

For the participants' safety, the maximum temperature applied to the hand is 50 °C, which is applied for a maximum of 5 s, after which the device will turn OFF. This temperature was chosen by the firefighter commander to be high enough to elicit a sense of danger while still being multiple orders of magnitude below the temperatures encountered in a real fire scenario [37]. Fig. 4 shows how the device is controlled.

The cell was firmly secured to each participant's hand with the help of a glove, leaving the hand free to interact with other equipment. The additional thermal device cable was attached to the existing cable connecting the HMD to the computer in order to reduce the impact of the cabling on the usability of the equipment (see Fig. 5).

C. Independent Variable

In this study, we varied the sensory cues provided for the participants to identify the neutral plane.

1) *Haptic (thermal):* In the real-life training scenario, firefighters use their hands to assess the temperature of a door. For this reason, this condition is the baseline that is most representative of the real world, since here, the participants also assessed the temperature with their hands using only the thermal device (Section III-B).

设备的控制方式
看不懂

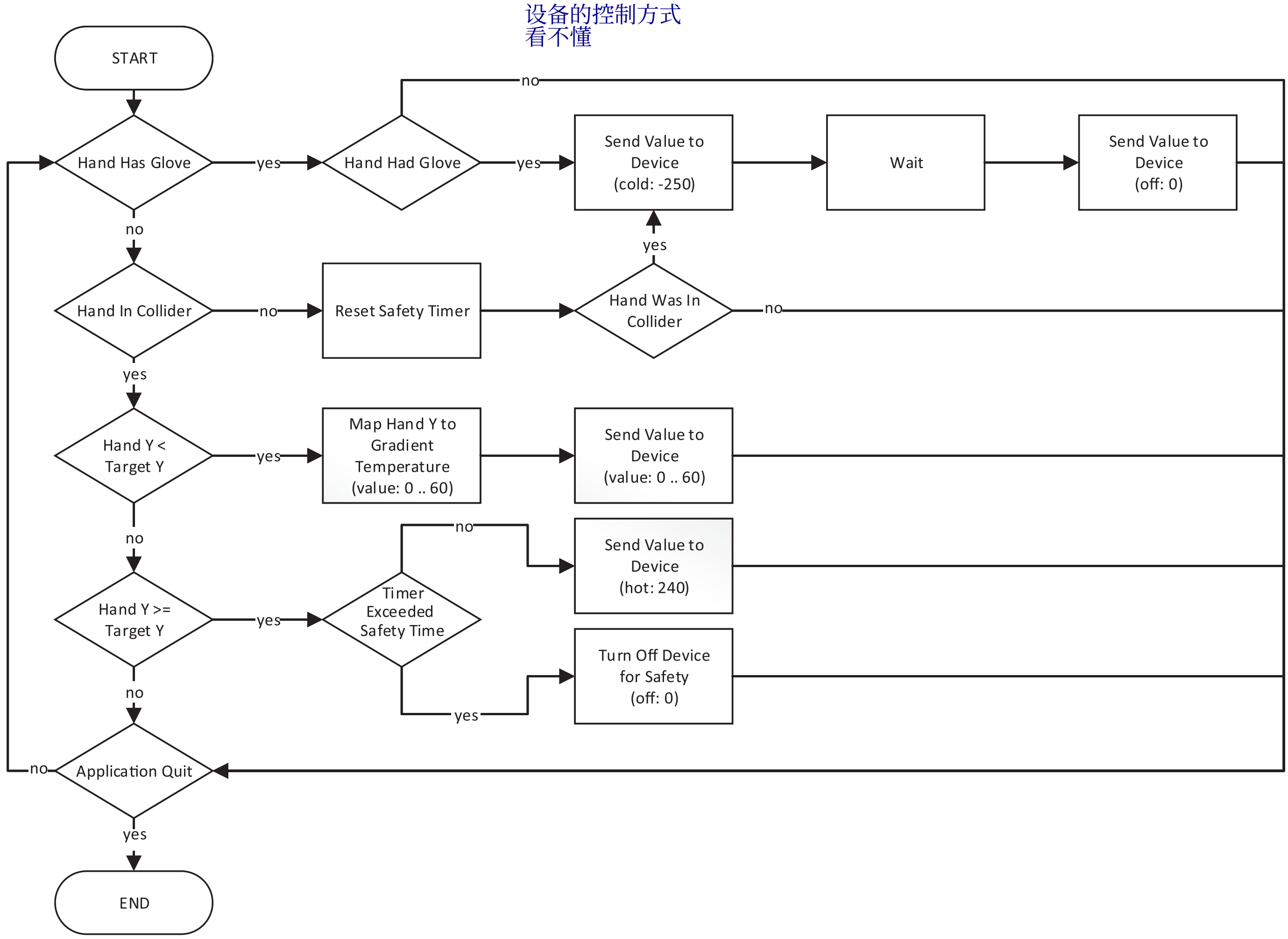


Fig. 4. Application logic for controlling the device temperature. The application checks various conditions to determine the temperature to be applied. If the hand is below the target height, a warmer temperature is applied the closer the hand is to the target height, with ambient temperature at the bottom of the door and a temperature of 30 °C close to the neutral plane height. Once above the target height, the maximum temperature (50 °C) is applied.

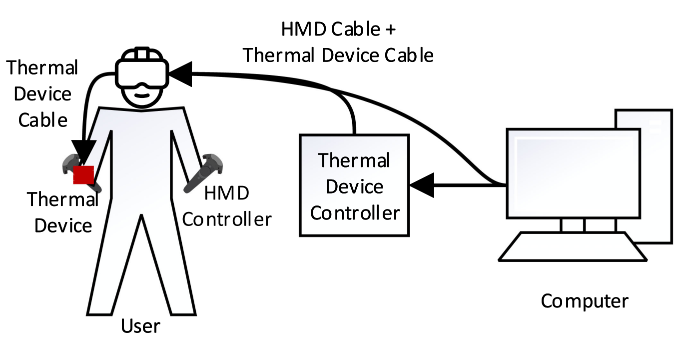


Fig. 5. Overview of the VR system, including the thermal device controller, the device itself (red) and the HMD. The HMD controllers allow hand tracking and are part of the VR setup, connected via bluetooth. The arrows represent physical cables, being that the cable that goes from the thermal device controller to the thermal device is managed through the HMD cable.

2) *Visual:* Because most VR applications only use audio-visual stimuli, for this condition, we used an equivalent means of checking the temperature in a visual manner—A visual temperature bar (virtual thermometer) from which the participants could check the temperature of the door (see Fig. 1).

3) *Visual + Haptic (V and H):* Here, participants could use both methods to check the temperature of the door at the same time. This condition allows us to understand which stimulus the participants will tend to lean on through comparison with the other conditions. Moreover, from a training standpoint, it is important to provide new trainees with a visual way to associate the neutral plane (easily seen from the thermometer) with a change in temperature (felt on the hand).

D. Dependent Variables and Instruments

The VR application recorded the time each participant took to complete the procedure for each door and the deviation between the height where the participant identified the neutral plane and its correct location. It also recorded how many doors were correctly identified and for how many doors the participant performed the correct procedure. These metrics contributed to evaluating the performance of the participants.

A simple sociodemographic questionnaire was used to determine the sample characteristics. To assess each participant's levels of presence, cybersickness, system usability, and user satisfaction, we used scientifically validated questionnaires.

1) *Presence:* The Portuguese version of the Igroup Presence Questionnaire (IPQp) [38] was used to provide insight into presence in terms of spatial presence (the sense of being physically in the VE), involvement (the attention devoted to the VE), and experienced realism (the subjective realism of the VE). The values range from 1 to 5. Higher values are better.
2) *Cybersickness:* The Simulator Sickness Questionnaire (SSQ) [39] was used to provide information about the overall cybersickness values related to nausea, oculomotor discomfort and disorientation symptoms. The final scores start from zero, and the closer they are to this value, the better. The different subscales have maximum values of 200.34, 159.18, 292.32, and 2437.88 for nausea, oculomotor discomfort, disorientation, and overall cybersickness, respectively.
3) *Usability:* The system usability scale (SUS) [40] was used to measure the system's usability. The final SUS score is assessed on a scale of 0 to 100 and is interpreted in terms of percentiles, as described in [41].
4) *User satisfaction:* The after-scenario questionnaire (ASQ) [42] was used to assess the users' satisfaction with the experiment. The final score ranges from 0 to 10, with a higher value representing higher user satisfaction.

A knowledge transfer questionnaire (KTQ) was elaborated to evaluate the knowledge retained by the participants. This questionnaire was developed in cooperation with the *Cruz Verde* Firefighters Brigade and consists of six multiple-choice questions regarding the main concepts of neutral plane identification. This questionnaire was used to assess whether exposure to different stimuli would change how users assimilated the procedure concepts, not to approve or certify the users.

E. Procedure

The procedure and experimental design were validated beforehand by the Cruz Verde firefighters brigade commander, who is also a certified firefighting trainer. All experiments were carried out in a laboratory environment with the required apparatus, where external variables were controlled.

The participants were received at the experimental site and completed a consent form and sociodemographic questionnaire. Then, the participants were informed about what they were about to do, and the full experimental procedure was demonstrated in the real world by the researcher, who also explained the concept of the neutral plane. All participants were informed that the thermoelectric device could accumulate heat as the thermal stimulus was delivered and that if it became too hot, they should inform the researcher to trigger a brief pause in the experiment to cool the device. After this briefing, the participants were equipped with the developed device, HMD, headphones, and controllers. To avoid bias that could be introduced using the developed device, regardless of the experimental condition, all participants were equipped with the complete apparatus.

The VR application was started, and the participants were presented with a tutorial in which a virtual firefighter guided them to learn the required buttons and try performing the procedure at two doors before we started collecting data (see Fig. 1). In this tutorial, the participants could also ask the researcher for more precise instructions if needed. At the end, the researcher asked the participants if they had any questions, and if not, they were ready to start the experiment and begin data gathering.

The users were then relocated in the VE to a two-story building with eight doors, in which they had to correctly identify the neutral plane by checking each door's temperature (see Fig. 2). This test environment was the same for all participants. The procedure for each door was as follows:

1) Move next to the door;
2) Move to the side of the door, facing it sideways;
3) Remove the virtual glove on the nondominant hand, by touching the glove with the other hand and pressing a button;
4) Start sensing the temperature of the door, with the back of the nondominant hand, from bottom to top without touching it;
5) Identify at which height the temperature rises significantly, by pressing a button with the hand used to check;
6) Put on the glove again and move to the next door.

The eight doors the participants had to check were grouped in pairs, corresponding to four different types of doors regarding the heights of their neutral planes. The neutral plane heights for the first three door types were $T1_{\mathrm{height}} = 0.7$ m, $T2_{\mathrm{height}} = 1.3$ m, $T3_{\mathrm{height}} = 1.7$ m; since the fourth type corresponded to safe doors (with no neutral plane), we defined the associated height as the maximum door height, $T4_{\mathrm{height}} = 2.2$ m. The participants were allowed an error tolerance of 30 cm (a distance validated by the firefighter trainers) within which the identified height would be considered correct (except for $T4$), with a third of this tolerance lying below the correct height and two thirds lying above because of the device activation delay. The neutral plane height schematics for each door type are shown later in the results section, where we explore the participants' performance. Additionally, we considered the procedures to be correctly executed only when the participant started checking the door from the very bottom, as required by the training procedure.

After checking all the doors, the participants had finished the experiment, and the equipment was removed. The participants completed the IPQp, SUS, ASQ, and KTQ and then proceeded to leave the experimental site. The whole experimental procedure had an average duration of 25 min.

F. Statistical Procedures

The Shapiro–Wilk test was used to assess the data normality. It was verified that the data did not follow a normal distribution ($p < 0.05$), and as such, nonparametric tests were used to compare the different conditions. The level of significance was maintained at 95% (alpha level of 0.05) for all statistical tests, namely, Kruskal–Wallis H tests to compare the mean ranks of the dependent variables for each experimental condition and pairwise comparisons using Dunn's [43] procedure with Bonferroni correction for multiple comparisons (adjusted p-values

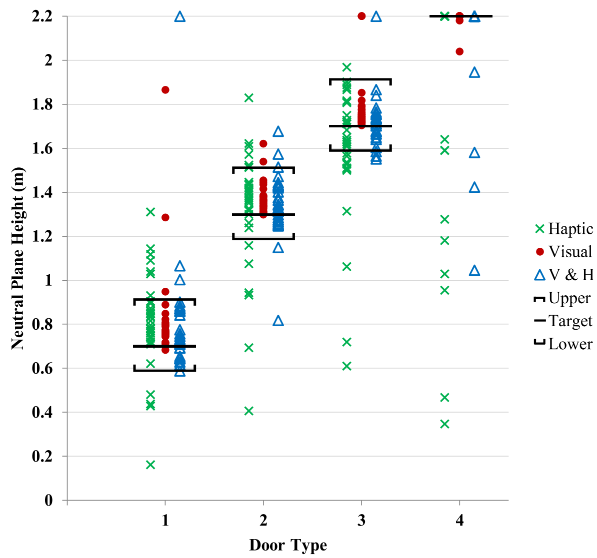


Fig. 6. Plot of the neutral plane heights identified by the participants in each condition by door type. For each door type, the target height (actual neutral plane height) and corresponding lower and upper bounds of the error tolerance are shown. A value of 2.2 represents a nonexistent neutral plane.

are presented), when the test revealed statistically significant differences. All statistical procedures were performed using the IBM SPSS Statistics (version 24.0) software.

IV. Results

Regarding the total time (in seconds) the participants spent performing the experiment, we found statistically significant differences among the three conditions, $\chi^2(2) = 10.240,\ p = 0.006$. Subsequently, a post hoc analysis revealed statistically significant differences in execution time between the visual (mean rank = 15.94) and haptic (mean rank = 31.56) ($p = 0.005$) conditions, but not between the V and H group (mean rank = 26.00) and the other two groups. The means show that the participants spent more time completing the experiment in the haptic condition ($M = 195.911$ s), followed by the V and H condition ($M = 154.114$ s), with the visual condition ($M = 121.292$ s) being the fastest.

From the neutral plane heights identified by the participants (see Fig. 6), we calculated the mean deviations between those heights and the target heights. When comparing these mean deviations among the experimental conditions, statistically significant differences were found, $\chi^2(2) = 8.958,\ p = 0.011$. A post hoc analysis showed statistically significant differences in mean deviation between the haptic group (mean rank = 33.00) and both the visual group (mean rank = 19.44, $p = 0.018$) and the V and H group (mean rank = 21.06, $p = 0.048$). No significant difference was found between the visual and V and H groups ($p = 1.000$). The mean values show that the participants deviated more from the target heights in the haptic group ($M = 0.271$ m), followed by the V and H group ($M = 0.097$ m), with the visual group having the lowest deviation ($M = 0.071$ m).

When comparing the numbers of times participants correctly identified the neutral plane (see Table I), statistically significant differences were found among the study conditions, $\chi^2(2) = 10.372,\ p = 0.006$. A post hoc analysis revealed statistically significant differences in the correct identification of the neutral plane between the visual and haptic conditions ($p = 0.007$), but not when comparing the V and H group with the visual group ($p = 1.000$) or the haptic group ($p = 0.051$).

As expected by observing the values (see Table I), no statistically significant differences were found when comparing the study conditions regarding the number of times the procedure was correctly executed, $\chi^2(2) = 0.511,\ p = 0.774$.

Similarly, no statistically significant differences were found when comparing the numbers of correct answers on the KTQ, $\chi^2(2) = 1.887,\ p = 0.389$.

Regarding presence levels (see Table II), the visual group had generally higher scores than the other groups. Despite this, no significant differences were found in any of the IPQp subscales, with the statistical test results being $\chi^2(2) = 1.509,\ p = 0.470$ for spatial presence, $\chi^2(2) = 0.340,\ p = 0.844$ for experienced realism, $\chi^2(2) = 0.639,\ p = 0.727$ for involvement and $\chi^2(2) = 1.166,\ p = 0.558$ for overall presence.

For cybersickness symptoms (see Table III), the values on each SSQ subscale were zero or close to zero, and no statistically significant differences were found in terms of nausea ($\chi^2(2) = 2.621,\ p = 0.270$), oculomotor discomfort ($\chi^2(2) =$

TABLE I
Means, Medians, and Mean Ranks of the Identified Dependent Variables for Each Experimental Condition

	Condition	*M*	*Mdn*	*Mean Rank*
# of Correctly Identified Neutral Planes	Haptic	5.000	5.000	15.940
	Visual	7.313	8.000	30.380
	V & H	6.875	7.500	27.190
# of Correctly Executed Procedures	Haptic	7.688	8.000	25.690
	Visual	7.688	8.000	23.280
	V & H	7.688	8.000	24.530
# of Correct Answers on the KTQ	Haptic	5.313	6.000	21.500
	Visual	5.750	6.000	25.500
	V & H	5.750	6.000	26.500
SUS	Haptic	74.375	78.889	22.220
	Visual	82.778	86.112	27.340
	V & H	80.278	77.223	23.940
ASQ	Haptic	5.979	6.333	18.810
	Visual	6.583	7.000	30.720
	V & H	6.336	6.357	23.970

TABLE II
IPQp Subscale Results for Each Experimental Condition

IPQp Subscale	Condition	*M*	*Mdn*	*Mean Rank*
Spatial Presence	Haptic	4.025	4.167	23.280
	Visual	4.198	4.167	27.940
	V & H	4.071	4.000	22.280
Experienced Realism	Haptic	2.875	3.000	23.250
	Visual	2.953	3.000	24.190
	V & H	3.000	3.125	26.060
Involvement	Haptic	3.375	3.625	24.880
	Visual	3.453	3.625	26.250
	V & H	3.188	3.250	22.380
Presence	Haptic	3.511	3.500	23.440
	Visual	3.629	3.679	27.530
	V & H	3.513	3.495	22.530

TABLE III
SSQ SUBSCALE RESULTS FOR EACH EXPERIMENTAL CONDITION

SSQ Subscale	Condition	*M*	*Mdn*	*Mean Rank*
Nausea	Haptic	7.751	4.770	28.340
	Visual	4.770	0.000	22.940
	V & H	5.366	0.000	22.220
Oculomotor Discomfort	Haptic	10.896	0.000	28.250
	Visual	3.316	0.000	23.720
	V & H	3.316	0.000	21.530
Disorientation	Haptic	11.310	0.000	27.970
	Visual	3.480	0.000	23.410
	V & H	3.480	0.000	22.130
Cybersickness	Haptic	11.454	3.740	29.780
	Visual	4.441	0.000	23.160
	V & H	4.675	0.000	20.560

$2.850,\ p = 0.241$), disorientation ($\chi^2(2) = 2.858,\ p = 0.240$), or overall cybersickness ($\chi^2(2) = 4.275,\ p = 0.118$).

When comparing the system usability scores (see Table I) among the conditions, no statistically significant differences were found in the SUS questionnaire scores, $\chi^2(2) = 1.113,\ p = 0.573$. However, the ASQ scores (see Table I) for user satisfaction were significantly different among the conditions, $\chi^2(2) = 6.144,\ p = 0.046$. A post hoc analysis of the ASQ scores showed that user satisfaction was significantly higher in the visual group than in the haptic group ($p = 0.040$). No significant differences were found between the V and H group and either the haptic group ($p = 0.853$) or the visual group ($p = 0.484$).

V. Discussion

This study aimed to assess how different sensory cues in VR impact the performance and acquired knowledge of trainees, especially when comparing the results obtained in different conditions to the baseline condition (with the most coherent stimulus). We compared the performance of trainees during a firefighter procedure in which, the thermal stimulus is critical. Moreover, we studied the users' mental state (sense of presence) and physical condition (cybersickness symptoms) as well as the system usability and user satisfaction to determine the applicability of the developed VR application for training.

The performance results show the visual condition yielding a better performance than those involving the haptic stimulus. These results can be explained as follows: 1) checking the door temperature visually is an objective task, in which users are watching the temperature bar for the immediate moment at which it reaches a certain threshold value, making the task easier, whereas; 2) checking the door temperature based on physical heat on the hand is a subjective task, in which users will feel the heat to different extents, thus increasing the focus required and the cognitive load of the task and, making it more difficult to identify the temperature threshold. The combination of these factors led to users being both faster and more precise in identifying the neutral plane in the visual condition, while also explaining the false positives for the Type 4 doors.

Participants presented with combined visual and haptic stimuli outperformed those in the haptic condition in terms of their accuracy in identifying the neutral plane. Moreover, except for the variables in which no absolute differences were found, the V and H condition yielded results closer to those in the visual condition, corroborating the conclusion that the visual stimulus was dominant [20]. The V and H participants were slower than the visual participants because they spent more time confirming what they were seeing on the thermometer against what they were feeling on their hands.

Although studies such as [1], [4], [3], [29], and [44] have shown a good performance of VR training simulators using only visual and auditory stimuli, we believe that these stimuli alone are not sufficiently representative of many real-life scenarios. This is in line with studies such as [45] and [46], which refer to the inclusion of coherent stimuli as essential for trainees' performance and the applicability of VR to real-world simulations, and was also corroborated by the commander of the Cruz Verde firefighters brigade who is experienced in firefighter training. This means that even though the conditions with visual cues showed better performance results, they are not truly representative of the real-world condition and can introduce some false confidence in the ability to perform a procedure such as that addressed here. Due to the nature of this exercise, it is critical to ensure correct execution in the real-world environment, and a sense of false confidence can have disastrous results, as it can lead to incorrect decisions and endanger not only the physical integrity of a firefighter but also that of his/her team and of possible victims at the site. In this particular case, rather than simply judging the conditions by directly comparing which yielded best performance results, we should verify how they compare to the baseline condition (the thermal stimulus condition); ideally, there should not be significant differences between the visual or V and H conditions and the baseline condition. If we consider our performance results with the haptic thermal stimulus condition as the baseline, we can assume that training without this stimulus will not adequately prepare firefighters for real-life situations.

Regarding the simulator itself, we registered no significant differences in presence or cybersickness. Based on the observed results and the score ranges for the presence questionnaire, we registered higher than average levels of spatial presence, involvement, and overall presence, which is in line with what is expected of a VR application compared with traditional training tools [3], [4]. The experienced realism results showed the importance of this factor in training applications, with the registered values being lower than those on the other presence subscales. Despite this, these lower values were not strong enough to negatively impact the experiment's training potential, which suggests that even though high realism is necessary [1], the combination of all factors and their coherence can compensate for some deficiencies. Additionally, the presence values contradict the idea that adding coherent multisensory stimuli always correlates with an increase in presence [47], [48]; similar findings were recorded in our preliminary tests [25].

Moreover, the median cybersickness values were zero or very low [15], indicating that the simulator did not induce sensory conflicts [15], even though the users needed to perform many rotations during the experiment to reorient themselves. Since no significant differences were found in the simulator characteristics, we can assume that the system and equipment used did not interfere with the VR experience of the users, indicating that this VR system is a viable tool for training. If differences had been found, this would mean that the VR equipment was inadequate for training.

Our results showed that there were no significant differences in the system usability scores, which were good to excellent [41]. These results are also corroborated by the fact that the participants understood the interaction methods and by the number of correctly executed procedures, since the participants were able to perform the procedure at most doors correctly. Our results did show significant differences regarding user satisfaction, with the visual condition being preferred, although the values for all conditions were relatively high. We theorize that this can be attributed to the increased cognitive load of assessing the temperature from the haptic stimulus and because of the inherent nature of the haptic stimulus, i.e., feeling a significant amount of heat on the most sensitive part of the hand is not a satisfactory experience.

Finally, we should note that our sample consisted mainly of students who were eligible to enroll in a firefighter course, rather than certified firefighters, since it would not make sense to teach this concept to a professional who already knows it. Therefore, the good KTQ results show that the simulator can successfully teach concepts and procedures. This knowledge was correctly used during the experiment (the number of correctly executed procedures was the highest possible) and was retained after the experiment ended (the number of correctly answered questions on the KTQ was the highest possible). These results corroborate those found in [3], with the engagement factor of the VR simulator being positively correlated with the retention of knowledge.

VI. Conclusion

The work presented in this article studied the performance of trainees in VR when exposed to different stimuli, as well as the applicability of a haptic thermal device and a VR simulator for firefighter training. Depending on the experimental condition, participants could use their vision to check the hand temperature or feel the temperature on the backs of their hands to identify the neutral planes of a series of doors.

One major conclusion of this work is that a certain level of authenticity is required so that VR training tools can prepare trainees for the real-world performance. By analyzing the performance results, especially how accurately the participants correctly identified the neutral planes, we can observe that both conditions with the visual stimulus produced significantly different results compared to the reference haptic condition. No differences were found when comparing visual + haptic stimulus with the reference haptic condition, namely in the time taken to complete the experiment and the number of times that they identified the neutral plane correctly. This finding indicates that the visual condition is not able to deliver the same training results that the baseline condition that is closer to the real word, which indicates that the visual condition may be misleading for the training evaluation as training results would give a false sensation of readiness. This false sensation is expected because, in the real world, they would not determine the neutral plane as they did in the training condition. Thus, we believe that for VR simulators to be valid alternatives to real-world training, it is critical to present trainees in VR with the same critical stimulus they will face in a real-world scenario.

Although the results also suggest that using both sensory cues simultaneously can provide different training capabilities than using only the real stimulus alone, more studies should be carried out to better understand this possibility, and as such, we recognize the usefulness of augmenting the coherent stimuli with other cues, especially when teaching concepts. Additionally, further studies should be conducted in other areas of application and using other stimuli to better understand how far from the real-world performance VR applications are when multisensory stimuli are needed for decision making.

Moreover, we believe that in order to certify trainees using only VR applications, it is necessary for such applications to produce high-fidelity real-world conditions and stimuli that will enable the suitable measurement of the trainees' performance. Otherwise, the measured performance will not be representative of the performance that trained professionals should achieve in real-life scenarios.

The presence, user satisfaction, and knowledge transference results of this study are in line with those of previous studies showing that VR is a useful tool for training professionals, with the users feeling they were present in the VE, correctly understanding the knowledge the simulator presented, and using that knowledge during the training exercises. These results indicate that VR can be successfully used as a training tool to reduce the costs of creating real scenarios, as once a VE has been created and validated, it can be used to train as many people as desired. Additionally, one major advantage of VR training tools such as the one presented here is that they are always available, unlike practical training scenarios. In fact, this is a limitation of firefighter training, as organizations must coordinate the availability of trainers, training materials, and consumables, and a certain minimum number of trainees such that the costs associated with training are justified.

Acknowledgment

All the works were conducted at INESC TEC's MASSIVE VR Laboratory.

References

[1] A. Mossel, M. Froeschl, C. Schoenauer, A. Peer, J. Goellner, and H. Kaufmann, "VROnSite: Towards immersive training of first responder squad leaders in untethered virtual reality," in *Proc. IEEE Virtual Reality*, 2017, pp. 357–358.

[2] A. Alaraj *et al.*, "Role of cranial and spinal virtual and augmented reality simulation using immersive touch modules in neurosurgical training," *Neurosurgery*, vol. 72, no. suppl_1, pp. A115–A123, Jan. 2013.

[3] R. Sacks, A. Perlman, and R. Barak, "Construction safety training using immersive virtual reality," *Construction Manage. Econ.*, vol. 31, no. 9, pp. 1005–1017, Sep. 2013.

[4] S. Colombo, S. Nazir, and D. Manca, "Immersive virtual reality for training and decision making: Preliminary results of experiments performed with a plant simulator," *SPE Econ. Manage.*, vol. 6, no. 04, pp. 165–172, Oct. 2014.

[5] D. Manca, S. Brambilla, and S. Colombo, "Bridging between virtual reality and accident simulation for training of process-industry operators," *Adv. Eng. Softw.*, vol. 55, pp. 1–9, Jan. 2013.

[6] J. Psotka, “Immersive training systems: Virtual reality and education and training,” *Instructional Sci.*, vol. 23, no. 5/6, pp. 405–431, 1995.

[7] L. P. Berg and J. M. Vance, “Industry use of virtual reality in product design and manufacturing: A survey,” *Virtual Reality*, vol. 21, no. 1, pp. 1–17, Sep. 2016.

[8] W. R. Sherman and A. B. Craig, *Understanding Virtual Reality: Interface, Application, and Design*. San Mateo, CA, USA: Morgan Kaufmann, 2002.

[9] P. Fuchs, G. Moreau, and P. Guitton, *Virtual Reality: Concepts and Technologies*, 1st ed. Boca Raton, FL, USA: CRC Press, Jul. 2011.

[10] G. Sankaranarayanan *et al.*, “Immersive virtual reality-based training improves response in a simulated operating room fire scenario,” *Surgical Endoscopy*, vol. 32, pp. 3439–3449, Jan. 2018.

[11] S. L. Farra, E. T. Miller, and E. Hodgson, “Virtual reality disaster training: Translation to practice,” *Nurse Educ. Pract.*, vol. 15, no. 1, pp. 53–57, Jan. 2015.

[12] R. M. Clifford, S. Jung, S. Hoermann, M. Billinghurst, and R. W. Lindeman, “Creating a stressful decision making environment for aerial firefighter training in virtual reality,” in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, Mar. 2019, pp. 181–189.

[13] M. Slater and S. Wilbur, “A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments,” *Presence: Teleoperators Virtual Environ.*, vol. 6, no. 6, pp. 603–616, Dec. 1997.

[14] T. B. Sheridan, “Musings on telepresence and virtual presence,” *Presence: Teleoperators Virtual Environ.*, vol. 1, no. 1, pp. 120–126, Jan. 1992.

[15] J. J. LaViola, “A discussion of cybersickness in virtual environments,” *ACM SIGCHI Bull.*, vol. 32, no. 1, pp. 47–56, Jan. 2000.

[16] S. Davis, K. Nesbitt, and E. Nalivaiko, “A systematic review of cybersickness,” in *Proc. Conf. Interactive Entertain.*, 2014, pp. 1–9.

[17] S. V. G. Cobb, S. Nichols, A. Ramsey, and J. R. Wilson, “Virtual reality-induced symptoms and effects,” *Presence: Teleoperators Virtual Environ.*, vol. 8, no. 2, pp. 169–186, Apr. 1999.

[18] P. G. de Barros and R. W. Lindeman, “Performance effects of multi-sensory displays in virtual teleoperation environments,” in *Proc. 1st Symp. Spatial User Interact.*, 2013, pp. 41–48.

[19] M. Bessa, M. Melo, A. A. de Sousa, and J. Vasconcelos-Raposo, “The effects of body position on Reflexive Motor Acts and the sense of presence in virtual environments,” *Comput. Graph.*, vol. 71, pp. 35–41, Apr. 2018.

[20] G. Bruder, F. Steinicke, B. Bolte, P. Wieland, H. Frenz, and M. Lappe, “Exploiting perceptual limitations and illusions to support walking through virtual environments in confined physical spaces,” *Displays*, vol. 34, no. 2, pp. 132–141, Apr. 2013.

[21] D. Narciso, M. Bessa, M. Melo, A. Coelho, and J. Vasconcelos-Raposo, “Immersive 360° video user experience: Impact of different variables in the sense of presence and cybersickness,” *Universal Access Inf. Soc.*, vol. 18, pp. 77–87, Oct. 2017.

[22] M. Feng, A. Dey, and R. W. Lindeman, “An initial exploration of a multi-sensory design space: Tactile support for walking in immersive virtual environments,” in *Proc. IEEE Symp. 3D User Interfaces*, Mar. 2016, pp. 95–104.

[23] F. Hülsmann, N. Mattar, J. Fröhlich, and I. Wachsmuth, “Wind and warmth in virtual reality—Requirements and chances,” in *Proc. Workshop Virtuelle Erweiterte Realität*, M. E. Latoschik, O. Staadt, and F. Steinicke, Eds. Herzogenrath, Germany: Shaker Verlag, 2013, pp. 133–144. [Online]. Available: https://pub.uni-bielefeld.de/publication/2624780

[24] H. Engelbrecht, R. W. Lindeman, and S. Hoermann, “A SWOT analysis of the field of virtual reality for firefighter training,” *Frontiers Robot. AI*, vol. 6, Oct. 2019, Art. no. 101.

[25] L. Barbosa, P. Monteiro, M. Pinto, H. Coelho, M. Melo, and M. Bessa, “Multisensory virtual environment for firefighter training simulation: Study of the impact of haptic feedback on task execution,” in *Proc. IEEE 24° Encontro Português de Computação Gráfica e Interação*, Oct. 2017, pp. 1–7.

[26] P. Backlund, H. Engstrom, C. Hammar, M. Johannesson, and M. Lebram, “Sidh—A game based firefighter training simulation,” in *Proc. 11th Int. Conf. Inf. Vis.*, 2007, pp. 899–907.

[27] Z. Xu, X. Z. Lu, H. Guan, C. Chen, and A. Z. Ren, “A virtual reality based fire training simulator with smoke hazard assessment capacity,” *Adv. Eng. Softw.*, vol. 68, pp. 1–8, Feb. 2014.

[28] M. Cha, S. Han, J. Lee, and B. Choi, “A virtual reality based fire training simulator integrated with fire dynamics data,” *Fire Saf. J.*, vol. 50, pp. 12–24, May 2012.

[29] C. Li, W. Liang, C. Quigley, Y. Zhao, and L.-F. Yu, “Earthquake safety training through virtual drills,” *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 4, pp. 1275–1284, Apr. 2017.

[30] M. Bergamasco, A. A. Alessi, and M. Calcara, “Thermal feedback in virtual environments,” *Presence: Teleoperators Virtual Environ.*, vol. 6, no. 6, pp. 617–629, Dec. 1997.

[31] M. Guiatni and A. Kheddar, “Theoretical and experimental study of a heat transfer model for thermal feedback in virtual environments,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 2996–3001.

[32] J. Citerin, A. Pocheville, and A. Kheddar, “A touch rendering device in a virtual environment with kinesthetic and thermal feedback,” in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2006, pp. 3923–3928.

[33] F. Hülsmann, J. Fröhlich, N. Mattar, and I. Wachsmuth, “Wind and warmth in virtual reality,” in *Proc. Virtual Reality Int. Conf.*, 2014, pp. 1–8.

[34] J. Dionisio, V. Henrich, U. Jakob, A. Rettig, and R. Ziegler, “The virtual touch: Haptic interfaces in virtual environments,” *Comput. Graph.*, vol. 21, no. 4, pp. 459–468, Jul. 1997.

[35] J. Dionisio, “Virtual hell: A trip through the flames,” *IEEE Comput. Graph. Appl.*, vol. 17, no. 3, pp. 11–14, May 1997.

[36] J. Tewell, J. Bird, and G. R. Buchanan, “Heat-Nav: Using temperature changes as navigation cues,” in *Proc. Conf. Human Factors Comput. Syst.*, 2017, pp. 1131–1135.

[37] J. M. Willi, G. P. Horn, and D. Madrzykowski, “Characterizing a firefighter’s immediate thermal environment in live-fire training scenarios,” *Fire Technol.*, vol. 52, no. 6, pp. 1667–1696, Jan. 2016.

[38] J. Vasconcelos-Raposo *et al.*, “Adaptation and validation of the igroup presence questionnaire (IPQ) in a portuguese sample,” *Presence: Teleoperators Virtual Environ.*, vol. 25, no. 3, pp. 191–203, Dec. 2016.

[39] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, “Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness,” *Int. J. Aviation Psychol.*, vol. 3, no. 3, pp. 203–220, Jul. 1993.

[40] J. Brooke, “SUS: A quick and dirty usability scale,” in *Usability Evaluation in Industry*. New York, NY, USA: Taylor & Francis, 1996, pp. 189–194.

[41] J. Brooke, “SUS: A retrospective,” *JUS: J. Usability Stud.*, vol. 8, no. 2, pp. 29–40, Feb. 2013. [Online]. Available: https://uxpajournal.org/sus-a-retrospective/

[42] J. R. Lewis, “IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use,” *Int. J. Hum.-Comput. Interact.*, vol. 7, no. 1, pp. 57–78, Jan. 1995.

[43] O. J. Dunn, “Multiple comparisons using rank sums,” *Technometrics*, vol. 6, no. 3, pp. 241–252, Aug. 1964.

[44] Y. Lang, L. Wei, F. Xu, Y. Zhao, and L.-F. Yu, “Synthesizing personalized training programs for improving driving habits via virtual reality,” in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, Mar. 2018, pp. 297–304.

[45] O. A. J. van der Meijden and M. P. Schijven, “The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: A current review,” *Surgical Endoscopy*, vol. 23, no. 6, pp. 1180–1190, Jan. 2009.

[46] D. Harter, S. Lu, P. Kotturu, and D. Pierce, “An immersive virtual environment for varying risk and immersion for effective training,” in *Proc. ASME World Conf. Innov. Virtual Reality*, 2011, pp. 301–307.

[47] N. Cooper, F. Milella, I. Cant, C. Pinto, M. White, and G. Meyer, “The effects of multisensory cues on the sense of presence and task performance in a virtual reality environment,” in *Proc. EuroVR Conf.*, Lecco, Italy, 2015, pp. 1–6.

[48] N. Cooper, F. Milella, C. Pinto, I. Cant, M. White, and G. Meyer, “The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment,” *Plos One*, vol. 13, no. 2, pp. 1–25, Feb. 2018.