Training soldiers to calibration procedures in Virtual Reality, the FELIN IR sight use case

Jean-Daniel Taupiac, Nancy Rodriguez, Olivier Strauss, Martin Rabier

To cite this version:

Jean-Daniel Taupiac, Nancy Rodriguez, Olivier Strauss, Martin Rabier. Training soldiers to calibration procedures in Virtual Reality, the FELIN IR sight use case. 13emes Journées de la Réalité Virtuelle (j.RV), Oct 2018, Evry, France. hal-02091030

HAL Id: hal-02091030
https://hal.science/hal-02091030

Submitted on 5 Apr 2019

Training soldiers to calibration procedures in Virtual Reality, the FELIN IR sight use case

Jean-Daniel Taupiac*
LIRMM - Universit de
Montpellier, Montpellier
Capgemini Technology
Services, Bayonne

Nancy Rodriguez†
LIRMM - Universit de
Montpellier, Montpellier

Olivier Strauss‡
LIRMM - Universit de
Montpellier, Montpellier

Martin Rabier§
Capgemini Technology
Services, Bayonne

ABSTRACT

The infantrymen of the French army are equipped today with a combat system called FELIN, which includes an infrared sighting device: the IR sight. One of the first manipulations learnt by the soldier is the IR sight calibration. Currently, calibration training is a two-step procedure. The first step consists in practicing on a 2D WIMP software, and virtually practicing until making no mistakes. Then, if the soldier succeeds without mistakes, he can apply his knowledge in real situation on the shooting range. In this paper, we present a learning method that includes a prototype in Virtual Reality for training on the FELIN IR sight calibration procedure. We experimented it on real infantrymen learners in an infantry school. Results showed interesting added value of Virtual Reality in this specific use case. It improved the learners' motivation, learning efficiency and helped to identify specific mistake types not detected by the traditional learning software.

1 INTRODUCTION

French infantrymen are equipped with an infrared sight device called IR sight, within their combat system FELIN.

One of the first main tasks a new infantryman has to learn is the IR sight calibration. This procedure requires him to apply his knowledge about the sight system and calculate the appropriate correction, while staying focused on his senses during the control shooting phases. Learning is supported by a 2D WIMP software called EAO FELIN. Althought the tool allows a simultaneous training of a large number of learners, this software nevertheless shows some learning limits. It seems that learners must make a lot of replays before making no mistakes, which can let us suppose that the procedure understanding might be distorted by the mechanical replaying of the tasks. It also brings out potential efficiency lacks on skills transfer when learners are on the shooting range.

As part of works between the French Land Army, Capgemini and the LIRMM (Montpellier Laboratory of Informatics, Robotics and Microelectronics) , we are interested in measuring the contribution of Virtual Reality for the IR sight calibration training procedure. Capgemini is also currently carrying out the long-term application management of the EAO FELIN 2D software.

This paper is oragnized as follows. In the first section we will briefly present related works. Afterwards, we will briefly present the FELIN combat system, its IR sight, the procedure of calibration and the *EAO FELIN* learning software traditionally used. After that, we will present the VR prototype we implemented, and then the experiments we drove, before communicating results. We will finish by discussing them and underlying the directions of the future prototype.

2 RELATED WORK

Virtual Reality had already been used for training in many military areas: fight training [26], security procedures [40], social skills [19], technical procedures [18], etc. It also had been used for technical procedures training in numerous fields: maintenance procedures [16], medical procedures [33], emergency procedures [27], etc.

Learner's motivation is one of the main factors of the learning efficiency [39] and it appears that Virtual Reality can foster it largely [13,35]. [36] mention that *VR enhances trainees' intrinsec motivation* and bind it to freedom degrees and *"high-tech" nature* of this technology.

As relayed by [36], [9] suggested positive effects of Virtual Reality for training performances. They compared the performance of trainees which were exposed to a VR assembly line learning set-up, to those who trained with real equipment. Their results suggested that VR trainees learned faster and made few errors. In the military field, [18] demonstrated significant improvements in learning for Virtual Reality on naval equipement and procedures. They found improvements practically independently regardless of the trainees' experience levels.

Simulation's fidelity level tend towards training efficacy. [11] showed that, for driving learning simulators, the training transfert effectiveness varied with simulation fidelity. Results showed that learners which faced less accidents when they used to practice before on the highest fidelity simulator.

We produced a Virtual Reality prototype based on the *EAO FELIN* software functionalities, and we experimented it on infantrymen learners. We wanted to see if using Virtual Reality in this use case could improve learning performances and motivation and mistake awareness. We also wanted to determine how learning simulation fidelity could impact the learning efficiency.

3 FELIN

3.1 The FELIN combat system

The infantrymen of the French army are equipped with a combat system called FELIN (***F**antassin à **É**quipement et **L**iaisons **IN**tégrés*: Integrated Infantryman Equipment and Communications). This system, developed by Safran Electronics Defense [7], improves a modified FAMAS rifle with a set of clothes, pouches, body armors along with electronic, optic and optronic devices.

One of these devices is the IR sight, an infrared vision sight device which is fixed over the infantryman's FAMAS rifle (Fig. 1). It contains optronic components allowing night and un-camouflaged vision. When the soldier looks by the sight eyepiece, he can display an interface overlaid and navigate within some specific menus. To interact with this interface, the infantryman uses a remote control fixed on the rifle's handle (Fig. 1, soldier's left hand). He can also use other buttons, placed on each side of the IR sight (on the side of the sight on Fig. 1).

---

[*] e-mail: jean-daniel.taupiac@lirmm.fr
[†] e-mail: nancy.rodriguez@lirmm.fr
[‡] e-mail: olivier.strauss@lirmm.fr
[§] e-mail: martin.rabier@capgemini.com

Figure 1: FELIN solider equipped with the IR sight on the FAMAS rifle [3]

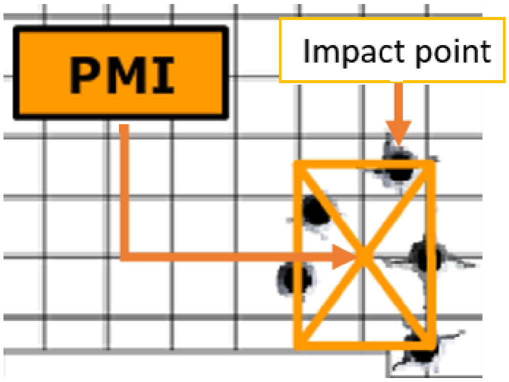

Figure 2: Medium Impact Point (PMI)

3.2 FELIN IR sight's calibration procedure training

To guarantee the operational maintenance of his weapon, the soldier must be efficient in the calibration of his IR sight. This specific procedure is realized on the shooting range. By navigating within the sight interface, the soldier has to merge the sight and the barrel axis. He controls if they are aligned by shooting five cartridges on a target. He moves next to the target and calculates the Medium Impact Point (Fig. 2), by using the position of the five impacts. After that, he must enter the correction on the sight interface and repeat the process as many times as needed. Two different corrections are needed: height correction (vertical shift between rifle and barrel axis) and direction correction (horizontal shift between rifle and barrel axis).

3.3 *EAO FELIN* software

Today, the French infantryman train to the FELIN combat system using a software called *EAO FELIN*, developed by Capgemini. This software includes a specific module for the IR sight calibration training. To practice the procedure, learners use a 2D interface showing the remote control interface, the sight buttons and the trigger of the rifle (Fig. 3).

Currently, for pedagogical and safety reasons, soldiers practice on the software until making no mistakes. When they succeed, they receive a procedure certificate and they are allowed to experience the real situation on the shooting range.

4 VIRTUAL REALITY PROTOTYPE

4.1 Principle

The VR prototype has been developed to assess the potential improvements on learning that Virtual Reality can provide on the IR sight calibration procedure. The goal was to focus on the training procedure only, so we ported the *EAO FELIN* software procedure on Virtual Reality (Fig. 4). We exploited the possibilities of Virtual Reality by extending some procedure steps of the software, to make the experience closer to the procedure execution in real situation. We modelled a shooting range with a target placed at 25 meters. The trainees had to take the prone shooting posture and move between the rifle and the target position. The following steps were implicit or facilitated on the 2D software :

Figure 3: Sight calibration training module on the *EAO FELIN* software

- *Takes the shooting posture*. This step is absent in the 2D software, but we gave the users the possibility to take the prone shooting position in Virtual Reality. The goal was to improve fidelity closest to the real operating conditions.
- *Shoots 5 cartridges*. While on the 2D software the user had to click only once on the trigger button, we let the user start each shot by a trigger pressure. The shooting behavior then became more faithful in Virtual Reality.
- *Goes to the target*. After the shots, the 2D software automatically zoomed on it. We extended this step in Virtual Reality by allowing the user to move to the target.
- *Measures the PMI*. On the 2D software, the PMI is displayed on the target for 5 seconds after the zoom. In Virtual Reality, we also chose to display it on the target, but there were no time restrictions.

4.2 Implementation

We developed this prototype with Unity 3D [1], and we used an HTC Vive [8] headset. The virtual environment was displayed in stereoscopy and the sight view was displayed on the right eye when the user approached his right eye to the lens filter.

For the first version, we used an off-the-shelf rifle support [5] for the Vive controllers. The controllers fastenings were magnetized on the structure and we adjusted their position to match with the FAMAS handle positions (Fig. 5). We also fixed an off-the-shelf bipod [2]. This material combination was intended to allow the trainee to take the prone shooting position. The goal was to determine if these off-the-shelf supports could bring about a satisfying level of fidelity.

The trigger of one controller had been assigned to the trigger of the rifle, and the trackpad of the other one to the remote control interface. In order to be able to find the button position when equipped with the headset, an adhesive mask taking the shape of the remote control button was placed over the trackpad (Fig. 5). Indeed, the trackpad zone was divided in four areas and each one matched a remote control button.

A teleportation zone was placed on the side of the shooting range (Fig. 6), to allow the learner to move to the target. When the user

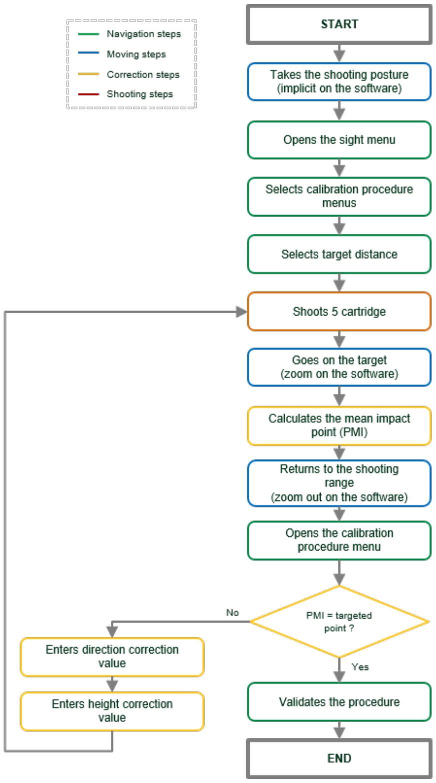

Figure 4: Training procedure, applied on the 2D software and the VR prototype

entered in the zone, he was instantly teleported in front of the target. To go back to the rifle, a teleportation zone was also placed near the target.

The VR prototype used the same rules present in the 2D software for each procedure step. If the user's action is unexpected, it is impossible to navigate in the wrong menus of the interface, shoot and move to the target. If the learner tried to do an unexpected action, the system indicated that he had made an mistake by playing a sound, just like in the *EAO FELIN* software. These mistakes and all the actions done were also timed and traced in a log file. Mistakes were classified in three different types:

- Procedure mistakes: the learner tried to go to a wrong menu or make an action at the wrong time (shooting, moving to the target, etc.).
- Correction mistakes: the learner didn't enter the right correction value or he mistook the height and/or direction corrections.

Figure 5: VR rifle support vs FAMAS rifle (left) and adhesive mask added on the Vive controller (right)

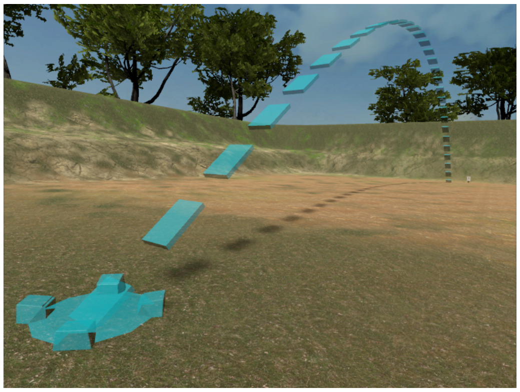

Figure 6: Teleportation zone

- Button mistakes: the learner used the wrong button or didn't press it for enough time if a long push was needed.

As the goal was not to train on shooting, the bullet impacts were randomly computed. There was a tolerance concerning the shooting position, as well as aiming and shooting the target.

Finally, for this first prototype version, we did not let the user to see his hands in the virtual environment, keeping this feature for future improvements.

5 EXPERIMENTS

During the experiments, we compared the two learning methods through to two angles. First, we conducted user tests on the VR prototype. Afterwards, we wanted to see if the Virtual Reality showed potential benefits for this specific use case. We wanted to observe if allowing the learner to live an experience closer to the real one would impact the learning efficiency of the procedure.

5.1 User tests

For user tests, we wanted to see if the users interactions with the different elements were pleasant, enjoyable and realistic. We focused ourselves particularly on the rifle, the sight, the remote control, the menus and the moving method. We also look at motion sickness and eyestrain. We also wanted to see if there was a difference between the feelings of the learners and those of the instructors. We believe that comparing these results could let us spot potential lacks only identifiable by experienced users.

5.2 Learning improvement hypothesis

For this use case, we formulated the following hypotheses on the Virtual Reality method effectiveness on learning:

H1 Learners will be more motivated when using Virtual Reality rather than the 2D software

H2 Learners will need less replays until making no mistakes on the 2D software, after the use of the Virtual Reality prototype

H3 Learners will make less mistakes after the use of Virtual Reality

H4 Learners will have a better awareness of their mistakes in Virtual Reality

5.3 Experimentation process

5.3.1 Groups

To check these hypotheses, we opted for the experimentation process shown in Fig. 7. At present, learners practice on the *EAO FELIN* software until they make no mistakes, after following a set of lessons about the FELIN system and a summary on how to handle the equipment. So, we chose to split the learners in two groups:

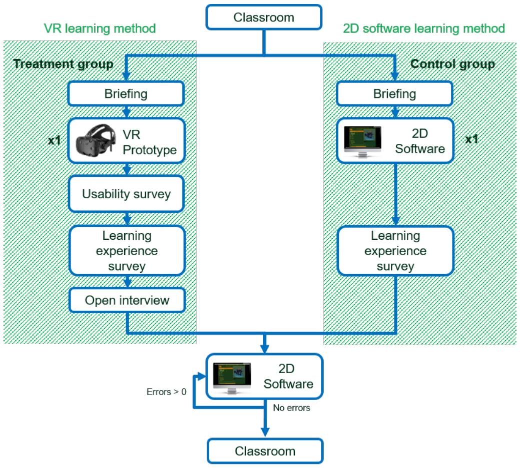

Figure 7: Experimentation process

- a control group, on the 2D software learning method, including iterations on the 2D software until making no mistakes. Before using the 2D software, they were given a short set of instructions.

- a treatment group, on the VR learning method, making their first iteration on the VR prototype before iterating on the *EAO FELIN* software until making no mistakes. Before using the VR prototype, they also were given a short set of instructions that presented the prototype and its functioning.

In the treatment group, learners used the VR prototype independently from one another, only accompanied by an instructor and an experimenter. Learners of the control group were in the traditional learning conditions: in the same computer room, equipped with headphones. It was not possible for us to isolate subjects using the 2D software. Nevertheless, instructors supervised the groups and kept a close watch on them, to avoid exchanges between learners.

5.3.2 User tests

After prototype use, learners from the treatment group were invited to answer a set of Likert scale [25] questions (values from 1 to 5). The purpose of this survey was to evaluate the VR system, focusing on different elements like the rifle, the sight, the remote control and the menus. This survey was also aimed to detect if the moving method and metaphor were efficient. It was also intended to discern if the learner felt cybersickness or eyestrain. We also included the IPQ [6] presence survey too [32], which allows assessing presence through three subscales: Spatial Presence (sense of being physically present), Involvement (attention on virtual environment, involvement experienced) and Experienced Realism (subjective experience of realism). Finally, one question was aimed to encourage learners to express their satisfaction level after the prototype use.

After answering these survey, learners from the treatment group were encouraged to express a feedback about their experience, during an open interview. The goal was to potentially let appear elements which weren't detectable by the surveys.

5.3.3 Learning benefits experiments

After their first iteration, both groups also had to answer a Likert survey which was focused on their learning experience. Some questions analysed the user motivation. Motivation is an interesting indicator for learning, so three Likert scales focused on it: *"Performing the procedure was felt as a pleasure."*, *"The used support was attractive"* and *"Performing again the procedure would be a pleasure."*. We also added a question where learners had to estimate approximately the number of mistakes they made. This information was used to analyse the awareness of their mistakes by comparing it to the log file.

5.3.4 Results exploitation

We calculate a score (arithmetic mean) from answers to questions for each element analyzed. For the negative questions (e.g. *"The element use was unpleasant"*), the value taken for the mean was the answer value subtracted from the maximum value (5).

5.4 Subjects

We led experimentations on 76 learners, lieutenants of the Draguignan Infantry School [4], from 6 different squads. Due to time and logistic constraints, 22 learners were assigned to the treatment group and the 54 others to the control group. Randomization was not used, but instructors were asked to select randomly learners for the treatment group.

The prototype implemented was only adapted for right-dominant eye users and it was the unique constraint required before the treatment group random selection.

The average age of subjects was 27(±4) years old, and they were 99% male. In the treatment group, 74% of them had never used a Virtual Reality headset before. We also experimented it on 6 instructors of the military school. They were all male, their average age was 41(±5) years old and 60% of them had never used a Virtual Reality headset before.

6 RESULTS

6.1 Methodology

We performed a statistical analysis for each result, by comparing the results of each group as two independent samples.

For data from the Likert scale values, we proceeded with the Mann-Whitney U test [28] to see if there was a statistically significant difference between treatment and control groups. In the same way, for other data like the number of mistakes and iterations, we check if they had a normal distribution (Shapiro-Wilk [34]) and a homogeneous variance (Brown-Forsythe test [12]). If it was the case we proceeded with the Student's t-test [38]. If not, we also used the Mann-Whitney U test.

Finally, for data like the number of mistakes reduction on each replays, we used the same test (Shapiro-Wilk) to check normal distribution, and if it succeed we also used the Sudent's t-test. If not, we used the Wilcoxon signed-rank test [42].

6.2 First iteration results

6.2.1 User tests

We compared the results of the prototype assessment from the surveys answered by learners and instructors (Fig. 8).

The rifle (2.94 ±0.68) and the remote control (3.18 ±0.83) were moderately rated. Three questions were related to the rifle and four were related to the the remote control. As shown in Table 2, the lack of fidelity of the rifle seemed to make its handling quite unpleasant. For the remote control, users felt significant difficulties to find the buttons. This effect seemed to have affected them by making its handling quite unpleasant too.

General Presence (3.87 ±1.15) and Spatial Presence (3.72 ±0.64) were well rated by subjects (Fig. 9). But Involvement (3.06 ±0.60) and Experience realism (2.89 ±0.43) received moderate scores. For each scale of presence, there wasn't a statistically significant difference between learners and instructors (Table 1, PRDIFF).

The sight (3.93 ±0.71), the menus (4.17 ±1.05) and the moving method (4.76 ±0.44) were rated satisfactorily. Users seemed to have felt practically no cybersickness (1.04 ±0.2) and eyestrain (1.8 ±0.83). Finally, users expressed a good satisfaction (4.17 ±0.70) after using the VR prototype.

Table 1: Statistic tests results

		Test	Results	Significant?	Cohen's d
	First iteration results				
PRDIFF	Presence differences learners-instructors	Student t-test	$t(27) = -0.61$, $p > 0,05$	Not significant	
UTDIFF	User tests differences learners-instructors	Mann-Whitney U test	$p > 0.05$	Not significant	
MSDIFF	Motivation scores	Mann-Whitney U	$U = 404$, $z = -2.52$, $p = 0.012$	Significant	0.28
EEDIFF	Approximation of mistake estimations	Shapiro-Wilk	Normality test failed		
		Mann-Whitney U	$U = 556,5$, $p > 0.5$	Not significant	
	Replays results				
PMSRED	Procedure mistakes-Soft. errors reduction	Wilcoxon	$T = 16$, $z = -4.342$, $p < 0.001$	Significant	0.69
PMVRED	Procedure mistakes-VR proto. errors reduction	Wilcoxon	$T = 0$, $z = -3.461$, $p < 0.001$	Significant	0.89
PMCOMP	Procedure mistakes-Soft./proto. comparison	Shapiro-Wilk	Normality test failed		
		Mann-Whitney U	$U = 246.5$, $z = 1.22$, $p = 0.024$	Significant	0.16
CMSRED	Correction mistakes-Soft. errors reduction:	Wilcoxon	$T = 68.5$, $p < 0.001$	Not significant	
CMVRED	Correction mistakes-VR proto. errors reduction	Wilcoxon	$T = 0$, $z = -2.414$, $p = 0.016$	Significant	0.60
CMCOMP	Correction mistakes-Soft./proto. comparison	Shapiro-Wilk	Normality test failed		
		Mann-Whitney U	$U = 229.5$, $z = 0.92$, $p = 0.011$	Significant	0.12
BMCOMP	Button mistakes-Soft./proto. comparison	Shapiro-Wilk	Normality test failed		
		Mann-Whitney U	$U = 292.5$, $z = 3.45$, $p = 0.033$	Significant	0.46
NBITER	Number of iterations needed	Shapiro-Wilk	Normality test failed		
		Mann-Whitney U	$U = 45$, $z = -5.62$, $p < 0.001$	Significant	0.67

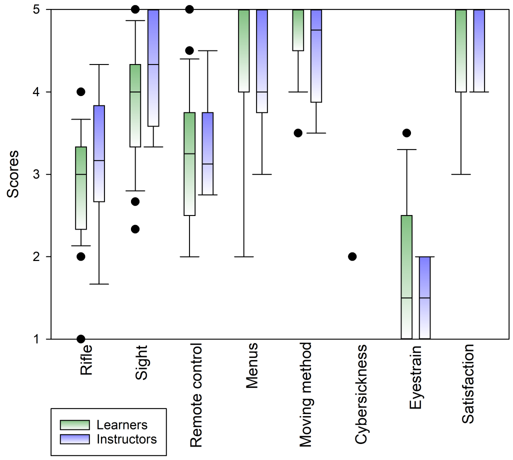

Figure 8: Prototype assessment results

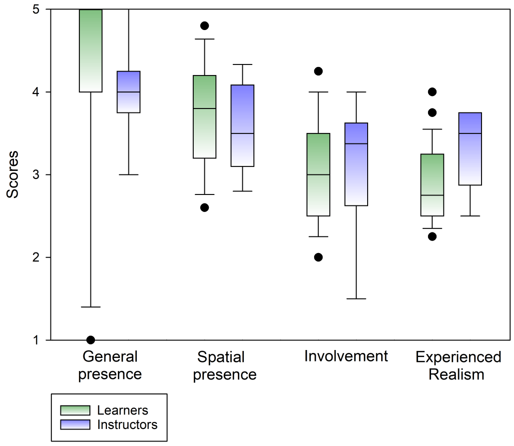

Figure 9: Presence results

For each item (rifle elements, motion sickness, satisfaction, presence), there wasn't a statistically significant difference between learners and instructors (Table 1, UTDIFF).

6.2.2 Motivation results

Motivation scores showed a statistically significant difference (Table 1, MSDIFF) between the treatment and the control groups (Fig. 10). Users who used the VR prototype seemed to feel more motivated (Med: 4.00) than the *EAO FELIN* group (Med: 3.67).

6.2.3 Mistakes and error awareness

For procedure and correction mistakes, results tended to show that learners made less mistakes on their first iteration if they used the VR prototype instead of the *EAO FELIN* software, but the difference between the two groups wasn't statistically significant (Table 1, MIDIFF).

In order to analyze the learners' error awareness, we calculated the approximation errors of the mistake estimations from both groups (Fig. 12). We compute it from their number of mistakes made (saved in the log files) and their estimated amount (answered in the learning experience questionnaire). These results didn't show a statistically significant difference between the control and the treatment group (Table 1, EEDIFF).

6.3 Replays results

Results from replays let us analyse the impacts of the VR learning method on mistakes committed and number of replays needed. 6 subjects of the treatment group were excluded from the results. 4 of them were faced with time issues and did not have the time to finish the 2D software iterations. The 2 others were confronted with technical issues with the 2D software, which could have impacted their results.

14 subjects of the control group were excluded from the results. 11 of them was faced time issues and did not have the time to finish the 2D software iterations. During the replays, 2 others did not respect the instructions and restarted the software when they were making errors. Finally, the last one faced technical issues with the 2D software during replays, which could have impacted his results.

Consequently, the final population studied for the replay results was 41 subjects (15 in the treatment group and 36 in the control

Table 2: Rifle and remote control questionnaire results

Rifle		
Questions (Likert [1-5])	Learner means	Instructor means
Handling it was easy	3.17 (±0.87)	3.67 (±0.47)
Handling it was disagreeable	2.91 (±1.17)	3.33 (±1.11)
Handling it was realistic	2.64 (±0.88)	3.17 (±1.34)
Remote control		
Questions (Likert [1-5])	Learner means	Instructor means
Finding buttons was easy	2.48 (±1.25)	2.83 (±1.07)
Finding way to use it was easy	3.30 (±1.04)	4.00 (±0.82)
Using it was disagreeable	2.52 (±1.10)	3.00 (±1.00)
Using it was realistic	3.42 (±1.06)	3.33 (±0.94)

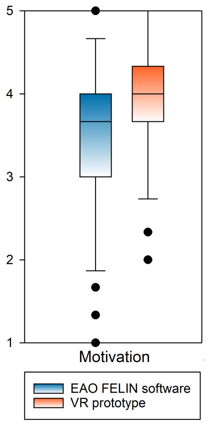

Figure 10: Motivation questionnaire results

group).

For the mistake estimation, 1 subject from the treatment group and 4 subjects from the control group were excluded because they didn't answer to the mistake estimation question.

6.3.1 Mistakes

Experiments showed interesting results about committed mistakes between the treatment and the control group (Fig. 11). To exclude eventual bias, we only compared mistakes which were possible in both environments (VR prototype and *EAO FELIN* software). For the treatment group, we rejected mistakes which were due to a manipulation error (e.g. accidental push on a button while catching the rifle).

There was a statistically significant difference for the mistakes made on replays. Learners who used the VR learning method didn't make any procedure and correction mistakes, while 2D software method users still continued to make some (Table 1, PMSRED, PMVRED, PMCOMP, CMVRED, CMCOMP).

Nevertheless, for the correction mistakes of 2D software learners, the difference is not statistically significant (Table 1, CMSRED). However, these effects are more flagrant when we analyse the percentage of learners making mistakes. Procedure mistakes: 80% (Software learners) vs 65% (VR learners) on the first try, 26% (Software learners) vs 0% (VR learners) on replays. Correction mistakes: 50% (Software learners) vs 48% (VR learners) on the first try, 31% (Software learners) vs 0% (VR learners) on replays.

The opposite effect appeared for button mistakes. Most probably due to the lack of fidelity of the remote control, on their first iterations, the treatment group made more mistakes (Med: 0) than the control group (Med: 0) (Fig. 11). This result was statistically significant (Table 1, BMCOMP). Another interesting result came from the fact that users who made their first iteration on the 2D software didn't make any button mistakes when VR learners still continued to make some.

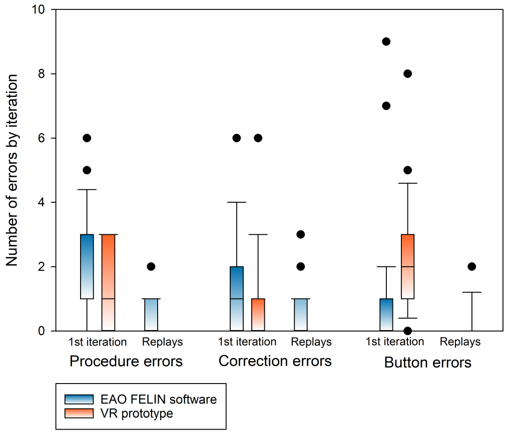

Figure 11: Mistakes made on the first iteration and mean of mistakes made on each iteration of the replay

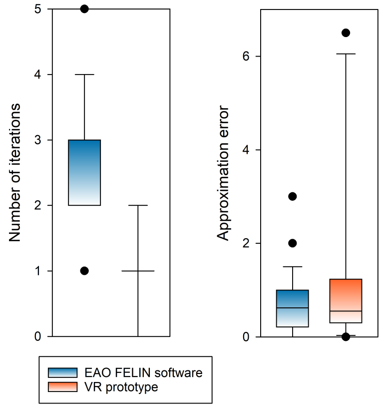

Figure 12: Replays needed until making no more mistakes (left) and mistakes estimations' approximation errors (right)

6.3.2 Replays needed

VR learners needed less replays until making no mistakes than the software ones (Fig. 12, left). Learners from the treatment group needed significantly less iterations (Med: 1) until making no mistakes, than learners from the control group (Med: 3) (Table 1, NBITER).

7 DISCUSSIONS AND FUTURE WORK

7.1 Result interpretation

7.1.1 Fidelity lacks

User tests clearly underlined fidelity lacks from the rifle and the remote control. Open interviews tended to confirm these results. A significant number of users freely indicated they were uncomfortable in finding the remote buttons (65%) and most of them declared they wanted to see their hands to place correctly their thumb on buttons (60%). The rifle was also pointed out for its lack of fidelity. A lot of learners underlined that the gun weight was too light (48%) and then felt bother. Another criticized aspect was the magnetic system of the support. Infantryman tend to strongly grab the handles of the rifle on their shooting position, and with this system, the controllers slightly moved. Finally, some users confessed being bothered by the fidelity lacks of the trigger (22%) and the absence of recoil after a shot (13%).

Learners expressed different opinions about these lacks. Several of them expressed the need of "*recovering real sensations*" during the experience, and some others mentioned that the absence of recoil let them be "*more focused on the calibration*".

These fidelity lacks seemed have mixed effects on our results. That's why we suppose it's relevant to tend towards a high level of physical fidelity when the training purpose concerns a procedure for a device where its manipulation requires a high level of focus on senses. The most important improvement axes of our future works, must consequently consist in reducing these gaps and comparing the effects on the learners' behaviours and performances. Some works of the litterature [21, 22, 37] suggest that *haptic feedback in an early training phase may improve the trainee's performance, by enhancing the trainee's sensoric perception capabilities and thus facilitating transfert of skill* [41]. One of the possible improvements could consist in using demilitarized equipment as a tangible user interface, ideally modified with a pneumatic system. It could then be interesting to compare the learners' concentration and performance, by varying on these different fidelity levels. In this purpose, we plan to lean on frameworks from literature, like the AFFECT [29] one.

Finally, we supposed that the users' desire of visualizing their hands in the virtual environment was due to the tactile differences between the trackpad-adhesive assembly and the real remote buttons. Real remote buttons are indeed large and accompanied by a tactile cue to easily help finding their positions. Some litterature works suggest that effects of body visibility in the virtual world can create an adaptation time for the learners, which can affect significantly positively their learning performance [20]. It could so be interesting to equip some users with data gloves on a high fidelity version of the prototype and then analyse if allowing learners to see their hands impacts their efficiency and their committed mistakes.

All these fidelity aspects had also been brought up by instructors during their open interview. They additionally underlined that it could be interesting to enlarge the system to assess some good shooting practices during the experience (breath control, pressure levels on the trigger, etc.). This could be a good prospect as part of future works. It could indeed be interesting to analyse if learners are as focused and attentive to the procedure when they also have in mind their good shooting practices.

Finally, results underlined a good assessment for menu visibility and ergonomics. The sight functioning was well rated too, confirming our implementation choice to render the sight camera on the dominant-eye when the user approach it, close to the lens filter.

7.1.2 Sense of presence

As noticed in the "Results" section, Involvement and Experienced realism were moderately rated. It's highly likely that they could have been impacted by the rifle fidelity lacks. It could then be interesting to compare these values with new values after improving the fidelity levels of the rifle.

We didn't try to compare the sense of presence between the control and the treatment groups because we naturally presume that the *EAO FELIN* software would have less impact than the VR prototype. Nevertheless, it could be an interesting point to confirm in future works.

Finally, we could have supposed that the presences of the instructor and the experimenter in the same room could have impacted the Involvement scale of presence, but that seems doubtful because they paid attention to be particularly discreet and act only in last resort.

7.1.3 Cybersickness

As explained by [15], sensorimotor discrepancies are mainly behind cybersickness issues. Users expressed to have not felt any disturbing effect. We explain this result by correlating it with the moving method assessment which was very good. Indeed, the teleporting method easily limits sensorimotor discrepancies.

They could also be linked to other factors, like the refresh rate issues. This result consequently encourages us about the performances of the system.

7.1.4 Eyestrain

Eyestrain was also very lightly reported but noticed during or after the prototype use by the half of users (55%). This result can be explained by issues in the headset adjustment which was sometimes reported in open interviews. It could also come from the breakdown between accommodation and convergence [24]. The possibility that some of users were stereo-blind could be taken into account. More experiments are needed to determine the origins of this effect and find solutions to reduce it. We believe this is important, to avoid biasing results or excluding a substantial proportion of learners in the long term.

7.1.5 Users' experience

As noticed previously, there was no statistically significant differences between learners and instructors on presence and system assessments. These results let us suppose that the prototype assessment is probably not affected by the users' experience and that the system didn't deviate from actual field conditions. Nevertheless, as the sample of instructors was thin, we can't draw any conclusions about it.

7.1.6 Learning improvement hypothesis

H1: *Learners will be more motivated in Virtual Reality.* Results showed a positive impact on the motivation of the learners. It validated the assumption that living an experience in Virtual Reality for a military device calibration procedure should enhance the motivation of the learners. Although this result is significant, the effect size is low (Cohen's $d = 0.28$). To confirm this impact and clarify the size of its effect, we should explore literature and apply a more advanced motivation assessment method (than a 3 Likert scale) in our future works. The fact that 2D software learners were performing in the same room at the same time could also have impacted the motivation. More experiments, with isolated subjects could help disprove this hipothesis.

H2: *Learners will need less replays until making no mistakes on the 2D software, after the use of the Virtual Reality prototype.* Linked to these results on committed mistakes, we noticed that there was an important significant effect (Cohen's $d = 0.67$) on the number of needed replays. Learners in VR needed less iterations than software users.

H3: *Learners will make less mistakes after the use of Virtual Reality.* After its use, learners using the *EAO FELIN* software made no procedure or correction mistakes. On the other hand, the 2D software learners, continued to make this kind of errors on the following trys. These results showed that living an experience within an immersive simulated situation facilitates the activation of cognitive levers, necessary to assimilate the procedure.

However, these results only partially validate the hypothesis because we noticed the opposite effect on button mistakes. We suppose that these results are most probably linked to the fidelity lacks previously discussed. In future works, we plan to verify it after having carried out fidelity improvements on the VR prototype. We also noticed that for button errors, the software users didn't make any errors of this kind. When fidelity issues will be corrected, it could be interesting to analyze if it is also the case on the VR prototype.

Table 3: Hypothesis results

#	Hypothesis	Results
H1	More motivated	Validated
H2	Less replays needed	Validated
H3	Less mistakes after use	Partially validated
H4	Better awareness of mistakes	No significant results

H4: *Learners will have a better awareness of their mistakes in Virtual Reality.* The mistakes estimation didn't show significant results. We observe an important data scattering for the approximation of errors of the VR learners about the estimation of their number of mistakes. It brought us to suppose that these results could highly come from the fidelity of the remote control. Consequently, in future works, it could be interesting to do this experiment again, when the lacks of fidelity will be corrected. We also plan to distinguish the different kind of mistakes on their estimation. If there is a significant difference, it could be interesting to analyse if the effect changes according to the type of mistake.

The litterature in other fields shows that using repetitive VR training can improve the learning curve before practicing on real situations [10, 14, 31]. In our future works, it would so be interesting to let each group iterate on the same support (software or VR prototype), then analyse the total errors made, time spent and iterations needed to assess if it implies training time gains and effectiveness.

As summarized in Table 3, experiments allowed us to validate hypotheses **H1** and **H2**. **H3** was partially validated, due to fidelity lack issues. **H4** didn't showed significant results and require more improvements and experiments.

7.2 Observations

It was interesting to notice that Virtual Reality can help detect errors which can hardly be detected on the 2D software. Indeed, experiments helped to highlight the limits of the *EAO FELIN* software. Some errors made by learners which were impossible on the software were made by users. For example, some of them tried to navigate in menus instead of moving to the target to check the PMI. In the software, this stage was automatically done by zooming on the target after the shot.

We also observed that in Virtual Reality, even if learners were informed that shooting accuracy wasn't necessary for the training situation, they still paid a lot of attention on aiming at the target, stopping their breath at shooting time and controlling their pressure on the trigger. We interpret this fact by the sense of presence provided by the VR prototype, which allowed learners to perform their procedure conscientiously.

Finally, during open interviews, some learners noticed a *"sensation of calm"* through the VR procedure execution. We think that it could be interesting to focus on this potential effect as part of future works, in order to try to identify reasons and analyse how it could be an interesting learning lever for training through Virtual Reality.

7.3 Limits

We identified two potential limits for our results. The first one is the simplicity of the procedure. Indeed, the IR sight calibration procedure is quite simple, and we suppose that some insignificant results could be more pronounced for other procedures, depending on their difficulty.

The second potential limit comes from the population of subjects on which we made experiments. They all were lieutenant trainees, which mean that some of them had just recently graduated from the officer school, and others had been promoted internally. We suppose that their educational level could have reduced some of our results. As part of future works, we think that it could be interesting to experiment the VR prototype on learners having a lower military rank. As part of future works, if the population military rank is the same, we plan to ask learners to indicate their experience in order to analyze if there are different results between these two different lieutenant profiles.

7.4 Other prospects

Additionally, we also plan to develop a left dominant-eye prototype version. Comparing results from left and right dominant-eye subjects could be interesting.

To focus on the learning procedure, we made the choice to generate shooting impacts randomly on the target and to indicate the PMI and the correction values required on the target. We think it could be interesting to make a more faithful prototype version and compare it to a new 2D software version. In these versions, the aimed point could be the shot one, and the PMI or correction values could be deliberately absents. This faithfulness improvement could allow analysing if Virtual Reality uses impacts sight behavior and value estimations.

The litterature shows that, in other fields such as surgery area, Virtual Reality simulators have proven effectiveness on the skills transfert [17, 23, 30, 31, 33]. In our experiments, we analysed the learning improvements by comparing the performances of trainees through replays on the software. We presume that we could also notice some improvements when VR learners perform their first real procedure on the shooting range. We strongly suppose that VR learners will spend less time in making their first calibration procedure in real situation. Experiment it could be an interesting way to analyse the learning transfers effects of Virtual reality for this use case. We plan to use our results in order to be able to study the efficiency of skills transfert.

8 CONCLUSIONS

In this paper, we presented a Virtual Reality prototype used for training on a military sight calibration procedure. We experimented it on real infantry learners from an infantry school.

Results showed that for this kind of procedures, Virtual Reality can improve the motivation of the learner. The prototype also showed interesting results in learning efficiency. VR learners made no errors when 2D software learners continued to make some. Experiments let us to identify some fidelity lacks on the prototype. This issue brought us to suppose that it's relevant to tend towards a high level of fidelity when the training purpose concerns a procedure for a device for which its manipulation requires a high level of sensorial focus. Finally, results also helped identify some procedure mistakes which are currently not detectable by the traditional software.

In our future work, we will focus on improving the fidelity of the rifle and of its remote control, to verify if fidelity lacks could have impacted some of our results. We also plan to experiment a less restricted version of the prototype, to analyse if the freedom degrees impact the learning process of a procedure in Virtual Reality. Finally, we want to confirm all the results by making some additional experiments when the soldiers are on the real shooting range, to assess the effects of Virtual Reality on the learning transfers for a military sight calibration procedure.

ACKNOWLEDGMENTS

The authors wish to strongly thank the Human Resources Department of the French Land Army (DRHAT) and the Infantry School of Draguignan (EMD) for allowing us to experiment the prototype on real infantryman learners and also publishing this results.

REFERENCES

[1] `http://unity3d.com/`.
[2] `http://www.amazon.fr/gp/product/B00M9RF0ZM/`.

[3] `http://www.defense.gouv.fr/terre/equipements/materiels-generiques/equipement/felin-fantassin-a-equipement-et-liaisons-integres`.

[4] `http://www.emd.terre.defense.gouv.fr/`.

[5] `http://www.etsy.com/listing/491103695/htc-vive-magnetic-dual-controller-rifle`.

[6] `http://www.igroup.org/pq/ipq/index.php`.

[7] `http://www.safran-electronics-defense.com/`.

[8] `http://www.vive.com/`.

[9] N. Adams. A study of the effectiveness of using virtual reality to orient line workers in a manufacturing environment. *Motorola University, unpublished dissertation*, 1996.

[10] R. Aggarwal, T. P. Grantcharov, J. R. Eriksen, D. Blirup, V. B. Kristiansen, P. Funch-Jensen, and A. Darzi. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Annals of surgery*, 244(2):310, 2006.

[11] R. W. Allen, G. D. Park, M. L. Cook, and D. Fiorentino. The effect of driving simulator fidelity on training effectiveness. *DSC 2007 North America*, 2007.

[12] M. B. Brown and A. B. Forsythe. Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367, 1974.

[13] G. C. Burdea and P. Coiffet. *Virtual reality technology*. John Wiley & Sons, 2003.

[14] P.-J. Fager and P. von Wowern. The use of haptics in medical applications. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 1(1):36–42, 2004.

[15] P. Fuchs. *Virtual Reality Headsets-a Theoretical and Pragmatic Approach*. CRC Press, 2017.

[16] N. Gavish, T. Gutiérrez, S. Webel, J. Rodríguez, M. Peveri, U. Bockholt, and F. Tecchia. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6):778–798, 2015.

[17] T. P. Grantcharov, V. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch-Jensen. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British Journal of Surgery*, 91(2):146–150, 2004.

[18] R. T. Hays, D. A. Vincenzi, A. G. Seamon, and S. K. Bradley. Training effectiveness evaluation of the vesub technology demonstration system. Technical report, NAVAL AIR WARFARE CENTER TRAINING SYSTEMS DIV ORLANDO FL, 1998.

[19] R. W. Hill Jr, J. Gratch, S. Marsella, J. Rickel, W. R. Swartout, and D. R. Traum. Virtual humans in the mission rehearsal exercise system. *Ki*, 17(4):5, 2003.

[20] R. V. Kenyon and M. B. Afenya. Training in virtual and real environments. *Annals of Biomedical Engineering*, 23(4):445, 1995.

[21] H. K. Kim, D. W. Rattner, and M. A. Srinivasan. The role of simulation fidelity in laparoscopic surgical training. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 1–8. Springer, 2003.

[22] P. Lamata, E. Gómez, F. Sánchez-Margallo, F. Lamata, F. Del Pozo, and J. Usón. Tissue consistency perception in laparoscopy to define the level of fidelity in virtual reality simulation. *Surgical Endoscopy and Other Interventional Techniques*, 20(9):1368–1375, 2006.

[23] C. E. Lathan, M. R. Tracey, M. M. Sebrechts, D. M. Clawson, and G. A. Higgins. Using virtual environments as training simulators: Measuring transfer. *Handbook of virtual environments: Design, implementation, and applications*, pp. 403–414, 2002.

[24] L. Leroy, P. Fuchs, and G. Moreau. Visual fatigue reduction for immersive stereoscopic displays by disparity, content, and focus-point adapted blur. *IEEE Transactions on Industrial Electronics*, 59(10):3998–4004, 2012.

[25] R. Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[26] M. Macedonia. Games soldiers play. *IEEE Spectrum*, 39(3):32–37, 2002.

[27] D. Manca, S. Brambilla, and S. Colombo. Bridging between virtual reality and accident simulation for training of process-industry operators. *Advances in Engineering Software*, 55:1–9, 2013.

[28] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pp. 50–60, 1947.

[29] R. P. McMahan and N. S. Herrera. Affect: Altered-fidelity framework for enhancing cognition and training. *Frontiers in ICT*, 3:29, 2016.

[30] R. M. Satava. Virtual reality surgical simulator. *Surgical endoscopy*, 7(3):203–205, 1993.

[31] M. Schijven, J. Jakimowicz, I. Broeders, and L. Tseng. The eindhoven laparoscopic cholecystectomy training course-improving operating room performance using virtual reality training: results from the first eaes accredited virtual reality trainings curriculum. *Surgical Endoscopy and other Interventional Techniques*, 19(9):1220–1226, 2005.

[32] T. W. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, 15(2):69–71, 2003.

[33] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. Obrien, V. K. Bansal, D. K. Andersen, and R. M. Satava. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery*, 236(4):458, 2002.

[34] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[35] E. M. Sims. Reusable, lifelike virtual humans for mentoring and role-playing. *Computers & Education*, 49(1):75–92, 2007.

[36] R. Stone. Virtual reality for interactive training: an industrial practitioner's viewpoint. *International Journal of Human-Computer Studies*, 55(4):699–711, 2001.

[37] P. Ström, L. Hedman, L. Särnå, A. Kjellin, T. Wredmark, and L. Felländer-Tsai. Early exposure to haptic feedback enhances performance in surgical simulator training: a prospective randomized crossover study in surgical residents. *Surgical endoscopy and other interventional techniques*, 20(9):1383–1388, 2006.

[38] Student. The probable error of a mean. *Biometrika*, pp. 1–25, 1908.

[39] A. Sutcliffe. *Multimedia and virtual reality: designing multisensory user interfaces*. Psychology Press, 2003.

[40] D. L. Tate, L. Sibert, and T. King. Virtual environments for shipboard firefighting training. In *Virtual Reality Annual International Symposium, 1997., IEEE 1997*, pp. 61–68. IEEE, 1997.

[41] O. A. 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*, 23(6):1180–1190, 2009.

[42] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.