Testing Usability of a Medical Device Using Virtual Reality.

Laura Douze, E. Netter, L. Wallard, F. Occelli, T. Paget, P. Charlier, Jessica Schiro

Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240593

Testing Usability of a Medical Device Using Virtual Reality

Laura DOUZE[a,1], Elie NETTER[b], Laura WALLARD [c], Florent OCCELLI [d], Thomas PAGET [e], Pierre CHARLIER[f] and Jessica SCHIRO[a]
[a] *Inserm, CIC-IT 1403, Univ. Lille, CHU Lille, ULR 2694 - METRICS : Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France*
[b] *Univ Paris Cité, F-75006 Paris, France*
[c] *Univ Polytechnique Hauts-de-France, LAMIH, CNRS, UMR 8201 –, F-59313 Valenciennes, France*
[d] *Univ. Lille, UFR3S-ILIS, Laboratoire d'expérience Immersive (LEXIM), F-59000 Lille, France*
[e] *Univ. Lille, UFR3S-Médecine, Centre de simulation en santé PRESAGE, F-59000 Lille, France*
[f] *BRAINDEX, F-59000 Loos, France*

Abstract. Before commercialization of a medical device, it is important to evaluate its usability. Traditional methods such as user testing to evaluate usability of medical device encountered difficulties to put participants in simulation conditions that are sufficiently realistic to be representative of real life. Virtual reality can be used to immerse participants in a high-fidelity simulation at a lower cost, but is not widely used today. This feasibility study aims to compare the results obtained between user tests in a real simulated environment and in a virtual reality environment, with feedback on the advantages and disadvantages of both conditions.

Keywords. Usability, Medical Device, Evaluation, Virtual Reality, Methods

1. Introduction

Assessing usability is a regulatory requirement for the marketing of medical devices (MDs) in several countries. For carrying out these evaluations the most widely used method, and one of the most effective, is user testing [1]. This method consists of simulating the use of the device by following scenarios representative of the tasks to be performed by participants representative of the end-users [2]. One of the methodological challenges often raised is to put the participants in simulation conditions that are sufficiently realistic for the results obtained to be transposed to use in real life. To be as close as possible to a real-life situation, the ecological validity of the simulation must be considered. Ecological validity is defined as "the extent to which the test environment mirrors the environment in which a product would be used in 'real life'" [3]. It is easy to simulate a simple environment, but more difficult to simulate complex environments with many actors, medical equipment and noise, such as a hospital setting. Reproducing

[1] Corresponding Author: Laura Douze; E-mail: laura.douze@univ-lille.fr.

real-life environmental characteristics in a simulation can be costly and time-consuming [4].

The use of virtual reality (VR) is a possible opportunity to overcome the difficulties to simulate high fidelity environments with a good level of efficiency [5]. VR is already used in several fields that require the creation of simulations, such as training or iterative design in industry [6]. However, in the medical field, VR is still not widely used to design and evaluate MD usability [7], possibly because VR is not currently validated as a standard evaluation method for MDs, nor is it mentioned in guidance as a potential evaluation method. As the internationally recognized IEC 62366-1v2015 standard requires manufacturers to demonstrate that the method used produces objective evidence for the validation of safety of use [2]. The performance of VR to evaluate the usability of MDs has yet to be demonstrated.

We proposed a feasibility study comparing user tests in a real simulated environment and in a VR simulated environment. Comparison criteria were defined and measured on a small panel of participants to evaluate the feasibility of testing a MD in VR, and to understand the strengths and weaknesses of both conditions.

2. Methods

The device tested in this usability evaluation was a multiparametric monitor combining NIRS and EEG technologies used by the anesthesia team during surgery. Three scenarios were defined to test the monitor's main use cases: calibration during the installation of the monitor (sc1), monitoring and detection of sedation problems during induction (sc2), and monitoring and detection of cerebral perfusion problems during surgery (sc3).

The user tests were carried out in a usability laboratory. To create the real simulated environment, a low fidelity operating room was set up (screen displaying scope data, bed with mannequin, sterile table, etc.). The various interactions with other professionals (the operating-room nurse, the nurse anesthetist, the patient and the surgeon) usually present in real environment and necessary to play the scenarios were played by the moderator. The virtual environment was created from 360° videos of a simulated operation in the PRESAGE simulation center. The simulated operation was filmed once. The roles played by the moderator in the real simulated environment were played by different actors in the VR simulation. In the real simulated environment, the participants could directly interact with the monitor and other elements of the environment. In the VR environment interactions with controllers were added in post-production using the Speedernet Sphere software, based on point-and-click principle: screen zoom, actions on the various elements, modification of the scene according to the actions of the participant… (Figure 1).

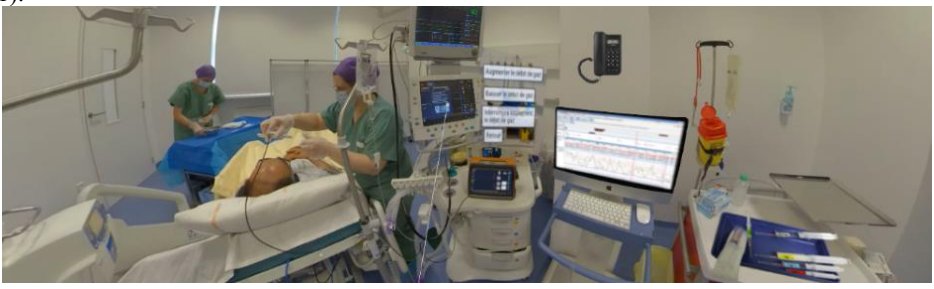

Figure 1. View of a participant in the virtual scene, interacting with the scope in the environment

Four participants took the test individually: two anesthetists and two anesthesia residents. Each participant completed all the scenarios twice, once in the real simulated environment and once in the VR condition using a Meta Quest 2 headset. The order between the two conditions was counterbalanced between participants.

To guide results analysis, criteria for comparing the two experimental conditions were defined. These criteria aimed to assess the ecological validity of the simulations, the performance of the method to gather relevant usability data (use errors, usability problems…) and limitations and advantages to use each method (Table 1). The analysis was carried out afterwards using video capture.

3. Results

There was no difference between real simulated and VR environments on several criteria (response to scenarios, number of artefacts, number of use errors, perceived usability score) (Table 1). In terms of ecological validity, the feeling of presence was higher in VR than in real simulated environment (virtual: 4.7/5; real: 3.6/5). The participants were positive about the virtual environment but felt limited in the possible interactions. Regarding performance, the number of problems detected, and potential improvements was higher in real simulated environment tests (n=6) than in the VR tests (n=4), particularly as the participants interacted directly with the moderator and verbalized more in the real environment. On the limitations and advantages of both conditions, VR requires expensive one-shot hardware cost (camera, software, VR headset, etc.) and human time to create VR scenes (filming, editing videos, adding interactions, etc.), but it facilitates the test-taking process, as only a headset and a laptop were required. The real simulated environment is easier to set up, but requires a usability lab and the environment as to be reproduced for each test.

Table 1. Comparison criteria and results by criteria between traditional user test versus virtual reality user test

	Criteria		Results	
	Effective criteria	**Measurement indicators**	**Real simulated environment**	**VR environment**
Ecological validity	Response to scenarios	Description of participants' response to scenarios during tests	In both simulations, the participants responded in the expected way to the scenarios: they launched the calibration (scenario 1) and identified induction or perfusion problems (scenarios 2 and 3).	
	Artefacts[2]	Number and type of artefacts	n=1 Confusion between a scope alarm and a time countdown	n=0 But several technical problems
	Feeling of presence	Standardized questionnaire (ITC-SOPI [8])	mean=3,6/5 min=2,8, max=4,4	mean=4,7/5 min = 4,4, max=5
	Subjective feedback	Interviews with participants about their experience with both conditions	Fidelity sufficient for the tasks required in the scenarios But less realistic More passive	More immersion in the scene Impression of being less passive and having access to more information But feeling of vertigo Limited number of possible interactions

[2] Behavioral response of a participant caused by the simulation and not representative of real behavior

Performance	Use errors	Number of use errors detected	n=0	n=1 Confusion between the patient's left and right hemispheres
	Usability problems/ ideation	Number of usability problems and improvements detected	n=6 Cable length Launch calibration process Buttons readability Information readability Warning message Minimum actions to navigate the interface	n=4 Flexibility of interface for different situation Differentiation between hemispheres Color code Absence of alarms
	Perceived usability	Standardized questionnaire (French SUS [9])	mean=75/100 min=58, max=93	mean= 79/100 min=58, max=100
Limitations and advantages	Material and human costs	Estimate of material and human costs for implementation, execution and analysis	Cost of a usability laboratory Cost of reproducing the test environment	High cost of equipment for an initial study (camera, software, VR headset, etc.) Significant human time to create virtual environment
	Limitations and advantages	Estimation of the constraints and advantages associated with implementing each environment	Possibility of modulating the scenarios depending on participants' responses Easier interaction between participants and moderator during the test run But need for an available simulation environment Need to bring participants to the usability laboratory	High-fidelity simulation environment that can be transported to participants (no need for a simulation room) Easy repetition of a high-fidelity environment But internet access required Greater risk of bugs or technical difficulties Risk of motion sickness

4. Discussion and Conclusions

The aim of this feasibility study was to test and to assess the usability of a MD in VR, compared with a real simulated environment, which was the gold standard. The results showed that VR tests can be used to collect the same data as in real simulated environment condition. In particular, we could observe and collect use errors and detect usability problems and potential improvements to a device under development. The ecological validity scores and verbalizations collected showed that the participants felt immersed and enjoyed the VR experience.

The use of VR facilitates several aspects that are sometimes difficult or blocking for user test in a real environment. Creating a realistic environment can be costly in terms of time and money [4], and can limit manufacturers in iterative evaluations and/or the implementation of high-fidelity tests. In VR condition, the simulated environment can be reused without resetting the hardware and the manpower between each user test. This allows to achieve a greater realism without increasing simulation costs. Moreover, the environment can be brought to the participant without the need to get him/her to a usability lab, which facilitates access to rare or remote user profiles.

However, some differences have been observed. As participants are immersed in the situation and are cut off from the real world, they verbalize less readily and interact less with the moderator, as observed in other experiments [5]. In this way, user tests in a real simulated environment may be better suited to usability evaluations during the design phase, where the expected level of ecological validity is less important and where

interaction between participant and moderator is needed to identify potential improvements. VR may be better suited to summative, end-of-design validation evaluations because it is not recommended to verbalize during the evaluation so as not to disturb the participant during the test [10]. Finally, evaluation in VR is made easier when the device is an information-taking device, which requires few interactions. Hardware devices may require handling that are more difficult or costly to implement in VR, for example through the use of 3D models. Consideration must be given to determining the cost/benefit of using or not using VR according to the different constraints of the study.

This study has several limitations, including a small number of participants (four) and different levels of ecological validity in the two experimental conditions. It was not possible to access the simulation center and mobilize several actors for each user test; it was easier to film it once for the virtual environment. The aim of this feasibility study was also to identify the advantages and disadvantages of applying the two conditions, which led to different design choices for each experimental condition. To validate the differences identified in this study and understand their underlying causes, a performance study is needed. This study will need to include a larger panel of participants and compare several levels of fidelity in each condition (high and low fidelity in real simulated environment and in VR) to understand whether the differences are caused by the technology used or the participant's level of immersion.

References

[1] Sethumadhavan A, Sasangohar F (eds). *Design for Health: Applications of Human Factors*. 1st edition. London: Academic Press, 2020.

[2] IEC. IEC 62366-1:2015 Medical devices — Part 1: Application of usability engineering to medical devices, https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/31/63179.html (2015, accessed 19 October 2021).

[3] Jordan PW, Thomas DB. Ecological Validity in Laboratory Based Usability Evaluations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 1994; 38: 1128–1130.

[4] Bias RG, Mayhew DJ (eds). *Cost-Justifying Usability: An Update for the Internet Age*. 2nd edition. Amsterdam ; Boston: Morgan Kaufmann, 2005.

[5] Pettersson I, Karlsson M, Ghiurau FT. Virtually the Same Experience?: Learning from User Experience Evaluation of In-vehicle Systems in VR and in the Field. In: *Proceedings of the 2019 on Designing Interactive Systems Conference*. San Diego CA USA: ACM, pp. 463–473.

[6] Liao T, She J. How does virtual reality (vr) facilitate design? A review of vr usage in early-stage engineering design. *Proc Des Soc* 2023; 3: 2115–2124.

[7] Lorenz M, Neupetsch C, Rotsch C, et al. Early Virtual Reality User Experience and Usability Assessment of a Surgical Shape Memory Alloy Aspiration/Irrigation Instrument. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Osaka, Japan: IEEE, pp. 1056–1057.

[8] Lessiter J, Freeman J, Keogh E, et al. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments* 2001; 10: 282–297.

[9] Gronier G, Baudet A. Psychometric Evaluation of the F-SUS: Creation and Validation of the French Version of the System Usability Scale. *International Journal of Human–Computer Interaction* 2021; 37: 1571–1582.

[10] ISO IEC. IEC/TR 62366-2 Medical devices — Part 2: Guidance on the application of usability engineering to medical devices, https://www.iso.org/standard/69126.html (2016, accessed 14 March 2023).