



Revista Brasileira de Computação Aplicada, November, 2019

DOI: 10.5335/rbca.v11i3.9477 Vol. 11, Nº 3, pp. 28-38

Homepage: seer.upf.br/index.php/rbca/index

ORIGINAL PAPER

A Usability Evaluation Experiment of 3D User Interfaces for Elderly: a pilot study

Manoela Rogofski Brum Postal¹ and Rafael Rieder^{10,1}

¹Universidade de Passo Fundo

*manoela.brum@hotmail.com; rieder@upf.br

Received: 2019-05-24. **Revised:** 2019-08-15. **Accepted:** 2019-09-28.

Abstract

Health professionals have used 3D user interfaces as support tools for the elderly rehabilitation, offering fun and beneficial resources for the practice of physical and cognitive activities to them. In this context, it is necessary to establish mechanisms to evaluate the usability of these interfaces, in order to achieve a balance between functionality, ease of use and sense of well-being. This paper aims to report a pilot usability study for a virtual reality game developed specifically for the elderly, as a means to identify the needs of this public regarding 3D user interface evaluation. An initial methodology was tested exploring two points of view in the game – first–and third–person, showing good results for seniors. However, the need to include training periods was noted, and an evaluation with a heterogeneous group of seniors to consolidate and optimize the proposed approach, as well as readjust the instruments used.

Keywords: Usability Testing; Seniors; User Studies; Virtual Reality.

Resumo

Profissionais de saúde têm utilizado interfaces 3D como ferramentas de apoio à reabilitação de idosos, oferecendo recursos divertidos e benéficos para a prática de atividades físicas e cognitivas. Nesse contexto, é necessário estabelecer mecanismos para avaliar a usabilidade dessas interfaces, a fim de alcançar um equilíbrio entre funcionalidade, facilidade de uso e sensação de bem-estar. Este artigo tem como objetivo relatar um estudo piloto de usabilidade para um jogo de realidade virtual desenvolvido especificamente para idosos, como forma de identificar as necessidades desse público considerando a avaliação de interfaces 3D. Uma metodologia inicial foi testada explorando dois pontos de vista no jogo – primeira e terceira pessoa, mostrando bons resultados para os idosos. No entanto, observou-se a necessidade de incluir períodos de treinamento e uma avaliação com um grupo heterogêneo de idosos como forma de consolidar e otimizar a abordagem proposta, além de reajustar os instrumentos utilizados.

Palavras-Chave: Avaliação de usabilidade; Estudos de usuário; Idosos; Realidade virtual.

1 Introduction

Usability is the variety and the degree to which system features can be used efficiently so that the user can accomplish tasks effectively and intuitively (Karray et al., 2008). The interactive systems only can be useful and practical if they have good usability, and the real efficacy of a system is obtained when there is balance between the functionality and the usability. Therefore,

it is important to evaluate the usability of a system to reach best results.

According to Nielsen (2012, 1993), the main usability characteristics to evaluate are the easily and the efficiency during the task performance, the easily to reuse resources, the reestablishment of the services after system faults, and the satisfaction experienced by the participant during the use of the system.

Evaluate the usability is fundamental to establish

the relation between the quality of the interactive system and the quality of the interaction (Cockton, 2012). This author mentions that the methods and the metrics contribute for determinate the usability extension, measuring the robustness, the goals and the reliance – when the usability evaluation points the utility of a system or device. Because of this, it is necessary the utilization of a method or protocol that include and can evaluate reliably all these issues.

Tridimensional user interfaces (3DUI), as Virtual Reality (VR) applications, are becoming popular in the game area, and requires of a usability evaluation protocol to test its interaction process. Serious Games for elderly, for example, have been used in clinical intervention of rehabilitation (Fiorin et al., 2014). This kind of games can help to stimulate the practice of beneficial activities to the human body and increase the interest of the patient for the treatment, because the traditional intervention usually is slow and painful (Broeren et al., 2008).

Applications for seniors require, as any other system, of an evaluation method to test the quality of the interaction, in order to demystify the lack of access, practice or fear of this public (Carvalho and Ishitani, 2013). Besides that, it is important to ensure that the solution be proper to elderly profile. This can motivate a greater use and greater production of these solutions, including by the tendency that the population, in a near future, will be older and prepared for the technological news.

With this in mind, this paper aims to report a pilot usability study for a virtual reality game developed specifically for the elderly, as a means to identify the needs of this public regarding 3DUI evaluation. Therefore, it is possible to obtain subsidies for further development of a specific evaluation methodology for 3DUI used only by elderly. We chose to apply a methodology proposed by Simor (2016), used with seniors in 2D interfaces, making some adjustments to apply it in 3DUI context. It consists of a sequence of steps to identify which interface features meet the specifications foreseen and which need review, contributing to the future project of an evaluation methodology.

This article is organized as follow: Section 2 presents the related work; Section 3 describes our approach; Section 4 demonstrates the approach validation; Section 5 presents the results of the experiment; Section 6 presents the discussions; and Section 7 shows our conclusions about this study.

2 Related Work

Sheu et al. (2015) address issues of how to project a serious game based in gestures so that elderly can play in a safe, conveniently and nice way. The study used two games developed by the authors, EG I and EG II (being the second an optimized version of the first), using the Kinect device. Participants with ages between 60 and 77 years interacted with the applications, executing all the tasks. Preliminary,

the authors showed that, in mean, the tasks were performed more quickly in EG II, comparing with EG I. That is why the users also obtained a better score, suggesting that the improvements in the EG II interface were satisfactory. According to the authors, the work also shows that the technique of selection adopted for the games is tiring and not proper to the elderly public – once that was needed to use the movement of both arms to move the cursor across the screen.

Fang et al. (2015) present a game for training of balance in elderly, and seek to verify the experience positive or negative of the user in relation to the environment developed. Therefore, the authors used the EFS (Evergreen Fitness System), a prototype game that adds six exercises selected by health specialists with the goal to train the balance and strengthen the elderly lower limbs. This study determined that the elderly participants (with ages between 60 and 80 years) appreciated the exercises based in game and showed a positive experience using the EFS. Tests provided important feedbacks about the improvement of the conception of the system, about the adequacy of the six exercises, of the system operation, game design and demonstrated the willing to use. They also verified that the system might include the navigation requiring less learning, corrective feedback and timely warnings when idle.

Harrington et al. (2015) approach the usability challenges of the device Kinect based in exergames for elderly, pointing out which aspects of these programs are of difficult assimilation by elderly people. Tests with ten elderly with ages between 60 and 69 years and ten elderly with ages between 70 and 79 years used two prototypes of games that stimulate physical activities. The research showed the satisfaction of the participants, which admitted that the exergames are beneficial to health and are useful for the incentive to the practice of exercises. However, the approach also showed that there are usability problems in these applications, conforming the age advances. Most of participants in the group 60-69 years old agreed that the interface is user-friendly, while most people of the group 70-79 years old disagreed about the ease to use.

Palacio et al. (2017) evaluate the usability perception of elderly about the use of games with different control devices. Twenty-four elderly participated in the study (12 women, 12 men; mean of 69 years old) and eight children (mean of 8 years old). The configuration included two Kinect motion sensors, three computers, three projectors, three video cameras, two audio devices, the games Angry Birds and Happy Sky, one Xbox 360, one device Nintendo Wii, and one touch screen. Equipment was configured to play in pairs, being one elderly and one child. The data were collected according to the user's experience and social interaction, through individual and group interviews. In the interviews, questions were used to evaluate the perception and the apprehension of the participants while playing. During the tests, the authors evaluated the characteristics of efficiency and number of errors. This study suggested that the game devices for the elderly might be adapted to balance their functional,

sensorial and cognitive limitations.

Considering the related work, we noted that the authors do not apply an evaluation method to the elderly particularities. Since these studies are recent researches, we understand that offer a suitable evaluation process is interesting for designers and developers to comprehend and attend the particularities of elderly, providing better experiences to them. In addition, the systematic review presented by Simor et al. (2016) points to similar results, showing the lack of human-computer interface evaluation mechanisms designed specifically for the elderly users.

3 Methodology

Our approach consists in apply the methodology proposed by Simor (2016), making some adjustments to apply it in 3DUI context: use a VR game and new questionnaires.

We defined three conventional stages: firstly, select candidates and collect important information according to the aim's research (pre-test); secondly, execute tests with participants using a 3DUI and collect performance data (test); and thirdly, evaluate the 3DUI for elderly, considering user preferences by questionnaires and/or interviews, and user performance and system performance data (post-test). In all these stages, data could be collected using registers in paper, software, audio or video.

It is important to highlight that this work is an initial project to study if this approach is useful to evaluate 3DUI interfaces, and to identify possible changes or needs observed in the experiment. Thus, we can contribute to the design of a methodology specifically for evaluations with the elderly subjects.

For the experiment, we used a VR serious game in two moments: firstly, a preliminary evaluation with voluntaries, in May 2016; and, after, an evaluation with the elderly, in August 2016. The local research ethics committee (protocol number CAAE 53589116.8.0000.5342) approved this project.

3.1 Pre-test

During this stage, the researcher must orient each participant to read and fill:

- A Informed Consent Form ICF;
- A Sociodemographic and Background Questionnaire;
- The Mini Mental State Examination MMSE (Bertolucci et al., 1994);
- The Geriatric Depression Scale GDS-15 (Yesavage et al., 1982, Sheikh and Yesavage, 1986).

GDS-15 is a superficial evaluation to verify if any person has some mild degree of depression. According to Sheikh and Yesavage (1986), a depressive subject tends to provide unreliable data because there is a possibility that one's psychological state interferes in the results. Therefore, this test is useful to select participants presenting severe degrees of depression, not indicated to continue the experiment.

questionnaire informs about the satisfaction of the participants with their life and themselves, answering always only "yes" or "no". We used a cut-off point of \geq 5 to indicate clinically important depressive symptoms (Almeida and Almeida, 1999).

MMSE is a quickly test (≤ 10min) to evaluate the cognitive function of the person. It does not require specific material and uses a point-Like the GDS 15, it is useful to select outlier participants, not indicated to continue the experiment. The questionnaire deals with spatial and temporal orientation, immediate and evocate memory, calculation, language naming, repetition, understanding, writing and drawing copy. We used a cut-off point of \geq 25 to literate and \geq 19 to illiterate elderly (Lourenço and Veras, 2006).

sociodemographic and background questionnaire aim to characterize the sample. The questions approach about education, physical and cognitive disabilities, and familiarity with technologies used during the tests.

ICF is a form to explain about the study that you are considering, and for getting permission before conducting an intervention on a person. It also preserves the individual integrity and the collected information only for research studies.

3.2 Test

At this stage, the participant firstly may receive and read an overview document of the experiment, explaining the aims of the test, how to use the application resources (devices and interfaces), and the description of the user tasks. This document also serves to reinforce that the study will evaluate, only and exclusively, the software and the equipment – and not the participant.

The document also informs that the user tasks need to be made naturally. Besides that, it must inform that interaction process registers would be collect by the application itself and by a software of capture of the computer's screens and filming, preserving personal image.

Each participant needs of instructions to verbalize and externalize their actions and thoughts during the interaction process, using the Think Aloud Protocol (Nielsen, 2012). This method helps the observer to perceive some user's difficulties or facilities during the interaction process. After, the observer must ask to the participant if there are any remaining doubts, because questions cannot be answered during the test.

Afterwards, it is time to start the experiment with the elderly. Each participant will test each level of application using VR devices during thirty seconds. This time was defined in order to avoid mental and physical exhaustion (Simor, 2016) - but it can be changed according to the approach. We adopted twominute time interval for rest between each level and the use of the each equipment. This time can be changed too, considering the effort over time.

Table 1: Proposed Questionnaire for 3DUI Evaluation for elderly

N°	Question
	When I used the Oculus Rift
1	I felt comfortable during the interaction in the game, using different equipment
2	I felt immersed, the interaction with the game was transparent, arresting my attention
3	I felt present within the virtual scene of the game, as if I were part of it
4	I felt good, the interaction with the game and the equipment did not cause discomfort, such as motion sickness,
	headache, dizziness or nausea
5	I felt oriented because the equipment provided a better visual perception of the 3D space
	When I used the Smart TV 3D
6	I felt comfortable during the interaction in the game, using different equipment
7	I felt immersed, the interaction with the game was transparent, arresting my attention
8	I felt present within the virtual scene of the game, as if I were part of it
9	I felt good, the interaction with the game and the equipment did not cause discomfort, such as motion sickness,
	headache, dizziness or nausea
10	I felt oriented because the equipment provided a better visual perception of the 3D space
	About the 3D User Interface, it allowed
11	To use and interact easily in the game
12	Clarity on the steps to be followed to perform the tasks in the game
13	Adequate and sufficient time for the execution of the tasks
14	Naturally to perform the tasks of the game, without difficulties
15	Easily to visualize, interpret and understand the interactive elements of the game (visual aspects)
16	To listen and assimilate easily the sound elements of the game
17	Easily to pick up objects in 3D space
18	To have a fun experience
	Another questions
19	The theme of the game is associated with your age
20	Rest intervals during the experiment were sufficient

3.3 Post-test

In this stage, the participant receive a usability evaluation questionnaire about the experiments.

Initially, the usability evaluation contemplates questions about the visualization devices considering these aspects: comfort, immersion, presence, welfare and visual perception. The next part of the questionnaire considers the evaluation of the 3DUI: easily in use the application, easily of the task performance, clarity about the procedure, adequate time to execute the task, quality of the visual and aural elements, connection between scene and task, adequate time interval. The questionnaire uses 5-point Likert Scale.

Table 1 shows the statements. Our questionnaire considers the main usability characteristics defined by Nielsen (1993), and use as basis the questionnaire developed by Simor (2016).

During the filling, the participant can comment about the test openly, allowing a collect of complementary information by the observer. In the end of the session, the observer thanks him/her for the participation.

The time required for the application of our protocol is less than 40 minutes. Table 2 presents the order of steps to apply it.

4 Experiment

In order to validate our approach, we realized an pilot study evaluation using an exergame named Motion

Table 2: Sequence of steps of the proposed protocol

Step	Step Description				
Pre-test	Filling out forms	~20 min			
Test	Experiment instructions and wear the equipment	∼6 min			
	Experiment using device 1	~30 s			
	Rest	\sim 2 min			
	Experiment using device 2	~30 s			
	Rest	\sim 2 min			
Post-test	Usability evaluation questionnaire	∼8 min			

Rehab AVE 3D (Trombetta et al., 2017). According to Fiorin et al. (2014), it is a software to help health professionals in activities of motor and cognitive rehabilitation of elderly. Fig. 1 shows the game interface, first-person (left) and third-person (right). To interact, the user can wear a head-mounted display or use a Smart TV 3D, plus a Kinect motion sensor.

For this evaluation, each participant interacted twice in the same level of the game, using two different visualization devices: an Oculus Rift DK 1 (HMD), and a Smart 3D TV 46" (TV). The goal was to evaluate the usability difference between the experiments with a group of people using the same game scenario on two different display devices. We also considered the methodology used in this study.

We determined to select a small group of elderly



Figure 1: First-person (left) and third-person views (right) of the Motion Rehab AVE 3D game.

volunteer subjects, balanced by gender. Based on Benyon (2014), the reduced number of participants is justified because the evaluation is destined to a group relatively homogeneous. As inclusion criterion, the individuals must be literate, with 60 years old or more, without cognitive or motor commitment and without severe depressive symptoms. All the participants received orientations about the goals of the research and signed the ICF.

4.1 Subjects

For the experiment, we obtained the voluntary adhesion of twenty subjects (60+ years) of a center of reference and attention to the elderly (CREATI). This center offers programs and services to the elderly, with varied educational, physical, technical, mental, cultural, social, civic and affective activities.

In order to balance the comparison we defined that half of the participants (ten subjects) would test the game using the first-person version, and the other half (ten subjects) the third-person version. In both cases, the two display devices were used (Oculus Rift and Smart TV 3D).

We also defined the counterbalance of the participants to use the visualization devices. For each version tested of the game, half of the participants of each group (five subjects) tested firstly with the HMD, and secondly with the TV, and the other half (five subjects) inversely: TV and HMD. Table 3 shows the distribution of groups.

In this context, we described the following groups for future analysis:

- HMD1 and HMD2: participants using the HMD, testing it as the first or second device during the experiment, respectively;
- TV1 and TV2: participants using the TV, testing it as the first or second device during the experiment, respectively;
- HMDTV: participants using the HMD firstly, and the TV secondly;
- TVHMD: participants using the TV firstly, and the HMD secondly.

For the statistical analysis, we applied the Shapiro-Wilk test to verify the normality, and the Mann-Whitney U to compare the samples and test the hypotheses.

Table 3: Study Design.

Distribution of groups	HMD1 – 5 subjects	HMD2 - 5 subjects
TV1 - 5 subjects	-	TVHMD - 10 subjects
TV2 - 5 subjects	HMDTV – 10 subjects	_

4.2 Task

The task considers exercises of the game level 1. They encourage the use of the upper and lower limbs and memorization of the objects. During the interaction process, the subject remains standing. Fig. 2 illustrates the use of game with the two devices.

4.3 Test Environment

Material and configurations, as well as the instructions for a good progress of the experiment are described in Table 4.

Results

The experiment was attend by 23 participants. From the sociodemographic and background data, we noted that three users did not meet the inclusion criteria, and their results were excluded.

Therefore, the sample analyzed considered 20 participants with ages between 60 and 81 years old, 16 females and 4 males. Of these, 25% (five subjects) already have played some computer game and 10% (two subjects) have known one of the devices used in the

A ten people group firstly played the game in first-



Figure 2: Two participants testing the game in first-person (left, HMD) and third-person (right, TV).

person version, and secondly in third-person; and the other ten inversely. Within each group, half of the participants (five) started wearing the HMD and after using the TV; and the other half inversely too.

We elaborated four alternatives hypotheses to analyze the results:

- There is difference between using the HMD and the TV to play the Motion Rehab AVE 3D (HA, General);
- There is difference between using the HMD in the first experience, and using the HMD in the second experience (HB, HMD1 x HMD2);
- There is difference between using the TV in the first experience, and using the TV in the second experience (HC, TV1 x TV2);
- There is usability difference between the groups of the experiment (HD, HMDTV x TVHMD).

The following subsections present the results of the statistical analysis, posteriorly considered in the Discussion Section.

5.1 Results independent of version for Motion Rehab AVE 3D

The tests applied to evaluate HA, HB and HC did not present results statistically significant, rejecting the alternative hypotheses (Table 5). The evaluation did not consider the game version (first- or third-person). For the HA (U-critical value = 127), the comparison considered the 20 participants. For the HB and HC

hypotheses (U-critical value = 23), the comparison considered the order of the devices used (between groups of 10 participants).

Table 6 presents the results of the HD hypothesis (U-critical value = 23), checking for some usability difference at the evaluated game. It considers elements about the 3DUI interface (e.g. visual and aural feedbacks and game theme) using different devices. The tests did not present statistically significant results. The evaluation compared HMDTV and TVHMD groups of 10 participants.

5.2 Results for Motion Rehab AVE 3D - First-Person Version

For the first-person version, the analysis did not point to statically significant results, shown in Table 7. For the HA (U-critical value = 23), the comparison considered 10 participants. For the HB and HC hypotheses (U-critical value = 2), the comparison considered the device used during first-person experiences (groups of 5 participants).

Table 8 presents the results of the HD hypothesis (U-critical value = 2), checking for some usability difference at the evaluated game during the first-person experience – without statistically significant results.

Name

Smart TV 3D Kinect

Oculus Rift

Calibration

Illumination

Chair **Tables**

Film Cam

Orientation

equipment.

comfortably.

TV positioned for the participant comfortably visualize the game interface. Motion sensor must be between 0.6 m and 1.8 m of floor height and at a minimum distance of 1.2 m from the participant, without obstacles between participant and sensor. HMD must wear the participant in a comfortable way. Participants using eyeglasses may use it with the It is essential to calibrate the Kinect sensor and the Oculus Rift for each user, in order to provide a natural and enjoyable interaction.

Table 4: Experiment Material

Table 5: Results for HA, HB and HC hypotheses

It must be close to the participant for rest and support, at least 1.2m from the motion sensor.

It is necessary one table to put the TV plus Kinect, and other to the participant fill out the requested forms,

To recognize the user movements correctly, the local must have a good illumination.

Cam positioned diagonally, \sim 1.5m from the participant (horizontal angle: 30 degrees).

Hypothesis: Questions	U-value	Z-Score	p-value	Mean \pm Std. Deviation
HA: Q1 x Q6	182.0	-0.47	0.64	$4.30 \pm 1.17 \text{ x } 4.60 \pm 0.75$
HA: Q2 x Q7	192.5	-0.19	0.85	$4.15 \pm 1.14 \text{ x } 4.20 \pm 1.11$
HA: Q3 x Q8	183.5	0.43	0.67	$4.65 \pm 0.67 \text{ x } 4.55 \pm 0.76$
HA: Q4 x Q9	177.0	-0.61	0.54	$4.55 \pm 1.05 \text{ x } 4.90 \pm 0.31$
HA: Q5 x Q10	181.0	-0.50	0.62	$4.70 \pm 0.73 \text{ x } 4.60 \pm 0.75$
HB: Q1 - HMD1 x HMD2	41.5	0.60	0.55	$4.60 \pm 0.70 \text{ x } 4.00 \pm 1.49$
HB: Q2 - HMD1 x HMD2	45.5	-0.30	0.76	$4.00 \pm 1.33 \text{ x } 4.30 \pm 0.95$
HB: Q3 - HMD1 x HMD2	45.5	-0.30	0.76	$4.60 \pm 0.70 \text{ x } 4.70 \pm 0.67$
HB: Q4 - HMD1 x HMD2	48.5	0.07	0.94	$4.70 \pm 0.67 \text{ x } 4.40 \pm 1.35$
HB: Q5 - HMD1 x HMD2	30.0	1.47	0.14	$5.00 \pm 0.00 \text{ x } 4.40 \pm 0.97$
HC: Q6 - TV1 x TV2	48.5	-0.07	0.94	$4.70 \pm 0.48 \text{ x } 4.50 \pm 0.97$
HC: Q7 - TV1 x TV2	39.0	-0.79	0.43	$4.30 \pm 1.25 \text{ x } 4.10 \pm 0.99$
HC: Q8 - TV1 x TV2	47.0	-0.19	0.85	$4.50 \pm 0.97 \text{ x } 4.60 \pm 0.52$
HC: Q9 - TV1 x TV2	40.0	-0.72	0.47	$5.00 \pm 0.00 \text{ x } 4.80 \pm 0.42$
HC: Q10 - TV1 x TV2	39.0	0.79	0.43	$4.40 \pm 0.97 \text{ x } 4.80 \pm 0.42$

5.3 Results for Motion Rehab AVE 3D - Third-Person Version

The analysis also did not suggest statistically significant results for third-person version (Table 9). The methodology analysis is the same of the Section 5.2, and the U-critical values remain unaltered.

Table 10 presents the results of the HD hypothesis (U-critical value = 2). In the same way that the previous analyzes, the results are not statistically significant.

Discussion

Next sections discuss the results about our methodology initial proposal (Section 6.1) and about the usability of the Motion Rehab AVE 3D (Section 6.2).

6.1 Methodology Evaluation

As stated in Section 5, there are not statistically significant results for any hypothesis presented. These results demonstrate that, for elderly, both the devices (HMD Oculus Rift and Smart TV 3D) can offer sense of comfort and well-being during the interaction process. The use of different devices does not interfered in the

ease of use, in the ease of execute the task, in the procedures' clarity, in the time to rest, and during the user interface interaction. Palacio et al. (2017) also report for this panorama, indicating that elderly might not percept any difficult during their first experience with new equipment. They take more time to adapt to new technologies.

In related work (Section 2), the authors used metrics as speed to execute the task and ranking points (Sheu et al., 2015), ease of use (Harrington et al., 2015), and accuracy (Palacio et al., 2017) to evaluate the system usability. All these metrics refer to the task performance and user preference during the use of system, but the authors do not determinate a default evaluation method to these requirements for elderly. For this reason, we propose a protocol to consider these and other usability metrics noted by Nielsen (2012).

According to the tests' observers, they perceived that the elderly were not comfortable in a 3D space because they are not accustomed with this technology. This indicate that our protocol needs a training moment, before each experiment, in order to familiarize users with the equipment. It is important to offer an adaptation period, because the elderly can have a critical view about the experiment during the evaluation process. This behavior can contribute to improve the user task performance too.

Table 6: Results for HD hypothesis

Hypothesis: Questions	U- value	Z- Score	p- value	$\begin{array}{cc} \textbf{Mean} \\ \pm & \textbf{Std.} \\ \textbf{Deviation} \end{array}$
HD: Q11	45.5	0.30	0.76	4.60 ± 0.75
HD: Q12	45.0	0.34	0.73	4.95 ± 0.22
HD: Q13	42.0	0.57	0.57	4.40 ± 1.27
HD: Q14	47.0	-0.19	0.85	4.50 ± 0.95
HD: Q15	50.0	0.04	0.97	4.90 ± 0.31
HD: Q16	44.5	0.38	0.70	4.60 ± 0.75
HD: Q17	39.0	0.79	0.43	4.60 ± 0.75
HD: Q18	46.0	0.26	0.79	4.60 ± 0.75
HD: Q19	43.5	-0.45	0.65	4.60 ± 0.75
HD: Q20	45.0	-0.34	0.73	4.60 ± 0.75

Another situation that highlights the need of a training session is that, to the large majority of the participants, the game was not intuitive (although its simplicity). The elderly do not understand the game in the first contact with the interface and they request tips about how to interact with it.

In particular, 40% (eight elderly) of the participants commented that, if they could test the game before using it, they could have better performance during the test. This happens because of the lack of intimacy with the devices, and the lack of practice and access for this type of technology (Carvalho and Ishitani, 2013). In addition, a training session can make the new users have a gradual contact and curiosity with the equipment, avoiding cybersickness and some discomfort like dizziness and tiredness (cited by some

During the experiments, we perceived that it is possible to understand better the feedback of this kind of public through of interviews. Palacio et al. (2017) also used interviews (collective and individual) in order to obtain data, because this instrument allows social interaction to identify the lived experiences of the users.

We also identified that the use of Likert Scale for the elderly is complicated, because it is needed to guide how they could evaluate considering the points of scale. In this context, it is very important to detail each statement in a simple, straightforward way, in order to keep the participant focused. A glossary can also be offered, containing expressions and less familiar words to this age group (e.g., immersion, virtual reality, interface) to easily understand the context. Revisiting the related work, we noted that no specific questionnaire for elderly is used.

During the experiment, we observed that the abstraction of terms as game, 3D and virtual

environment, is complicated to the elderly. They were simply lost when they read expressions like "I felt immersed", "3D user interface", "steps to be followed in order to execute game tasks", "interactive game elements" and "I felt oriented because the equipment provided a better visual perception of the 3D space".

These expressions sound easy in the entertainment On the other hand, for elderly is important to take simple approaches in questionnaires and interviews. Explicitly, only two participants commented that the post-test questionnaire and Likert-type Scale are hard to understand. However, the observer indicated that 60% (12 participants) had difficulties in the comprehension of the questionnaire.

To solve this problem, we suggested the use of a text description for each point of the scale and the use of a glossary about the terms of the game. Another possibility to improve the descriptions of the questionnaire is to present them before the experiment. So, doubts about words and expressions could be explained by the researches verbally or using videos.

In our experiment, none user was excluded by GDS-15 and MMSE, becoming questionable the importance of these instruments in Pre-test stage. However, we considered the use of them, because our sample were a homogeneous group of healthy elderly, participating of specific programs for their age. Anyway, it is important to validate the importance of MMSE and GDS-15 in this protocol with a heterogeneous group. Related work no elucidated the use of these kind of instruments.

For this reason, we suggest for future evaluations that the experiments must realize in two sessions: a session only to Pre-test stage and training, and another exclusive to Test and Post-test stages.

6.2 Usability Evaluation

Although we have been identified points to improve the proposed protocol, the process also allowed analyzing the quality of the game, considering the usability evaluation of the 3DUI for two different visualization

We identified that 20% (four participants) presented some difficulties in terms of spatial orientation. They did not capture the objects proposed during the task because they did not perceive that it was necessary to open their arms in a slightly larger angle (physically) to reach the objects. Another 20% (four participants) presented this same difficult in the beginning of the interaction process – but, along of the experiment they perceived what should be done and they were able to execute the test normally.

The tests also showed, subjectively, considering the user opinion and the observations of the observer, that visual and aural feedbacks must be more intuitive to help in the user understanding and cause more immersion. Moreover, also subjectively, we viewed that the people that tested the game in first-person had more difficulty using the TV and more facility using the HMD, whilst those who tested in third-person had more facility using the TV than using the HMD.

Regarding our observations, 25% (five participants)

Hypothesis: Questions	U-value	Z-Score	p-value	Mean \pm Std. Deviation
HA: Q1 x Q6	42.0	-0.57	0.57	4.00 ± 1.33 x 4.40 ± 0.97
HA: Q2 x Q7	48.0	0.11	0.91	$4.00 \pm 1.49 \text{ x } 4.20 \pm 1.23$
HA: Q3 x Q8	37.5	0.91	0.36	$4.70 \pm 0.67 \text{ x } 4.50 \pm 0.53$
HA: Q4 x Q9	44.5	-0.38	0.70	$4.70 \pm 0.67 \text{ x } 4.90 \pm 0.32$
HA: Q5 x Q10	40.0	0.72	0.47	$4.80 \pm 0.42 \text{ x } 4.60 \pm 0.52$
HB: Q1 - HMD1 x HMD2	12.5	0.10	0.92	$4.20 \pm 0.84 \text{ x } 3.80 \pm 1.79$
HB: Q2 - HMD1 x HMD2	8.0	-0.84	0.40	$3.60 \pm 1.67 \text{ x } 4.40 \pm 1.34$
HB: Q3 - HMD1 x HMD2	12.0	0.00	1.00	$4.80 \pm 0.45 \text{ x } 4.60 \pm 0.89$
HB: Q4 - HMD1 x HMD2	12.5	0.10	0.92	$4.80 \pm 0.45 \text{ x } 4.60 \pm 0.89$
HB: Q5 - HMD1 x HMD2	7.5	0.94	0.35	$5.00 \pm 0.00 \text{ x } 4.60 \pm 0.55$
HC: Q6 - TV1 x TV2	11.5	-0.10	0.92	$4.60 \pm 0.55 \text{ x } 4.20 \pm 1.30$
HC: Q7 - TV1 x TV2	8.5	-0.73	0.47	$4.40 \pm 1.34 \text{ x } 4.00 \pm 1.22$
HC: Q8 - TV1 x TV2	10.0	-0.42	0.67	$4.60 \pm 0.55 \text{ x } 4.40 \pm 0.55$
HC: Q9 - TV1 x TV2	10.0	-0.42	0.67	$5.00 \pm 0.00 \text{ x } 4.80 \pm 0.45$
HC: Q10 - TV1 x TV2	7.5	0.94	0.35	$4.40 \pm 0.55 \text{ x } 4.80 \pm 0.45$

Table 7: Results for HA, HB and HC hypotheses, first-person experiences.

Table 8: Results for HD hypothesis, first-person experiences

Hypothesis: Questions	U- value	Z- Score	p- value	$\begin{array}{c c} \textbf{Mean} \\ \pm & \textbf{Std.} \\ \textbf{Deviation} \end{array}$
HD: Q11	12.0	0.00	1.00	4.70 ±0.67
HD: Q12	12.5	0.10	0.92	5.00 ±0.00
HD: Q13	10.0	0.42	0.67	4.80 ±0.63
HD: Q14	12.5	0.10	0.92	4.60 ±0.52
HD: Q15	10.0	0.42	0.67	4.90 ±0.32
HD: Q16	9.5	0.52	0.60	4.40 ±1.26
HD: Q17	5.0	1.46	0.14	4.20 ±1.48
HD: Q18	10.0	0.42	0.67	4.90 ±0.32
HD: Q19	11.0	-0.21	0.83	4.20 ±1.03
HD: Q20	10.0	-0.42	0.67	4.80 ±0.63

presented the best performance (hit everything) during the task, choosing all the selectable objects and ignoring the distractors objects. other hand, 20% (four participants) got terrible performance (missed everything), not being able to hit any of the selectable objects. However. independently of the users' task performance, all the participants demonstrated very interest in exergames and technologies projected for elderly. They classified the experience as interesting and beneficial to the age.

In addition, although it is not part of the protocol, we suggested collect user performance data (e.g. hits, errors, total time) in order to contribute for the protocol validation. We believed that this data could reduce the subjectivity of the user satisfaction, considering situations as feeling of immersion, spatial orientation

and 3DUI quality.

Conclusion

The evolution of our approach in a specific evaluation methodology for 3 DUI can be useful for the creation of new VR solutions, like games or simulators for elderly, exclusively. We also understood, because of the results of our experiment, that this kind of approach guarantee that the interface at least meets the accessibility needs of this age group, once that became explicit the elderly user experience with the interface. Researchers and developers also may direct efforts to improve the quality of the technology, and offer more comfort, welfare and satisfaction to the user. In addition, a specific evaluation methodology for 3DUI used only by elderly may be useful to evaluate new projects of same purpose, either academic or professional.

We still pretend to do the suggestions indicated in our protocol and evaluate the necessity of the GDS-15 and MMSE instruments. Regardless of these two instruments, it is necessary an evaluation protocol differentiated to elderly for age-related issues, as speed of understanding of technological terms and use of more recent equipment (HMD, motion sensors, smartphones, etc.).

It is important to highlight that, at the end of each experiment, all participants classified verbally as a great and wonderful experience for the age. In order to identify elderly users' preferences, we suggested that the inclusion of the interview in Post-test stage.

We concluded that this kind of experiment is not exhausting to the elderly and it can trigger the curiosity for technological news. Even without understanding or being able to play and achieve a good performance, the participants showed interest to use games for fun, for physical and mental exercise, and for a healthy lifestyle. As future work, we intended to validate the protocol with different elderly groups, in order to present a final instrument for the academic and professional communities.

Hypothesis: Questions	U-value	Z-Score	p-value	Mean \pm Std. Deviation
HA: Q1 x Q6	49.0	-0.04	0.97	4.60 ±0.97 x 4.80 ±0.42
HA: Q2 x Q7	49.0	-0.04	0.97	$4.30 \pm 0.67 \text{ x } 4.20 \pm 1.03$
HA: Q3 x Q8	46.0	-0.26	0.79	$4.60 \pm 0.70 \text{ x } 4.60 \pm 0.97$
HA: Q4 x Q9	44.0	-0.42	0.67	$4.40 \pm 1.35 \text{ x } 4.90 \pm 0.32$
HA: Q5 x Q10	50.0	0.04	0.97	$4.60 \pm 0.97 \text{ x } 4.60 \pm 0.97$
HB: Q1 - HMD1 x HMD2	7.5	0.94	0.35	5.00 ±0.00 x 4.20 ±1.30
HB: Q2 - HMD1 x HMD2	9.5	0.52	0.60	$4.40 \pm 0.89 \text{ x } 4.20 \pm 0.45$
HB: Q3 - HMD1 x HMD2	9.5	-0.52	0.60	$4.40 \pm 0.89 \text{ x } 4.80 \pm 0.45$
HB: Q4 - HMD1 x HMD2	12.0	0.00	1.00	4.60 ±0.89 x 4.20 ±1.79
HB: Q5 - HMD1 x HMD2	7.5	0.94	0.35	5.00 ±0.00 x 4.20 ±1.30
HC: Q6 - TV1 x TV2	12.5	0.10	0.92	$4.80 \pm 0.45 \text{ x } 4.80 \pm 0.45$
HC: Q7 - TV1 x TV2	11.0	0.21	0.83	4.20 ±1.30 x 4.20 ±0.84
HC: Q8 - TV1 x TV2	12.0	0.00	1.00	$4.40 \pm 1.34 \text{ x } 4.80 \pm 0.45$
HC: Q9 - TV1 x TV2	10.0	0.42	0.67	$5.00 \pm 0.00 \text{ x } 4.80 \pm 0.45$
HC: Q10 - TV1 x TV2	12.0	0.00	1.00	$4.40 \pm 1.34 \times 4.80 \pm 0.45$

Table 9: Results for HA, HB and HC hypotheses, third-person experiences.

Table 10: Results for HD hypothesis, third-person experiences.

Hypothesis: Questions	U- value	Z- Score	p- value	$\begin{array}{c c} \textbf{Mean} \\ \pm & \textbf{Std.} \\ \textbf{Deviation} \end{array}$
HD: Q11	12.0	0.21	0.83	4.60 ±0.70
HD: Q12	12.5	0.42	0.67	4.90 ±0.32
HD: Q13	10.0	0.31	0.76	4.00 ±1.63
HD: Q14	12.5	-0.21	0.83	4.40 ±1.26
HD: Q15	10.0	-0.42	0.67	4.90 ±0.32
HD: Q16	9.5	0.10	0.92	5.00 ±0.00
HD: Q17	5.0	-0.52	0.60	4.10 ±1.20
HD: Q18	10.0	0.00	1.00	4.70 ±0.67
HD: Q19	11.0	-0.31	0.76	3.60 ±1.58
HD: Q20	10.0	0.10	0.92	5.00 ±0.00

References

Almeida, O. P. and Almeida, S. A. (1999). Confiabilidade da versão brasileira da escala de depressão em geriatria (gds) versão reduzida, *Arq Neuropsiquiatr* **57**(2B): 421–6.

Benyon, D. (2014). Designing interactive systems: A comprehensive guide to hci, ux and interaction design.

Bertolucci, P. H., Brucki, S. M., Campacci, S. R. and Juliano, Y. (1994). O mini-exame do estado mental em uma população geral: impacto da escolaridade., *Arquivos de Neuro-psiquiatria*.

Broeren, J., Bjorkdahl, A., Claesson, L., Goude, D., Lundgren-Nilsson, A., Samuelsson, H., Blomstrand, C., Sunnerhagen, K. S. and Rydmark, M. (2008). Virtual rehabilitation after stroke, *Studies in health technology and informatics* **136**: 77.

Carvalho, R. N. S. d. and Ishitani, L. (2013). Fatores motivacionais para desenvolvimento de mobile serious games com foco no público da terceira idade: uma revisão de literatura, ETD-Educação Temática Digital 15(1): 16–32.

Cockton, G. (2012). Usability evaluation, 2 edn, The Encyclopedia of Human-Computer Interaction. Available at http://www.interaction-design.org.

Fang, W.-C., Sheu, F.-R., Lin, Y.-L., Lee, Y.-L. and Chen, N.-S. (2015). Interactive physical games: Improving balance in older adults, *Smart learning environments*, Springer, pp. 159–174. https://doi.org/10.1007/978-3-662-44447-4_9.

Fiorin, M., Marchi, A. D., Colussi, E., Rieder, R. and Trombetta, M. (2014). Motion rehab: Um jogo sério para idosos com sequelas de acidente vascular encefálico, Anais do XIV Workshop de Informática Médica.

Harrington, C. N., Hartley, J. Q., Mitzner, T. L. and Rogers, W. A. (2015). Assessing older adults' usability challenges using kinect-based exergames, International Conference on Human Aspects of IT for the Aged Population, Springer, pp. 488-499. https://doi.org/10.1007/978-3-319-20913-5_45.

Karray, F., Alemzadeh, M., Saleh, J. A. and Arab, M. N. (2008). Human-computer interaction: Overview on state of the art.

Lourenço, R. A. and Veras, R. P. (2006). Mini-Exame do Estado Mental: características psicométricas em idosos ambulatoriais, *Revista de Saúde Pública*40: 712 - 719. https://dx.doi.org/10.1590/S0034-89102006000500023.

Nielsen, J. (1993). Iterative user-interface design, Computer 26(11): 32-41. https://doi.org/10.1109/2.241424.

- Nielsen, J. (2012). Usability 101: Introduction to Available at https://www.nngroup.com/ articles/usability-101-introduction-to-usability.
- Palacio, R. R., Acosta, C. O., Cortez, J. and Morán, A. L. (2017). Usability perception of different video game devices in elderly users, Universal Access in the Information Society 16(1): 103-113. https://doi.org/ 10.1007/s10209-015-0435-y.
- Sheikh, J. I. and Yesavage, J. A. (1986). Geriatric recent evidence and depression scale (gds): development of a shorter version., Clinical Gerontologist: The Journal of Aging and Mental Health.
- Sheu, F.-R., Lee, Y.-L., Yang, S.-J. and Chen, N.-S. (2015). User-centered design of interactive gesturebased fitness video game for elderly, Emerging Issues in Smart Learning, Springer, pp. 393-397. https:// doi.org/10.1007/978-3-662-44188-6_54.
- Simor, F. W. (2016). Método de avaliação da usabilidade para "serious games" com interação gestual para idosos, Master's thesis, Universidade de Passo Fundo.
- Simor, F. W., Brum, M. R., Schmidt, J. D. E., Rieder, R. and De Marchi, A. C. B. (2016). Usability evaluation methods for gesture-based games: a systematic review, JMIR serious games 4(2): e17. https://doi.org/ 10.2196/games.5860.
- Trombetta, M., Bazzanello Henrique, P. P., Brum, M. R., Colussi, E. L., De Marchi, A. C. B. and Rieder, R. (2017). Motion rehab ave 3d, Comput. Methods Prog. Biomed. 151(C): 15-20. https://doi.org/10.1016/j. cmpb.2017.08.008.
- Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M. and Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: a preliminary report, Journal of psychiatric research 17(1): 37-49.