Explore 360° VR to Improve the Ecological Validity of Screening Tests on Cognitive Functions

Francesco Vona
francesco.vona@polimi.it
Politecnico di Milano
Milan, Italy

Luca Pieri
l.pieri3@campus.unimib.it
Università degli Studi di Milano-Bicocca
Milan, Italy

Alberto Patti
alberto.patti@polimi.it
Politecnico di Milano
Milan, Italy

Simone Tafaro
simone.tafaro@mail.polimi.it
Politecnico di Milano
Milan, Italy

Sara Saccoccio
sara.saccoccio@mail.polimi.it
Politecnico di Milano
Milan, Italy

Daniele Romano
daniele.romano@unimib.it
Università degli Studi di Milano-Bicocca
Milan, Italy

Franca Garzotto
franca.garzotto@polimi.it
Politecnico di Milano
Milan, Italy

ABSTRACT

Cognitive impairment is a condition that results in a person's inability to remember, learn, concentrate or make decisions that affect his/her everyday life. The assessment of these deficits is usually performed using standardized paper and pencil or computerized tests within a controlled clinical setting. Many traditionally designed tools show only low to moderate levels of ecological validity, limiting the reliability of the collected measures. The proposed system adapts existing screening tests within an immersive virtual reality environment with 360° video, recreating a familiar setting for the patient. This faithful reproduction of everyday environments and situations can enhance the ecological validity of the assessment procedure while maintaining a standardized stimuli delivery, all in a controlled and safe setting. As a computerized system, virtual reality technology allows an error-free computation of the test scores, here collected by means of accuracy for each task. The system involves many technologies aimed at capturing any kind of user input provided by the patient. Additionally, using a visor with integrated eye-tracker sensor, the system can register the visual exploration pattern adopted by the patient during the task execution, providing information concerning the attentional and visuo-spatial functioning which are not obtainable using traditional assessment procedures. Finally, the results of an exploratory study that was conducted with 11 users on the reliability and usability of the system are presented.

AVI 2022, June 6–10, 2022, Frascati, Rome, Italy

ACM ISBN 978-1-4503-9719-3/22/06…$15.00
https://doi.org/10.1145/3531073.3531171

CCS CONCEPTS

• **Human-centered computing** → *Virtual reality*; Empirical studies in interaction design.

KEYWORDS

Cognitive Impairment, 360video VR screening test, Pico Neo 3 Pro Eye, Eye tracking

ACM Reference Format:
Francesco Vona, Luca Pieri, Alberto Patti, Simone Tafaro, Sara Saccoccio, Daniele Romano, and Franca Garzotto. 2022. Explore 360° VR to Improve the Ecological Validity of Screening Tests on Cognitive Functions. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces (AVI 2022), June 6–10, 2022, Frascati, Rome, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3531073.3531171

1 INTRODUCTION

Cognitive impairment refers to a deterioration of intellectual abilities often related to the occurrence of acute brain lesions (e.g., a stroke) or neurodegenerative diseases (e.g., Alzheimer's disease). This condition includes different levels of severity: with mild impairment patients report changes in their behaviour and cognitive efficiency but are still able to carry out daily activities, while severe impairment can result in a total inability to live independently. The evaluation of a suspect cognitive impairment relies on a first-level formal assessment, usually performed through screening tests in which the patient is asked to perform behavioural tasks related to the cognitive domains tested (e.g., words recall as a memory measure). The results collected during the testing phase are then compared with normative samples to determine whether they are within the normative range or not. Currently, there is an ongoing debate concerning the ecological validity of these kind of assessment tools [4]. In fact, classical neuropsychological tests seems to be limited in describing the actual everyday cognitive functioning of patients. As a result, some of them may show a good performance during the test, but still report difficulties in real life and vice versa. Recently, Virtual Environments (VEs) have been explored as a way to improve the ecological validity and the experimental control of these screening tests by recreating real-world scenarios through the use of immersive media (e.g., head-mounted displays). VEs are usually designed using 3D models, but there is a particular category of virtual environments that is created from 360° images or videos. These environments are quick and easy to create but internal user interaction is limited to 3 degrees of freedom (d.o.f.). In this paper we present a system that integrates 360° virtual reality screening tests with a data collection system specifically designed for each task. 360° VR Cognitive Examination enriches the users' experience allowing them to interact with the surrounding environment. Furthermore, it automatically collects both quantitative and qualitative data to support the evaluation of cognitive performances, maintaining strong experimental control and obtaining enhanced ecological validity. We performed one exploratory study N=11 users to test the system usability and reliability. The results suggest that 360° VR Cognitive Examination is a usable tool for collecting reliable data concerning neuropsychological performance, encouraging a future use in clinical contexts.

2 STATE OF THE ART

In recent years the number of studies concerning novel VR-based neuropsychological tests grew rapidly [14]. Most of these tools were implemented using 3D VEs [10], while 360°VEs emerged more recently with the widespread availability of 360° cameras. The use of 360° videos or images as virtual environments has attracted the attention of researchers interested in designing more ecological solutions for neuropsychological tests. This is due to the fact that virtual environments with 360° video are more photorealistic compared to 3D VEs and are also easier to implement, not requiring any special programming skills. Few works in literature have already showed that 360°VEs can provide a more ecological assessment of executive functions [13] and memory [12] . However, these approaches lacked of an integrated system that collects measures of cognitive functioning, such as those related to the patient's behaviour in the simulated scenario. In fact, 360° VEs are only used as a background to deliver stimuli, not allowing an active interaction. Furthermore, the user's responses are collected separately on a paper protocol.

3 APPLICATION DESIGN

The 360°VR Cognitive Examination (CE-360°) is an innovative neuropsychological assessment tool developed using 360° immersive photo and video as virtual environments. VEs were recorded using the Insta360 One X, an omnidirectional video camera which can record spherical photos with a resolution of 6080x3040 pixels and spherical videos with a resolution of 5760x2880 pixels. The final version of the system consists of a custom Android application targeting Pico headsets, in particular we used a Neo 3 Pro Eye for the implementation.

3.1 Design of the Experience

The CE-360° experience consists of two phases: a familiarization phase and an assessment phase. The screening tests used in the assessment phase were taken from existing tests in the literature and translated in the corresponding virtual activities.

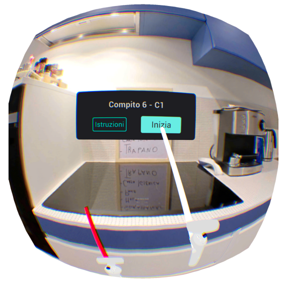

Figure 1: User Interface

Figure 2: Target selection

3.1.1 Familiarization phase. The Familiarization phase is not taken into account in the calculation of the test final score but it is part of the experience as it instructs the users on the exploration modality, the use of controllers and eventually detect any side effects related to VR exposure (e.g., motion sickness).

Starting the experience, users find themselves in a neutral environment where they are given auditory instructions. They can choose whether to listen again or to start the task by selecting one of the two buttons on the user interface. In each task, the user can always make a choice as the one represented in Fig.1. The instructions invite them to explore the environment and select different buttons. To interact with the elements inside the scenario users have to trigger a lever on the headset controller. Throughout the experience they are seated on a swivel chair to make exploration with 3 degrees of freedom more smooth. When this phase is completed, they are invited to proceed to the assessment phase.

3.1.2 Assessment phase. The assessment phase consists of two tasks. Each of them evaluates different cognitive functions. The first is the visual exploration task. Users are placed at the centre of a living room and when the task starts, they are required to explore the environment and name all the objects they identify within a minute. This test is an adaptation of free visual exploration (FVE) task, which has been shown to be more sensitive when using eye-tracking technology [8]. For these reasons, two background processes have been added to the system that run in parallel while the user is speaking. The first one consists in recognising the pronounced words, transcribing and labelling them: using an external service to translate words in text, the system checks their correctness and eventually register the timestamp. The second one, exploiting the embedded eye-tracking sensor in the visor, creates a mapping of the visual exploration producing an heatmap. The second task is the target stimulus selection task. Also in this setting users are in the centre of the same living room, but yellow and green cups and glasses have been added to the scene. When the task starts, users must scan for the yellow cups and use the controller ray caster to point and select them (Fig.2).

The system computes the final score of the task relying on the concept of the Signal Detection Theory (SDT) [6]. The starting point for SDT is that nearly all reasoning and decision making takes place in the presence of some uncertainty [3]. To recreate this condition we used 36 stimuli in the scene. Of these, only 11 are defined as target (yellow glasses), the remaining 25 (green cups and glasses) are distractors. The user must be able to recognise the target objects and select only the yellow glasses. As we need to translate the users' action into quantitative data we used same metrics found in the SDT, namely:

- Hit Rate (Hit): the number of target trials selected over the total number of them.
- False Alarm Rate (FA): the number of non target trials selected over the total number of green glasses.
- Sensitivity (d'): the difference between the z-transforms of the Hit Rate and the False Alarm Rate. It represents the final score of the task.

4 EXPLORATIVE STUDY

Verifying the accuracy of the system results in the assessment process of cognitive impairment is a very long and complex task. Therefore, before conducting a rigorous study along these lines, we wanted to test the features implemented in the system, focusing on the proper functioning of the system itself and the collection of data. For these reasons we carried out an explorative study with N=11 users.

4.1 Sample

The study involved N=11 healthy participants subjects, age 21 to 26. Each subject had little or no experience with VR. All participants were contacted in advance and were recruited voluntarily. Exclusion criteria and study procedure, goals and data treatment were provided through appropriate documentation. Prior to the beginning of the study informed consents were collected. Furthermore, each participant could leave the study at any time and that they could request the deletion of any sensitive data. The Ethical Committees of Politecnico di Milano approved the study protocol and authorised its execution. The conditions and restrictions dictated by Covid19 negatively impacted the user recruitment phase. For this reason, it was decided to recruit a sample consisting of healthy subjects and not the ideal target of the application. This does not diminish the results obtained as the usability and acceptability of Virtual Reality in elderly patients has already been investigated in several studies [15] [7].

4.2 Variables and metrics

Data gathering methods include:

- *Quantitative measure* administrated to the subjects at the end of the session:
  - SUS: System Usability Scale questionnaire[2].
  - Keyword recognition: Number of pronounced keywords
  - Target counter: Performance evaluation
- *Qualitative measures* as direct observations of users behaviours and comments during the performances.

The gathering of all necessary information, as well as relevant behaviours and comments was managed by manually reporting observations on a standardized form.

Being a preliminary study, the testing was focused on obtaining results that could be evaluated and collected in a short time. Then SUS was chosen considering the limited number of items compared to other evaluating methods and the evidence supporting its use in assessing VR systems [15] [11] [5] [12].

4.3 Procedure

The protocol consisted of three phases with a total session duration of about 20 minutes. The first phase served as an introduction to explain the context of the project, and to start the tutorial of the experience (familiarization phase). In the second phase the user started the real test. During the assessment the participant performed both tasks, the visual exploration task and the target stimulus selection task. Participants sat on a turning chair, allowing a 3-degrees-of-freedom exploration of the virtual environments. Throughout the duration of the study a team member was close to the participant to offer assistance, if necessary. The last phase was dedicated to the user's completion of the questionnaire. All answers were collected with a Google form.

4.4 Data Analysis and Results

4.4.1 SUS Questionnaire. The questionnaire consists of 10 items in which even-numbered questions are posed in negative form. All the items need to be rated on a Likert scale ranging from 1 ("Completely Disagree") to 5 ("Completely Agree").

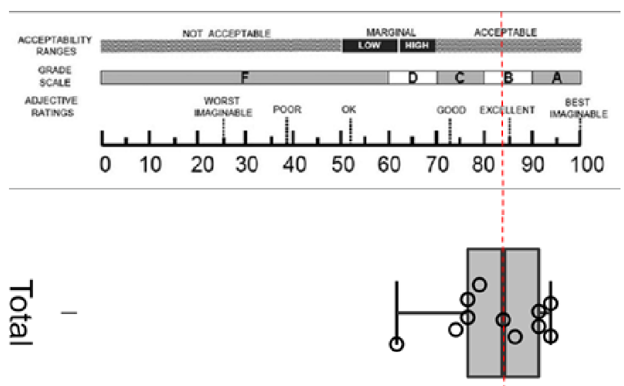

Figure 3: Boxplot SUS

The SUS total median value (Fig.3) is 85 (IQR=15) with a minimum rating score of 62.5 and a maximum rating score of 95. The mean score of 83,6 (SD=10,2) shows that the system is perceived as "good" and "acceptable" according to the adjective scale developed by Bangor et al. [1].

4.4.2 Heatmap. During each session, the system captured the visual exploration of the participant, producing an heatmap for each of them (Fig.4).

4.4.3 Keyword recognition. Out of a total number of 188 keywords, pronounced during the 11 demos, the system correctly recognised 177 keywords. The failed transcription of 11 keywords has probably been caused by the noisy environment where the study took place.

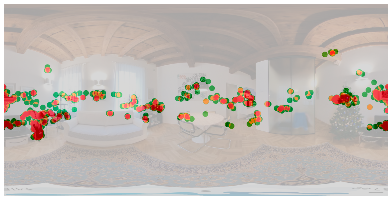

Figure 4: Heatmap produced during the explorative study

By further analysing the transcribed text, we also added some new keywords, so reiteration after reiteration the system gains accuracy. Figure 5 shows the aggregated data concerning the keyword recognition. As the number of pronounced keywords increases, the error rate remains low. These data, however, should be interpreted with caution, as the sample of the study is very small.

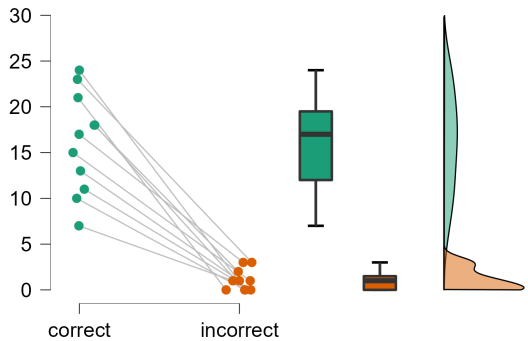

Figure 5: Keywords Recognition Activity

5 DISCUSSION

The analysis of the results is based on the assumption that the external services used for the functioning of the scripts are accurate and reliable. We did not focus on testing the latter, but on making sure that the system was capturing data continuously, without bugs and errors.

5.1 SUS questionnaire

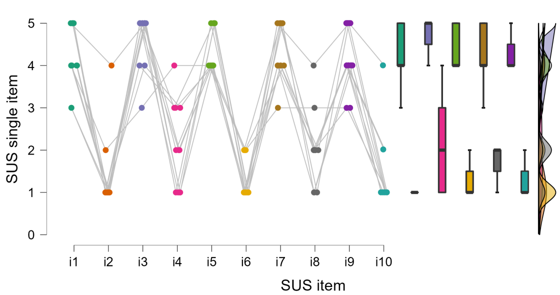

Figure 6: SUS user ratings on single SUS items

SUS results related to the single items are represented in Figure 6. The higher scores were given to items i3 ("I thought the app was easy to use") and i7 ("I would imagine that most people would learn to use this app very quickly"), both concerning the ease of use, with 91% of respondents reporting a score equal or higher than 4. Noteworthy is the data from i6 ("I found inconsistency between the various functionalities of the system"). This suggests that the way in which the user interacts with the system is consistent along all the experience. These results are promising considering that the sample participating in the study was not the target of the application.

5.2 Heatmap and keyword recognition

By manually comparing the mapping of visual behaviour described by the heatmap and the transcribed speech for each tester, we empirically observed that there is a correlation between what they saw and what they said. Not wanting to dwell on the actual skills of the person but only on the performance of the system, we can affirm that, when the user's eyes point to a certain coordinate of the VE, the custom script correctly computes the corresponding pixel in the output image.

5.3 Target counter

Since the system automatically calculates the Hit, FA and d' scores from the target selection task, it would have been unnecessary to add further analysis. The main problem that came out during this task was the lack of feedback after the "select" action. We noticed that users tried multiple times to select the same glass since there weren't any responses coming from the system. Therefore, we decided to add non intrusive haptic feedback whenever the trigger is pressed during the entire duration of the scene, in order to proceed with the task with confidence.

6 CONCLUSION AND FUTURE WORK

The purpose of this exploratory study was to test the usability of a system integrating 360°VEs with an automated data collection tool for the assessment of cognitive functions, based on the ongoing debate regarding the ecological validity of classical tests and on the evidence that supports the use of VR in neuropsychological assessment [9]. The results, although preliminary, are encouraging and show that the aforementioned technology can be investigated in more details as an assessment tool in neuropsychological tests. This is, to our knowledge, the first step towards the development of a new system whose ultimate goal is to assess multiple cognitive functions within a single tool, while collecting and analyzing qualitative and quantitative data to support the assessment. Today's technology is ready to be fully exploited and tomorrow's head mounted displays will be increasingly powerful and precise.

6.1 Future Work

Once the usability of the system has been ascertained, the next steps will be to evaluate its effectiveness as a neurocognitive assessment tool. First of all, it will be necessary to administer the test to a wider sample of users in order to validate the tool and obtain normative data. Furthermore, it will be necessary to involve a sample more similar to the target group, i.e. people with suspected cognitive impairment. This would also allow us to gather additional feedback on the usability of the system from elder people.

REFERENCES

[1] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[2] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[3] Sonia Maria Dozzi Brucki and Ricardo Nitrini. 2008. Cancellation task in very low educated people. *Archives of Clinical Neuropsychology* 23, 2 (2008), 139–147.

[4] Naomi Chaytor and Maureen Schmitter-Edgecombe. 2003. The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology review* 13, 4 (2003), 181–197.

[5] Eva Dulau, Chrisna R Botha-Ravyse, Mika Luimula, Panagiotis Markopoulos, Evangelos Markopoulos, and Kimmo Tarkkanen. 2019. A virtual reality game for cognitive impairment screening in the elderly: a user perspective. In *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 403–410.

[6] D. M. Green and J. A. Swets. 1966. *Signal detection theory and psychophysics*. John Wiley & Sons, Inc., New York, NY, USA.

[7] Hanne Huygelier, Brenda Schraepen, Raymond Van Ee, Vero Vanden Abeele, and Céline R Gillebert. 2019. Acceptance of immersive head-mounted virtual reality in older adults. *Scientific reports* 9, 1 (2019), 1–12.

[8] Brigitte C Kaufmann, Dario Cazzoli, Tobias Pflugshaupt, Stephan Bohlhalter, Tim Vanbellingen, René M Müri, Tobias Nef, and Thomas Nyffeler. 2020. Eyetracking during free visual exploration detects neglect more reliably than paper-pencil tests. *cortex* 129 (2020), 223–235.

[9] Thomas D Parsons. 2015. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in human neuroscience* 9 (2015), 660.

[10] Thomas D Parsons, Susan McPherson, and Victoria Interrante. 2013. Enhancing neurocognitive assessment using immersive virtual reality. In *2013 1st Workshop on Virtual and Augmented Assistive Technology (VAAT)*. IEEE, 27–34.

[11] Elisa Pedroli, Silvia Serino, Pietro Cipresso, Federica Pallavicini, Irene Alice Chicchi Giglioli, Erika Guastafierro, Andrea Gaggioli, and Giuseppe Riva. 2015. Neglect App. Usability of a new application for assessment and rehabilitation of neglect. In *Proceedings of the 3rd 2015 Workshop on ICTs for improving patients rehabilitation research techniques*. 139–143.

[12] Luca Pieri, Silvia Serino, Pietro Cipresso, Valentina Mancuso, Giuseppe Riva, and Elisa Pedroli. 2021. The ObReco-360°: a new ecological tool to memory assessment using 360° immersive technology. *Virtual Reality* (2021), 1–10.

[13] Silvia Serino, Francesca Baglio, Federica Rossetto, Olivia Realdon, Pietro Cipresso, Thomas D Parsons, Giacomo Cappellini, Fabrizia Mantovani, Gianluca De Leo, Raffaello Nemni, et al. 2017. Picture Interpretation Test (PIT) 360: an innovative measure of executive functions. *Scientific reports* 7, 1 (2017), 1–10.

[14] Ngeemasara Thapa, Hye Jin Park, Ja-Gyeong Yang, Haeun Son, Minwoo Jang, Jihyeon Lee, Seung Wan Kang, Kyung Won Park, and Hyuntae Park. 2020. The effect of a virtual reality-based intervention program on cognition in older adults with mild cognitive impairment: a randomized control trial. *Journal of clinical medicine* 9, 5 (2020), 1283.

[15] Cosimo Tuena, Elisa Pedroli, Pietro Davide Trimarchi, Alessia Gallucci, Mattia Chiappini, Karine Goulene, Andrea Gaggioli, Giuseppe Riva, Fabrizia Lattanzio, Fabrizio Giunco, et al. 2020. Usability issues of clinical and research applications of virtual reality in older people: A systematic review. *Frontiers in Human Neuroscience* 14 (2020), 93.