

Accepted Manuscript

Cognitive Evaluation for the Diagnosis of Alzheimer's disease based on Turing Test and Virtual Environments

Juan Manuel Fernandez Montenegro, Vasileios Argyriou

PII: S0031-9384(17)30026-4
DOI: doi: [10.1016/j.physbeh.2017.01.034](https://doi.org/10.1016/j.physbeh.2017.01.034)
Reference: PHB 11654

To appear in: *Physiology & Behavior*

Received date: 2 April 2016
Revised date: 20 October 2016
Accepted date: 19 January 2017



Please cite this article as: Montenegro Juan Manuel Fernandez, Argyriou Vasileios, Cognitive Evaluation for the Diagnosis of Alzheimer's disease based on Turing Test and Virtual Environments, *Physiology & Behavior* (2017), doi: [10.1016/j.physbeh.2017.01.034](https://doi.org/10.1016/j.physbeh.2017.01.034)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Cognitive Evaluation for the Diagnosis of Alzheimer's disease based on Turing Test and Virtual Environments

Juan Manuel Fernandez Montenegro^a, Vasileios Argyriou^{a,*}

^aKingston University, Penrhyn road, Kingston Upon Thames KT1 2EE, UK

Abstract

Alzheimer's screening tests are commonly used by doctors to diagnose the patient's condition and stage as early as possible. Most of these tests are based on pen-paper interaction and do not embrace the advantages provided by new technologies. This paper proposes novel Alzheimer's screening tests based on virtual environments and game principles using new immersive technologies combined with advanced Human Computer Interaction (HCI) systems. These new tests are focused on the immersion of the patient in a virtual room, in order to mislead and deceive the patient's mind. In addition, we propose two novel variations of Turing Test proposed by Alan Turing as a method to detect dementia. As a result, four tests are introduced demonstrating the wide range of screening mechanisms that could be designed using virtual environments and game concepts. The proposed tests are focused on the evaluation of memory loss related to common objects, recent conversations and events; the diagnosis of problems in expressing and understanding language; the ability to recognise abnormalities; and to differentiate between virtual worlds and reality, or humans and machines. The proposed screening tests were evaluated and tested using both patients and healthy adults in a comparative study with state-of-the-art Alzheimer's screening tests. The results show the capacity of the new tests to distinguish healthy people from Alzheimer's patients.

Keywords: Alzheimer's disease, Dementia, Early detection,

*Corresponding author

Email address: Vasileios.Argyriou@kingston.ac.uk (Vasileios Argyriou)

Screening tests, HCI, Virtual environments, Turing Test
2010 MSC: 00-01, 99-00

1. Introduction

Dementia is a term used to encompass a number of symptoms such as the decline of memory, reasoning, language or perceptual interpretation. Specific combinations of these symptoms are used to define different types of dementia such as Alzheimer's or Parkinson's disease. Alzheimer is a disease that usually affects elder people, with the number of cases increasing over the last decades. The number of people with early onset dementia (people under 65 years old) has also increased in the UK and worldwide [1]. Alzheimer's is the most common type; it accounts for 60 to 80 percent of cases, and one of the most noticeable symptoms is the difficulty in learning new information. In addition, when the disease advances, there are other symptoms such as disorientation, mood and behaviour changes; confusion about events, date and place; being suspicious about family, friends and caregivers; and difficulty in speaking, writing and walking [2].

According to the World Health Organization, 1.37% of the estimated deaths around the world in 2030 will be caused by Alzheimer's disease and other dementias. People with medium dementia live on average 8 years [2]. Currently, when Alzheimers disease (AD) is diagnosed, the neuronal damage is spread enough to make it irreversible [3]. When neurons die, the other neurons do not divide and replace them, as other cells do, so the damage cannot be reversed [2]. Therefore, it is important to detect dementia at its very early stages in order to reduce the deterioration speed [4, 5].

The cost of dementia is another issue to take into consideration. The worldwide cost of dementia is \$818 billion, and it will become a trillion by 2018[6]. The cost of dementia in the UK estimated in 2013 was 26.3 billion pounds; 4.3 billions were spent on healthcare costs, of which around 85 millions were spent on diagnosis [1]. Therefore, it is essential to develop affordable diagnosis and

support tools, limiting the increasing cost of dementia. One of the proposed initiatives is focused on the implementation of e-health (the use of Information Communication Technology (ICT)) solutions to reduce the cost and to make the health systems and solutions universally accessible [7]. Arief et al presented in [6] the strengths of e-health tools such as the improvement in accessing health-care services by senior citizens, their cost-effectiveness and their efficiency in managing health resources.

Furthermore, Alzheimer's symptoms are studied to improve the results of previous approaches or create novel and more accurate methods relying on new affordable and publicly available technologies. Alzheimer's detection methods can be classified in two different categories: invasive and non-invasive. Invasive methods require obtaining data from the interior of the patient's body through procedures such as lumbar puncture or blood extraction. These tests are not always safe and comfortable for the patient, while some of them are unbearably painful. On the other hand, non-invasive tests are harmless and more convenient during the diagnosis process.

Focusing on non-invasive methods, cognitive tests are the most commonly used Alzheimer tests for diagnosis. These are question/task/problem based tests trying to measure patients' cognition, while cognition is defined as the use of the information that has been previously collected by a person to make behavioural decisions [8]. Cognitive or aptitude tests have high accuracy when it comes to Alzheimer's detection. Nevertheless, one of their weaknesses is the evaluation of brain's capacity to compensate brain damage (cognitive reserve) [3, 5]. Another well-known problem is the adaptability of the tests according to the patient's IQ, since most of the tasks that integrate an Alzheimer's detection cognitive test usually are too simple to evaluate intelligent patients. Therefore, it is necessary to create tests, whose results are not correlated with the IQ of the patients. The use of computerized tests helps to create intelligence adaptable cognitive tests [9, 3]. Moreover, based on the available technology, it is possible to design new types and more effective tests using for example virtual environments (VEs). Virtual environments provide additional advantages to cognitive tests, since it

is possible to immerse the patient in a controlled situation [10, 11, 12, 13].

Parsons et al. have demonstrated in [14, 15, 16] the ecological validity of VEs and their benefits for neurocognitive assessment, such as the precision of data retrieved by the computer or the better control of the environment. They also prove the advantages of using virtual reality over the traditional pen and paper methods.

When it comes to the Alzheimer's symptoms, the American Psychiatric Association has gathered in [17] the cognitive domains affected by Alzheimer's disease: complex attention, executive function, learning and memory, language, perceptual-motor and social cognition. The existent tests try to examine specific cognitive areas. For example, the Mini-Mental State Examination (MMSE) studied different cognitive areas such as the learning and memory, the complex attention or the language cognitive domains.

Analysing the cognitive domains affected by Alzheimer's, a relation can be found with the areas that Alan Turing uses to define intelligence. Alan Turing proposed in 1950 a test related with human and computers cognition. It tests the degree of intelligence of a machine in relation with its ability to impersonate a human [18]. In order a computer is considered intelligent should possess the next capabilities [19]: natural language processing, knowledge representation, automated reasoning and machine learning. Comparing these capabilities with the correspondent cognitive domains it is possible to consider the use of existent Turing Test for evaluating human cognition. In addition, according to Warwick et al [20], the success of Turing test does not depend only in the quality of computer's AI; it also depends on the intelligence of the human that is judging the machine. Therefore, reversing this test, it is possible to obtain a cognitive test that checks the degree of intelligence or the existence of a cognitive impairment.

In this paper, we propose novel cognitive and executive function based, non-invasive screening tests for early Alzheimers diagnosis implemented as an e-health tool. These new tests are conducted in a virtual room due to the flexibility to create different tasks and the capacity of the involved users, [10, 12, 13]. The developed test has been chosen following the next principles: it should evaluate

at least one of cognitive domain, it should take advantage of VEs flexibility and it should be easily upgraded. As a result, an innovative, cost efficient and flexible e-health Alzheimer's detection application is introduced.

The remainder of this paper is organized as follows: section 2 describes previous related work on Alzheimers diagnosis. Section 3 analyses the proposed methodology and in section 4 details on the evaluation process and the obtained results are presented. Section 5 gives some conclusion remarks.

2. Previous Work

Apart from the invasive/non-invasive classification, the Alzheimer's detection methods can also be classified as non-cognitive and cognitive tests. Cognitive tests encompass the methods that evaluate patients' cognition and these procedures are non-invasive and easy to implement. Non-cognitive tests are regarded all other methods, invasive or non-invasive, used to detect and diagnose dementia.

Regarding the non-cognitive approaches, in [2, 21] some of the methods used to detect dementia are based on defining potential biomarkers that are acquired through invasive techniques, such as cerebrospinal fluid tau protein and beta-amyloid peptide. Lau et al proposed in [22] salivary trehalose as a reliable Alzheimer detector biomarker obtained through non-invasive methods. Those biomarkers have been proved an accurate indicator of the presence of Alzheimer, but they are not validated yet in large groups as an accurate and reliable tool for Alzheimers disease detection, so they cannot be used in medical clinics [2].

There are other non-invasive methods that require the use of external devices during a certain period of time to analyse patients behaviour [24]. For example, Aztiria et al in [24] placed a sensor on the patients foot in order to analyse their gait (step length and step height), since it reflects patients dementia level. However, the gait measurement method still has some drawbacks, such as the patient has to wear a device for long periods and the results obtained, despite being promising, are still not useful, since they have not been tested in

Alzheimer patients. Abe et al present in [4] another detection method using sensors in patients' homes to identify certain events. This non-cognitive approach is non-invasive and it requires only the consent of the patient to install the sensors. The main issue with these methods is that the obtained results are not precise enough (less than 75% detection rates) to provide an accurate technique for dementia diagnosis at early stages.

Magnetic Resonance Imaging (MRI) is another non-cognitive technique commonly used for alzheimer detection that provides accurate results since it is possible to study visually the deterioration of the brain. In the works presented in [25, 26] the head MRI data of the patient is compared with the corresponding data of patients with Alzheimer's using machine learning techniques. Nevertheless, since acquisition of MR images involves the use of medical equipment that is not easily accessible and the process is unpleasant (claustrophobic and noisy), these methods are not suitable for testing large groups of people due to cost and time, neither mental disease patients' for safety and discomfort reasons.

When it comes to cognitive methods, the tests based on problem solving tasks and questions, in order to detect a cognitive impairment, are the most commonly used by doctors. Cordell et al presents in [27] a comparative study of these methods, including Mini-Cog test or Saint Louis University Mental Status (SLUMS) [28]. Also it should be mentioned that the Mini Mental State Examination (MMSE) is one of the most well-known and used tests [29]. Other examples of cognitive tests not included in [27] are the memorization of interacting objects in the Visual Association Test (VAT) [30] or the detection of the prevalence of the right ear when sounds are memorized by Alzheimer's patients during the Dichotic Listening test [31]. The results provided by this test demonstrate high accuracy and specificity. Nevertheless, the results show a ceiling effect, since the test is not complex enough for high IQ patients [29].

Other non-invasive methods are focused on analysing visual impairments since visuospatial functions and visual processes decline due to Alzheimer [32, 33]. These methods evaluate the mental state of a patient through the measurement of the reaction time in certain stimulus, the assessment of attention or

the evaluation of patients visual memory [33, 34]. Pereira, et al. [33] analysed different methods that use eye-movement to determine visual impairments in Alzheimer patients. These methods compared eye movements of healthy people with Alzheimer patients in terms of fixation duration, refixations or saccade orientation. As a result, most of the studies revealed an increment of saccades, defects in fixations and slow pursuit movements. Nevertheless, as suggested by Pereira et al in [33, 34], these studies fail to consider attentional impairments as a multi-domain concept and their results still need to be corroborated by future studies.

Recently, due to the new advances in computer technologies, Virtual Environments (VEs) started to become part of medical tests and rehabilitation therapies. The study conducted by García-Betances et al in [13] shows the advantages of virtual environments for Alzheimer's disease and point out the lack of immersion or interaction in most of the current virtual reality applications. The work presented by Tarnanas, et al. in [11] demonstrates that the use of VEs is beneficial, when it comes to early dementia detection. It is possible to improve the results from previous cognitive tests, since it reduces the floor and the ceiling effects, creating tests that adapt to the patients' IQ; and it is possible to increase immersion in the task that is in progress. Their work uses large screens to display the environment and some depth sensors to recognize the gestures of the patients' body. This approach requires the patient to move to the location where the facilities for this test are available, due to the fact that the components are not portable and cost efficient making it inappropriate for e-health applications.

Recently approaches were introduced to validate the use of virtual reality for neurocognitive assessment. The work presented in [16] asses complex attention on healthy and military subjects and examine the ecological validity of virtual reality using head mounted VR devices in comparison with less immersive conditions. In [35] the authors evaluate learning and memory cognitive abilities on healthy controls while navigating through a virtual environment using a head mounted VR device. The test is based on traditional neurocognitive methods

proving the correlation between their and the traditional tests' results. Cushman et al. in [36] also present a comparative study between a navigation system on VEs and the real world while evaluating Alzheimer's participants navigation aptitudes and proving the ecological validity of VR environments. Parson et al work in [14] analyse the appropriate way to create neurocognitive interfaces in VEs. In this paper they consider the possible lost of the experimental control related with the VE fidelity. An increase of the VE fidelity means a loss in the control so a balance should be found in relation with the test type. Another interesting concept is the psychophysiological computing; this term implies the human monitoring by the computer being able to create user adaptive applications.

2.1. Relation with the proposed work

In our work a novel non-invasive Alzheimer's detection screening test based on VEs is proposed. The Virtual Environment created for this test is a room that the patients can recognize and feel comfortable, such as a doctor's office, a living room, etc. The use of new immersive technologies, such as virtual reality glasses, provides a huge improvement in immersing the subjects in a VE and its combination with depth cameras makes possible the improvement of interaction with the VE. The proposed and developed tool is the first Alzheimer's detection screening test that provides a full immersion in Virtual Environments implying that the patients are able to be more focused on the tasks, while additional information related to the environment-patient interaction is obtained. Also, it is possible to adapt some of the tasks to the patient's IQ level during the test, avoiding the ceiling effect. The adaptability to the participant's IQ is based on user's options election during the procedure and the test supervisor criteria, therefore it is not related with the psychophysiological computing concept presented in [14].

These test are based on the examination of cognitive abilities. Therefore, the tests are question and task based, trying to provide a user friendly and comfortable environment for the patients. Furthermore, the proposed tests do



Figure 1: Virtual Reality glasses (Oculus Rift DK2 or Samsung Gear VR) and Depth sensor (Microsoft Kinect 2) are the devices required to perform the tests.

not consider only memory tasks, but introduce novel approaches based on the concept of differentiating between reality and a virtual world; and on a new cognitive test that aims to identify normal and abnormal events or objects present in the virtual scene. Also, novel checks are proposed based on a reversed Turing test, where the patients' cognition is evaluated according to their ability in distinguishing absurd from correct information or a machine from a human.

3. Proposed Methods

The proposed methods for Alzheimer's detection were developed to be conducted in Virtual Environments. Thus, VR glasses are used in order to maximize the patients' immersion in the VE and a depth camera is utilised to track the patients' movements and animate their avatars in order to increase the interactivity (see figure 1). These tools are affordable and they can be reused by other patients, resulting to a low cost Alzheimers detection application. The main scene was designed such as to recreate a doctor's office or any other real room that the patient can recognize (office, living room), see figure 2. Then, the tests were designed taking advantage of the virtual environment versatility, such as the absence of physical limits.

Regarding the real environment, since Alzheimer's patients are mainly el-

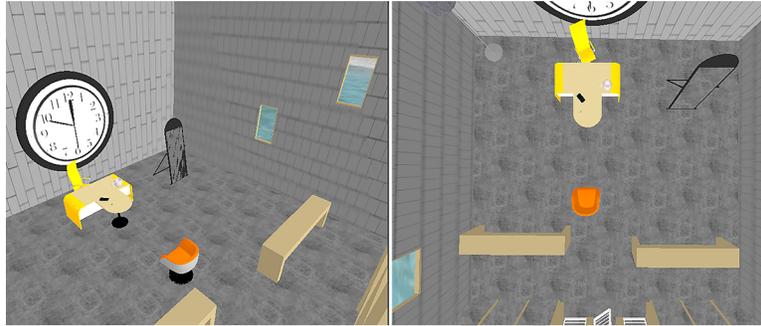


Figure 2: Virtual Reality Room. Corner and top down view

der people; all tasks are performed seated on a chair requiring only minimal movements. The patient is sitting on a comfortable chair about 2 meters in front of the depth sensor and is helped to place on the Virtual Reality glasses (see figures 3 and 4). Once the patient feels comfortable with the glasses, we proceed with the tests. The doctor (or test supervisor) has to be in front of the computer locally or remotely, in order to control the software application and is advised to take notes during the process typing details that may be useful later, since it is possible to obtain in that way extra information about the patient's behaviour. Despite most of the tests are automatic and the patient only has to follow the instructions, some actions, such as starting a new task or playing sounds, requires the interaction of the doctor/instructor in order to provide enough time to the patient to perform the actual tasks or to add a degree of randomness and adaptability.

All the proposed tests should be performed under the supervision of a healthy person following the instructions on the screen or the test's information sheets. Despite the fact that the test could be monitored by any person without mental impairments, it is recommendable that the supervisor during the tests to be a doctor, either locality or remotely, since additional information may be obtained during the whole process. In this work, four tests are introduced: Virtual Objects Memorization, Abnormal Objects Recognition, Virtual vs. Real Sounds and Bot Doctor Turing Tests.

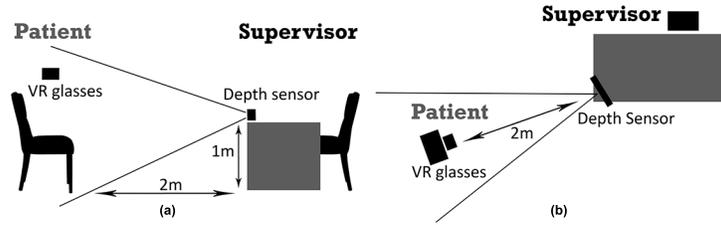


Figure 3: (a) Top down view showing the position of the devices, and the location of the patient and the supervisor. Around the patient should be enough space to avoid collision with any real objects. (b) Side view showing the position of the devices, and the location of the patient and the supervisor. The depth sensor should be located approximately 1 meter above the floor and the patient has to be approximately 2 meters away from the depth sensor, in order to be in the view range.

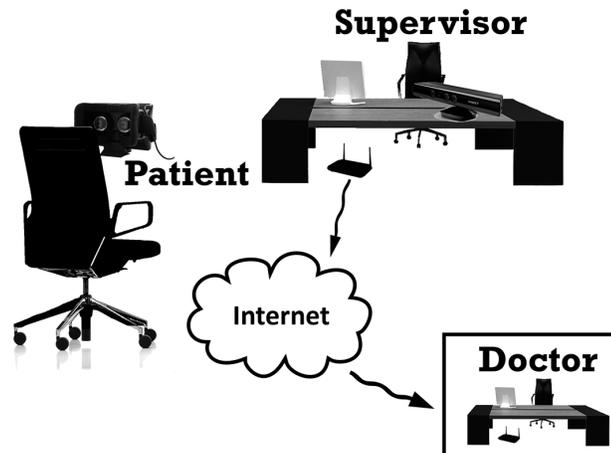


Figure 4: Early detection Test setup. The patient is sat in front of the depth sensor, wearing the VR glasses. The supervisor/doctor monitors the test from the computer locally or remotely.

3.1. Virtual Objects Memorization test (VOM)

VOM is a test focused on memory and its main objective is the analysis of patients learning and memory cognitive domain [17]. It is split into three steps. The first step of this memory test is the recognition of six different virtual objects. The objects are commonly used during everyday activities making them easily identified and recognized. The objective of this first task is to check

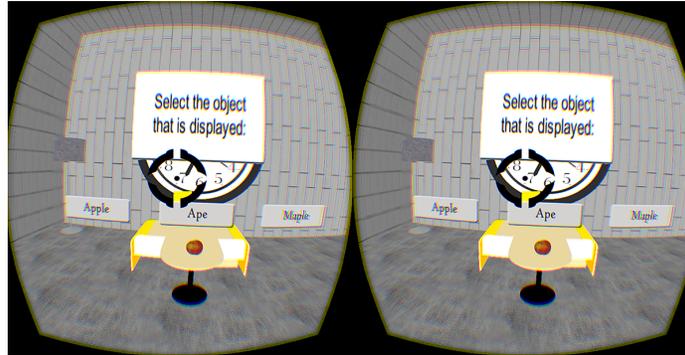


Figure 5: Virtual Object Memorization tasks. Object Recognition task where the patient has to identify and select the correct name of the objects on a table.

if the patient has problems in recognizing quotidian objects (see figure 5). Once all the objects are recognized, a virtual monitor displays all of them in specific positions (the positions of these objects are the ones where they normally are located on a desk). Therefore, the patient is asked to memorize the objects but nothing is mentioned about their position.

During the second step, the subject is asked about the objects' locations, in order to check the visual and the associative memory of the patient. In more details, the desk is shown in front of the patient and a number appears in the position where an object was located, while a list of the six possible objects is provided (see figure 6). The subjects are asked about the object that was located in the position indicated by a number and they have to select the name of the correct one. The fact that the objects were located in logical positions should help the user to select the right answers.

Finally, in the last task, interacting objects are placed in the virtual room and have to be located and memorized by the patient. Then, one of the objects is shown and the subject has to recall its interacting couple and their functionality. Once the third stage is completed, there is an extra task associated with the first step, where the patient is asked to recall the objects that were shown. The aim of the last two tasks is to evaluate the patient's short-term memory incorporating interacting objects and their functionalities. So, the novel part here is focused

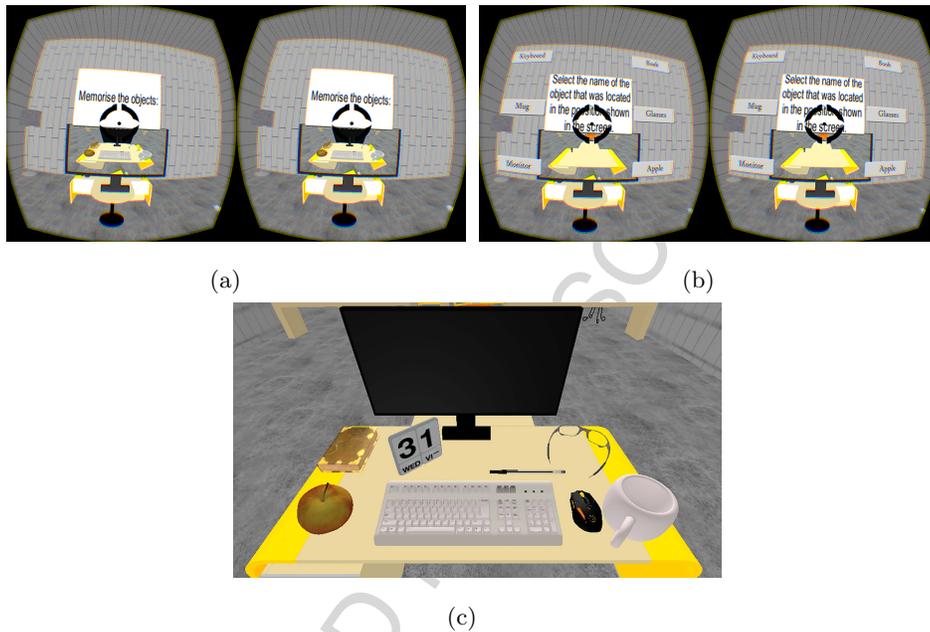


Figure 6: Virtual Object Memorization tasks. (a) The objects in their respective location are shown in order to be memorised by the patient. (b) Objects' location task where the patient has to select the name of the object that was shown on the screen. (c) A close view of the objects and their locations.

mainly on the combination of interacting objects with their functionalities.

Also, it should be mentioned that in order to proceed through the test, the subject has to select amongst different options. This selection is performed moving the head to aim towards the desired option using the direction of view (scope), as we can see in Figures 5 and 6. Therefore, the patients have to move slightly their head to fulfil the tasks.

This test if all tasks are completed successfully provides in total 21 points. The points are associated with the recognition of the objects, the recall of their position and the recall of the objects themselves and their functionality.

3.2. Abnormal Objects Recognition test (AOR)

AOR is a test that evaluates if the patient is able to discern if something is abnormal or not within the virtual environment and it is divided into three

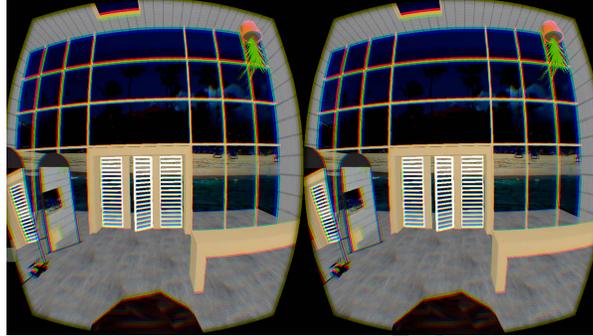


Figure 7: Abnormal Objects Recognition test preview. Some abnormalities such as the upside-down pot and the toy car going through the mirror can be observed in this figure.

Table 1: Abnormalities in the room ordered by identification difficulty

Object	Abnormality
Chair	Wanders around the room
Mug	Spins
Car	Goes through the wall Change shape when goes through the mirror
Potted plant	Upside down
Mirror	Does not reflect

separate stages. During the first stage, the patient is asked to detect any abnormalities present in the room in order to analyse their perceptual-motor cognitive skills [17]. Table 1 describes the abnormalities that can be find in the room. The subject in the second stage is requested to read the clock that is located on one of the walls in the room. Shortly afterwards the patient is asked if the illumination (i.e. ambient light) of the room is in accordance with the time, (see Figure 7).

Finally, the objective of the third stage is to evaluate the coordination of the patient when the physics of a mirror reflection are changed. An avatar, whose movements are linked to the patients movements, is displayed in front of them. Therefore, it will look like a mirror; however, the movements of the avatar are

Table 2: Movements that the patient has to perform to move the avatar and their evaluation

Task	Answer and Evaluation
Raise your right hand. Is the avatar behaving as a mirror?	No = 1 point
Move the body of the avatar towards the mug (there is a mug on the right of the avatar)	First attempt = 1 point
Grab the mug	First attempt = 1 point

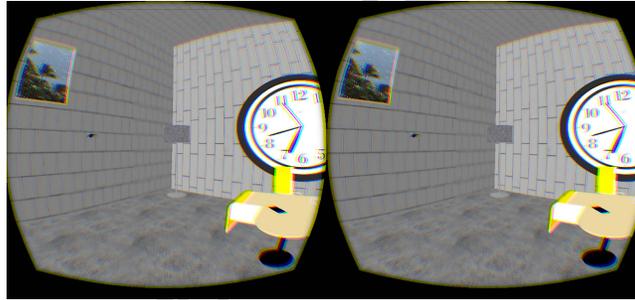


Figure 8: Virtual vs Real Sounds test preview. Some of the visual clues such as the fly, the clock and the mobile phone appear on this figure.

reversed. Once the patients detect how the avatar moves, they are asked to perform specific movements with the avatar (see table 2). This task allows the study of perceptual-motor and executive function impairment [17].

This test can provide a maximum of 10 points. Seven points are associated with the correct detection of the abnormalities and three are associated with the correct movement of the avatar.

3.3. Virtual vs Real Sounds test (VRS)

The main objective of VRS test is the detection and recognition of objects/events through audio and to check if the patient is able to discern between real and virtual sounds. As a result this test analyses patients' mental and cognitive flexibility [17].

Different sounds are played during the test, such as a clock ticking, a fly,



Figure 9: Selection Based - Bot Doctor Turing Test preview. The two avatars represent a doctor. The closest one is the one asking the questions in the screen, once it finishes the questions the other avatar will move to the position of this avatar and ask it will ask its questions. The participant has to read the information above and choose one of the options below moving the scope over the desired choice using head movements.

rain at the background, etc., (see Figure 8). The patient has to recognise and name all the sounds that they listen during the test. In addition, the virtual environment will provide some visual clues to help them with the identification. Once the subjects have identified all the sounds, they are asked about the origin of the sounds. During this stage, the source of some sounds is coming either from the virtual environment or the real world reproduced by the supervisor. This second part requires the patient to recall all the sounds that were played and to identify which ones were reproduced by the supervisor. This test gives a maximum score of 6 points. One point is obtained for each object that is recognized and categorized properly as real or virtual.

3.4. Bot-Doctor Turing Test (BDTT)

The proposed Turing test is based on the ability of a patient to evaluate if a computer is able to impersonate a human. In order to perform the test, a human and a computer are located in a room, while the patient is in another location trying to communicate with the human and the computer in the first room without knowing who is responding, (see Figure 9). The patient in the second room asks questions aiming to distinguish who is the human and who is the

computer [18]. We propose two different versions of the Alzheimer Turing test to evaluate patients' executive function, such as the ability to hold information and manipulate it and assess their process of making decisions [17].

3.4.1. Selection Based - Bot-Doctor Turing Test

In this test two virtual doctors have been designed to provide answers about a specific topic to the patient based on a simple AI architecture. The conversation is formatted using a tree structure and is based on the same discussion and interaction system used in classic adventure video games. In more details, the discussion system provides a set of questions to the patients and based on their choice the AI-Bot-Doctors will provide an answer. The patient is informed that will have two discussions with two different Bot-Doctors and each AI system provides related but different answers. One of them will provide logical and correct answers, whereas the other will provide incorrect and sometimes absurd responses.

The whole discussion takes place in a virtual room, and two different human like avatars are used for the Bot-Doctors. The patients are immersed using the virtual reality glasses in the Virtual Environment, allowing them to see the avatars of the virtual doctors. Different variations of this test can be designed depending on the avatars used to represent the AI-Bots. In more details the Doctors' avatar can have either human or robotic appearance. Therefore, there can be 2 human avatars, 2 robotic avatars or a mixture of them. The objective of using non-human-like avatars is to check if the patient will have a different interaction if the avatar that provides the information does not look as a human. So, with this approach we can evaluate if the visual characteristics of the avatars can affect the provided answers. The test terminates either when the patients decide that they have asked enough questions to the avatars or if all the provided questions have been answered. This test gives in total 3 points. One point if the answer is that both doctors provided real answers and three if they distinguish them correctly.

3.4.2. Script Based Bot-Doctor Turing Test

The second Alzheimer Turing test that is suggested is script based using chatbots. Chatbots are defined as computer programs that can maintain a conversation with a person [37]. ELIZA, IBM's Watson and Apple's Siri are some of the most famous chatbots. ELIZA was the first chatbot and was created by Joseph Weizenbaum; IBM's Watson is a chatbot that won in Jeopardy TV show in 2011; and Apple's Siri is the iPhone's personal assistant application. Also, many open source chatbots are available such as ALICE that can be used in this Bot-Doctor Turing test.

In order to interact with chatbots, people start a chat with the AI, they write the messages on the computer chat screen and the AI or human will engage in conversation. If voice recognition and synthesis are supported then this option can be selected. The proposed test uses a chatbot (e.g. ALICE) and a human to have a discussion with the patient. Therefore, all the entities that are involved in this test are the same as the used in a Turing Test, two humans and one computer. The patients are asked to chat for five minutes with each entity and, at the end, they are asked to decide which entity is a human. The intelligence of the chatbot is settled in accordance with patient's IQ in order to avoid the ceiling effect. It is expected that Alzheimer's patients fail distinguishing a simple AI based chatbot from a human. Regarding the scoring system, this test provides exactly the same points as the previous one.

4. Results

4.1. Participants

The 20 participants of this test are aged between 23 and 82 year old including both healthy people and Alzheimer's patients diagnosed with mild dementia less than year ago. Furthermore, Alzheimer's patients in advanced conditions were tested but, the process was not possible to be completed. Half the subjects were male and the other half female of whom eleven had college or higher educational

background and nine did not reach high school. None of the Alzheimer's participants' education level reached high school. The participants' nationalities included French, Spanish, Vietnamese and Greek so the non English speakers were tested in their own language. It has to be mentioned that one of the patients was illiterate so the test was performed orally.

4.2. Data

The main data collected during the experiment are the individual scores for each test. As it was explained in section 3, each test has a maximum score and each task provides a different number of points. In more details, regarding the recognition and memory tasks, each recalled and recognized object scores 1 point.

In addition, patients have been also tested using other well-known state of the art Alzheimer screening tests, in order to perform a comparative study and allow the evaluation of the proposed novel non-invasive diagnosis screening tests. All the tests were computerized so the patients have to use a computer through the whole process retaining also the same conditions among all tests. The state-of-the-art screening tests used in our comparative study are Dr. Oz Memory Quiz [28]; Visual Association Test [30]; and Dichotic Listening Test (DLT)[31].

The Dr. Oz Memory Quiz is an alzheimer detection screening method, that is based on SLUMS test [28]. It is formed by 17 questions and tasks that score 1 point each.

In the case of Visual Association Test (VAT) [30], six pairs of objects are shown consecutively, and after the patients memorize them, one object of each pair is shown and they have to recall its couple. This test gives a maximum score of 6 points, one per recalled couple.

During the Dichotic Listening Test (DLT), the patients have to memorise 6 digits played to each ear separately and the process is repeated 8 times [31]. As the results are stored by ear, the maximum score is 24 points per ear. This test is based on the fact that the difference of the recalled digits between the left and right ear in a healthy person is minimal compared with the predominance of

the right ear for patients with Alzheimer's. Therefore, the data collected from this test is the difference of the left ear score minus the right ear score. The maximum score for each ear (24) is added in order to obtain positive values. Finally, the maximum possible value is delimited to 24, since left predominance or both ear equivalence are considered as healthy results (DLTre - Dichotic Listening Test right ear). In this way, the results decrease when the right ear predominance is high. In this way, these results and the ones obtained by the other tests can be compared. In addition, the number of digits recalled are also collected. This value will represent the memorization capacity.

4.3. Evaluation

The mean results for each test in relation to the subject's health status are shown in Table 3. It is observed that healthy patients have achieved results close to the maximum score of each test. For example, the normalised averaged score of the Abnormal Objects Recognition test, for healthy subjects, is 0.826316, which is close to the maximum score of 1. On the other hand, the Alzheimer's patients have obtained significantly lower scores in comparison with the healthy participants. For instance, healthy subjects have obtained a mean of 0.924821, whereas the Alzheimer's subjects scored 0.238100 in the Virtual Objects Memorization test. This separation between results is also apparent in the other state of the art tests, such as the relation 0.986842-0.375000 (Healthy Alzheimer) on the Dichotic Listening test (DLTre). Table 4 demonstrates that the p-value of the novel tests proves that their results are correlated with the healthy and Alzheimer's status of the patients. When it comes to sensitivity and specificity in relation to the detection of healthy patients, the results are perfect. This suggest that more subjects need to be tested in order to corroborate the results of the novel Alzheimer's screening tests.

Figure 10 compares the worst healthy participant results and the best Alzheimer's participant ones. There is a clear difference in the results that allows to separate Alzheimer's from healthy participants. For example, the current state of BDTT's score system relies on few marks (3 points) so even if the given marks

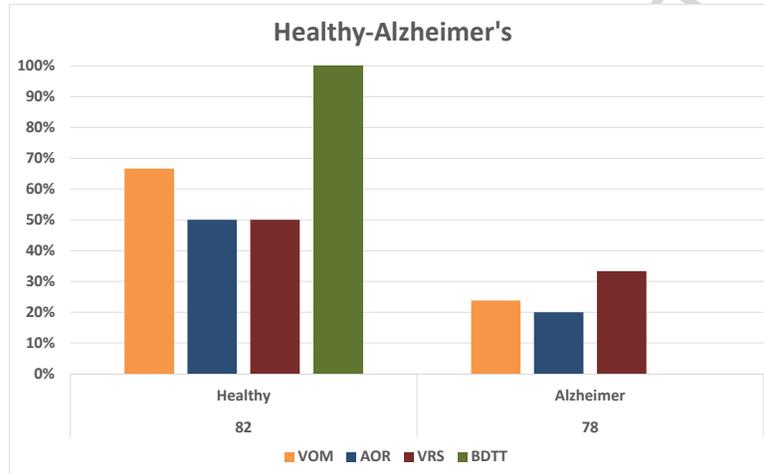


Figure 10: Worst healthy control result vs. best Alzheimer's result comparison. The scores of each test are shown as percentages where the 100% correspond to the maximum result possible of each test.

Table 3: Normalised mean value of the tests regarding the Healthy and Alzheimer's patients

	Age	DrOz	VAT	DLT	DLTre
Healthy	42.47	.900953	.894732	.889253	.986842
Alzheimer	78.00	.529400	.000000	.437500	.375000
	Age	VOM	AOR	VRS	BDTT
Healthy	42.47	.924821	.826316	.833311	.736832
Alzheimer	78.00	.238100	.200000	.333300	.385600

are based on a strong criteria, the addition of new ones will allow to increase the robustness of the test.

Table 4 also demonstrates that most of the results of the test are not correlated with the age and the educational level of the patients, since the p -values do not reject the *Null* hypothesis. Regarding the novel test AOR results a p -value of 0.003 (age) and 0.004 (educational level), indicating a relationship

Table 4: T-test for each test according to the healthy and Alzheimer's cases. The p-value (Sig. (2-tailed)) is less than 0.01 so the Null Hypothesis is rejected

		Sig.(2-tailed)			
		Dementia	Age	Educational Level	Gender
DrOz	e.v.a ^a	.000	.033	.067	1.000
	e.v.n.a ^b	.	.088	.104	1.000
VAT	e.v.a	.001	.325	.462	.615
	e.v.n.a	.	.395	.492	.615
DLT	e.v.a	.000	.038	.111	.616
	e.v.n.a	.	.076	.142	.619
DLTre	e.v.a	.000	.198	.262	.400
	e.v.n.a	.	.314	.326	.410
VOM	e.v.a	.000	.052	.045	.908
	e.v.n.a	.	.128	.084	.909
AOR	e.v.a	.000	.003	.004	.433
	e.v.n.a	.	.020	.013	.435
VRS	e.v.a	.000	.061	.078	.818
	e.v.n.a	.	.113	.104	.818
BDTT	e.v.a	.000	.131	.041	.438
	e.v.n.a	.	.178	.060	.445

^aEqual Variances Assumed

^bEqual Variances Not Assumed

with those factors. Therefore, tasks that are more difficult can be added for patients with high educational background in order to eliminate this correlation. When it comes to the relationship between the results and the gender of the participants, the results of the male participants and the females one are not correlated (see table 4).

The correlation between the DrOz, VAT and DLT tests, and the novel ones

Table 5: Correlation between state of the art tests and novel ones

		DrOz	VAT	DLT	DLTre
VOM	Pearsons correlation	.589**	.666**	.741**	.832**
	Sig. (2-tailed)	.006	.001	.000	.000
AOR	Pearsons correlation	.676**	.526*	.723**	.724**
	Sig. (2-tailed)	.001	.017	.000	.000
VRS	Pearsons correlation	.486*	.602**	.720**	.688**
	Sig. (2-tailed)	.030	.005	.000	.001
BDTT	Pearsons correlation	.388	-.060	.228	.249
	Sig. (2-tailed)	.100	.807	.347	.304

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

is shown in Table 5, where the Spearman's correlation has been calculated, including the correlation coefficient and the p -value. Most of the novel tests result high correlation value with at least two of the previous tests.

In addition to the numerical results, some notes were collected during the testing process that could be used by the doctors to make a decision. The most interesting of these notes is the lack of attention of Alzheimer's patients due to the animated virtual environment. For example, during VOM test there are some animated objects close to the questions area that divert the attention of the patients making them to forget about the task in progress. This reveals a problem in the Complex Attention cognitive domain.

4.4. Qualitative evaluation of the proposed system

At the end of the process, patients were asked to evaluate the application. They had to fill a form evaluating the quality of the application in terms of interaction; simplicity of instructions/processes and the comfort of the VR glasses. Additionally, the patients had the option to write any additional comments about the proposed cognitive tests.

Table 6: Qualitative evaluation of the proposed cognitive tests by the participants. Four characteristics were evaluated from 0 (Very bad) to 5 (Very good)

Type	Mouse Inter	Kinect Inter	Instructions	VR Glasses
≥ 50	4.17(0.983)	4.00(1.155)	4.43(0.787)	3.86(0.9)
< 50	4.33(0.778)	3.33(0.985)	4.17(0.835)	3.50(1.0)

The average evaluation of the proposed approach was positive. Some of the subjects commented that the glasses were not comfortable at the end of the test and that the graphics could be improved, which is something that will be available in future versions of the hardware devices. Nevertheless, most of the participants evaluated the application as motivating and interesting; and the interaction and instructions as simple and easy to understand.

Table 6 shows the overall evaluation of the proposed cognitive tests provided by the participants. The interaction with the application (i.e. keyboard, mouse, depth sensor and virtual reality glasses), the quality of the instructions (i.e. written and/or oral provided by the supervisor) and the comfort of the VR glasses were evaluated from 0 (really bad) to 5 points (really good). The overall evaluation obtained is positive, since the average result in all these categories is above 3 points, with elder people being more enthusiastic with new technologies, probably due to the completely new to them features and capabilities.

5. Conclusion

Alzheimer is a disease without foreseen cure, that is affecting more people every year, with the related research and associated services (care) cost to increase rapidly. Our objective is to provide affordable approaches that would help improving the current cognitive based detection systems.

Virtual Environments (VEs) start to be part of medical treatments or diagnosis methods, and they have been proved to be considerably useful by applying cognitive tests in secure and controlled environments; helping to increase the effectiveness of the related methodologies. These facts have been further vali-

dated with the proposed novel tests, since the results obtained show that it is possible to create an accurate e-health screening Alzheimer's diagnostic system. In this work, new methods for detection of Alzheimer's disease based on Virtual Environments were introduced. The proposed tests are focused on the evaluation of memory loss related to common objects, recent events, the diagnosis of problems in expressing and understanding language and the ability to recognise abnormalities, and to differentiate between virtual worlds and reality. In addition, two novel tests related with Alan Turin Imitation Game were proposed, where the human's intelligence is evaluated instead of the machine's one. The proposed approaches were evaluated in a comparative study with well-known state-of-the-art cognitive tests. Finally, the obtained results indicate that the proposed methodologies and tests can provide accurate indications of the presence of Alzheimer and the potential to further improve them by providing tests more adaptive to the patients.

References

- [1] M. Prince, M. Knapp, M. Guerchet, P. McCrone, M. Prina, A. Comas-Herrera, R. Wittenberg, B. Adelaja, B. Hu, D. King, A. Rehill, D. Salimkumar, *Dementia uk: Update*, Alzheimers Society.
- [2] A. Association, 2015 alzheimers disease facts and figures, *Alzheimer's & Dementia* 11(3) (2015) 332–384.
- [3] I. Tarnanas, M. Tsolaki, T. Nef, R. M. Mri, U. P. Mosimann, Can a novel computerized cognitive screening test provide additional information for early detection of alzheimer's disease?, *Alzheimer's & Dementia* 10 (2014) 790798.
- [4] Y. Abe, M. Toya, M. Inoue, Early detection system of senile dementia by behavior sensing, 17th International Symposium on Consumer Electronics (ISCE), IEEE (2013) 67–68.

- [5] L. J. Whalley, J. C. S. Breitner, *Fast facts: Dementia*, Oxford: Health Press Ltd.
- [6] M. Prince, A. Wimo, M. Guerchet, G. Ali, Y. T. Wu, M. Prina, *World alzheimer report 2015. the global impact of dementia. an analysis of prevalence, incidence, cost and trends*, Alzheimer's Disease International, London.
- [7] T. Lewis, C. Synowiec, G. Lagomarsino, J. Schweitzer, *E-health in low- and middle-income countries: findings from the center for health market innovations*, *Bulletin of the World Health Organization* 90(5) (2012) 332–340.
- [8] C. Rowe, S. D. Healy, *Measuring variation in cognition*, *Behavioral Ecology* 25(6) (2014) 1287–1292.
- [9] K. Wild, D. Howieson, F. Webbe, A. Seelye, J. Kaye, *Status of computerized cognitive testing in aging: a systematic review*, *Alzheimer's & Dementia* 4(6) (2008) 428–437.
- [10] J. M. Taekman, K. Shelley, *Virtual environments in healthcare: immersion, disruption, and flow*, *International anesthesiology clinics* 48(3) (2010) 101–121.
- [11] I. Tarnanas, W. Schlee, M. Tsolaki, R. Mri, U. Mosimann, T. Nef, *Ecological validity of virtual reality daily living activities screening for early dementia: longitudinal study*, *Journal of Medical Internet Research* 15 (2013) 114.
- [12] D. Weibel, B. Wissmath, *Immersion in computer games: The role of spatial presence and flow*, *International Journal of Computer Games Technology*.
- [13] R. I. Garca-Betances, M. T. Arredondo-Waldmeyer, G. Fico, M. F. Cabrera-Umpierrez, *A succinct overview of virtual reality technology use in alzheimers disease*, *Aging Neurosci* 7 (80).

- [14] T. D. Parsons, C. Courtney, Neurocognitive and psychophysiological interfaces for adaptive virtual environments, *Human centered design of e-health technologies* 9 (2011) 208–233.
- [15] T. D. Parsons, C. G. Courtney, B. J. Arizmendi, M. E. Dawson, Virtual reality stroop task for neurocognitive assessment, *Medicine Meets Virtual Reality (MMVR)* 18 (2011) 433–439.
- [16] T. D. Parsons, L. Cosand, C. Courtney, A. Iyer, A. A. Rizzo, Neurocognitive workload assessment using the virtual reality cognitive performance assessment test, In *International Conference on Engineering Psychology and Cognitive Ergonomics* (2009) 243–252.
- [17] A. P. Association, *Diagnostic and statistical manual of mental disorders (dsm-5)*, American Psychiatric.
- [18] R. French, Moving beyond the turing test, *Communications of the ACM* 55 (2012) 7477.
- [19] S. J. Russell, P. Norvig, *Artificial intelligence: A modern approach*, EUA: Prentice Hall.
- [20] K. Warwick, H. Shah, Human misidentification in turing tests, *Journal of Experimental & Theoretical Artificial Intelligence* (2014) 113.
- [21] S. D. Han, J. Gruhl, L. Beckett, H. H. Dodge, N. H. Stricker, S. Farias, D. Mungas, Beta amyloid, tau, neuroimaging, and cognition: sequence modeling of biomarkers for alzheimers disease, *Brain imaging and behavior* 6(4) (2012) 610–620.
- [22] H. C. Lau, I. K. Lee, P. W. Ko, H. W. Lee, J. S. Huh, W. J. Cho, J. O. Lim, Non-invasive screening for alzheimer’s disease by sensing salivary sugar using drosophila cells expressing gustatory receptor (gr5a) immobilized on an extended gate ion-sensitive field-effect transistor (eg-isfet) biosensor, *PLoS One* 10(2).

- [23] K. L. de Ipia, J. B. Alonso, N. Barroso, M. Faundez-Zanuy, M. Ecay, J. Sol-Casals, C. M. Travieso, A. Estanga, A. Ezeiza, New approaches for alzheimers disease diagnosis based on automatic spontaneous speech analysis and emotional temperature, *Ambient Assisted Living and Home Care* (2012) 407414.
- [24] A. Aztiria, M. Ugarte, A. Izaguirre, Gait analysis for identifying parameters related to dementia in intelligent environments, *Intelligent Environments (IE)* (2013) 3–6.
- [25] D. Unay, A. Ekin, Dementia diagnosis using similar and dissimilar retrieval items. in *biomedical imaging, From Nano to Macro* (2011) 1889–1892.
- [26] C. B. Akgul, A. Ekin, A probabilistic information fusion approach to mr-based automated diagnosis of dementia, *Pattern Recognition (ICPR)* (2010) 265–268.
- [27] C. Cordell, S. Borson, M. Boustani, J. Chodosh, D. Reuben, J. Verghese, W. Thies, L. B. Fried, Alzheimer’s association recommendations for operationalizing the detection of cognitive impairment during the medicare annual wellness visit in a primary care setting, *Alzheimer’s & Dementia* 9(2) (2013) 141–150.
- [28] D. M. Cruz-Oliver, T. K. Malmstrom, M. Roegner, N. Tumosa, G. T. Grossberg, Cognitive deficit reversal as shown by changes in the veterans affairs saint louis university mental status (slums) examination scores 7.5 years later, *Journal of the American Medical Directors Association* 15(9) (2014) 687.e5687.e10.
- [29] K. N. Fountoulakis, M. Tsolaki, H. Chantzi, A. Kazis, Mini mental state examination (mmse): a validation study in greece, *American Journal of Alzheimer’s Disease and Other Dementias* 15(6) (2000) 342–345.
- [30] J. Lindeboom, B. Schmand, L. Tulner, G. Walstra, C. Jonker, Visual as-

- sociation test to detect early dementia of the alzheimer type, *Journal of Neurology, Neurosurgery & Psychiatry* 73(2) (2002) 126–133.
- [31] J. M. Duchek, D. A. Balota, Failure to control prepotent pathways in early stage dementia of the alzheimer's type: evidence from dichotic listening, *Neuropsychology* 19(5) (2005) 687695.
- [32] N. B. M. Quental, S. M. D. Brucki, O. F. A. Bueno, Visuospatial function in early alzheimers disease-the use of the visual object and space perception (vosp) battery, *PLoS ONE* 8(7).
- [33] M. L. F. Pereira, A. C. M. von Zuben, I. Aprahamian, O. V. Forlenza, Eye movement analysis and cognitive processing: detecting indicators of conversion to alzheimers disease, *Neuropsychiatric disease and treatment* 10 (2014) 1273–1285.
- [34] M. D. Crutcher, R. Calhoun-Haney, C. M. Manzanares, J. J. Lah, A. I. Levey, S. M. Zola, Eye tracking during a visual paired comparison task as a predictor of early dementia, *American journal of Alzheimer's disease and other dementias* 24(3) (2009) 258–266.
- [35] T. D. Parsons, A. A. Rizzo, Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test, *CyberPsychology & Behavior* 11(1) (2008) 17–25.
- [36] L. A. Cushman, K. Stein, C. J. Duffy, Detecting navigational deficits in cognitive aging and alzheimer disease using virtual reality, *Neurology* 71(12) (2008) 888–895.
- [37] S. M. Shieber, Inverting the turing test [review of the most human human by brian christian], *American Scientist* 99 (2011) 14.

HIGHLIGHTS

- Computerized cognitive tests for early detection of Alzheimer's using Virtual Environments
- Virtual Reality devices for adaptive cognitive tests
- Computerized cognitive test that incorporates a reverse version of Turing Test