

A Multiplayer Voice-Enabled Game Platform

Francisco Oliveira
francisco.campos@ist.utl.pt
Instituto Superior Técnico

Abstract— As elder people become an increasingly large portion of the general population, mental health problems are becoming more common. Diseases such as dementia are the more common in the elder population, and lead to the gradual deterioration of various mental processes. As there is no cure for dementia, there is a big focus on its early detection, as well as Cognitive Stimulation which aims to delay the loss of cognitive ability. However, there are not yet many technological solutions that allow elders to undergo cognitive stimulation exercises in a group context without supervision.

This work aims to develop a multiplayer game to be applied in the context of cognitive stimulation. The game, based on the format of a quiz, serves as a platform for elders to exercise their memory in a more interactive way. The multiplayer component allows players to compete with each other, promoting social interaction which is essential for cognitive stimulation. This project also implements Speech and Language Technologies, such as speech synthesis and recognition, present in the VITHEA platform (Virtual Therapist for Aphasia Treatment) that served as an inspiration for this project. The architecture and design of the game aimed to simulate the experience of participating in a question and answer gameshow, using elements like a virtual agent to act as a gameshow host.

Generally, the results of the user satisfaction survey were positive, with all participants enjoying the game. The multiplayer component was particularly well received, reinforcing the idea that this type of system has potential.

Index Terms— Cognitive Stimulation, Elder games, Speech and Language Technology, Speech Recognition, Virtual Agent

I. INTRODUCTION

The world's population is ageing. Statistics indicate that in the present day a little over 20% of the Portuguese population is elderly, with estimates that in 40 years elders, will comprise over a third of Portugal's overall population [1]. Major advances in medicine in the past decades lead not only to there being a greater amount of old aged people, but also to a greater life expectancy. This means mental diseases such as Dementia are increasingly common. Dementia is a type of brain disease, in which there is a gradual deterioration of various mental processes such as memory and thinking, to a point that severely affects everyday activities [2].

With no cure in sight, a big effort has been put into the early identification of dementia, as despite it being such a common disease, Mild Cognitive Impairment (MCI) can often go unnoticed. MCI is one of the earliest stages of cognitive impairment that does not have noticeable symptoms, making it difficult to diagnose. For this reason, a lot of attention is now on not only identifying cognitive impairment, but also on cognitive stimulation as a means to maintain or even improve cognitive capabilities. The main area of application of cognitive stimulation for elders is in the administration of cognitive stimulation therapy (CST). This commonly consists of group sessions where elders discuss past events and play games that stimulate the brain, under the supervision of a therapist. There is not yet an easily accessible technological solution that can allow cognitive stimulation exercises to be performed in a group dynamic, which is important for it to have the largest impact possible.

This project aims to develop such a solution by developing a baseline system that would consist of a multiplayer game being applied in the context of cognitive stimulation of the elderly population. Taking the form of a quiz game, the multiplayer component allows elders to play against each other or in teams, which encourages cooperation and competition, a group dynamic which is important for cognitive stimulation to have a greater effect. This application would have Automatic Speech Recognition (ASR) and Text-to-speech (TTS) technology embedded in the system. The essence of this idea borrows heavily from VITHEA (Virtual Therapy for Aphasia Treatment), an award-winning online platform that has been used in several health-care institutions that treat patients with aphasia. The main characteristics of the VITHEA platform that are desirable for this project are its robustness and flexibility. Its robustness is a consequence of the method used to validate patients' answers [3], which is based on a keyword spotting (KWS) approach that allows the detection of specific words through the use of a model of background speech competing with the expected target keyword model. The flexibility of the platform is the other characteristic which lends itself well to the idea of a multiplayer quiz game. VITHEA essentially allows the therapist to add whatever variants of existing exercise models by simply uploading images, video or sound files depending on the desired exercise. This allows for the game to host questions that cater to the preferences of the players' preferences. Adopting these VITHEA inspired components to the game could provide a more stimulating experience for elderly.

II. BACKGROUND

In this section, a brief overview of the essential aspects that are going to be tackled in this work is presented, in the form of a description of important background information relating to the scientific areas that are most pertinent to this work. In addition to this, the relevant state of the art for each respective area is also presented, detailing the most recent studies and developments carried out by others on the various subjects. Firstly, the VITHEA platform is described briefly. Cognitive impairment and the development of cognitive stimulation as well as Social Agents is also explored. Finally, the related work discusses the use of Speech and Language Technologies (SLT) in automated systems and cognitive stimulation in the context of computer-based systems and video games

A. VITHEA

VITHEA (Virtual Therapist for Aphasia treatment) is a web-based software designed to act as virtual therapist for patients that suffer from Aphasia, a disability that consists of the inability to comprehend and formulate language as a result of brain damage [3]. VITHEA is composed of two main modules, which are the administration module and the client module. The first is designed for therapists to create new exercises, continuously monitor a patient's performance over time, and generally manage patient information. The second module is designed for the execution of the tests with direct input of the patients, allowing the user to pick from a number of exercises that range from audio-visual to text-based. Despite having started as a platform for aphasia treatment, the robust nature of the software and the language-oriented exercises already implemented, motivated VITHEA's adaptation for other purposes. One of the earlier efforts concerned tasks that are typical of screening tests for cognitive impairment [4]. The new exercises that were added belong to categories like naming objects and fingers, word recognition tasks, word recall tasks among many others that are commonly used in neuropsychological tests.

A generic utilization of this system would consist firstly of the virtual therapist presenting a set of different visual or auditory stimuli at the beginning of the session, depending on the choice of the patient and medical assistant. After completing whichever task was requested by the exercise, the patients have to verbally respond to these stimuli in a way that is appropriate for the exercise. The system records the utterance and sends the audio file via network to a server that has a web application that then processes it through an ASR module in a keyword spotting mode to generate the respective textual representation. The resulting text is then compared with a variety of predetermined textual answers for that given question so as to validate the patients' answer. Finally, the system provides feedback via a graphical cue, as well as a predetermined spoken message that is produced by the virtual therapist using TTS synthesis. The in-house systems used by the VITHEA platform to perform ASR and TTS are AUDIMUS and DIXI respectively, and were both developed by INESC-ID's Spoken Language Processing Lab (L²F).

B. Cognitive Stimulation

Cognitive stimulation has been defined as the engagement in a range of activities and discussions with the objective of generally enhancing cognitive performance and social behaviour [5]. Despite the platform developed in this work not being intended for medical use as a form of therapy for patients that suffer from diagnosed cognitive impairment, the most important research aspects of cognitive stimulation applied to elders is found in CST. Exploring this side of cognitive stimulation is essential as conclusions can be drawn from it that influence the design of the game platform.

The idea of "exercising" the brain and using cognitive training, which is the supervised execution of a set of standard tasks targeted at stimulating particular cognitive functions, originated from the treatment of elder patients that suffered from dementia [6]. These cognitive stimulation therapy sessions are executed in a group context, individually or even by family members with therapist's support with the exercises being applied with pencil and paper or on computers [5]. It was found that when applying this therapy to patients that suffered from mild dementia, there was a positive correlation between therapy and cognitive functioning [6]. These results were promising, and showed that, in conjunction with early diagnosis, CST is effective for patient's maintenance or even temporary recovery of cognitive capabilities.

1) *Effect on Quality of Life*: When considering CST, the main purpose of the treatment is to help the person recover from the symptoms caused by dementia and retain the quality of life they had previously. It has also been found that many times elders that suffer from dementia also suffer from depression. For this reason, QOL has been identified as an important measure of the success of any treatment applied to a patient. A study, [7], found that there was a positive correlation between QOL and improved cognition in memory, social functioning, activity level and communication.

2) *Group vs. Individual cognitive stimulation*: One of the vital elements of CST is that it should be done primarily in a group context. A small study on Individualised CST (iCST) [8] reported that the dementia patients who received individual sessions and group sessions preferred the group sessions as it made them feel positive and confident, with most looking forward to continuing with the group. Another study, [9], looked into the effect of cognitive stimulation therapy on two separate groups of elderly people, one comprising mostly of elders staying in nursing homes, while the others stayed in day centres. Results indicated that the day centre group had lower levels of depression compared to those who stayed in nursing homes. The implication here being that in day centres there are more activities and people have greater social contact compared to nursing homes where elders are usually isolated and lack psychological stimulation.

3) *Social agents*: One of the main areas of science that has increasingly been associated with elder care is social agents, both embodied and virtual. Many studies indicate that upon interacting with social robots, older people have a positive reaction in respect to mood and social interactions [10]. Projects such as Furhat and PARCEIRO, developed by the INESC-ID's research group on Intelligent Agents and Synthetic Characters

(GAIPS), have developed embodied, physical robots that focus on social interactions with users. PARCEIRO was developed as a social robot that played tabletop card games like Sueca with other players. The robot could communicate in several different contexts like encouraging the teammate, discouraging the opponents or saying the current score of the game. This made it quite a social environment, which encouraged many positive and enjoyable exchanges between the various players and the robot.

However social agents do not require a physical embodiment as social interaction can be done virtually through dialogue. In an experiment performed with elder users that were presented with a virtual agent and physical robot, the difference in terms of social presence was not great. This suggest that perhaps the most important factor is not so much the physical embodiment or not of the social agent, but rather the design of the social interactions that give the agent social presence [11]. The use of social agents in applications such as the card games described above or potentially quizzes and word puzzles also opens up a strong possibility that social agents could be used in the context of cognitive stimulation.

C. Related Work

Traditionally, cognitive stimulation is a practice that requires direct contact between elders and a trained instructor, which entails setting up an appointment that commonly takes place away from the elders' place of residence. This process can be time consuming as it requires that whenever an elder person wishes to participate in cognitive exercises they must leave their house. Furthermore, this traditional implementation of cognitive stimulation training can be expensive which, in conjunction with all the other issues previously identified, can make it an unappealing process for elders to undergo. It has been widely accepted that a computer-based approach to cognitive stimulation is the solution that best resolves many of the issues that come with traditional applications of cognitive stimulation training [12].

1) *Computer-Based Cognitive Stimulation:* Computer-based implementation of cognitive training can make it no longer necessary for elders to arrange appointments and interact directly with a specialized therapist, as they can be developed for online distribution and be readily available when accessed through a client PC connected to the internet, regardless of location. A digital platform also allows training programs to be more flexible as exercises can be added or removed at any point by whoever is over-looking the cognitive training process.

In addition to simple computer-based solutions, much attention has been given to the use of Speech and Language Technologies (SLT) in these systems, more specifically, speech recognition and speech synthesis. The VITHEA platform described earlier in this report is a good example of a computer-based implementation of cognitive exercises with the use of SLT. This platform shows how the use of SLT provides an adequate alternative for patient interaction when performing word-based exercises such as the ones that could potentially be used in the game platform developed for this thesis. It also has the desired characteristics of customizability and ease-of-access. It is for these reasons that it was used as the main source of inspiration for the program developed in this thesis.

2) *Video Games in Cognitive Stimulation:* It is increasingly common to see games being introduced into cognitive training as they are designed to be dynamic and entertaining, features that are seen as potential motivators for elders to stick with cognitive stimulation programs [12].

Traditional video games, such as strategy or role-playing games, have also been applied in the context of cognitive stimulation, in order to see potential cognitive benefits. A systematic review of computerized cognitive training [12] looked into the effects of traditional video games and found results that indicated this kind of training has the most impact on elders' reaction time and processing speed. The results for measures of executive function and memory were less successful, suggesting that perhaps video games are not an effective method of improving these aspects of cognition [12]. However, it is also possible that these poor results happened since the video games that have been used for these studies are not adequate for cognitive training as they were not designed for this purpose specifically [13].

One study that supports this idea [14] that compared the cognitive benefits of regular computer games and personalized, computerized training. It concluded that, despite small improvements using regular video games, individuals that suffered from cognitive impairment greatly benefited from personalized cognitive training. Part of the reason for this being that a regular computer game engages a wide range of cognitive process at once which can be overbearing for someone that suffers from deteriorated cognitive abilities. Given that, a personalized game that targets more specific cognitive processes can prevent the experience from becoming so overwhelming.

Naturally, the main focus of cognitive stimulation is directed towards obtaining measurable improvements in the various neuropsychological tests, in order to evaluate how successfully the exercises prevented cognitive decay. However, improving the quality of life of people that could be feeling the negative impact of cognitive impairment is of an almost equal importance. As such, qualitative improvements on elders' well-being also have to be taken into account when making this kind of evaluation.

When confronted with a large amount of free time upon reaching the age of retirement, seniors occupy a large portion of their time with leisurely activities. What motivates them to engage in these activities is simply a desire to be mentally and socially stimulated, or to simply enjoy the experience. There is a misconception that elders do not play games, however this is simply because games are now more associated with digital videogames, which are generally played by people younger than 65 years old. In fact, playing games in a group context is a very common activity amongst elders, specially card games, such as Sueca, Bisca or Canasta. These gaming activities are not only a means to socially interact, which is important as loneliness is a common problem amongst elders, but they are also generally seen as entertaining and engaging experiences [15].

This indicates that even when outside the context of cognitive stimulation, simply playing games can bring many benefits to the quality of life of elders. When investigating the qualitative effects of video games in cognitive stimulation training, one study [16] found that most elderly showed signs of happiness

and enjoyment, often exclaiming that the games were fun and entertaining to play. In addition to this, the elders' showed their contentment by often smiling and laughing as they played, particularly when certain objectives were completed. Despite taking pleasure in playing the games, participants found it difficult to interact with the computer using the mouse and keyboard, often avoiding them as much as possible. It is also important to note that the participants gave preference to games that did not require speed and precision, as well as those that had a time limit as it added pressure to the experience [16].

Taking into account the state of the art and all the research that has been done previously on the themes relevant to this thesis it is clear that computer-based solutions are the future of cognitive stimulation. The use of SLT in automating cognitive testing has also shown to be very promising, as it can provide accurate results and in doing so allows elders to interact with the computer in a manner that is more intuitive to them, using speech. Additionally, applying these technologies in the context of a video-game that attempts to provide cognitive stimulation by recreating exercises common in neuropsychological testing can make cognitive stimulation exercises more entertaining and engaging activity for elders. Finally, the research on group and individual cognitive stimulation also strongly indicates that a group experience provides a more enjoyable experience, as well as encouraging social interactions which is very important to elders that are lonely or suffer from depression. This highlights the importance of developing this game platform as a multiplayer experience. For these reasons, the implementation of the platform proposed by this thesis has the potential to successfully provide cognitive stimulation to elders.

III. MULTIPLAYER QUIZ GAME

As stated in the previous chapters, this thesis proposes to develop a multi-player videogame to be used in the context of cognitive stimulation. This application of a video-game implies that every aspect, from the game design to the way the users interact with the system, requires a careful study and analysis so that the players can benefit as much as possible from the experience.

A. System Architecture

The architecture of the system, which essentially dictates how the users are going to interact with the system, defines the quality of their experience. This section introduces each main component of the overall system architecture separately, starting with the choice on the type of game, and finishing with the game development software that was used to create the game.

1) *Game Choice*: The intention from the beginning of this project was to develop a game in the vein of what the VITHEA system did, which was to present a stimuli to the user, either text, visual, or auditory, and then to ask the user to respond to it. Therefore, a quiz game was considered to be a good option, as it would allow the integration of a set of question and answer interactions in the format of a game, in a way that could mimic the style of exercises used in VITHEA. It is worth noting that despite this type of game not being particularly designed to stimulate cognitive abilities, a quiz is still common practice in cognitive stimulation therapy sessions. However, an important

aspect to consider in the development of a quiz game for the targeted context is that the questions cannot be of multiple choice answers like many quiz games do. The questions need to be of the open ended type so players have to use their voice to answer the questions.

Moreover, because being able to play in a group context is an important aspect of cognitive stimulation, the game Buzz! was identified as one of the main sources of inspiration as it consists of a multi-player quiz game, incorporating many players in a "gameshow" context. This aspect of allowing the players to compete against each other plays an important role as it not only makes the experience more engaging, but it has also been found that elder users prefer games that encourage social interaction [15].

2) *Player Interface*: This aspect of the system architecture is extremely important, since a poor interface design that makes the users struggle to play would undermine the potential usability and benefits of the game itself.

Once again, VITHEA was used as a reference with its use of Speech and Language Technology. Similarly to that case, it was fundamental to allow the users to respond verbally to the quiz game questions. Not only is this a more user-friendly interface, but it also gives players the feeling that they are responding to questions as if they were participating in a gameshow, which helps with their engagement and enjoyment of the game experience.

As VITHEA was only designed for individual sessions, the user was supposed to be placed close to the computer. This way, a built-in computer microphone, or any close-talking microphone could be used to record a patients' answer with a reasonable quality for posterior speech recognition. In a group dynamic, however, where multiple users can be answering a question relatively far from the game display, the problem of distant speech acquisition, processing and recognition needs to be considered.

Despite initially attempting to mitigate the distant speech problem by attributing individual audio channels for each player, it was decided that a distant speech based solution should be implemented, very likely at the cost of decreasing the speech recognition performance. Thus, an array of microphones was required as they could provide acoustic localization to map the utterances to the different speakers in addition to being able to extract voice input from ambient noise better than regular microphones. A development kit for the Amazon Echo, a state-of-the-art device in distant speech that had recently been made available, was used for this purpose.

This device includes built-in voice signal processing which produces a very clear output with a very low noise-floor that is ideal for voice-controlled systems, as well improving ASR accuracy which is integral for the correct detection of players' answers.

In terms of interacting with the game and controlling the start and end of a recording, physical pressure buttons were chosen due to their simplicity of use. These kinds of pressure buttons are commonplace in daily TV gameshows that elderly populations have watched over the years, meaning that not only would they be more comfortable with them, but the buttons also enhance the experience of playing a competitive quiz game.

The buttons that were used as the physical pressure buttons were the Buzz! Game controllers. This was ideal as it not only has a main red buzzer, but it also contains additional buttons that allow for a wider variety of uses, like navigating a menu. Figure 1 shows the microphone array and the Buzz! Controller, the two input devices that serve as the main sources of interaction between the user and the system.

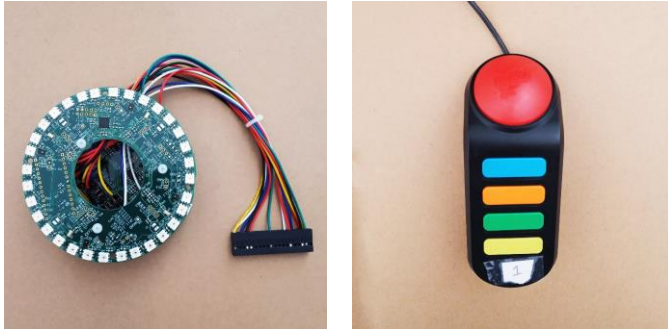


Fig 1 – Input devices used to interact with the system.

With the two chosen input devices, the way in which users interact with the system to answer the quiz questions is by pressing a button and then saying the answer aloud using their voice. This mechanism provides quite a natural user interface which almost mimics the experience of participating in a gameshow, further enhancing the experience and making it more engaging.

3) *Virtual Agent*: Virtual social agents can play an important role in providing meaningful interactions with users and often encourage social behaviour. Therefore, they are especially useful in systems such as the one being developed for this project, which intend to provide a social environment for elders. One of the virtual social agents developed for the VITHEA platform was used for this project. The physical representation of the character used is a low poly 3D model generated using the Unity game engine, with the ability to show facial emotions and simulate speech. Despite the character being built in a 3D environment, the game developed for this thesis was built in a 2D environment, which meant that this character was simply projected onto a 2D plane.

This character serves as the interface with the speech synthesis, and has a component that can forward and receive text from a TTS server. The speech synthesis is generated in a remote TTS server after having received text from the character. Once the character receives the server reply, its lips move synchronized with the speech synthesis, simulating that the character is talking. The TTS engine used in the game environment is DIXI.

Additionally, this character is also the interface with the speech recognition element, and has a component that sends audio to a remote server and receives a result in text form. The speech recognition is performed in a remote ASR server after receiving an audio clip from the character. The server then returns the words that were spoken in text format, or, in the case that no word was recognized in the audio clip, the server returns “_REPEAT_”. Depending on the requirements of the game,

however, the ASR system can also work in a Keyword Spotting mode (KWS) where in addition to sending the audio to the server, the character also sends the words that need to be recognized. If the recorded audio contains any of the keywords, these words are sent back to the character. In the case none of the audio contains a keyword, the server returns “_REPEAT_”. The ASR engine used in the game environment is AUDIMUS.

This implementation of the social agent allows him to take the role of the gameshow host by using speech synthesis to ask questions or give feedback, while also using speech recognition to “listen to” users’ answers. This use of SLT in the game is inspired by the VITHEA platform’s virtual therapist which had a social agent playing the role of the health worker. This gives users a more unique and engaging experience as it almost feels like a two-way interaction between user and virtual host.

4) *Game Development Platform*: During early planning stages of this thesis project it was decided that PyGame would be used to develop the game. PyGame is a free, open source Python programming language library used to make games and other multimedia applications. This choice was made taking into account that Python is a relatively simple programming language to use, and in this sense, it was seen as a convenient option to develop a quiz-like game, which is in some sense quite simplistic.

Despite PyGame being used quite far into the development process, when it became necessary to add the more complex components of the system, such as the integration of the Unity-based virtual agent, it proved to be a very limited platform. Considering the virtual agent is an indispensable component for the project, the development of the game using Python and PyGame was abandoned.

Unity, a game engine that uses common libraries and programming languages such as C#, was chosen to start developing the game from the beginning again. One aspect that greatly improved upon the transition to Unity was the game’s graphical interface (GUI), due to Unity’s editor that has the tools to create more complex visual elements without resorting to coding.

Another aspect in which switching to Unity proved to be beneficial for the development of this thesis was in its ease of compatibility when integrating elements such as the controllers and the social agent. The Buzz! controllers were very simple to introduce into the game, as Unity recognized it instantly as a joystick controller once they had been connected via USB.

The 3D virtual agent developed using Unity and the SLT that came with it were fully embedded with the game and could interact with the other elements.

The microphone was also very simple to introduce to the system as the script that recorded audio simply used the computer’s default microphone. This meant that once the microphone array was connected to the computer via USB, the game used this instead of the computer’s microphone.

5) *Final System Architecture*: With these components, the minimum requirements defined for the first iteration of the game, which were that it should be able to host up to four players that could play a round of randomly selected questions with a functioning scoring system and game host, were met.

In terms of distribution of the game, as the game was developed on a Windows operated computer, it was decided

that the easiest option was to simply build the game for Windows and play it locally on the laptop. This was a safe choice as all the simulations with the Buzz! controllers and microphone had been done on a Windows computer, and so there was no risk of having issues with the input devices. The game also requires a connection to the internet due to the 3D character that holds the speech recognition and speech synthesis elements of the game that are performed in remote servers.

Lastly, in terms of how the game was going to be presented or displayed to the players, the choice fell upon using a TV screen connected to the laptop, in order to provide a louder source of audio and clearer visual output for the players. This would allow players to be arranged in a more natural setup, as if they were sitting to watch TV, with the only limit being the length of the wires from the Buzz! controllers. Figure 2 shows the final architecture of the system developed for this thesis.

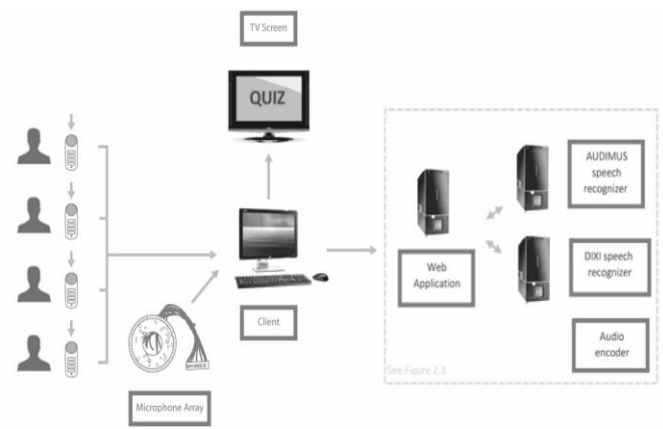


Fig 2 – Illustration of the final system architecture

Given this configuration, elderly users do not have to interact with a keyboard, a mouse or even a touch screen, making this a natural user interface. Naturally, supervision would be advised to ensure the experience runs smoothly and to aid the users in case something was not clear. However, the social agent acts as the gameshow host, interacting with the users and controlling the flow of the game, which means that the need for the participation of a caretaker to dynamize the game is very reduced. This is in line with the initial objectives of this thesis that aimed to provide a multiplayer cognitive stimulation platform in the format of an engaging game with SLT elements present in systems like VITHEA, so as to offer and engaging and entertaining cognitive stimulation experience.

B. Game Design

The flow of the game, the design, and the in-game mechanics are all elements that have been developed with eldercare and cognitive stimulation in mind. It is important that the game design caters to the preferences of the target users

1) *Gameflow*: The flowchart shown in Figure 3 describes overall structure of the game and serves as a means to see the order in which events happen in-game, as well as the different options players have.

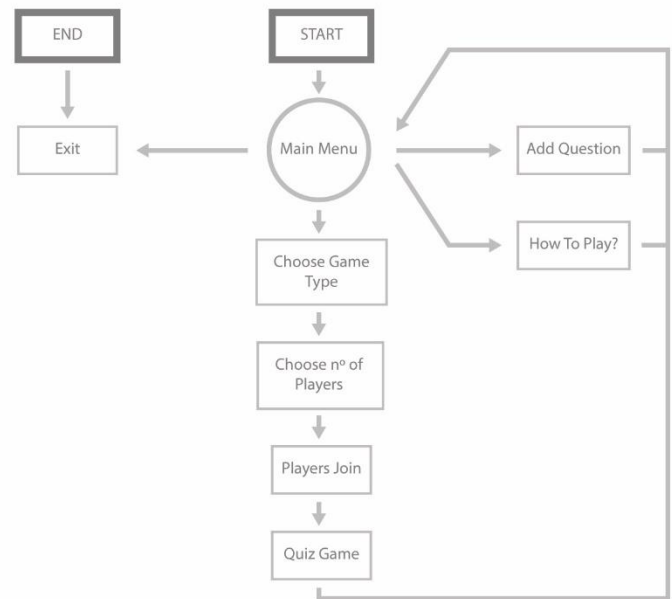


Fig 3 – Simplified flowchart of the quiz game

Upon starting the game, players are shown the main menu screen where they are faced with four options. They can start a new game, read instructions to learn how to play in case they have not played the game previously, add new questions to the game's question database and quit the game.

The instructions screen details how the main game works in very succinct points. In case the user wishes to add a new question to the system, they can use the visual interface developed for this purpose, where users can type new questions and add them to the question database in-game.

Figure 4 shows the main menu screen described above.



Fig 4 – Main Menu screen

In order to avoid that more than one player might interact simultaneously with the menu, the main navigation controls are limited to be used only by Player 1. If the player chooses to start the game, a screen that offers the choice between two game types. Here users can choose between playing the regular “Speed” mode, as well as a “Turns” mode.

After having chosen the game mode they wish to play, the user is given the choice to select how many players are going to participate in the game. Players can choose from one player, for solo play, to up to four players, for a multiplayer game. Then

users are faced with a player “join” or waiting screen. At this point in the game players must indicate that they are ready to play by pressing their buzzers. After all players have joined, each of them is asked to say their name, in order to create a player profile that displays each player’s name and score. The game begins after all the players are ready, and depending on the game mode that was chosen previously, the mechanics of the game differs.

The “Speed” game mode consists of a round of 10 questions where the first player to press their buzzer gets to answer the question. The player has 30 seconds to attempt to give the correct response before the timer runs out. At the end of the 10 questions, the player with the most points is the winner. In case there is a draw, i.e. when more than one player has the same score after 10 questions, players go on to a final stage where the first player to answer a question correctly wins.

The “Turns” game mode comprises of 4 rounds of questions, and in each round a different question is asked to each player individually. The player with the most points after the 4 rounds wins the game. Similarly to the “Speed” game, players have 30 seconds to respond to the question, or they can skip it if they don’t know the answer. Once again, if there is a draw after 4 rounds, the players with the top score move on to extra rounds of questions until a winner is found.

A single playthrough of the game ends after completing either of the games and returning to the main menu, at which point the whole process can begin again.

2) *Game elements and mechanics*: Game design and the mechanics of different elements of the game is, along with the system architecture, what defines how appropriate this project is for the desired audience. As such, it is important these game elements be implemented while keeping in mind the application of the game for elders.

The process of allowing the users to have their own names as their player name, despite not being necessary to the functioning of the system, was seen as an important element of the game. This allowed the experience to become more personal to the specific individual as they could relate to seeing their name and hearing it being said by the host, rather than simply having a generic Player X name.

This process is done using AUDIMUS in keyword spotting mode. This is to guarantee that nothing besides a name will be returned by the ASR server, filtering out any sounds or words mistakenly recorded and sent for speech recognition. The KWS mode requires *Extensible Markup Language* (XML) code that contains the keywords to be sent to AUDIMUS alongside the recorded audio. In this case, the keywords consist of as many common Portuguese first names as possible. This was implemented by storing a comprehensive database of 789 Portuguese names in a .csv file that was automatically imported as a table into the game once the user opened the application. This way, whenever the process of recording the player saying their name occurred, the program would scan the table of names and insert them in the XML code that was to be sent with the audio.

As determined previously, the questions used for this game had to be open-ended so users had to use their voice to answer. In order to record players’ response, the use of ASR in KWS mode is required to check if the response matches any of the

answers sent as keywords. To implement this, a database of around 160 Portuguese questions followed by their respective answers was stored as a .csv file and loaded as a table into the game environment upon starting the game. Before the game starts, the necessary set of questions and answers are retrieved from the table so they can be accessed during the quiz.

The questions chosen consist mainly of Portuguese historical and older pop culture questions to cater to the target users. Inspired by VITHEA, this game also has an interface that allows users to add custom questions and their respective answers. With this tool, a supervisor at a day centre could add questions that cater specifically to the tastes of the individuals that play the game in that environment.

The implementation of a score system is an important element of this project, as it serves as a motivating factor for players to engage with the game and answer questions correctly. Additionally, having the players’ scores constantly displayed on screen allows users to keep track of each other’s progress, promoting healthy competition. This competitiveness encourages the players to interact with each other, something that is greatly appreciated by elders when engaging with video games [15]. In the case of this game, players are given 250 points whenever they answer a question correctly and 0 points if they answer incorrectly or the time runs out.

As stated previously, players have the possibility of choosing between two different game modes. However, initially, there was only a “Speed” game mode. This game mode forced players to have quick reactions in order to press the buzzer first, despite research indicating elders usually avoided playing games that required speed as it put more pressure on them. In addition to this, it was possible for a player to never get the chance to participate and answer a question if they were not quick enough. This game dynamic could potentially put too much pressure on the players as they would have to worry about how quickly they could respond, rather than taking their time to use their memory and think about the question.

In order to avoid this problem, the “Turns” game mode was developed. In this game mode each player took a turn answering questions for a number of rounds. This removed the element of speed from the game as players no longer had to press their buzzer faster than other players in order to answer. Making the game turn-based also meant that all players can answer an equal number of questions, allowing each player to exercise their memory. The idea behind this game mode was to remove the sense of urgency players could have playing the “Speed” mode, whilst still maintaining the element of competitiveness as the players are still required to answer more questions correctly to win.

The graphical interface that players are faced with when playing a game strongly dictates how the experience flows and how easy users can interact with a game. The visuals were approached in a very simplistic fashion, in order to only display what was absolutely necessary. It was important that the visual interface be simple enough that players would be able to understand how to interact with the game without any previous experience.

In order to achieve this, whichever Buzz! controller buttons could interact with the game at a given stage were represented on-screen by respectively similar looking buttons, as seen in Figures 4 and 5. This makes the interface and navigation of the

game very intuitive, as players can recognize what buttons they have to press by comparing what they see on-screen with the controller they are holding.

The 3D character implemented in this game plays the role of the gameshow host. The host's main function is to drive the game, and provide the players with the basic instructions of what they have to do, as well as reading the questions aloud before giving feedback on whether the players answered correctly or not. However, this could prove to be quite tedious if the host only ever said the same specific instructions and feedback.

In an attempt to make interactions seem more friendly and less robotic, a series of more casual and light-hearted comments for the host to say were added throughout various different stages of the game. Not only this, but multiple comments were added for the same event, so that one would be picked randomly for the host to say in order to avoid becoming repetitive.

One of the most significant elements of the speech recognition process is defining the strategy used to start and stop recording the audio intended to be sent to the ASR server. When embedding ASR in the context of a game, it is important that this process feels as natural as possible, so that the game may flow more smoothly.

Inspired by the VITHEA system, the first method used to record users was based on a Push-To-Talk (PTT) approach, whereby the recording starts once a button has been pressed, and ends when the same button is pressed again. This solution worked well in the quiz section when playing in the "Speed" game mode. Once the host finishes reading out the question, the players need to be the first to press their buzzers in order to be able to answer the question. Therefore, this buzzer press signal can be simultaneously used as the trigger to start recording the player's answer.

However, the PTT approach was not ideal for the other stages of the game that required speech recognition. In both the player profile section and the "Turns" game, the system already knows who is going to speak before they start recording. This means that despite the player knowing that it is their turn to speak, they still have to press the button to record their speech. In order to avoid requiring user interaction to start recording, a solution that automatically starting to record once the host stopped speaking was developed. This method of triggering the recording allowed players' responses to begin more intuitively.

How users ended the recording could still be improved by using automatic endpoint detection. This is the process by which a system can automatically detect if a person's speech has ended by analysing the audio signal's properties. Unity provides tools to analyse spectrum data of audio clips, one of which is the amplitude of the Fast Fourier Transform (FFT). The average amplitude was used as the measure of sound level in this implementation. If the sound level went above a reference value for speech and then went below it again, the game detected that the speech had finished, which would trigger the recording would stop.

The combination of the various different strategies for voice recording used in the proposed system, meant that minimal user interaction was required. This allows for interactions between the players and the game to be more natural, resembling real-life dialogue more closely.

IV. RESULTS

In this chapter, the developed game is evaluated in two main aspects. The first is an evaluation of the user's experience when playing the game. The second part of the evaluation is an analysis of the KWS module used in the game to detect whether players' answers are correct or not.

A. User Experience

The approach used to assess the success of the game was that of using user satisfaction surveys.

This evaluation consisted of various sessions where users could play the game. Each session required 3 participants and was divided into two parts. The first part was an individual session playing the game on a laptop. In the second part, the 3 participants played the game together using a TV or monitor as a display. The players were asked to play both the Speed game mode and the Turns game mode.

Seven sessions were held for this evaluation, resulting in a total of 21 participants ranging from 24 to 88 years of age. 5 sessions were held at participants' houses in a more casual living room setting, which led to different conditions for each session. The remaining 2 sessions occurred at a senior University in a classroom setting, held under the same conditions. Figure 5 shows an image taken from one of these sessions.



Fig 5 – Photo of a group session at the senior University

The questionnaire developed for this evaluation was made up of two main sections. The first section contained 8 questions related to the overall game playing experience while the second section contained a further 6 questions that aimed to evaluate the players' preference in terms of individual or group experience, and in terms of Speed game mode or Turns game mode.

In order to make a comparison between the experience of elder users and younger users playing the game, the responses to the multiple-choice questions were separated into two groups. This distribution is shown in Table I.

TABLE I
DISTRIBUTION OF PARTICIPANTS IN THE TWO DIFFERENT AGE GROUPS

Group	Number of Participants
Young (24 - 28 years old)	4
Elder (59 – 88 years old)	17

In order to facilitate the grouping together of the results for analysis, users between 59 and 65 were grouped with the Elders. For the first set of questions this distinction was not made as results did not vary with age.

In general, the results were very positive. When asked to rate their enjoyment of the game, results indicate that everyone who participated had a fun time playing the game. This was also observable during the sessions as most players were eager to answer the questions and often smiled and laughed throughout the game.

Similarly positive was the rating of the game’s navigation system with most participants responding that it was easy to navigate the game.

The simplistic design of the game also meant most players found that playing the game was easy and intuitive. Most players understood the simple question and answer dynamic of the game without any difficulty.

Another very successful result obtained from this evaluation is the performance of the virtual host and player’s reaction to it. Players felt that the interactions with the virtual character were natural and that the host improved the experience of playing the game. This could be due to the casual, and often sarcastic attempts at humour made by the host at various stages in the game that helped create a fun and relaxed environment for the players.

Finally, the aspects of the game that received the least positive reviews were the recording and speech recognition elements. In one of the home sessions, the internet conditions were significantly poor which leads to ASR elements like checking the answer to be very slow, interrupting the flow of the game. The other element which received some negative feedback was the voice recording element. Despite in theory it making sense to have different methods of starting to record given the context of the game, having more than one method ended up confusing the players.

In terms of the second portion of the questionnaire, evidently, regardless of the age of the participants, players prefer to play the game as a group experience rather than individually. Interestingly, the main reason why participants preferred playing as a group was differed between young and older players.

Most elder players gave a lot of importance to the social interactions that came with playing with other people. Something that happened often with elder participants, was players helping each other answer the question, promoting healthy social interactions between players. Younger participants, however, favoured the competitiveness of playing against other players rather than just the social interactions. This not only led to the players being more engaged with the game, but it also brought enjoyment to the experience, as they would tease each other for getting questions wrong or losing.

These results strongly reinforce the importance of the multiplayer aspect of the developed game, as a gaming platform that can provide cognitive stimulation for multiple users, allows them to interact and engage with each other.

In terms of the “Speed” game vs. the “Turns” game, all of the younger participants preferred to play the Speed game mode.

Again, younger players valued the more competitive nature of the Speed game. Some elder participants shared this preference, and also enjoyed the more competitive aspects of this game as it encouraged more back and forth comments between players. However, there were some cases when playing the speed game where one of the participants did not manage to answer any question as they were not quick to press their buzzer.

More than half of the elders preferred playing the Turns game mode, with the main reason being that it gave everyone a chance to participate. This was an important result as the game’s intent is to ask questions and provide cognitive stimulation to all the participants and not to test people’s reaction time. Additionally, this game mode did not pressure players as much, which gave them time to discuss and think aloud before reaching a final answer. The atmosphere during the Turns games was also more casual and friendly, and the interactions between users often involved them helping each other reach the answer.

B. Keyword Spotting

A great portion of the system relies on the correct functioning of the KWS, and so its evaluation is important when analysing the success of the developed game. The metric used to evaluate KWS was the Word Verification Rate (WVR), which is used to determine how reliable the game is in identifying players’ answers as correct or incorrect. The WVR has the following equation

$$WVR = \frac{T_{positives} + T_{negatives}}{T_{positives} + T_{negatives} + F_{positives} + F_{negatives}} \quad (1)$$

with T representing True, and F representing False. The results for the WVR were separated according to the type of session: individual sessions, group sessions at home (Group I) and the group sessions that took place at the senior University (Group II). This is because the three categories had very different acoustic conditions under which the participants’ answers were recorded. The average WVR for each of these categories, as well as information on the correct and incorrect identification of the answers is presented in Table II.

TABLE II
AVERAGE WVR FOR THE DIFFERENT CATEGORIES

Session Category	Average WVR	T _{Positive}	T _{Negative}	F _{Positive}	F _{Negatives}
Individual	0.88	86	45	12	6
Group I	0.84	82	42	11	12
Group II	0.75	13	17	2	8

As expected, the WVR performed best in the individual sessions due to the players being close to the microphone, which resulted in a good quality audio recording of the answers. The Group I sessions had a slight decrease in WVR which was also to be expected as players were further from the microphone in the group game set-up. This introduced problems related to distant-speech which made the keyword spotting process more challenging. Finally, the result for Group II was particularly poor, mainly due to very poor acoustic conditions.

Overall, the performance of the KWS was underwhelming as the WVR for good acoustic conditions should be closer to 0.96. This was mainly due to a large number of false positives, as seen in Table II. The cause of this was answers that contained dates and years, because if a player said “1920” when the keyword was “1921”, this was considered correct as they are very similar. In order to avoid this, a specific grammar for answers which consisted of a year should have been implemented. This grammar would detect and return the specific year the user said which could then be compared to the answer. If these false positives were removed from the calculation of the WVR for the individual and home group sessions, their respective WVR would be 0.96 and 0.95, which is a considerably better result.

The other element of the game that used KWS was when players created their profiles and said their name. Analysing this element strictly through the lens of KWS performance this could be seen as a negative result. However, the players being attributed wrong names usually led to players to laugh and making fun of each other. These positive social interactions are a very positive result it is one of the main goals of creating this multiplayer experience.

V. CONCLUSIONS

This thesis proposed to develop a multiplayer voice-enabled game platform inspired by the VITHEA platform, to be used for cognitive stimulation of elders. Due to the intended audience being the elder population, the main focus of the project was on creating a platform that was designed to be as user-friendly and intuitive as possible. In addition to this, an extensive study of cognitive stimulation and how video games can be used as a tool for this purpose was done, which informed most of the game’s design elements. The game developed in this work succeeded in creating a central platform that could have a wide variety of components such as a virtual host, SLT, Buzz controllers and a microphone array interacting with each other, and all in the context of a quiz game.

One aspect of this thesis which slightly hindered the development of the project was the choice of the game development platform. Despite having determined that using Python and Pygame was the right choice because it was an easy programming language to learn, this proved to be a misstep. Having to restart from the beginning with Unity, which uses C#, after having developed a large portion of the game using Python proved to be a big challenge and delayed the progress of the project. Additionally, the lack of experience programming with Python and C# implied a considerable effort and time into learning these languages., at the cost of dedicating less time to certain elements of the game such as the speech recognition, which was not used to the best of its capabilities. Despite this, the developed game platform serves well as a baseline system, which could see these elements that were given less attention be further developed in the future.

The evaluation of the game itself produced quite promising results, with most players enjoying the main aspects of the game, such as the virtual host and the user interface. The main objective of creating a multiplayer experience that encouraged

elders to interact with each other was achieved, with many valuing the group experience.

Finally, it was gratifying to witness the great interest raised by the platform in the senior university, where some of the sessions were performed, with many elders showing how much they enjoyed playing, and wishing to play future iterations of the project.

Regarding future work, the main area of improvement is clearly the speech recognition and distant speech elements. Many of the false negatives that occurred when answering a question occurred with audio that was clear enough to be correctly classified. Given more time, an extensive study into the KWS could have been performed, in order to adjust the parameters of the ASR server in a way that would improve the results. Additionally, a more comprehensive analysis of the microphone array and how it processes audio could be developed, as perhaps that could help the game take better advantage of its capabilities.

REFERENCES

- [1] United Nations, Department of Economic and Social Affairs, Population Division (2017). World Population Prospects: The 2017 Revision, custom data acquired via website.
- [2] World Health Organisation, ‘Dementia’, World Health Organisation. Available online at: <<http://www.who.int/mediacentre/factsheets/fs362/en/>>, 2017
- [3] Abad, Alberto, et al. "Automatic word naming recognition for treatment and assessment of aphasia." Thirteenth Annual Conference of the International Speech Communication Association. 2012.
- [4] Pompili, Anna, et al. "Speech and language technologies for the automatic monitoring and training of cognitive functions." 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT). 2015.
- [5] Clare, Linda, and Robert T. Woods. "Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer's disease: A review." *Neuropsychological rehabilitation* 14.4 (2004): 385-401.
- [6] Woods, Bob, et al. "Cognitive stimulation to improve cognitive functioning in people with dementia." *Cochrane Database Syst Rev* 2.2 (2012).
- [7] Spector, Aimee, Charlotte Gardner, and Martin Orrell. "The impact of Cognitive Stimulation Therapy groups on people with dementia: views from participants, their carers and group facilitators." *Aging & mental health* 15.8 (2011): 945-949.
- [8] Orgeta, Vasiliki, et al. "Individual cognitive stimulation therapy for dementia: a clinical effectiveness and cost-effectiveness pragmatic, multicentre, randomised controlled trial." *Health Technology Assessment* 19.64 (2015): 1-108.
- [9] Vaz, Sérgio Filipe Alves. "A depressão no idoso institucionalizado: estudo em idosos residentes nos lares do distrito de Bragança." (2009).
- [10] Broekens, Joost, Marcel Heerink, and Henk Rosendal. "Assistive social robots in elderly care: a review." *Gerontechnology* 8.2 (2009): 94-103.
- [11] Heerink, Marcel, et al. "Influence of social presence on acceptance of an assistive social robot and screen agent by elderly users." *Advanced Robotics* 23.14 (2009): 1909-1923.
- [12] Kueider, Alexandra M., et al. "Computerized cognitive training with older adults: a systematic review." *PLoS one* 7.7 (2012): e40588.
- [13] Goldstein, Jeffrey, et al. "Video games and the elderly." *Social Behavior and Personality: an international journal* 25.4 (1997): 345-352.
- [14] Peretz, Chava, et al. "Computer-based, personalized cognitive training versus classical computer games: a randomized double-blind prospective trial of cognitive stimulation." *Neuroepidemiology* 36.2 (2011): 91-99.
- [15] Vasconcelos, Ana, et al. "Designing tablet-based games for seniors: the example of CogniPlay, a cognitive gaming platform." *Proceedings of the 4th International Conference on Fun and Games*. ACM, 2012.
- [16] Torres, Ana Carla Seabra. "Cognitive effects of video games on old people." *International Journal on Disability and Human Development* 10.1 (2011): 55-58.