Imperial College London

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Virtual reality platform for Self-Attachment therapy, assisted by a virtual agent with emotion recognition capabilities

Author:
Neophytos Polydorou

Supervisor:
Professor Abbas Edalat

Submitted in partial fulfillment of the requirements for the MSc degree in Computing (Artificial Intelligence and Machine Learning) of Imperial College London

September 2020

Abstract

Mental disorders have been a serious and growing problem for people around the planet in recent years and efficient ways of treatment are yet to be developed. Self-Attachment is a recently introduced psychotherapeutic method in which the patient creates a compassionate and affectional bond with their childhood-self using their childhood photos. The underlying theory and the promising results of the Self-Attachment therapy, in combination with its self-administrable nature, motivate further research in technologies that can enhance the procedure of the therapy. Virtual reality is a technology that has experienced rapid evolution over the last years, yet its ability to replace existing psychotherapeutic techniques remains fairly unexplored. The aim of this thesis is to create a user-friendly and highly interactive virtual reality platform that delivers the Self-Attachment therapy in an efficient way by creating an avatar of the patient's childhood-self based on a 2D childhood photo. For the fulfillment of this goal, the implemented platform has been equipped with the latest features of the head-mounted Oculus Quest device and interactable components. In addition, the virtual environment features a virtual agent who acts as an assistant and is able to predict the emotional state of the user. The aforementioned achievements have led to the creation of a complete virtual platform, capable of autonomously delivering the Self-Attachment therapy in a personalised experience for the user. During an impact evaluation trial, a version of the platform was tested by seven healthy individuals and the results were promising. The high levels of immersiveness of the platform and its ability to attract and maintain the user's engagement played a crucial role in obtaining positive feedback from the participants. Furthermore, the platform was successful in triggering participants' emotions, who in the end preferred the virtual experience over the original therapy which is based on the use of their childhood photo. Consequently, we believe that this project is the beginning of the creation of a tool that will be able to autonomously and effectively treat people with mental disorders in the near future.

Acknowledgments

This Independent Study would not have been possible without all the guidance and support of some very important people, to whom I would like to express my sincere gratitude for all their help.

First of all, I am very thankful to my supervisor Professor Abbas Edalat at Imperial College London, who has been there for me since the beginning, showing understanding and support. Thank you for all your constructive feedback and guidance over the course of this project.

Secondly, I would like to thank Dr Ronald Clark at Imperial College London for providing useful feedback on the interim report and offering suggestions on how to progress in this demanding project.

I would also like to express my sincere thanks to Syed Ibrahim Ghaznavi at Imperial College London for all their help regarding Unity and the trials that are carried out during this project.

I am also grateful to Adrian Millea at Imperial College London for providing mental support and valuable help with all the technical issues I encountered in the course of this study.

Equally important is to express my deep thanks to Lucia Simkanin at Imperial College London for the productive collaboration and the useful ideas we shared together during this journey.

Last but not least, my gratitude goes to my family for all their support and understanding.

Thank you all for your support and encouragement.

Contents

Chapter 1

Introduction

The 21st century is characterized as the century of technological revolution, where the widespread use and rapid development of technology have led to some of the biggest changes in our way of living, thus affecting all sectors of society, including the medical field. Medicine and technology converge together in patients' health care at faster rates and with innovative methods, in this way making the use of technology in medicine a career of the future.

Nowadays, we observe an increased number of cases of people suffering from mental disorders and therefore the use of technology for their effective treatment has become an urgent need. As a consequence, virtual reality constitutes a technological tool that is being developed at recent times and can be used to effectively and rapidly treat people suffering from mental disorders, such as anxiety and depression.

In fact, it is common for people with serious mental disorders not to receive regular treatment or to abandon their treatment before it is completed. Diverse factors help maintain a patient's commitment and willingness to engage with the treatment [10]. These factors usually have to do with the type of care provided to the individual for recovery, including key elements such as the accessibility of care and the patient's trust on the procedure of the therapy.

This paper will present and analyse a platform that has been created in order to deliver an innovative type of therapy, the Self-Attachment therapy. The platform is personalised on each patient, featuring their custom child avatar and emotion recognition capabilities. The flexibility of the platform to adapt to the needs of the patient is very important since researchers argue that by incorporating elements that meet the immediate needs of the patient, can positively contribute to the successful completion of the treatment [10]. An innovative approach, that does not only require open-mindedness and flexibility in the way the treatment is delivered, but also offers autonomy and respects the patient.

1.1 Motivation

Nowadays, the number of people that suffer from mental disorders is on the rise despite the advancement of technology and healthcare [28]. Mental disorders are a serious problem not only for the patients, but also for the society, because of their huge impact on it. Individuals face significant problems in their everyday life due to their mental disorder, which can also lead to physical pain or illness [19]. In addition, their problem affects both the economy and the society since the individuals are less productive, take a lot of sick leaves or have difficulties finding jobs [42]. It must be noted that the economic cost due to sick leave, as well as from the benefits and tax exemptions given to these individuals, is also high thus holding back the national growth [41].

A psychotherapy introduced by Edalat [13], called Self-Attachment therapy, is a self-administrable psychotherapeutic technique which aims to help people with chronic depression and anxiety. This therapy has promising positive effects on patients and has been enhanced by virtual reality (VR). A lot of research has provided strong evidence that VR can be used to treat people with various mental disorders as described with more details and examples in the Background chapter.

Furthermore, an initial implementation of Self-Attachment therapy into VR by Cittern et al. [7] shows positive results that give further motivation for research in this area. The VR can provide an immersive virtual environment that can be experienced in a very realistic way, allowing the user to provoke their emotions through real-time interactions. The costs and time required for a psychotherapy can also be dramatically reduced, in this way making it available to more people around the world. In addition to the numerous benefits that VR can offer to a patient, it will also be beneficial to the clinician who will assist the patient at the initial stages of the therapy. Hence, with the aid of a complete virtual psychotherapy, the clinician will be able to help more patients and still have more time to further engage with their research.

1.2 Aims and Objectives

The aim of the project is the development of a virtual reality platform that is going to replace the current methods of delivering the Self-Attachment therapy. We aim to create a complete VR application which is highly interactive and user-friendly, and can replace the human assistant in the process of the therapy.

In order to achieve this goal clear objectives need to be defined, which are listed below.

1. In depth understanding of mental disorders and Self-Attachment therapy.
2. Investigation of the current VR application in psychotherapy.
3. Review of the current VR platform designed for the Self-Attachment therapy.
4. Development of an appropriate virtual environment for the execution of the therapy on the-state-of-the-art Oculus Quest device.
5. Creation of new features and functionalities that enhance the experience of the user.
6. Incorporation of an emotion recognition algorithm that will make the platform personalised for each user.
7. Evaluation of the platform by collecting appropriate metrics, as well as by conducting trials on people.

1.3 Contribution

By the end of the project most of the objectives have been accomplished, some of which successfully and others less successfully. Nevertheless, the achievements of this project are very promising, indicating that in the near future a complete product that can be used to treat patients suffering from mental disorders around the world will become a reality. We also think that the work we have produced will motivate other people to further contribute to this research.

1.4 Outline

The rest of the report is organised in the following way. Chapter 2 introduces a lot of important information that is needed to provide a better understanding of the topic, as well as an extensive literature review on previous studies. Then, chapter 3 describes the implementation procedure of the platform, as well as the incorporation of an emotion recognition algorithm. In the following chapter, chapter 4, the evaluation methods along with the results obtained are analysed. Finally, chapter 5, will summarise the achievements and clarify the limitations of the project, before making suggestions for future work.

Chapter 2

Background

This chapter aims to inform the reader about any related background content, as well as some previous related works that will provide a better understanding of this project. Existing published works allow for a greater understanding of the progress made in the field and provide ideas and different approaches that can possibly contribute in the field. Therefore, after assessing the related literature, the project aims to find the most appropriate approach and adapt it to the Self-Attachment therapy. This chapter, firstly, introduces mental disorders before explaining the Self-Attachment therapy. Following, the chapter focuses on a review of research and meta-analysis papers related to the use of virtual reality in the treatment of mental disorders and the current implementation of the VR platform for Self-Attachment therapy. Finally, a small introduction on emotion recognition and on virtual agents is presented.

2.1 Mental disorders

The definition, as given by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) is that [1, p. 20]:

> A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning.

There are different mental disorders depending on the symptoms, as well as on the behaviour, thoughts and emotions of an individual. Common mental disorders are depressive and anxiety disorders with global prevalence of 264 and 284 million people respectively [28].

Anxiety disorders is a group of mental disorders where the patient experiences anxiety and fear with mild or severe symptoms. Some of the anxiety disorders are Generalized Anxiety Disorder, Separation Anxiety Disorder, Specific Phobia, Social Anxiety Disorder and Post-traumatic stress disorder (PTSD). This project is closely related to PTSD, which is a disorder that is caused by a traumatic event happened in the past of the patient and continues to provoke stress and anxiety in their present [1].

As given by DSM-5, depressive disorder is ‘the presence of sad, empty, or irritable mood, accompanied by somatic and cognitive changes that significantly affect the individual’s capacity to function’ [1, p. 155]. In addition to anxiety, this project targets patients with chronic and recurrent depression with the aim to treat the symptoms.

Well established and tested methods for treating mental disorders are the Cognitive Behavioral Therapy (CBT) and medication. The CBT is a psychotherapeutic technique that helps the patient to improve their mental health by affecting their behaviour, emotion and cognition. This therapy is widely researched with proven positive results [40].

Another therapy is the exposure therapy where, as the name suggests, the patient is exposed to a feared situation [2]. In this way, the avoidance behaviour is terminated and the patient learns how to deal with the situation in a controlled environment. The learning can be then transferred to a real-life situation. Some common variations of this therapy are the following:

1. In vivo exposure: exposure to a real-life feared situation.
2. Imaginal exposure: exposure by imagining a feared situation.
3. Virtual reality exposure: exposure in a virtual reality environment simulating a feared situation.

2.2 Self-Attachment therapy

This project is based on the Self-Attachment therapy proposed by Professor Edalat, which is a new psychotherapeutic technique, first reported at the Institute of Psychiatry in London in May 2013 [12] and introduced in the paper 'Introduction to Self-attachment and its Neural Basis' [13]. The Self-Attachment therapy is based on the Attachment Theory form John Bowlby [4] which suggests that from the first year of life, children create an emotionally attachment with their primary care-giver. The type of attachment determines the personality and emotional development of an individual in their adult life, as well as the way they perceive the world. There is one type of secure attachment and three types of insecure attachment known as Avoidant, Ambivalent and Disorganised attachments [18]. In the paper 'Self-attachment: A holistic approach to Computational Psychiatry' [14], Edalat explains the neurobiology of these types of attachment and provides evidence from ethology and religion where both can act as attachment objects. These attachment objects are utilised by a securely attached individual in order to feel secure in a stressful situation. On the other hand, an individual with an insecure attachment is vulnerable and is more likely to develop symptoms of depression and anxiety. This is where Self-Attachment therapy is employed in order to help the individual to feel the love they were deprived of in their childhood, and create a secure attachment that will allow them to control their emotions and cope with the distress situation. Thus, the individual [13, p. 6]:

> is imagined to consist of an inner child, representing the emotional self, rooted mostly in the right brain and the limbic system, which becomes dominant under stress, and an inner adult corresponding to the logical self, rooted mostly in the left brain and the prefrontal cortex, which is dominant in the absence of stress.

The aim of the therapy is the creation of an affectional bond between the childhood-self and the adult-self who takes the role of a new primary carer-giver. In this way, a secure attachment is created between the patient and the child that represents their emotional state. Edalat's paper [14] proposed that Self-Attachment protocol, that should be followed by the patient as part of the therapy and contained the following four stages:

1. The patient is introduced to the Self-Attachment therapy and its scientific and theoretical framework. This is an important stage as the patient must be dedicated and motivated in order to successfully complete the protocol.

2. In this stage the relationship between the patient and their childhood-self is initiated by looking at childhood photos. This is a visual stage emulating the child's early years when the vision was the main sense. Both a happy and sad photo are chosen by the patient for the therapy in order to remember the environment and relationship with their care-givers.

3. Here the bond is established as the patient falls in love with the childhood-self and vows to take care of and be a secure attachment object for the child. This process allows the patient to self-regulate their emotional state and subsequently to stay motivated for the rest of the protocol. In order to have a successful bonding process, the therapy employs techniques, such as singing and dancing, which are known for releasing dopamine in the brain.

4. At the final stage, the patient interacts with the childhood-self by embracing and reassuring it. Therefore, all negative emotions from past traumatic events are replaced by positive ones, while at the same time, insecure attachment objects are replaced by secure attachment objects.

The minimum practice of the protocol should be 20 minutes per session and must be repeated at least twice a day (once in the morning and once in the evening) for 8 weeks. It must be noted that the therapy will be successful when the created bond is maintained long after the therapy and thus every time the patient is in a distress situation they will be able to calm their childhood-self, meaning that they will be able to control their emotions.

In his work Edalat [14] also supported the argument that Self-Attachment can be modelled using game theory, where an interactive game between the adult-self and the childhood-self takes place. The game consists of two players, who each has two possible actions; the child can ask for support or not, and the parent can support or ignore the child. During the therapy, the aim is to make the transition from the avoidant equilibrium state, where the child does not ask for help and the parent is not willing to offer their help, to the secure attachment equilibrium state, where the child asks for the parent's support and the parent consents to offer it. The entire procedure can be modelled by Q-learning, which is a reinforcement learning algorithm, where the players are rewarded for performing the desired actions.

Edalat et al. [6] emphasize the importance of empathy in the Self-Attachment therapy, as the patient has to empathically connect with the emotional state of their childhood-self which will then create a bond between them. They present a neural model which is able to determine the level of empathy between one-self and others. This model is extended to apply to the Self-Attachment therapy in order to describe how the adult-self can increase empathy towards the child during the therapy.

Cittern [8] by using another computational model has showed how the patient can strengthen their bond and reduce stress in fearful or threatening situations. Based on the ideas of reinforcement learning, the patient can be trained to react differently to stimulus, by using a counter-conditioning procedure. The author claims that this procedure develops 'new social stimuli-reward associations in the OFC that drive an increase in OXT-modulated DA release, and a decrease in CRH release (which is involved in the stress response)' [8].

Self-Attachment therapy was successful in pre-clinical trials with promising results. Therefore, we think that virtual reality can enhance the procedure of the therapy to succeed in clinical trials as well.

2.3 Virtual Reality

Virtual Reality (VR) is a computer-generated environment that the user can experience with the aid of specific equipment, however it is not real. It uses graphics to create a virtual world where the user can have real-time interactions. At the moment, there are many VR solutions created by different tech giants like Facebook with the Oculus series, Google with Daydream series and Sony PlayStation with the PlayStationVR. This technology is commonly used for entertainment purposes, such as in the gaming industry, but also for healthcare, education and construction. Similar to VR is Augmented Reality (AR) technology where computer-generated images are jointly visible with the real world environment in the field-of-view of the user.

2.3.1 VR in psychotherapy

Virtual reality has been used in the resent years in psychotherapy since many studies have being suggesting the potential benefits of treating a range of mental disorders.

A study which proves that a virtual reality environment can produce a physiological and psychological stress response was conducted by Martens et al. [32]. This study measured the stress response of healthy individuals in a randomised controlled trial where the two groups were exposed in different VR scenarios. After several measurements, the two groups showed a significant difference in their stress response which suggests that VR environment can be 'felt real'. In this extent VR can be used to investigate the potential benefits of treating a range of mental disorders. However, this study also has its limitations, including the small number of only healthy young adult male participants. Nevertheless, the huge potential of VR in healthcare is noteworthy.

A feasibility study by Loucks et al. [30] suggests that the virtual reality exposure therapy (VRET) can help in the treatment of PTSD. They focused on the treatment of PTSD because of military sexual trauma and thus they developed an appropriate virtual environment that aims to safely deliver a treatment to 15 participants. Their approach, as the paper describes, was that [30, p. 59]:

> during VRE the patient recalls the events of the assault aloud while the therapist matches context and sensory details of the events in the virtual environment (e.g., location, sites, sounds) without showing an assailant in the VR.

They assessed the PTSD and stress symptoms before and after a series of VR sessions, as well as after three months from the completion of the therapy. The results showed significant clinical reduction in symptoms after the therapy and the reduction was also maintained after three months. In addition to the several study limitations mentioned by the authors, the problem we identify is that the procedure is not autonomous, thus meaning that the therapist needs to control the environment. As a result, in this project we have tried to solve this problem by creating a fully automatic procedure for the therapy.

The proof-of-concept study by Falconer et al. [16] investigated whether a virtual embodiment of a patient with depression within a virtual reality environment could have an improvement on their condition. The experiment included a scenario where in the first stage, the participants with depression were in the body of an adult trying to give compassion to a virtual child, while in the second stage, the participants were in the place of the child receiving compassion from themselves. The participants' feedback was based on previously established measures including: severity of depression, Self-compassion scale, Self-criticism scale, Fears of Compassion scale and a Virtual Reality Experience Questionnaire. By presenting all these collected data together, the study suggested that the immersive virtual reality can be beneficial to the participants. Thus, Falconer et al. concluded that [16, p. 77]:

> The most important finding was that there was evidence of significant reductions in depression severity and self-criticism, as well as increases in self-compassion.

However, the small number of participants and the existence of a single repeated scenario are limitations that might have had a big impact on their conclusion. Nonetheless, this study is closely related to this project and the feedback about the usability of the virtual reality must be taken into consideration as it was characterised as an 'enjoyable', 'interesting' and 'helpful' experience.

A study by Freeman et al. [20], also, showed promising results regarding the use of virtual reality to treat persecutory delusions. Their test hypothesis was that the threat beliefs of the patient could be reduced by dropping their safety behaviours. They suggested that the patient should live the feared situation without applying their behaviour which makes them feel safe. To implement this, the first step was to experience this situation through a virtual social environment and then apply their learning to real-life scenarios. Freeman et al. tried to prove the hypothesis by having two groups of patients interacting with the virtual environment where the first group was instructed to avoid their safety behaviours. By rating the patients before and after each experiment, those in the first group showed more improvement than the other group. They concluded that [20, p. 65]:

> In this study it has been shown that virtual reality can be used to present computerised versions of commonly feared situations to patients with persecutory delusions; that new learning can then take place; and, importantly, that the learning transfers into the real world.

However, implications, including the effort required by the patients to overcome the difficulties and enter the feared situations, combined with limitations, such as the assumption of perfect execution of the instructions by the patients, can have a huge impact on the presented results.

Freeman et al. [21] have conducted a systematic review of empirical studies that use VR in mental health and they realized that [21, p. 2394]:

> VR has extraordinary potential to help people overcome mental health problems if high levels of presence are achieved for situations that trouble them.

The authors investigated many studies which focus on different disorders like anxiety, schizophrenia and substance-related disorders and they concluded that treatments can be successfully implemented with VR.

We can understand that VR allows the people to enter difficult situations and learn how to deal with them through repeated exposures in the virtual environment. Although they know that the situation is simulated, their minds behave as if it is real and thus the learning can be transferred to a real-world situation. Another advantage of VR is the ability to simulate rare situations easily, as well as it can potentially reduce the need of therapists. Up to now, the VR equipment was more difficult to be obtained and therefore the research was moving slowly. This is not the case anymore as the technology is available outside the laboratories [21].

A quantitative meta-analysis by Opris et al. [35] investigated the effectiveness of VRET in anxiety disorders compared to classical interventions like the in vivo exposure therapy. The results from the 23 final selected studies revealed that VRET can be as efficient as classical evidence-based interventions. This result suggests that since VRET is effective at its initial stage and can treat patients with anxiety disorders as efficiently as the classical established methods, then it has the potential to be improved and therefore to become even more effective. However, the small number of studies and the limited types of anxiety disorders included in this meta-analysis suggests cautious interpretation of the results.

Ling et al. [29] carried out a meta-analysis in order to investigative 'the relationship between Self-Reported Presence and Anxiety in Virtual Reality Exposure Therapy (VRET) for Anxiety Disorders'. They hypothesized that anxiety is experienced by someone due to the sense of presence and thus the relationship between presence and anxiety must be explored so as to create a more effective VRET. After analysing 33 articles, Ling et al. concluded that self-reported presence and anxiety are correlated during VRET. However, this correlation was smaller or larger depending on the different characteristics of the participants in the experiment, such as age, gender, clinical status, disorder characteristics and technology characteristics. Related to this project is their finding on technology about which they concluded that [29, p. 7]:

> Advancements in immersive technology (i.e., higher degrees of freedom of the tracker and larger fields of view of the display) coincide with a higher correlation between presence and anxiety.

This finding justifies the need of a more realistic and user-friendly VR platform, which is one of the aims of our project as well.

A review by Garrett et al. [22] was carried out in order to evaluate the positive and negative points when using an immersive virtual reality (IVR) therapy for acute pain. They reviewed a number of carefully selected studies by comparing the IVR with other therapies. They concluded that the IVR therapy for pain control was 'fairly immature'. However, as the VR technologies are becoming more affordable and do not require lab-based equipment, the research in this area can provide more promising results.

A meta-analysis by Morina et al. [33] was conducted in order to investigate whether virtual reality exposure therapy (VRET) had an impact on the behaviour of patients with different kinds of phobias. They claimed that the advantage of VRET over other types of exposure therapy is that the virtual environments can 'resemble feared real-life situations' where is possible to control 'the quality, intensity, duration and frequency of the exposure' [33]. Their aim was to understand whether VRET helped the patients in a real-life situation where the conditions were not controlled. Based on some inclusion criteria, they selected suitable studies for the investigation and they assessed them based on the quality of their results. They concluded that the patients' scores improved after VRET and that VRET can change the behaviour of the patient in real-life scenarios.

The article by Hoffman et al. [26] reviewed evidence from studies that apply virtual reality (VR) to provide analgesia to patients with acute burn pain in order to reduce procedural pain. Pain reduction is achieved in all the studies mentioned in the article by using a VR system as a distraction mechanism, so that the patient paid more attention in the virtual environment rather than in the real world and thus receiving less pain. The logic behind this as given by the article is as follows [26, p. 184]:

> Pain requires attention. Humans have limited attentional capacity. Interacting with virtual reality uses a substantial amount of the patient's limited controlled attentional resources.

Patients from one of the reviewed studies 'reported a large, statistically significant and clinically meaningful reduction in pain during VR', while other studies also 'found that highly immersive VR systems are more effective at reducing pain than less immersive VR systems' [26]. The latest is very useful for the evolution of this project as it supports the need for an immersive implementation that includes auditory, visual and touch stimuli in order to have more effective results during the Self-Attachment therapy. In addition to the more immersive system, studies showed that by increasing the interactivity between the user and the environment, the 'analgesic effectiveness' can be increased. The article concluded that since VR practices are effective for very painful injuries, such as burn, it is possible that they will be effective and in other types of pain caused by depression and anxiety.

2.3.2 VR for self-attachment therapy

An immersive virtual reality mobile platform was proposed by Cittern et al. [7] which implements parts of the Self-Attachment therapy protocol. This platform aims to enable the user to create a secure attachment that was absent from their childhood and thus learn how to self-regulate their emotions. Their paper gave an overview of the Self-Attachment therapy and presented a prototype mobile app. Then, the authors suggested different development options for efficient implementation of the Self-Attachment protocol. The mobile app consists of videos that inform the user about the background and other aspects of the Self-Attachment therapy, as well as functionalities to manage the content of the platform. After each exercise the user can reflect on their experience by writing a diary and completing questionnaires related to anxiety, depression and attachment. These collected information are shown to the user at later stages to demonstrate their progress. Finally, the authors provided their future implementation plans, as well as their aim to conduct a randomised controlled trial on patients in order to assess the platform. Since the platform is under development there are many limitations, nonetheless, they plan to improve it. The aim of this project is to fix any issues related to the platform and try to enhance it with new user-friendly functionalities.

Their current VR platform consists of different protocols of Self-Attachment therapy. In the first protocol the user can create their own child avatar by using a photo from their childhood and then customise it by changing specific features of the face and the body. Then, follows the second protocol, where the user can select the emotional state of the child, thus projecting their emotion to the child. This protocol helps the user to conceptualise their childhood-self and to be compassionate towards the child. In the third protocol, the user tries to make an affectional bond with the child by approaching and embracing it. The final protocol is divided into different sub-protocols which the user can select from, depending on the cause of their current emotional state. The main limitation of the platform is that most of the protocols are not interactable and there are a lot of selections that must be manually done by the user, thus making the platform less user-friendly. Therefore, the user must take the initiative to select the appropriate sub-protocol and must correctly assess their own emotional state that will be projected to the child. However, sometimes, this is not possible because some patients face the problem of emotion regulation which requires emotion understanding and assessing. Consequently, the initiative that must be taken by the patients can complicate the progress of the therapy.

Ghaznavi et al. [23] provide an evaluation of the virtual reality platform for Self-Attachment. After a heuristic evaluation by two of the authors, problems were noticed, including non consistent navigation choices, non precise ray-casting controls and no description for some buttons' functionalities. After fixing these limitations, they performed a formative evaluation by tracking the progress of ten volunteers using the platform and by identifying common errors made by the participants. Furthermore, the paper presented the results from follow-up questionnaires completed by the participants, which constitutes a nice attempt to figure out the problems so as to improve the platform. In our project we will carefully consider their results and avoid similar errors.

2.4 Emotion recognition

Since this project aims to build an immersive VR platform which can actively adapt to each patient's needs emotion recognition plays an important role. A system that can recognise correctly the emotional state of the user is vital for an efficient and effective treatment. Automatic emotion recognition is a challenging task if we consider the different ways that emotions are expressed. Auditory and visual modalities include speech, facial expressions, and body movement, and can be combined to achieve a multi-dimensional analysis of emotions. Commonly used emotion dimensions are arousal and valence that are introduced by Russell in 1980 [39]. A lot of previous works with different approaches have remarkable results.

Tzirakis et al. [49] implemented an end-to-end deep neural network that takes as input audio and visual data and performs an emotion prediction. They used two networks to train the audio and visual data separately and the results were fused to train another network to produce the final prediction. The performance of their model was overall better than previous works on the same training dataset. Another application of neural networks for emotion recognition from Tripathi et al. [48] also presented good performance. The authors' approach was based on the training of different networks on speech, text and motion data, finding the best architecture for each modality and then fusing the results on the final layer to recognise the emotion. Interesting ideas are given by Rizos et al. [38] in their study for hate speech detection. Their contribution offered various ways to augment the data as a pre-processing step that allowed the extraction of additional information from the data and solved the problem of class imbalance.

By correctly detecting the emotional state of the patient, it will be easier to infer their anxiety and depression level and therefore keep track of the their improvement during the therapy. In addition, it will be possible to adapt the therapy to the needs of each patient based on their emotions. Finally, by collecting the emotional data from multiple users, further insights can be obtained that can improve the therapy itself.

2.5 Virtual agents

Stratou et al. [15] have created a virtual agent which is a complimentary support tool for healthcare providers. This virtual agent, called 'SimSense', conducts interviews in an automatic way by collecting and processing live data from different features of the interviewee. Features like people's gestures, voices, eye conducts and facial expressions are taken into consideration by the agent in order to successfully contact the interview. The main problem with this tool is the inability to make a psychotherapeutic intervention. The results of the interview must by analysed by an expert in order to understand the mental state of the interviewee and suggest a psychotherapy if needed.

Gordon et al. [25] have created a virtual environment with a virtual agent who conducts interviews. The animated virtual agent is able to adapt to the conversation based on the emotional state of the interviewee. Moreover, the agent is further able to adapt to their behaviour by changing the sitting postures. In order to accomplish this task, the agent collects and processes information on the person's gestures and expressions, as well as on their spoken utterances. The problem is that this particular emotion recognition model is only working on some specific utterances where the emotional state of the user is obvious.

2.6 Software and Hardware

The implementation of the platform requires both software and hardware. The software is implemented with Unity3D, which is a development platform created by Unity Technologies [45]. This tool allows the implementation of games and simulations in 3D, 2D as well as in VR and AR. Unity is a widely used engine for game development and is appropriate for the implementation of Self-Attachment therapy. The implementation requires a number of development tools, for example Android SDK, Oculus SDK and Avatar SDK.

Oculus Quest headset [34], which is a virtual reality head-mounted display (HMD) device, is used to test the implemented software. Oculus Quest is a standalone product of Oculus VR (division of Facebook Inc). It provides a six degrees of freedom system experience with the aid of two hand-controllers. This device is very easy to use and does not need any further equipment to work properly and for this reason is also suitable for patients carrying out the Self-Attachment therapy.

Chapter 3

Platform implementation

The focus of this project is the creation of a new immersive VR platform that functions with the latest version of Oculus Quest and implements the whole protocol of the Self-Attachment therapy. The graphic environment and the new functionalities in combination with the virtual assistant and the emotion recognition algorithm that are incorporated in the platform form a complete application. Furthermore, the use of a customised child avatar for each user, along with the real-time emotion recognition algorithm contribute to the creation of a personalised platform.

In this chapter, we are going to start by presenting the description of the design and the implementation of the platform with the new features and functionalities. Then we introduce the Self-Attachment therapy scenario which is based on the Self-Attachment protocol, and whose purpose is to deliver the therapy inside the platform. This scenario is the way of the virtual agent to communicate information and explanation about the platform and the procedure of the therapy to the user. What will follow is the description of the emotion recognition algorithm created by Lucia and James and how it is incorporated within the platform. Last, we aim to present the user guide that accompanies the application.

3.1 Platform design

3.1.1 User interface

User interface is the means that are available to the user and allow them to interact with the computer. Therefore, it is very important that these means are user-friendly and well designed in order to increase the user's involvement with the platform. Having the user engaged to the platform is vital for the successful completion of the therapy. As a consequence, this project lays emphasis on the design of the different elements and functionalities of the platform, with some of the most important listed below.

Hand tracking

Hand tracking is the latest feature of Oculus Quest headset which, as the name suggests, tracks the hand movement and gestures of the user in real-time. This feature can be used as a replacement or complementary to the provided Oculus controllers. Hands are the main input method for any VR application, while hand tracking allows a natural representation of the user's hands. What stands out is that hand tracking does not only track the palm, but also each finger in both hands. Hand tracking is a very powerful tool since it provides an increased sense of presence and is more user-friendly given that there is no need for the user to learn how to use the controllers. Figure 3.1 shows how the hands of the user are rendered in the platform.

Figure 3.1: Hand tracking feature

An experiment by Voigt-Antons et al. [50] also presented promising results about the immersiveness of the hand-tracking for different user interactions. In fact, they found out that the participants felt less aroused using hand tracking with higher valence rather than using the controllers.

The reason behind using hand tracking in this project is because it allows the implementation of all the features required by the therapy inside the platform. The most important feature is the user's interactions with the child avatar which become more realistic when using hand tracking. In combination with hand tracking, we have also used the interactable tool SDK, that is provided by Oculus, and which allows interactions with different elements to take place in the platform. This tool helps the developer to place a small spherical object on the finger tip of the virtual hand that allows the user to press buttons and enable or disable different elements in the scene.

The purpose of combining all these tools was to create customised hands which are adapted to the needs of the platform. In addition, all the interactable components in the platform are implemented in a way that are compatible with the virtual hands.

Navigation control

Due to the absence of controllers we have had to implement a functionality that allows the user to navigate and move inside the platform in order to visit the different rooms. As a result, we have used the hand tracking feature to implement a gesture that allows the user to move forward. By bringing the middle finger with the thumb together (pinch) at both hands simultaneously as shown in figure 3.2, the user is able to move forward. The forward direction is the one that the user is facing, which means that by turning the head to the desired direction, the user is able to change their path. Finally, the user is no more moving when they have stopped pinching. It must be noted that this moving functionality is very easy to use and very simple to understand.

In an attempt to make the environment more realistic, we have also separated the rooms with interactable doors. Therefore, in order to navigate from one room to another, the user has to open the doors by pushing them with their virtual hands. After the user passes through, the doors close automatically behind them.

Figure 3.2: Pinch gesture. The middle finger is touching the thumb at both hands simultaneously.

Buttons

The usability of the platform is vital for the successfulness of the therapy, thus there are different buttons that the user can press with the index finger so as to make a selection inside the platform. In each room there are buttons next to the doors that act as light switches and which can be used to turn the lights on and off independently in each room. It must be noted that there is also a button that allows the user to enter and exit the platform tutorial, as well as a button to turn on and off the television during the tutorial. Equally important is considered the button in the living room that can be pressed by the user in order to provide feedback on the platform.

At times, the user may also be asked questions by the virtual agent during the session. If the answer is simply yes or no, two buttons, as illustrated in figure 3.4, will appear in frond of the user to provide an answer. There will also be instances where the user is asked to confirm whether the emotional state predicted by the platform matches their current emotional state. In case the answer to both the first and second attempt is negative, then the emotion buttons, as shown in figure 3.5, will appear in order to allow the user to correct the predicted emotion. As a result, this will also give us an indication on how accurate is our emotion recognition algorithm.

Figure 3.3: The green cube on the table represents the interactable button that the user can touch and enter the tutorial. When the user enters the tutorial the label changes from "Enter Tutorial" to "Exit Tutorial". The red arrow points to one of the four light switches in the platform. It is a brown button that when it is pressed, the lights of the room are turned on or off respectively.

Microphone

The microphone of the headset is used to record the voice of the user during their conversation with the virtual agent. This input speech is used by the emotion recognition algorithm to make predictions about the emotional state of the user. The microphone is enabled only when the virtual agent asks a question to the user or during the stages of the protocol where the user has to speak or sing to the child avatar. The use of speech is an easy way of providing information and is very similar to psychotherapy, where the patient is involved in private sessions with their therapist.

Figure 3.4: 'Yes' and 'No' buttons allow the user to make a selection in order to answer a question.

Figure 3.5: Emotion buttons allow the user to make a selection of their current emotional state.

Interactions with the child avatar

During some stages of the Self-Attachment therapy protocol the user is required to interact with the customised child avatar. These interactions can be accomplished both by speech and touch. In cases where the user has to sing to the child, we have developed an algorithm that is able to detect the user's song and trigger an action. This action can be either a change in the emotional state of the child or some background music playing. In some other cases, the user has to approach and embrace the child avatar using their virtual hands. Similar to before, the platform is able to recognise the user's movements so as to trigger the necessary actions. Such abilities of the platform make the therapy procedure very realistic and simulate real-life situations where the user is encouraged to act normally; for example, to act in the same way as they were going to interact with a living child.

3.1.2 Virtual agent

The creation of the virtual agent avatar, who acts as an assistant to the user, is an important part of the platform. This agent uses speech in order to give clear instructions to the user on how to complete the Self-Attachment therapy. In addition to the instructions, the agent asks the user targeted questions that allow the emotion recognition algorithm to determine the emotional state of the user. The virtual agent stands in the living room and its avatar is illustrated in figure 3.6.

3.1.3 Child avatar

The Self-Attachment therapy is based on the interactions between the adult-self and the childhood-self of the patient. As a result, the existence of a representation of the child inside the platform is necessary. The best approach is the creation of an avatar that looks like the user in their childhood, which can also express emotions and respond to the actions of the user. For this project, we have decided to use the Avatar SDK [27] that allows the creation of a customised child avatar based on a photo. So as to improve the whole experience inside the platform we have then used different animations that are applied to the avatar and can change based on the user's actions.

Figure 3.6: Virtual agent avatar. The virtual agent is a humanoid character who holds conversations with the user inside the platform. In addition to the ability to speak, the agent can predict the emotional state of the user. The body avatar of the virtual agent is created by FAtiMA [17].

Customization

The child avatar is customised for each user based on a 2D photograph from their childhood. This representation of the child is more realistic as the users can actually see themselves and easily connect with the avatar. The head of the child avatar is created by the Avatar SDK as a 3D model and is attached to a body using a function that we have developed. The aim of this function is to allow the developer to upload a photo from the user's childhood in order to automatically create the 3D head of the avatar and attach it to a child's body. This avatar is then stored and can be used during the therapy. In an attempt to make the platform more user-friendly and effective, the created avatars also look very realistic, however they require some manual modifications, for example the style and color of the hair, as well as the colour of the clothes. Some example avatars are depicted in figure 4.8.

Figure 3.7: Cartoonish child avatar [47] that is used for demonstration purposes and will be replaced by the user's customised child avatar when the platform is customised for each user.

Animations

Animations are used in order to control the body and face of the child avatar. By using animations we managed to have specific emotions displayed by the child avatar and make the avatar move and dance. We have used animations so as to expose emotions like ‘happy’, ‘sad’ and ‘fear’, as well as an animation for dancing. The aim was to include animations for ‘disgust’, ‘surprise’ and ‘angry’ as well, however these were not provided by the Avatar SDK. Some of these animations are shown in figure 3.12. These animations are controlled by a script that determines how and when the animation has to change during the session. Equally important is that the decision is made based on the actions of the user.

Unity has a visual way to organise the animation which is called the Animator [46]. The Animator for this project is presented in figure 3.13 and consists of nodes and edges. Each node represents one state and it holds one animation clip alongside with different parameters, such as the speed of the clip. The nodes are connected with edges that represent the transitions between the states and these transitions are determined by some conditions. When these conditions are met, the animation state changes. As shown in figure 3.13, all states are connected to the idle (neutral) state in both directions. We have decided to make the connections in this way so as to achieve more efficiency and simplicity. This can be understood by taking an example where each state is connected with all other states, meaning that the number of edges that are needed would be much larger. In addition, the transition conditions would be more complex in this way slowing down other procedures in the platform.

During the different stages of the protocol the user should interact with the child avatar so as to improve the emotional state of both themselves and the child. Therefore, based on specific actions by the user, the animation of the avatar changes. For instance, when the user embraces the child, the happy animation is activated. Moreover, when the dance animation is initiated, the user’s favourite song starts playing in the background.

3.1.4 Environment

The environment of the platform consists of different areas that the user can visit inside the platform. There are four main areas which are described below:

1. An old house, as shown in figure 3.14, which is used at the initial stages of

Figure 3.8: Happy animation

Figure 3.9: Sad animation

Figure 3.10: Fearful animation

Figure 3.11: Dance animation

Figure 3.12: Animations that are used to express the emotional state of the child avatar are demonstrated on a cartoonish child avatar.

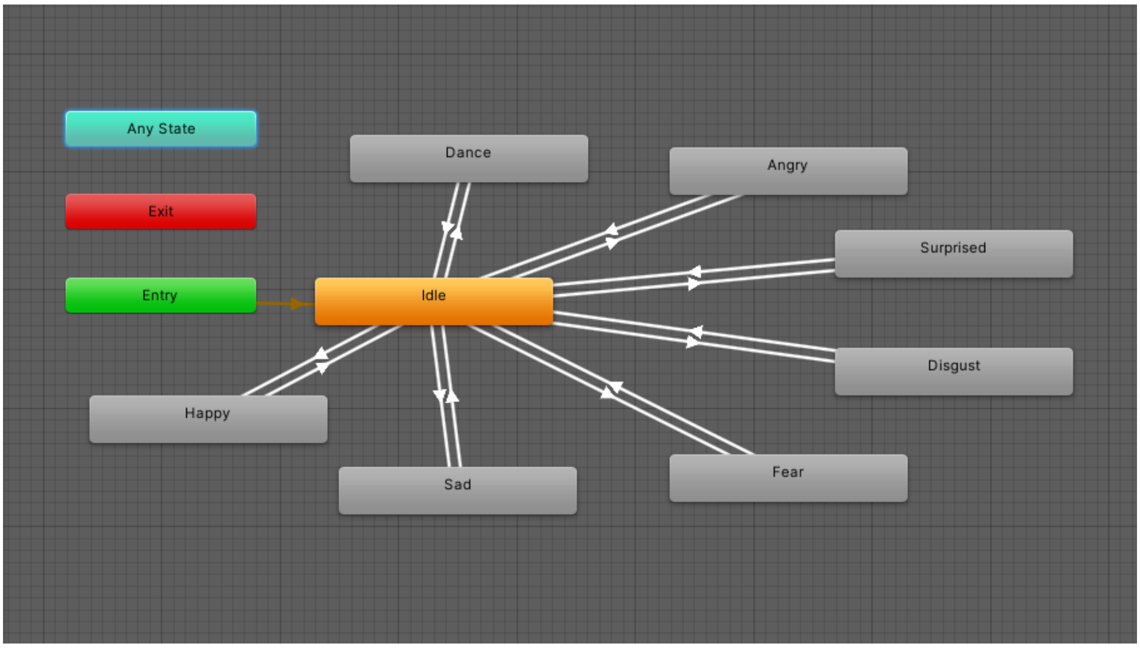

Figure 3.13: The custom animator that is used for this project. The nodes represent the animation state and the edges represent the transitions between the different states. The 'Entry' state represents the starting point of the animation which is immediately followed by the default state 'Idle'. The rest of the states are 'Happy', 'Sad', 'Fear', 'Disgust', 'Surprised', 'Angry' and 'Dance' that represent the different animations.

the Self-Attachment protocol. This is an empty building which is damaged and abandoned.

2. A room for the child avatar inside the new house, as depicted in figure 3.15, where the main part of the protocol takes place. This is the room where the user can interact with the child avatar. In the room, some children furniture and toys can be found.

3. A living room inside the new house, as illustrated in figure 3.16, which can be created during the different sessions of the therapy. In this room the user interacts with the virtual agent, who acts as an assistant and guides the user through the different stages of the protocol.

4. A house yard with trees, as shown in figure 3.17, which can be created after many therapy sessions in this way representing the progress of the user.

Figure 3.14: This is an old building which represents the old house that needs to be replaced by a new one during the Self-Attachment therapy. This asset can be found for free in the Unity Asset Store [47].

3.1.5 FAtiMA

FAtiMA [17] is a tool that allows the creation of virtual characters. This tool is able to handle the conversations between agents and players based on pre-determined set of dialogues. Therefore, when the agent asks a question, the player has to

Figure 3.15: This is the child's room inside a new house. This asset can be found for free in the Unity Asset Store [47].

Figure 3.16: This is the living room that we have created using different furniture assets that can be found for free in the Unity Asset Store [47].

Figure 3.17: This is the yard that surrounds the old and the new house and is covered with grass and multiple trees. The tree assets can be found for free in the Unity Asset Store [47].

choose between the listed answers. Based on the player's answer, the agent is able to emotionally react and reply to the player's question with an appropriate answer. The reason for using this tool in this project is to implement the scenario on the Self-Attachment therapy which is described in section 3.2. Consequently, Lucia [43] has implemented the dialogues between the agent and the user of the VR platform using the FAtiMA toolkit. Later, we have also adapted this tool in order to work with Unity and the other parts of the platform.

3.1.6 Code design and Functionalities

The primary API provided by Unity for coding is C#, which is a multi-paradigm programming language. The code is written in script files and it is organised in classes, where most of them are attached to Unity game objects. These scripts can modify the objects and allow interactions between them. Most of the scripts contain just one class and they share the same name as the class.

Most of the classes inherit from MonoBehaviour class [46], which is a mandatory class when working with Unity. MonoBehaviour contains useful methods that can be used for Unity specific operations such as the 'Start' and 'Update' methods [46]. The 'Start' method is called only once when a game object is created and can be used to initialize variables and to execute other procedures that must take place at the start. The 'Update' method is called once per frame and is very useful for procedures that must be executed at every frame, such as checking if a condition has been met.

Another important concept is coroutines [46]. A coroutine is similar to a method but with the ability to pause and continue execution based on some conditions. When a coroutine is called, it starts the execution in parallel to the main program that called the coroutine. In fact, this allows the coroutine to pause at anytime without affecting the procedures of the other methods. A coroutine can be paused for a specific time or until a condition is met. Reflecting on coroutines, we found them very helpful in the implementation of the platform and the Self-Attachment therapy scenario.

The different scripts that we have created are crucial to the program as they allow the implementation of the different functionalities of the platform. Below is presented a short summary of every script, sorted in alphabetical order.

ChildLookAtPlayer.cs

Let us begin with the first script which is using the world position vector of the player (user) and the world position vector of the child avatar to find the direction that the child needs to be facing in order to look at the player. Thus, the child's eyes are rotating in order to look to the calculated direction. This procedure is implemented in the 'Update' method in order to have real-time eye tracking. This small but important feature enhances the sense of presence for the user inside the platform and makes the child look as realistic as possible.

CustomAvatar.cs

The implementation of this script allows the automatic creation of the user's head avatar and must be run before the final application is built. The developer is responsible for making this happen, as they are the one working in the Unity editor, before creating the apk file for each user. The aim was to implement a procedure that can generate the avatar during the first session of the therapy and store it so as to be used in the next sessions. However, the Avatar SDK that we are using allows the creation of the avatar only in the Unity editor. If we had more time, we would definitely have attempted to understand and modify their implementation so as to work with the Oculus Quest.

Nevertheless, this script that we have created, combines different resources from Avatar SDK to automatically create, configure and save the customized avatar with the press of a button by the developer. What is more, the configuration of the avatar includes not only the attachment of the customised head to the body, but also the addition of the 'Animator' component to the avatar.

DisplayOnScreen.cs

The task of this script is to call a coroutine in order to display the time on the user's field of view. Given that only the hours and the minutes are displayed, instead of checking the system time at every frame, we have used a coroutine to implement checks on time every sixty seconds. In this way, unnecessary operations are reduced. The reason for displaying the time is to help the user not to lose the sense of time inside the platform, as well as to be able to check the time without removing the headset. Otherwise, the user would interrupt the therapy, in this way losing their concentration.

EmotionManager.cs

A method is contained in this script that can be called by other scripts and is responsible to set the emotion animations of the child avatar. Before setting the face and body animations, the method checks whether the given emotion is valid by querying the dictionary with the emotions. The dictionary contains the following emotions: 'happy', 'sad', 'fear', 'disgust', 'surprised', 'angry', as well as 'idle' and 'dance' for animation purposes. Whenever the method is called for the 'dance' animation, the user's favourite song starts playing in the background.

Fatima.cs

All the necessary procedures that allow the implementation of FAtiMA toolkit inside the platform can be found in this script. The script was initially written by Lucia but it was later modified to be compatible with the Unity and the other scripts.

It contains an initialisation method that loads all the necessary components and sets the starting state which may be different between sessions. Some of the components are loaded from files that are generated by FAtiMA. The problem we have encountered at this stage is that when we were building the application the files that were included in the Unity editor were not loaded into the application and thus were not imported during run-time. The solution to the problem was to create a folder with the name StreamingAssets and place the necessary files inside. In this way, Unity uploads these files to a server which can be then downloaded through a server request.

When the user speaks a method is responsible to update the current state of the protocol (FAtiMA specific variable). Finally, another method is used to return the next sentence that the agent should say to the user and update the current state.

FeedbackButton.cs / FeedbackScore.cs

These two scripts where implemented in order to allow the user to provide feedback inside the platform. When the feedback button is pressed then a number of questions will appear in frond of the user. In addition, another five buttons will appear with labels 1,2,3,4 and 5. For each question the user will have to select a score from 1 to 5 by selecting the corresponding button. When the user has made a selection the program continues with the next one. This is done with a coroutine. It must be noted the the user can provide feedback only once in each session.

GrabberHand.cs

The use of this script is to allow the user to grab objects using their virtual hands by pinching (touching the thumb with the middle finger). This is in experimental stage and it is not enabled in the final platform. The intended purpose is to increase the intractability of the platform as the user will be able to lift the child and place it in their arms. In addition, the user could potentially pick the sofa pillows which can represent an attachment object for the user as described in chapter 2.

HouseBuildingComponents.cs

This script is responsible for enabling the necessary visual components of the platform based on the session that the user is in. These components are the building parts of the house, such as the walls and the windows, as well as the different furniture. After each session, the platform builds a part of the new house to reward the user for completing another session and after multiple sessions the new house will be completed. This building progress is stored in order to be loaded in the next session.

LightSwitch.cs

In order to further enhance the realism in the platform, we have created a script that is attached on the different light switches inside the platform, in this way representing a real light switch. As a consequence, when the user touches the switch the main lights are turned on or off respectively. When the main lights are turned off, some back-up lights with very low brightness are turned on so that the platform will not be completely dark.

MicSettings.cs

This is an important script for the platform as it is responsible for recording the user's sayings, for sending the recorded data to the server and eventually returning the predicted emotional state. When a coroutine is called the microphone starts recording for a predetermined number of seconds (one utterance) and then stops. For this project the platform records utterances of 7 seconds length so as to match the way that emotion recognition algorithm was trained.

Then, the recorded clip is sent to another method that determines whether the user is talking or not during that 7 seconds period. We have implemented this functionality by taking the ogg vorbis sampling data from the audio clip, which are stored in an array of float numbers between -1 and 1, and then we have calculated the average of the absolute value of all samples. After that, the average is compared to a threshold value that we have set to determine if there is any input audio to the microphone. Therefore, if this procedure detects an audio input, then the coroutine makes a POST request to the server by sending the audio clip and enables the microphone again without waiting for the server's response. If the user has not started talking yet, then the microphone is enabled again so as to allow the user to input some speech. Finally, in the case where the user has been talking in previous recordings, yet in the last recording no audio was detected, that means that the user has stopped talking and the procedure is ended.

After the aforementioned procedure, the coroutine awaits all the server requests to finish so as to take the predicted emotion for each utterance. Finally, it calculates the frequency of each emotion and returns the most frequent one, which should be the user's current emotional state.

The procedure of sending a request to the server is executed by a method that can be described as follows. Firstly, the method receives the audio data which are stored in a float array and then converts them into a byte array. What follows is that the byte array is then sent to the server which uses the emotion recognition algorithm to predict and return the emotion. The method receives one utterance per time, however after having implemented it with asynchronous programming the method can make another request before the previous ones are completed. The use of asynchronous programming is very important and helpful as it ensures that the whole procedure is not delayed. After each request is successfully completed, the emotion for each utterance is returned.

MovingManager.cs

With the help of this script, the user is able to move inside the platform by pitching (touching the thumb with the middle finger) with both hands simultaneously. There is a method that checks whether the user is pitching in every frame and changes the position of the user if the conditions are met. The way the platform is built requires to cover a lot of ground by the user, which makes it unrealistic to demand from the user to physically have big enough spaces so that their real-world movement will be translated into virtual movement. For that reason, the above described functionality is an additional way for the user to move inside the platform and cover longer distances. Furthermore, another parameter that needs to be taken into consideration is that the movement speed is a variable that can be changed before building the application.

ProtocolManager.cs

The ProtocolManager script is the one that combines other scripts together and allows the implementation and control of the whole Self-Attachment therapy scenario.

This script is responsible for saving the information needed for the next session and for loading these information at the beginning of each session. Important information, such as the current protocol stage and the number of components built in the previous sessions are stored in a data file (.dat) on the device at the end of each session. The way these data are stored is by serializing them into a binary format to ensure that the user is unable to manually open the file and change its content. When the data are loaded into the program, they are deserialized in the same way.

After loading the data, the initialization of the different components, such as the graphic environment, the FAtiMA scenario and the child avatar, takes place. Then, the appropriate stage of the protocol is initialised, which is based on the stage the user ended the therapy in the last session. Each protocol stage requires a different implementation and as a result different coroutines are used.

Coroutines play a crucial role since they allow the program to pause for a given number of seconds or until another procedure is completed. For example, the program must wait until the user or the virtual agent has finished talking before continuing. In addition, when the user is prompted to enter the child's room to complete the therapy the program should wait until the user does so.

During the conversation between the user and the virtual agent, the user may have to answer yes or no to a question. When this happens, two buttons appear in frond of the user for them to make a selection. The program pauses until the user has made a selection.

The therapy scenario consists of the dialogues between the user and the virtual agent, and is returned by the fatima.cs script in text format. For this reason, we have decided to use a plugin that converts the text into speech in order to create a virtual agent who is able to speak. Thus, every time the agent needs to speak the script calls a method that starts the translation between the text and the speech.

SkyboxController.cs

The aim of this script is to simply make sure that the colour of the sky inside the platform changes based on the stage of the therapy. At some stages of the therapy the platform has dark colours that represent the unhappy emotional state of the user, while in others the sky colour changes to a brighter one, in this way representing happy emotions.

UserPosition.cs

During the therapy sessions, the position of the user inside the house must be known so as to detect whether they have entered a specific room. As a consequence, this script compares the vector position of the player with the position of the walls and returns information on the location of the user, whether they are in the child's room, the living room, or outside the house.

Buttons

There are different scripts that implement actions that must be triggered when a button is pressed. The **TutorialButton.cs**, for example, enables and disables the tutorial. Likewise, the **VideoPlayButton.cs** plays and pauses the self-massage video on the TV. The **YesButton.cs** and the **NoButton.cs** scripts indicate the user's decision of accepting or rejecting a choice and at last the **EmotionButton.cs** is responsible to inform about the emotion choices of the user.

3.2 Self-Attachment therapy scenario

The Self-Attachment therapy scenario that we have also introduced in the platform aims to translate the procedure of the therapy into clear instructions for the patient who is using the platform. It is worth noting that this scenario is based on the previous version of the Self-Attachment therapy and the reason for that is because the new updated protocol has not yet been translated into English. During one session the user must follow the instructions of the virtual agent in order to complete the therapy over multiple sessions.

There are four stages that will allow the patient to feel the love they were deprived of in their childhood and they are described below.

3.2.1 Stage I: Introduction to Self-Attachment therapy

In this first stage of the protocol the patient has to understand the scientific hypothesis of the Self-Attachment therapy and familiarize with the attachment theory in general. This will be achieved with the aid of the virtual agent who will explain the underlying theory of the therapy. Subsequently, the patient has to accept the terms and conditions of using the platform and undertaking the therapy.

Dialogue between the user and the virtual agent

Virtual agent: Welcome to the Self-Attachment therapy. My name is Ana and I will be your assistant throughout the therapy. Firstly, you will have to familiarize yourself with the platform.

[At this stage the user has to complete the platform's tutorial.]

Virtual agent: You can repeat the tutorial as many times as you wish as soon as we finish with the introduction. Now that you have familiarized yourself with the platform, I will give you a brief introduction to the Self-Attachment therapy.

Virtual agent: The Self-Attachment therapy is based on the Attachment Theory from John Bowlby, which suggests that from the first year of life children create an emotional attachment with their primary care-giver. The type of attachment determines the personality and emotional development of an individual in their adult life, as well as the way they perceive the world. Different types of attachment objects are utilised by a securely attached individual in order to feel secure in a stressful situation. Self-Attachment therapy is employed in order to help you feel the love you were deprived of in your childhood and to create a secure attachment that will allow you to control your emotions and cope with the distressing situation. In this therapy you are imagined to consist of a childhood-self, representing your emotional self, which becomes dominant under stress, and an inner adult corresponding to your logical self, which is dominant in the absence of stress. The aim of the therapy is the creation of an affectional bond between the childhood-self and the adult-self who takes the role of a new primary carer-giver. In this way, a secure attachment will be created between you and the child inside the platform. You will also be guided through all the stages of the Self-Attachment therapy.

Virtual agent: Would you like me to repeat the previous part?

User: [YES/NO]

[The previous description is repeated until the answer is negative.]

Virtual agent: Okay, let us move on.

Virtual agent: What do you think about the therapy?

User: [Here we give the user the chance to talk so we can apply emotion recognition.]

Virtual agent: It is important to believe in this therapy and be committed. Are you sure you can do this?

User: [YES/NO]

[The question is repeated until the answer is positive.]

Virtual agent: Very well.

Virtual agent: Do you accept the terms and conditions?

[The terms and conditions are written in the user manual (see 3.5).]

User: [YES/NO]

[The question is repeated until the answer is positive.]

Platform design

Initially, the user is found in an empty room with the virtual agent. This room is located in an old house that cannot provide safety and represents the inner world of the individual who suffers from a mental disorder. By the end of the therapy this old house must be replaced by a robust house which is a safe haven, representing an individual with secure attachments. The building process will reflect the progress of the patient who is trying to repair their damaged inner world and build a better life.

At this stage the introduction and tutorial take place. The aim of the tutorial is to allow the user to learn the navigation controls and how to interact with the platform. Moreover, a video that shows how someone can do self-massage is playing on the TV.

3.2.2 Stage II: Connecting with the childhood-self

At this stage the patient needs to create a psychological connection with the childhood-self. A 3D child avatar represents the childhood-self of the patient, with whom the patient can interact in order to establish a connection that is necessary for the rest of the therapy. The virtual agent will guide the user through the process.

Dialogue between the user and the virtual agent

Virtual agent: Now I will ask you to enter the child's room, where you have to conceptualise the childhood-self to develop empathy and then compassion towards it. Initially, the child will be happy, thus by looking at it try to recall happy moments from your childhood. After you do that you can turn off the light using the light switch next to the door. When you turn off the light the child will become sad. Therefore, try to recall sad memories from your childhood. You can turn the light on/off as many times as you want. Try to connect with the child. Take your time. I will be here.

[At this stage the virtual agent awaits the user to enter the room and complete the described exercise. At any time the user can exit the room and the virtual agent will continue with the dialogue.]

Virtual agent: Hello again. Did you manage to connect with the child?

User: [YES/NO]

Virtual agent: How are you feeling?

User: [User talking. Emotion recognition.]

Virtual agent: I see that you are [Detected emotion]. Is that right?

User: [YES/NO]

[If the answer to the above question is 'NO' then the agent asks whether the second predicted emotion is the correct one. If the answer is again 'NO' then the agent says:]

Virtual agent: Okay, which emotion best describes what you are feeling?

User: [The user makes a selection between 'happy', 'sad', 'fear', 'disgust', 'surprised' and 'angry' using the buttons in frond on them.]

Platform design

The user is in the same old room, as in Stage I, from which they must exit and enter the child's room. Inside the child's room there is the user's customised child avatar. Initially, the room is bright and the child feels happy. The user can switch off the light to make the child feel sad. Afterwards, the user can switch on the light to make the child happy again and repeat the procedure as many times as they wish. During this exercise, the individual tries to recall their happy and unhappy memories and tries to connect with the child. This initiates a bond between the two which is important for the rest of the therapy.

3.2.3 Stage III: Falling in love with the childhood-self

At this stage the patient needs to create an affectional bond and a loving relationship with the childhood-self. This will give them motivation and energy to complete the therapy.

Dialogue between the user and the virtual agent

Virtual agent: Now, I would like you to enter the room again and sing your favorite love song to the child in order to make it happy. This will help you to create an affectional bond and fall in love with the child. You can go whenever you are ready.

User: [The user enters the room and sings. We can probably predict arousal and valence level from the singing. User exits the room.]

Virtual agent: How are you feeling now?

User: [User talking. Emotion recognition.]

Virtual agent: Now you have to adopt the child and vow that you will support and protect it. Are you sure you can do that?

User: [YES/NO]

[The question is repeated until the answer is positive. The user should repeat the previous stage again if necessary.]

Virtual agent: You are ready for the main part of the therapy. You will have to interact with the child as a good parent in order to minimise the negative emotions and maximise the positive affects. This can take multiple sessions, which allows you to leave and enter the platform as many times as you want. I will keep track of your progress.

Platform design

The user enters the child's room and loudly sings a song which changes the emotional state of the child from sad to happy. If the user stops singing then the child gets sad again in order to show to the user that must keep singing. After the user exits the room, they will notice that the walls of their new house have been built and thus the development of the new house has initiated.

3.2.4 Stage IV: Developmental exercises for the childhood-self

This is the main stage of the therapy where the patient focuses on retraining and re-parenting the childhood-self. After completion of all the previous stages the user is ready for this stage. This stage consists of different sub-stages that will help the patient to minimise the negative affects and maximise the positive affects. This will be done over a period of time and after repeated sessions inside the platform. The aim of these exercises is to help the patient to learn how to regulate their emotions and apply that learning to stressful real-life situations.

Dialogue between the user and the virtual agent / Platform design

Virtual agent: Good morning/afternoon, how are you today?

User: [User talking. Emotion recognition.]

Virtual agent: I see that you are [Detected emotion]. Is that right?

User: [YES/NO]

[If the answer to the above question is 'NO' then the agent asks whether the second predicted emotion is the correct one. If the answer is again 'NO' then the agent says:]

Virtual agent: Okay, which emotion best describes what you are feeling?

User: [The user makes a selection between 'happy', 'sad', 'fear', 'disgust', 'surprised' and 'angry' using the buttons in frond on them.]

Virtual agent: Let's start with the different stages of the therapy.

Every time the user enters the platform, they will be in this main stage. At the beginning, the sky will be dark and after the first three sub-stages that will follow, it will become sunny with mountain views.

Type A: Sessions for processing the painful past

Virtual agent: Try to recall a traumatic episode in your childhood with as much details as possible. Try to remember a different event from previous sessions. How did you feel? Did you feel fear, helplessness, humiliation or rage?

User: [User talking. Emotion recognition. The predicted emotion is projected to the child avatar.]

Virtual agent: When you are ready, please enter the room and parent your childhood-self. You can loudly reassure and embrace the child to show support as a good parent. For example, say with a loud voice, “Why are you hitting my child?” and “My darling, I will not let them hurt you any more.” Also, you can cuddle your childhood-self by giving yourself a massage. All these will help to change the emotional state of the child and thus your emotional state.

User: [The user enters the room. We can do emotion recognition on the reassuring talk of the user. At any time the user can exit the room.]

Initially, the child will have negative emotions (‘anger’, ‘sad’, ‘fear’), depending on the predicted emotional state of the user. However, after being reassured the child will have neutral emotions and after being embraced the child will be happy.

Type B: Sessions to process the current negative emotions

Virtual agent: Now, what are your most recent negative emotions? Are they related to family, friends, work, education or social affairs? Try to feel these emotions right now.

User: [User talking. Emotion recognition.]

Virtual agent: How are you feeling? Do you feel anger, rage or fear?

User: [User talking. Confirmation of predicted emotions. The predicted emotion is projected to the child avatar.]

Virtual agent: Now enter the room and parent your childhood-self. You can loudly reassure and embrace the child to show support as a good parent. Also, you can cuddle your childhood-self by giving yourself a massage. All these will help to change the emotional state of the child, and thus, your emotional state.

User: [The user enters the room. We can do emotion recognition on the reassuring talk of the user. At any time the user can exit the room.]

Initially, the child will experience negative emotions ('anger', 'sad', 'fear') depending on the predicted emotional state of the user. Similarly to Type A stage, the emotional state of the child will change to neutral and then to happy.

Type C: Protocols for creating zest for life

Virtual agent: Now enter the room and sing your favourite love song to the child. Also, try shaking your head and shoulders and moving your eyes, eyebrows, hands and arms.

User: [The user enters the room, spends some time interacting with the child and then exits the room.]

Virtual agent: Singing will help you contain your negative affects and experience real joy. Thus, you have to repeat these exercises under many different circumstances, such as when you are walking or when you are working.

At this stage, the child's emotional state will be neutral, however, after the song the child will be happy and then will start dancing. The user's favorite song will be playing in the background.

Type D: Getting over the negative emotions

Virtual agent: On the wall you can see an image of the Gestalt vase. This vase represents the negative emotions and the more you look at it, the more you get drowned into its negativity.

User: [The user is starring at the image.]

Virtual agent: However, after successful completion of the previous exercises you have created a strong pattern of love and when you look at the image again, you will discover two white faces that represent your childhood-self and adult-self who face each other.

User: [The user is starring at the same image.]

The Gestalt vase will be on the wall on a white background so that the vase is more obvious. Gradually, the white background will be reduced so that the two faces will be more transparent to the user.

Type E: Socialising protocols for the childhood-self

Virtual agent: Giving yourself a massage will help you improve more your emotional state.

User: [Self-massage.]

Virtual agent: Do you feel better now?

User: [YES/NO]

Virtual agent: You have successfully competed all the exercises for this session. It is very important now to apply the things you have learned in your real life through your interactions with other people. Can you do that?

User: [YES/NO]

Type F: Creating a more optimal internal working model

Virtual agent: At the end of each session you will notice that a piece of your new beautiful house has been added. After many sessions you will have a complete house which will be your secure haven. This house represents your new optimal internal working model for interpreting and managing your relationships with others. Your developed internal working model will help you have peace with yourself and with other people.

Virtual agent: You have reached the end of this session. How are you felling now?

User: [The user is talking. Emotion recognition will reveal the improvement of the emotional state of the user.]

Virtual agent: Very well. See you next time. Goodbye.

After each session a piece of the house will be built, starting with the walls and then adding furniture piece by piece so that eventually a yard full of trees will be created.

3.2.5 Optimisation

Undoubtedly, the optimisation of the platform and the code are crucial factors so as to ensure the creation of an application with maximum performance. Oculus Quest headset is a stand-alone device with less performance than a computer and which cannot harness the power of a good performing computer. For this reason, a high quality implementation is very important for the overall efficiency of the platform.

During the implementation of the platform, we have been using the OVR Metrics Tool that has allowed us to view different statistics about the VR session. The statistics can be displayed in real-time inside the platform, as shown in figure 3.18.

Figure 3.18: Real-time metrics during the session [9]

Some of the metrics that we have been observing were the frames per second (FPS), CPU and GPU level, memory used and CPU and GPU utilization, as well as the App GPU time. Every time the metrics were not satisfying, we were trying to improve them by changing the way different objects were rendered in the scene. Also, we were trying to optimise the procedures in the code and remove unnecessary elements from the scene. For example, the performance of the platform improved a lot when we changed the lighting in the scene from real-time lighting to fixed lighting. In addition, we have removed the shadows and changed the objects in the scene that are not interactable to be static. Finally, we have removed the rendering and loading of components that are behind walls, so that they are only going to be enabled when the user enters the particular room, thus making the platform much more efficient.

3.3 Emotion recognition

Most importantly, the platform has the ability to recognise the emotional state of the user and thus it can provide a more personal experience and a more effective therapy to the user. The purpose of the emotion recognition algorithm is to predict the emotion of the user in real-time by making a prediction for every user's utterance. In this way, the predicted emotion can be used for a personalised therapy where the emotion is instantly projected to the child avatar. In addition to that, an empathic therapist can be created who responds to the user in real time based on their emotions. As a result, the benefit of having an instant emotion prediction is crucial not only for this project, but also for future applications.

Humans express emotions in many ways, for example using speech and facial expressions, which give us the auditory and visual modalities. During the therapy, the user wears the head-mounted display which covers a large part of the face in this way making it difficult to accurately use the visual modality to predict emotions. Therefore, the emotions are recognised using only the input speech of the user, which is converted to text, so as to create two modalities: audio and text. Using only these two modalities should not be a problem as studies have shown that audio and text outperform audio and video in different models [31, 37].

3.4 Model

James Tavernor has initially created the emotion recognition model [44], which is an end-to-end neural network. His model uses audio and text modalities in order to do multi-class emotion classification. The model was trained on the IEMOCAP dataset [5] and generated some satisfying results. The advantages of this model over other best performing models is its ability to operate on live input data and does not require data pre-processing. Later, the model was modified and improved by Lucia Simkanin [43]. Lucia has adapted the model in order to do multi-label emotion classification using the CMU-MOSEI dataset [3] for training. The model predicts the six basic emotions (happy, sad, fear, disgust, surprise and anger) and has demonstrated a really good performance.

The emotion recognition code is written in Python, therefore in order to use the model from the platform, which is written in C#, we need to run the model externally. There are different tools that allow you to execute python code from other languages, such as Ironpython. Nevertheless, these methods require Python to be installed on the device, thus presenting an obstacle to the procedure since Oculus Quest does not support Python. Consequently, a solution is to create a local server that executes the Python code when it is required and returns the results. The idea is to make a request to the server by sending the recorded audio from the user and which the server will use to execute the emotion recognition algorithm and return the classified emotion.

3.4.1 Flask server

In order to create the server, we have used Flask [36], which is a micro web framework that is written in Python and is simple to use. Using this tool we have created a local server that can be accessed from any device connected to the same network. At a later stage, this can be implemented to a remote server, thus allowing everybody around the world to have access to it. The emotion classification algorithm created by James and Lucia is hosted on this Flask server, which accepts POST requests to the address https://[host-IP]:5000/emotion, where [host-IP] is the IP address of the local server.

3.4.2 Data collection

The data is collected by the platform as described in section 3.1.6 under the MicSettings.cs script. Therefore, the microphone records utterances of 7 seconds long as the emotion recognition model was trained on CMU-MOSEI utterances which are on average 7.28 seconds long. Subsequently, each recording is processed to determine whether it contains any user speech input, otherwise it is discarded. All the valid utterances are stored and sent to the server one by one through different requests.

3.4.3 Procedure

The procedure described below is followed for each classification request:

1. The sampled data from an utterance are stored in a float array.
2. The float array is converted to a byte array.
3. The client (platform) makes a POST request to the server, where the main body of the request contains the data of the byte array.
4. The server receives the byte array and converts it back to a float array.
5. Then, the float array is converted to a wav audio file to be used as input to the model.
6. Speech recognition is applied to the audio file in order to obtain the text representation of the utterance.
7. The model uses both the audio and text to make a forward pass through the neural network and get the emotion classification results.
8. The server returns the predicted emotion back to the platform as a string variable.
9. The platform receives the predicted emotion and stores it for later use.

The above procedure is repeated for each recorded utterance and the returned emotions are stored. Assuming that the user replies to a question and the duration of the speech is two or more utterances, the returned emotion should be the same for all utterances. However, this in not always the case. As a result, the platform is designed to count the occurrences for each emotion and to take the most frequent one as the final predicted emotion of the user.

3.4.4 Benefits

Using a server can be very beneficial. Firstly, thanks to the server, the emotion recognition model is not running on the Oculus Quest device and as a consequence, it is not slowing down other procedures of the platform. Furthermore, the client can make multiple requests to the server without waiting for a response, which makes the emotion prediction procedure really fast and convenient. Finally, the server can be generalised and used in other tasks and projects that require emotion recognition capabilities.

3.5 User guide

The guide is created as a manual for the users to give them all the necessary information regarding the Oculus Quest VR platform application.

3.5.1 Intended usage

The purpose of this platform is to provide a more efficient way to deliver the protocol for the Self-Attachment therapy, aiming to help patients with chronic depression and anxiety in their adulthood. The therapy's purpose is to assist individuals to acquire the consistent and unconditional love and attention they were deprived of in their childhood. After completion of the therapy, the patients will create more optimal neural circuits that will allow them to self-regulate their emotions.

3.5.2 Safety warnings

As a user of this platform, you have to comply with Oculus Quest Health & Safety instructions provided by Oculus. For more information follow the link: (`https://www.oculus.com/safety-center/quest/`).

3.5.3 Installation instructions

1. Download the application (.apk file) on your computer.
2. Connect your Oculus Quest to your computer with the cable by following the instructions in the link: (`https://support.oculus.com/525406631321134/#setup`).
3. If you have an Android Debug Bridge (ADB), you can use it to install the application (in this case move to step 10), if you do not have ADB installed move to step 4.
4. Download and install SideQuest (`https://sidequestvr.com/setup-howto`).
5. Make sure your device has the developer mode enabled in case it does not follow this tutorial (`https://learn.adafruit.com/sideloading-on-oculus-quest/enable-developer-mode`).

6. Start SideQuest. A green dot must appear at the top left corner signifying that the Oculus Quest is connected.

7. Select INSTALL APK FROM FOLDER ON COMPUTER from the top toolbar.

8. Select the downloaded apk file.

9. At the bottom you will see ALL TASKS COMPLETED if the process is successful.

10. Put on your headset and enable HAND TRACKING (Settings $->$ Device $->$ Hands and Controllers $->$ Turn on Hand Tracking)

11. Open the application (Go to APPS (nine small squares symbol) $->$ Select UNKNOWN SOURCE from the top right drop down menu $->$ Select the installed application).

12. Make sure that you allow the application to use your microphone when you are prompted to do so.

13. To access tho Oculus main menu, look at your right hand palm at eye-level, then hold your thumb and index finger together until the Oculus icon fills up.

3.5.4 Usage instructions

The platform is easy to use and everything you need to know will be explained to you by a virtual agent while you are inside the platform. The virtual agent will be there the whole time giving you clear instructions and asking you questions. Please follow the instructions and answer to the questions in order to have the best possible experience. During your first session you will learn how to navigate inside the platform and how you can interact with different objects. Make sure that you have followed and understood the tutorial provided during the first session. You can exit the platform at anytime and your progress will be stored on the device. In order to get the most out of the therapy, please spend as much time as possible at each stage of the Self-Attachment protocol. After each session you will notice that a piece of a new house has been added which indicates that you are making progress in the therapy. After many sessions, a new complete house with furniture and a big garden will be created, meaning that you have completed all the mandatory sessions. However, you can continue to use the platform for as long as you want.

3.5.5 Tutorial

A user tutorial is provided inside the platform, which will help you to familiarise yourself with the different functionalities of the platform. You can repeat the tutorial several times by pressing the green button. The steps of the tutorial that are explained inside the platform are listed below:

1. Put down your controllers and look at your hands.
2. You can move forward by bringing your middle finger with your thumb together (pinch) with both hands simultaneously.
3. While you are pinching, turn your head left/right to change your moving direction.
4. Stop pinching to stop moving.
5. Use your index finger to interact with objects.
6. Touch the red cube with your index finger to turn on the TV and watch how you can do self-massage.

3.5.6 Terms and conditions

Please read all the following terms and conditions and make sure you agree with them before starting using the platform.

1. You must be at least 18 years old.
2. You must have the permission of a clinician to use the platform.
3. You must agree to use the platform for the intended purpose only.
4. If you think you may experience any medical emergency you must contact your doctor immediately.
5. If you have any problems with the platform contact us.
6. This platform is under copyright protection.

3.5.7 Feedback

Inside the platform you will find a button that allows you to provide feedback on your experience. Please provide feedback by answering the displayed questions after each session. We really appreciate your feedback which will help us improve our services.

3.5.8 Contact details

Should you need any further information or for any other questions and problems, please do not hesitate to contact Neophytos Polydorou at the email address: np1519@ic.ac.uk.

Chapter 4

Platform evaluation

Evaluating the platform is crucial, as it shows whether the aims of the project have been achieved and whether there are problems that need to be tackled. It is worth mentioning that part of the platform has been evaluated during an ongoing trial. Thereafter, we have also collected some metrics so as to evaluate the whole platform. Finally, we have created a user feedback procedure that will be helpful for evaluation when the platform is tested on a trial at a later stage.

4.1 Impact evaluation

4.1.1 Trial by Ghaznavi

Syed Ibrahim Ghaznavi is a PhD student at Imperial College London [24] who has recently carried out an impact evaluation in order to compare the classic Self-Attachment therapy with a simple implementation of this therapy into a mobile VR application. In the original Self-Attachment therapy, the patient must look at two of their childhood photos, one taken from a happy memory and one from a sad memory, in an attempt to connect with their childhood-self. On the other hand, in the immersive mobile VR platform the user can see the child avatar who looks like the child in the photo and try to create a connection. The child avatar changes its emotion based on the actions of the user.

The participants for this experiment were recruited via an advertisement, circulated on social media platforms. The selection was made based on some inclusion criteria; for example, healthy subjects, within the age range of 18 and 50 years old, who also have a decent quality photo of their childhood. They were also required to have a Google cardboard VR headset and an Android mobile device so that they could install and use the application.

The trial procedure evolved as follows. Firstly, the participants had to complete the consent form and then install the application. Furthermore, they were advised to have a $4m^2$ clear space with no obstacles when using the application so as to avoid injuries. They were also required to use both their printed images first and then the photo-realistic avatar inside the platform in order to try and recall their childhood memories. Thereafter, they had to try to relate with their childhood-self and report on the level of connection that they had experienced. Inside the platform, the participants were also given the opportunity to project different emotions to the child avatar by pointing the head pointer on the child avatar. Finally, they had to report if the were able to complete the objectives of the therapy and to what degree the whole experience was realistic.

The participants were asked to complete multiple sessions inside the platform for a week, for as long as they could. After a week, they also had to complete a questionnaire for evaluation purposes. The results were promising as the participants preferred the VR solution over the images as they managed to connect better with their childhood-self. However, they wanted greater customisation options; for example, in their opinion, the child avatar's clothes should match with the ones in the image, and also, they requested to have more variety of emotion animation options.

4.1.2 Contribution in the evaluation

The way we have contributed into the aforementioned impact evaluation is by offering our Oculus Quest implementation for trial. A version of our platform was used in the experiment as an additional method of delivering the Self-Attachment therapy. Given that not everyone taking part in the trial had access to an Oculus Quest headset, only a subgroup of seven individuals was asked to try our platform along with the other two methods. All the seven participants were provided with the installation instructions described in section 3.5.3. In fact, the use of the Oculus VR platform enhanced the trial and gave the opportunity to the participants to experience the platform's latest features and functionalities.

4.1.3 Oculus VR application description

All participants were provided with a customised application (.apk file) specifically for them, which can be run on the Oculus Quest headset. This application was created specifically for the evaluation and consist only a part of the entire platform created in this project. Some of the features are only created for the trial and are not included in the final version of the platform. The reason for that is that due to the need of direct comparison between the two platforms some of the mobile VR application features were reproduced in the Oculus VR application.

Figure 4.1: Oculus VR platform setting used for the impact evaluation trial.

The layout of the application is presented in figure 4.1. This includes the customised child avatar of the individual at the middle of the figure and the individual's 2D photo in the background. Since this version of the platform did not include the virtual agent, the instructions were displayed on a canvas for the user to read. The instructions are described below:

VR for Self-Attachment

1. Put down your controllers and look at your hands.
2. Approach the child avatar with your hands to make him/her happy and then embrace the child to make him/her joyful (dance).

3. Press the Neutral button to repeat the protocol.

In addition, the user had the choice to enable or disable the child's room, which is shown in figure 3.15, by pressing the red button. Finally, there were four blue cubes representing the buttons that trigger different emotions on the child avatar.

The above described setting is created in order to match the mobile VR environment. However, the Oculus VR application has new and improved functionalities that make the environment more appealing and easier to interact with. The addition of the hand tracking feature has been proved beneficial as it makes the platform more immersive and user-friendly. Instead of using the head pointer to make selections in the mobile VR application, with the Oculus VR application the user can now use their hands to make different selections. For instace, by touching the blue button with the label 'Happy' in figure 4.1, the user can change the emotional state of the child to 'happy'.

Nevertheless, the most substantial improvement is the way the user can interact with the child avatar. Using the hand tracking feature it is possible to detect whether the user has approached or touched the child avatar. Let us take for example the scenario where the child avatar is feeling sad. The user can move towards the child and make the child feel happy. If the user also touches the child, then the child will start dancing. During the trial only the 'happy', 'sad' and 'fear' emotions were used, as well as the 'dance' animation. All the animations are displayed in figure 4.8 on customised child avatars.

4.1.4 Evaluation results

In this section we aim to understand and analyse the reasons for the obtained results and how they contribute in this project. A questionnaire had to be completed by the participants after taking part in the experiment in order to reflect on their experience of using the three different methods; firstly with the images, then with the mobile VR platform and finally with the Oculus VR platform. The results revealed a successful implementation of a part of the Self-Attachment therapy into the VR as the participants managed to identify and sympathise with their child avatars inside the virtual environment. What stands out from these promising results is that the participants had managed to connect with and project their emotions to the avatar better than using their childhood images. These results suggest that an immersive VR platform can be more effective than images in the process of delivering the Self-Attachment therapy. Moreover, our results reveal that combining diverse functionalities in the platform improves the whole experi-

Figure 4.2: Happy animation

Figure 4.3: Fear animation

Figure 4.4: Sad animation

Figure 4.5: Sad animation

Figure 4.6: Dance animation

Figure 4.7: Dance animation

Figure 4.8: Emotion animations on the customised avatars of participants from the trial [24].

ence for the user, as well as it facilitates the entire procedure.

Another important aspect of the experiment is that during the sessions the participants were given the opportunity to use three different methods to interact with their childhood-self and indicate in which case they were able to better relate emotionally to the child. We note, based on the results obtained, that the participants managed to better relate to the mobile VR and the Oculus VR avatars than the images. However, the most important is that they had a bigger capacity to emotionally relate to the Oculus VR avatar than the mobile VR avatar. In addition, in their opinion, the new hand-tracking feature described in section 3.1.1 is of great value to the project. What is more is that the level of immersiveness was higher, indicating higher levels of user involvement in the Oculus VR environment than in the mobile VR environment. An example that proves this is when some of the participants during the trial started dancing alongside with the child avatar.

Nevertheless, we also acknowledge that the project is bound by a number of limitations, as the results themselves show that there is a lot of room for improvement of the platform. Despite this, new technologies and new features can greatly contribute to a finer experience for the user. It must be noted that even a simple part of the platform created in this project including only some of the features has had a bigger impact on the emotional state of the user. Therefore, we believe that the whole platform implementation that includes the virtual agent and the emotion recognition will show a massive improvement in the current way that the Self-Attachment therapy is delivered.

4.2 Platform metrics

Using the the OVR metrics tool described in section 3.2.5, we have decided to report the CPU and GPU utilization, as well as the average frame rate as an evaluation for the platform.

When building an application the developer has to set the number of frames per seconds (FPD) to display. We have chosen to use 72 FPS which is the maximum that Oculus Quest can support. Also, this high FPS rate, also known as refresh rate, is required by a VR application so that the images are changing fast and give the feeling of a real environment. In addition, low refresh rate can cause motion sickness to the user. The graph in figure 4.9 shows the average frame rate of the platform for a period of approximately 5 seconds. As you can see, some of the times the FPS drops well below the set amount of 72 FPS. This is an area where the platform needs improvement.

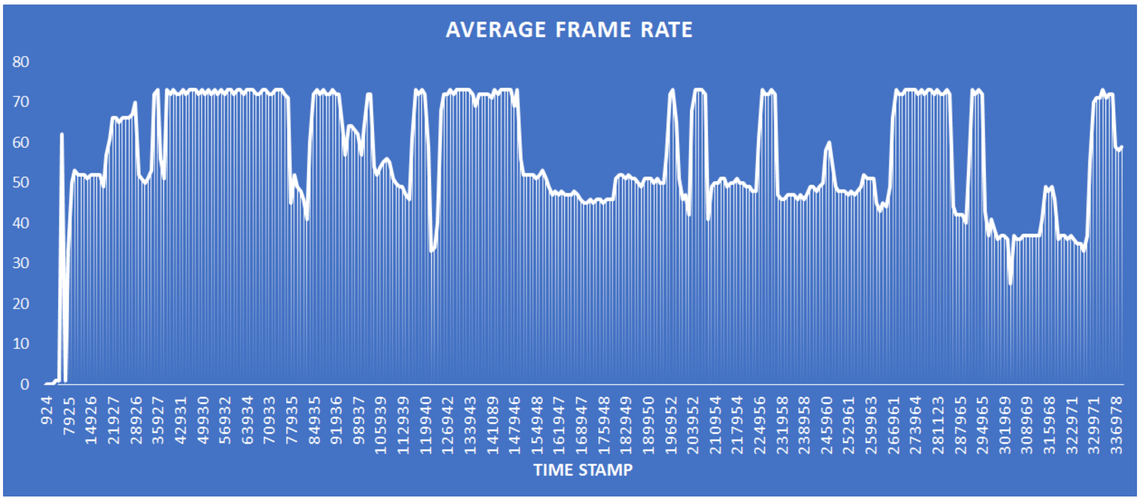

Figure 4.9: The x-axis shows the time stamp and the y-axis the numbers of frames per second (FPS) displayed on the screen. The recorded duration is about 5 seconds. The desired graph would be a constant line at 72 FPS, which is the preset value.

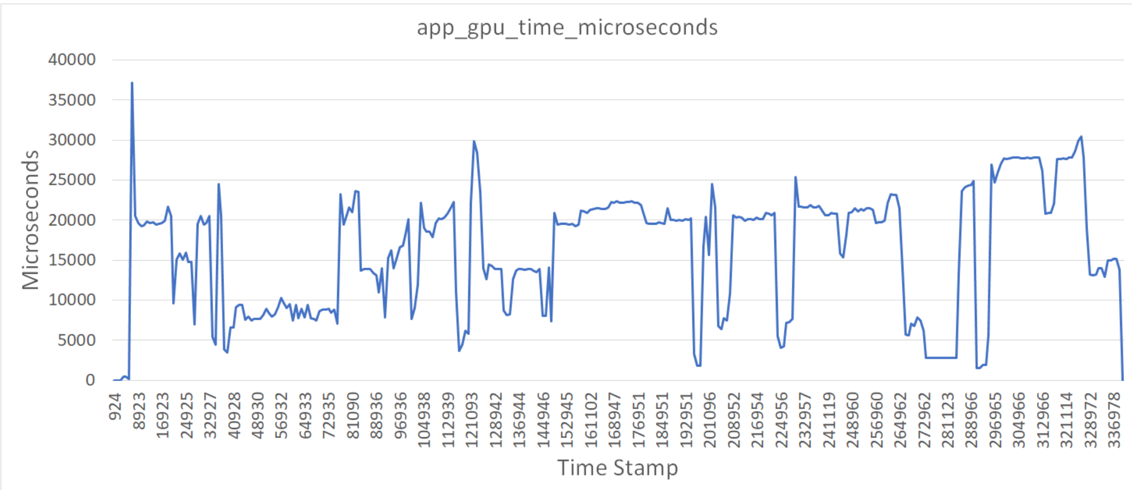

Figure 4.10: This graph shows the number of microseconds needed by the application to render on frame.

The same problem can be observed with the App GPU time metric, which is the number of microseconds (μs) the application needs to render one frame. By using the selected 72 FPS, the application should only need 13889μs to render each frame, however most of the time this is not the case as depicted in figure 4.10.

The graph in figure 4.11 shows the CPU and GPU utilization, thus signifying how much resources are used by CPU and GPU respectively. The graph illustrates the percentage of usage over a period of approximately 5 seconds. What we notice is that the GPU is used much more than the CPU which is reasonable because of the amount of graphics used for the 3D environment. The average percentage for the CPU is around 30%, whereas the average percentage for GPU is 80%. As long as these numbers do not approach 100% then the platform will be performing well.

4.3 User feedback

We have created an automatic way that the users can provide feedback on their experience through the platform after each session. This will allow to evaluate the platform during the therapy in a future trial on patients. A feedback button that is located on a coffee table in the living room can be pressed by the user in order to answer some questions by selecting a score from one to five. When the feedback button is pressed, then the questions will appear one after the other alongside with the score buttons. The setting is illustrated in figure 4.12, while the instructions and questions are displayed in the following order:

1. Please select a score between 1 and 5 for the following questions.
2. Do you feel better?
3. Was the platform easy to use?
4. Did the agent recognise your emotions correctly?
5. Did you build a connection with the child avatar?
6. Please rate your overall experience.
7. Thank you for the feedback provided.

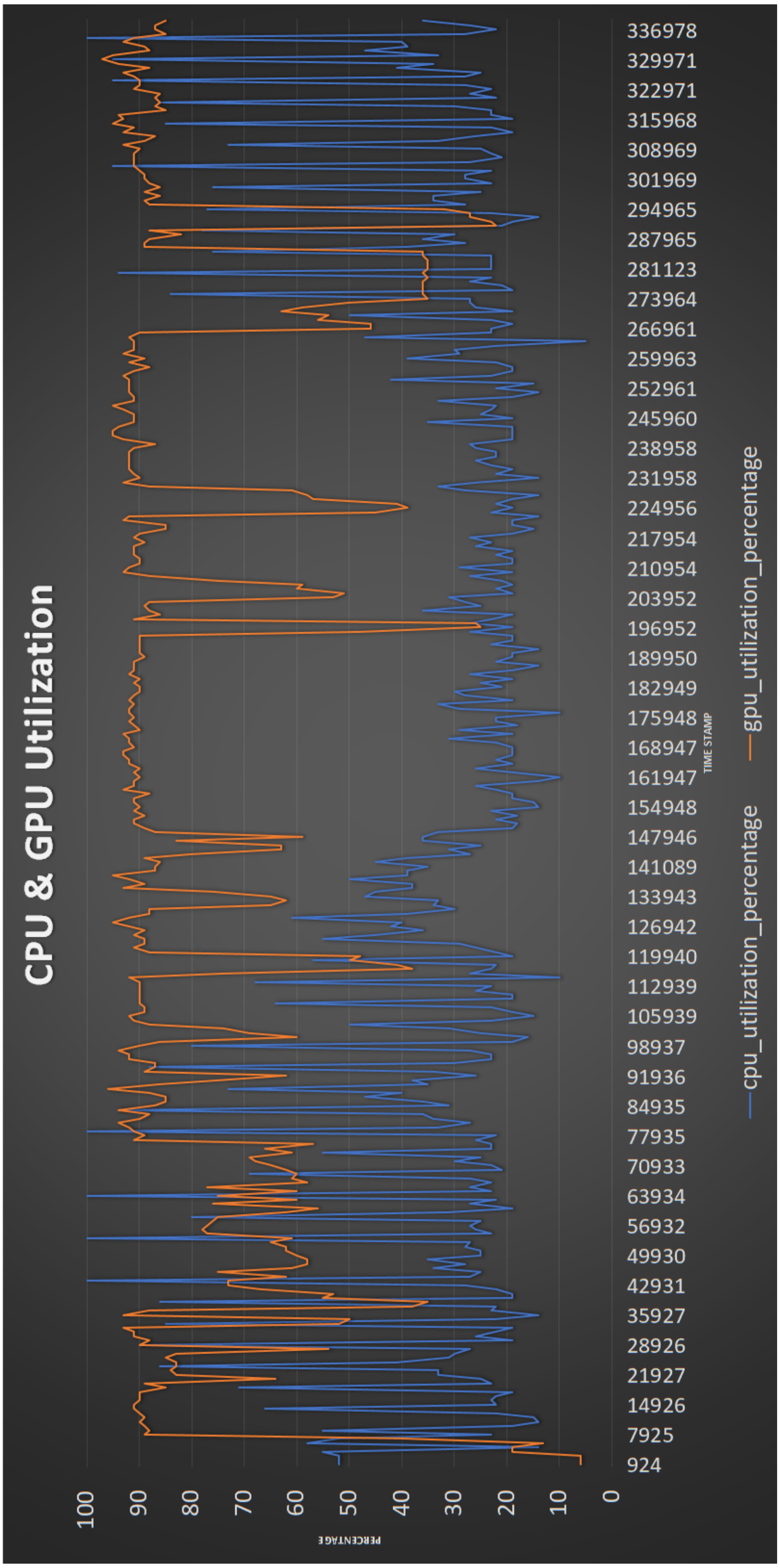

Figure 4.11: The x-axis shows the time stamp and the y-axis the percentage. The percentage of utilization for the CPU and the GPU are depicted by the blue and the orange line respectively. The GPU percentage is approximately constant for most of the time since the graphic environment is in constant use. On the other hand, the CPU percentage spikes since there are many small, but fast calculations that need to be carried out.

Figure 4.12: Platform setting when the green feedback button is pressed. Each blue button represents a score from 1 to 5. The questions appear in red colour text.

Chapter 5

Conclusion

The unprecedented experience that the whole planet has been undergoing for the past 9 months, due to the pandemic that COVID-19 has created, has made clear the need to use technology in order to deal with serious mental disorders. The forced confinement at home and social alienation, the diverse work difficulties, as well as the stress and insecurity about our physical health are all crucial outcomes of COVID-19 that have specifically exacerbated the problem of people with mental disorders. As a result, this aggravated situation has put a lot of pressure and strain on community mental health centres and psychiatric hospitals [11].

The Self-Attachment therapy which can be accessible through the platform described in this project, can be helpful especially during such difficult times, as it aims to facilitate the process of treating people at their own home, without further burdening hospital units and specialized clinicians. All the aforementioned highlight the need for such a platform which can offer help and treatment to people with mental disorders in a more user-friendly and effective way. It may be argued that technology alone cannot work wonders. However, we think that such a virtual reality platform can create a comprehensive and high-quality mental health care.

5.1 Achievements

The main achievement of this project is the creation of an immersive VR platform that can be used to deliver the Self-Attachment therapy to any patient. Importantly, the procedure can be carried out by anyone, without the assistance of a human clinician. In addition, the platform is personalised for each user which is a great advantage. The level of immersiveness and the sense of presence are very high. This comes as a result of the combination of a lot of small components that come together to create a complete product.

Initially, the design of the platform and the addition of different functionalities, like the virtual hands and the navigation controls, create a realistic environment that is very easy to use. Furthermore, other functionalities, such as the light switches and other buttons, construct an interesting to use platform that attracts the user. More features are also included in the platform that allow various interactions and urge the user to get actively involved. Some examples constitute the interactions with the child avatar and the virtual agent, where the user uses touch and speech to invoke different actions.

The platform personalisation is mainly achieved with a customised child avatar and the emotion recognition algorithm, which was successfully incorporated into the platform. In addition, the created scenario reflects almost exactly the objectives of the Self-Attachment therapy and is transmitted to the user with the aid of a virtual agent who acts as the main communication tool for the platform. The results of the impact evaluation are promising, however there is still a long way to go before the application can be used by patients, especially when taking into consideration the ultimate purpose of this platform.

5.2 Limitations

There is a number of different limitations related to this project, some of which are also mentioned above. Other limitations that should be noted are analysed below.

1. When the virtual agent speaks, its lips are not moving which is not realistic. The aim is to use a Lipsync tool that enables the synchronisation of the lips with the speech.

2. The text to speech tool that we have used does not provide a functionality to detect whether the speech has ended. We have tried to detect whether the Oculus Quest output device is in use so as to know when the speech has stopped, but with no success due to time limitations. Despite these obstacles, we came up with a possible solution to the problem that is less accurate however. By assuming that the text to speech conversion rate is about 12

characters per second, we were able to calculate an approximation for the time required by each sentence to finish. This is achieved by dividing the total number of characters in the text by 12 so as to get the number of seconds the program should wait for the speech to end.

3. The text to speech tool produces a computerised voice, in this way making the voice of the agent to sound unnatural.

4. The customised child avatar must be created before building the application. The aim is to create the avatar during run-time in the first session and then store it to be used in the next sessions.

5. The created scenario is based on an older version of the Self-Attachment therapy. As a result, it needs to be updated when the most recent version is available in the English language.

6. The final version of the platform needs to be evaluated through trials by more people.

5.3 Future work

The identified limitations pave the way for further work and improvement of the platform to take place in the future. Firstly, a virtual psychotherapist can be created, who is going to be controlled by a machine learning algorithm and who will be able to replace an actual psychotherapist. This virtual agent should be able to infer the patient's mental state and safely suggest the appropriate stage of the Self-Attachment therapy protocol that the patient should follow. Equally important will be the ability of the virtual psychotherapist to understand whether the mental state of the patient originated from any past trauma or from any recent problems that the patient may face at the time in order to recommend the most effective therapy.

Future studies should also focus on generating emotions that are not discrete, thus meaning that two emotions can coexist. Therefore, an algorithm can be created that can fuse different emotions and create the corresponding emotion animation. In addition, the customised avatar of the user can be improved and stored, along with other information, to an online server instead of using the device's memory. Finally, a functionality could be implemented which will allow the creation of more than one user accounts on the same device, so that multiple people can have access to the therapy on the same device.

Despite the aforementioned interesting additions, there are endless directions that someone can follow. For instance, the platform can be modified so as to be used in diverse psychotherapeutic procedures or maybe for conducting group therapy sessions in a virtual world. Also, multiple virtual agents with a different personality and therapeutic approach can be created to offer to the user the choice to select based on their preferences.

5.4 Legal and Ethical Considerations

This project takes into consideration all the relevant codes of ethical conduct and IT legislation. The final application created has not been tested on patients or normal population, however a part of it was included in the trial conducted by Ghaznavi [24] after obtaining ethical approval from Imperial College London. It was our duty to ensure that the application used in the trial complied with all the relevant rules and could not cause any harm to any of the participants, regardless their age, gender or ethnicity. What is more, the participants of the trial completed a consent form and agreed that the data collected about them could be used to support other research in the future. Nonetheless, none of the data collected included sensitive information and none of the data have been shared with others.

Finally, it must be noted that despite the fact that the application can be very helpful, yet it can become very harmful if it is misused. Military or other individuals can modify the application in order to serve a bad cause. For example, instead of using the platform to treat soldiers with PTSD, the application can be modified to torture people by making them recall only traumatic experiences. Nevertheless, the intended use of this project is clear and anyone that uses it for other purposes should face legal charges.

Bibliography

[1] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Association Arlington, VA, 5th edition, 2013. pages 5, 6

[2] American Psychological Association. What is exposure therapy? `https://www.apa.org/ptsd-guideline/patients-and-families/exposure-therapy`, accessed: 01/05/2020. pages 6

[3] Z A Bagher, P P Liang, S Poria, E Cambria, and L Morency. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2236–2246, Melbourne, Australia, 2018. Association for Computational Linguistics. pages 49

[4] J Bowlby. *Attachment: Volume One of the Attachment and Loss Trilogy*. Pimlico, second revised edition, 1997. pages 7

[5] C Busso, M Bulut, C Lee, A Kazemzadeh, E Mower, S Kim, J Chang, S Lee, and S Narayanan. Iemocap: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42(4):335, 2008. pages 48

[6] D Cittern and A Edalat. A neural model of empathic states in attachment-based psychotherapy. *Computational Psychiatry*, 1:132–167, 2017. pages 8

[7] D Cittern, A Edalat, and I Ghaznavi. An immersive virtual reality mobile platform for self-attachment. In *AISB*, 2017. pages 2, 13

[8] David Cittern. *Computational Models of Attachment and Self-Attachment*. Phd thesis, Imperial College London, 2016. pages 8, 9

[9] Oculus developer blog. Ovr metrics tool + vrapi: What do these metrics mean? `https://developer.oculus.com/blog/`

ovr-metrics-tool-vrapi-what-do-these-metrics-mean, accessed: 28/08/2020. pages 47

[10] L B Dixon, Y Holoshitz, and I Nossel. Treatment engagement of individuals experiencing mental illness: review and update. *World Psychiatry*, 15(1):13–20, 2016. pages 1, 2

[11] Benjamin G Druss. Addressing the COVID-19 Pandemic in Populations With Serious Mental Illness. *JAMA Psychiatry*, 77(9):891–892, 2020. pages 65

[12] A Edalat. Self-attachment: A new and integrative psychotherapy. Talk at the Institute of Psychiatry, London, 02/05/2013. pages 7

[13] A Edalat. Introduction to self-attachment and its neural basis. In *The 2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015. pages 2, 7

[14] A Edalat. Self attachment: A holistic approach to computational psychiatry. In E Peter, BS Bhattacharya, and A Cochran, editors, *Computational Neurology and Psychiatry Springer Series on Bio-/Neuro-informatics*. Springer, 2017. pages 7, 8

[15] G Stratou et al. A demonstration of the perception system in simsensei, a virtual human application for healthcare interviews. In *2015 International Conference on Affective Computing and Intelligent Interaction*, pages 787–789, Xi'an, 2015. ACII. pages 15

[16] C J Falconer, A Rovira, J A King, P Gilbert, A Antley, P Fearon, N Ralph, M Slater, and C R Brewin. Embodying self-compassion within virtual reality and its effects on patients with depression. *BJPsych Open*, 2(1):74–80, 2016. pages 10

[17] FAtiMA. Fatima-toolkit, accessed: 27/08/2020. https://fatima-toolkit.eu/1-what-is-fatima-toolkit/. pages 24, 29, 77

[18] P Fonagy. *Attachment Theory and Psychoanalysis*. Other Press, 2001. pages 7

[19] Mental Health Foundation. Physical health and mental health, accessed: 16/05/2020. https://www.mentalhealth.org.uk/a-to-z/p/physical-health-and-mental-health. pages 2

[20] D Freeman, J Bradley, A Antley, E Bourke, N DeWeever, N Evans, E Černis, B Sheaves, F Waite, G Dunn, M Slater, and D M Clark. Virtual reality in the treatment of persecutory delusions: Randomised controlled experimental study testing how to reduce delusional conviction. *British Journal of Psychiatry*, 209(1):62–67, 2016. pages 11

[21] D Freeman, S Reeve, A Robinson, A Ehlers, D Clark, B Spanlang, and M Slater. Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, 47(14):2393–2400, 2017. pages 11

[22] B Garrett, T Taverner, W Masinde, D Gromala, C Shaw, and M Negraeff. A rapid evidence assessment of immersive virtual reality as an adjunct therapy in acute pain management in clinical practice. *Clinical Journal of Pain*, 30(12):1089–1098, 2014. pages 12

[23] I Ghaznavi, U Jehanzeb, A Edalat, and D Gillies. Usability evaluation of an immersive virtual reality platform for self-attachment psychotherapy. In *ACM*, 2019. pages 14

[24] Syed Ibrahim Ghaznavi. *Immersive virtual reality platform for Self-attachment therapy*. Phd thesis, Imperial College London, 2020. pages 55, 59, 68

[25] C Gordon, A Leuski, G Benn, E Klassen, E Fast, M Liewer, A Hartholt, and D Traum. Primer: An emotionally aware virtual agent. In *IUI Workshops*, 2019. pages 16

[26] H G Hoffman, G T Chambers, W J Meyer III, L L Arceneaux, W J Russell, E J Seibel, T L Richards, S R Sharar, and D R Patterson. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. *Annals of Behavioral Medicine*, 41(2):183–191, 2011. pages 13

[27] itSeez3D. Avatar sdk, accessed: 27/08/2020. `https://avatarsdk.com/`. pages 23, 77

[28] SL James et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(101519):1789 – 1858, 2018. pages 2, 6

[29] Y Ling, H T Nefs, N Morina, I Heynderickx, and W Brinkman. A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. *PLOS ONE*, 9(5):1–12, 05 2014. pages 12

[30] L Loucks, C Yasinski, S D Norrholm, J Maples-Keller, L Post, L Zwiebach, D Fiorillo, M Goodlin, T Jovanovic, A A Rizzo, and B O Rothbaum. You can do that?!: Feasibility of virtual reality exposure therapy in the treatment of ptsd due to military sexual trauma. *Journal of Anxiety Disorders*, 61:55 – 63, 2019. pages 10

[31] N Majumder, D Hazarika, A Gelbukh, E Cambria, and S Poria. Multimodal sentiment analysis: Addressing key issues and setting up the baselines. *Knowledge-Based Systems.*, 161:124 – 133, 2018. pages 48

[32] M AG Martens, A Antley, D Freeman, M Slater, PJ Harrison, and EM Tunbridge. It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology*, 33(10):1264–1273, 2019. pages 9

[33] N Morina, H Ijntema, K Meyerbröker, and P MG Emmelkamp. Can virtual reality exposure therapy gains be generalized to real-life? a meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy*, 74:18 – 24, 2015. pages 12

[34] Oculus. Oculus Quest. `https://www.oculus.com/quest/`, accessed: 30/08/2020. pages 16, 77

[35] D Opriş, S Pintea, A García-Palacios, C Botella, Ş Szamosközi, and D David. Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. *Depression and Anxiety*, 29(2):85–93, 2012. pages 11

[36] Pallets organization. Flask's documentation. `https://flask.palletsprojects.com/`, accessed: 06/08/2020. pages 49

[37] S Poria, N Majumder, D Hazarika, E Cambria, A Gelbukh, and A Hussain. Multimodal sentiment analysis: Addressing key issues and setting up the baselines. *IEEE Intelligent Systems*, 33(6):17–25, 2018. pages 48

[38] G Rizos, K Hemker, and B Schuller. Augment to prevent: Short-text data augmentation in deep learning for hate-speech classification. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, page 991–1000, New York, NY, USA, 2019. Association for Computing Machinery. pages 15

[39] J A Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980. pages 15

[40] D Saldanha, L Bhattacharya, B Chaudari, and P Menon. Cognitive behavior therapy. *Medical Journal of Dr. D.Y. Patil University*, 6:132, 2013. pages 6

[41] D J Schofield, R N Shrestha, R Percival, M E Passey, E J Callander, and S J Kelly. The personal and national costs of mental health conditions: impacts on income, taxes, government support payments due to lost labour force participation. *BMC Psychiatry*, 11(72):1–7, 2011. pages 2

[42] Unite For Sight. Module 1: Introduction to global mental health: Effects of mental health on individuals and populations, accessed: 16/05/2020. `https://www.uniteforsight.org/mental-health/module1`. pages 2

[43] Lucia Simkanin. *Mutli-emotion recognition and dialogue manager for VR-based self-attachment therapy.* Master's thesis, Imperial College London, 2020. pages 31, 48

[44] James Tavernor. *Cross-corpus Speech and Textual Emotion Learning for Psychotherapy*. Master's thesis, Imperial College London, 2020. pages 48

[45] Unity Technologies. Unity, accessed: 15/05/2020. `https://unity.com/`. pages 16

[46] Unity Technologies. Unity scripting reference, accessed: 24/08/2020. `https://docs.unity3d.com/ScriptReference/`. pages 26, 31, 32

[47] Unity Technologies. Unity asset store, accessed: 27/08/2020. `https://assetstore.unity.com/`. pages 25, 29, 30, 31, 77

[48] S Tripathi and H Beigi. Multi-modal emotion recognition on IEMOCAP dataset using deep learning. *CoRR*, abs/1804.05788, 2018. pages 15

[49] P Tzirakis, G Trigeorgis, M A Nicolaou, B W Schuller, and S Zafeiriou. End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1301–1309, 2017. pages 15

[50] J Voigt-Antons, T Kojic, D Ali, and S Möller. Influence of hand tracking as a way of interaction in virtual reality on user experience. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–4, 2020. pages 19

Appendix A

Project README file

A.1 Installation

Please follow the procedure in order to install and build the application to your Oculus Quest device.

Install Unity version 2019.3.15f1 and add the 'Android Build Support' module to that version. Open the project in the Unity editor. You may face some issues if you use other versions of Unity for this project.

Go to 'Project Settings', download the 'XR Plug-in Management' and enable 'Oculus'.

In 'Project Settings' under the 'Player' tap, select the icon 'Android Settings' and then click on 'Other Settings' section and make sure that you have the following settings:

- Color Space: Gamma
- Graphics APIs: Make the 'OpenGLES3' to come first in the list
- Minimum API Level: Android 7.1 'Nougat' (API level 25)
- Target API Level: Automatic (highest installed)

In 'Project Settings' under the 'Quality' tap, select the 'Very Low' level option.

In 'Project Settings' under the 'Time' tap, change the 'Fixed Timestep' to 0.01388889 which is equal to $\frac{1}{72}$. This is done so as to have 72 frames per seconds (FPS) display.

In 'Build Settings', under the 'Platform' tap select 'Android' and then press the 'Switch Platform' button. Then change the following settings:

- Texture Compression: ASTC
- Run Device: Select the name of your Oculus Quest device. If it does not appear in the list make sure that the device is connected with a cable to your PC and then press the 'Refresh' button.
- Compression Method: LZ4

To build the application on the device, select the scene from 'Scenes In Build' list and click on the 'Build And Run' button.

A.2 Server

The emotion recognition algorithm is created by Lucia Simkanin, therefore if you need access please email her at lucia.simkaninova19@imperial.ac.uk and request access to it. After receiving the folder, follow the steps below:

- The folder you will receive contains the server file with the name 'server.py'.
- Install the necessary libraries needed by the model.
- Open a command prompt and direct to the location of the server file.
- Then type 'set FLASK_APP=server.py'.
- Next, run the file by typing 'python server.py'. Now, the server should be running.
- Before building the application, you have to change the IP address in the MicSettings class to match with your IP.
- The server should running during the VR session.

If you need to run this project without the emotion recognition algorithm, you can use the ‘server.py’ file located in this project folder. Please follow the steps above in order to create the server, which receives the data and always returns the ‘Happy’ emotion. The reason for doing this is to allow the application to run properly.

Link to the project folder:

`https://imperiallondon-my.sharepoint.com/:u:/g/personal/np1519_ic_ac_uk/EXXvTXk_rr1HhvavGtLuhlIBJs3Ar3OzlV4ZI5KbTLKFjg?e=5ZEzqr`

Appendix B

Unity Project Documentation

B.1 Files

All the files in the project are organised in different folders under the 'Assets' folder in the following way:

- Building tools: This folder contains all the components that we have used to create the environment. It includes all the different avatars, animations, furniture, skyboxes, buttons and materials that are used for the creation of all the graphics elements of the project. All of them are taken from the Unity Assets Store [47].

- Itseez3d: This folder includes the tools for the creation of a customised avatar. See the Avatar SDK [27] documentation for more information on how to use it.

- Oculus: This folder contains all the Oculus tools that help with the implementation of a VR application. See Oculus Integration [34].

- Plugins: This folder includes all the necessary plugins for the implementation of different functionalities. Plugins for Android, FAtiMA [17] and Text-To-Speech [47] are used in this project.

- Scenes: The folder contains the scene for the final version of the platform, as well as some other experimental scenes which were created during the development of the platform (some of them might not function properly).

- Scripts: This folder contains all the scripts used for the implementation of the platform, as well as some scripts that are not used in the final version.

- StreamingAssets: The name of this folder is case sensitive because it is used to transfer files to the target device. When the application is built on a device, not all the files from the project folder are transferred to the device but only the files inside this folder.

B.2 Classes

B.2.1 ChildLookAtPlayer

Methods	
LateUpdate()	Update the rotation of the object that is attached to it. Use the LateUpdate to override any animations running on the object.

B.2.2 CustomAvatar

Methods	
InstantiateChildAvatar()	Instantiates a saved avatar.
CreateChildAvatar()	Starts the process of creating an avatar.
PressUserPhotoButton()	Waits for the button to become available and then presses it.
FindAndSaveAvatar()	Locates and saves the created avatar.
ConfigureAvatar()	Adds necessaries components to the avatar.
SaveAvatar()	Saves the avatar.
LoadAvatar()	Loads an existing avatar.

B.2.3 CustomAvatarData

Data container class	Contains data about the child avatar.

B.2.4 DisplayOnScreen

Methods	
Start()	Starts the checking procedure.
CheckSystemTime()	Checks the time every 60 seconds.

B.2.5 EmotionButton

Methods	
OnTriggerEnter(Collider other)	When the component attached to it collides with another object, the current emotion changes.

B.2.6 EmotionManager

Methods	
Start()	Initialization of variables.
CurrentEmotion(string emotion)	Sets the animation to ‘emotion’.
IsValidEmotion(string emotion)	Checks if the ‘emotion’ is valid.

B.2.7 Fatima

Methods	
FatimaInitialisation()	Starts the initialization procedure.
DownloadFiles()	Sends download request for the files stored in StreamingAssets and initialises the state.
CreateRequest(string fileName)	Creates the request for the server.
ErrorCheck(UnityWebRequest uwr)	Checks for errors in the request.
PlayerSays(bool userAnswer)	Called when the user speaks.
AgentSays(string currentEmotion)	Called before the agent speaks and returns the sentence that it has to say.
UpdWrld(Name eventName)	Updates the world. Not in current use.

B.2.8 FeedbackButton

Methods	
OnTriggerEnter(Collider other)	When the component attached to it collide with another object, the ‘feedback’ starts.
Questions()	Displays questions in order.
Start()	Initializes variables and questions.

B.2.9 FeedbackScore

Methods	
OnTriggerEnter(Collider other)	When the component attached to it collides with another object, the score is noted.

B.2.10 GrabberHand

Methods	
Start()	Initialization.
Update()	Checks for grabbing, starts and ends grabbing.
CheckIndexPinch()	Checks if the user pinches.

B.2.11 HouseBuildingComponents

Methods	
ActivateRequiredBuildingComponents(int componentNumber)	Activates the components needed based on the ‘componentNumber’.

B.2.12 LightSwitch

Methods	
OnTriggerEnter(Collider other)	When the component attached to it collides with another object, activates/deactivates light objects.

B.2.13 MicSettings

Methods	
StartRecording()	Initializes recording procedure.
RecordUtterance()	Records for a number of seconds and then sends request to the server. After all requests are finished, returns the maximum predicted emotion.
CheckForSpeech(AudioClip audio)	Checks 'audio' for any input.
SendRequest(AudioClip utter)	Sends a request to the server and returns the response.

B.2.14 MovingManager

Methods	
Start()	Initialization.
MovePlayer()	Changes the transform of the player.
Update()	Checks for gesture and moves the player.

B.2.15 NoButton

Methods	
OnTriggerEnter(Collider other)	When the component attached to it collides with another object, sets the user's response to false.

B.2.16 ProtocolManager

Methods	
Start()	Starts initialization process.
Initialization()	Initializes components.
StartProtocolStageOne()	Executes first protocol.
StartProtocolStageTwo()	Executes second protocol.
StartProtocolStageThree()	Executes third protocol.
StartProtocolStageFour()	Executes fourth protocol.
StartTypeA()	Executes first sub-protocol.
StartTypeB()	Executes second sub-protocol.
StartTypeC()	Executes third sub-protocol.
StartTypeD()	Executes fourth sub-protocol.
StartTypeE()	Executes fifth sub-protocol.
StartTypeF()	Executes sixth sub-protocol.
UserAnswerYesOrNo()	Enables buttons.
UserSelectEmotion()	Enables buttons.
AgentSpeak()	Initializes agents' speech.
UserSpeak()	Starts the recording procedure.
Save()	Saves the data.
Load()	Loads the saved data.

B.2.17 UserData

Data container class	Contains the user's data for storage.

B.2.18 SkyboxController

Methods	
SetSkyboxTo(int newSkybox)	Sets the skybox to 'newSkybox'.

B.2.19 SpeakStart

Methods	
Start()	Initialization.

B.2.20 SpeakStart

Methods

Start()	Initialization.
Update()	Initializes text-to-speech tool once.
AgentSpeak(string sentence)	The agent speaks the 'sentence'.

B.2.21 TutorialButton

Methods

Start()	Initialization.
OnTriggerEnter(Collider other)	When the component attached to it collides with another object, activates/deactivates the tutorial.

B.2.22 UserPosition

Methods

CheckWhichRoomIsUserIn()	Checks the position of the user and returns the room that the user can be found.

B.2.23 VideoPlayButton

Methods

OnTriggerEnter(Collider other)	When the component attached to it collides with another object, plays/pauses the video.

B.2.24 YesButton

Methods

OnTriggerEnter(Collider other)	When the component attached to it collides with another object, sets the user's response to true.

Appendix C

Ethics checklist

Section 1: HUMAN EMBRYOS/FOETUSES	
Does your project involve Human Embryonic Stem Cells?	NO
Does your project involve the use of human embryos?	NO
Does your project involve the use of human foetal tissues / cells?	NO
Section 2: HUMANS	
Does your project involve human participants?	YES
Section 3: HUMAN CELLS / TISSUES	
Does your project involve human cells or tissues? (Other than from "Human Embryos/Foetuses" i.e. Section 1)?	NO
Section 4: PROTECTION OF PERSONAL DATA	
Does your project involve personal data collection and/or processing?	YES
Does it involve the collection and/or processing of sensitive personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	NO
Does it involve processing of genetic information?	NO
Does it involve tracking or observation of participants? It should be noted that this issue is not limited to surveillance or localization data. It also applies to Wan data such as IP address, MACs, cookies etc.	NO
Does your project involve further processing of previously collected personal data (secondary use)? For example Does your project involve merging existing data sets?	NO

Section 5: ANIMALS	
Does your project involve animals?	NO
Section 6: DEVELOPING COUNTRIES	
Does your project involve developing countries?	NO
If your project involves low and/or lower-middle income countries, are any benefit-sharing actions planned?	NO
Could the situation in the country put the individuals taking part in the project at risk?	NO
Section 7: ENVIRONMENTAL PROTECTION AND SAFETY	
Does your project involve the use of elements that may cause harm to the environment, animals or plants?	NO
Does your project deal with endangered fauna and/or flora /protected areas?	NO
Does your project involve the use of elements that may cause harm to humans, including project staff?	NO
Does your project involve other harmful materials or equipment, e.g. high-powered laser systems?	NO
Section 8: DUAL USE	
Does your project have the potential for military applications?	YES
Does your project have an exclusive civilian application focus?	NO
Will your project use or produce goods or information that will require export licenses in accordance with legislation on dual use items?	NO
Does your project affect current standards in military ethics – e.g., global ban on weapons of mass destruction, issues of proportionality, discrimination of combatants and accountability in drone and autonomous robotics developments, incendiary or laser weapons?	NO
Section 9: MISUSE	
Does your project have the potential for malevolent/criminal/terrorist abuse?	YES
Does your project involve information on/or the use of biological-, chemical-, nuclear/radiological-security sensitive materials and explosives, and means of their delivery?	NO
Does your project involve the development of technologies or the creation of information that could have severe negative impacts on human rights standards (e.g. privacy, stigmatization, discrimination), if misapplied?	NO
Does your project have the potential for terrorist or criminal abuse e.g. infrastructural vulnerability studies, cybersecurity related project?	NO

SECTION 10: LEGAL ISSUES	
Will your project use or produce software for which there are copyright licensing implications?	NO
Will your project use or produce goods or information for which there are data protection, or other legal implications?	NO
SECTION 11: OTHER ETHICS ISSUES	
Are there any other ethics issues that should be taken into consideration?	NO