DM

Increasing Empathy in Virtual Agent Interaction

MASTER DISSERTATION

Paulo Afonso Pinto Belim

MASTER IN INFORMATICS ENGINEERING

UNIVERSIDADE da MADEIRA

A Nossa Universidade

www.uma.pt

February | 2026

(This page intentionally left blank)

Increasing Empathy in Virtual Agent Interaction

MASTER DISSERTATION

Paulo Afonso Pinto Belim
MASTER IN INFORMATICS ENGINEERING

SUPERVISOR
Mónica da Silva Cameirão

CO-SUPERVISOR
Diana C. Gonçalves Mendes

(This page intentionally left blank)

UNIVERSIDADE da MADEIRA

A Nossa Universidade

www.uma.pt

FACULDADE DE CIÊNCIAS EXATAS E DA ENGENHARIA

MESTRADO EM ENGENHARIA INFORMÁTICA

Increasing Empathy in Virtual Agent Interaction

Paulo Belim

Supervised by:

Mónica S. Cameirão

Diana C. Gonçalves Mendes

(This page intentionally left blank)

Abstract

The need to socialize and connect is an essential part of human identity, and empathy, the capacity to perceive, share, and respond to the emotions of others, is at its core. In the case of virtual agents for self-disclosure, an extensive amount of work has studied the features that should be present to make them more empathetic. However, more studies are needed to improve how empathetic agents act and respond. The goal of this thesis is to study and observe how tracking non-verbal motion, like facial expressions, can help empathic agents detect human emotions and connect with their users. For this purpose, further development was done to an existing virtual agent for self-disclosure, an early prototype that aimed to depict empathetic conversation and active listening behaviors, to detect facial movement. We will study how interaction is affected by message quality, delivered responses, and how well the agent can adapt to non-verbal emotional changes. Our goal is to develop a package focused on improving empathetic features in virtual agents, contributing to the advancement in the design of virtual agents for mental health applications.

Keywords: virtual humans, personality, human-computer interaction, empathetic AI

(This page intentionally left blank.)

Resumo

A necessidade de socializar e criar conexões é uma parte essencial da identidade humana, e a empatia, a capacidade de perceber, partilhar e responder às emoções dos outros, está no seu cerne. No caso dos agentes virtuais para a auto-revelação, a literatura reporta extensivamente as características que devem estar presentes para torná-los mais empáticos. No entanto, são necessários mais estudos para melhorar a forma como os agentes empáticos agem e respondem. O objetivo desta tese é estudar e observar como o rastreamento de movimentos não verbais, como expressões faciais, pode ajudar agentes empáticos a detectar emoções humanas e interagir de melhor forma com os seus utilizadores. Para esse fim, foi incorporada deteção de movimentos faciais num agente virtual existente para auto-revelação, um protótipo inicial que tinha a intenção de retratar conversas empáticas e comportamentos de escuta ativa. Foram realizados testes de utilizador para analisar como a interação é afetada pela qualidade da mensagem, pelas respostas fornecidas e pela capacidade do agente se adaptar às mudanças emocionais não verbais. O nosso objetivo é desenvolver um pacote focado em melhorar as características empáticas em agentes virtuais, contribuindo para o avanço no design de agentes virtuais para aplicações de saúde mental.

Palavras-chave: humanos virtuais, personalidade, interação humano-computador, IA empática

(This page intentionally left blank.)

Acknowledgements

First and foremost, I want to express my gratitude to my parents, who made sure I stayed focused and supported the entire journey. I would also like to extend my sincere gratitude to my advisors, not only for making sure the project was not led astray, but also for providing suggestions along the way. I also extend my gratitude to the members of the NeuroRehabLab who facilitated both the necessary technology for this project and key information for my present and future work.

Such support is highly appreciated and I am grateful for your presence throughout this endeavor.

(This page intentionally left blank.)

Table of contents

List of Figures

List of Tables

Acronyms

Programming / Development Acronyms

AI - Artificial Intelligence

CA- Conversational Agents

ECA - Embodied Conversational Agents

FAU - Facial Action Units

HCI - Human-Computer Interaction

IDE - Integrated Development Environment

IPA - Intelligent Personal Assistants

LLM - Large Language Model

LCD - Liquid Crystal Display

NPC - non-Playable Character

SDK - Software Development Kit

SD - Standard Deviation

VR - Virtual Reality

Medical Acronyms

BMI - Body Mass Index

CBT - Cognitive Behavioral Therapy

DALY - Disability-adjusted Life Year

DSM-5 - Diagnostic and Statistical Manual of Mental Disorders Fifth Edition

HIPAA - Health Insurance Portability and Accountability Act

MDD - Major Depressive Disorders

PHQ-9 - Patient Health Questionnaire-9

Model Acronyms

BAS - Behavioral Activation System

BIS - Behavioral Inhibition System

NLP - Natural Language Processing

OCC - Ortony, Clore and Collins model

PAD - Pleasure-Arousal-Dominance model

RST - Reinforcement Sensitivity Theory

Package Architecture for Games

CiF - Comme il Fa

Questionnaire Acronyms

ITC-SOPI - ITC Sense of Presence Inventory

(This page intentionally left blank.)

1 Introduction

1.1 Background

Virtual Reality (VR) has undergone significant transformation in recent years, evolving from a niche technology into a powerful tool with widespread applications [1] This rapid advancement has opened the door for innovative approaches across a variety of sectors. One of the most significant areas to benefit from this growth is physical health and wellness [2] [3]. VR integration in this area has led to the creation of a new generation of applications focused on exercise, rehabilitation, and physical therapy. These immersive programs offer a dynamic alternative to traditional workouts, making fitness more engaging and accessible.

Beyond its physical applications, VR also offers immense potential for mental health and wellness. Its ability to be customized for a variety of users makes it a powerful tool for therapy and support. Given that many people face daily social and psychological struggles, a virtual agent that users could visually interact with could be a great resource to help them navigate these challenges [4][5]. When creating virtual agents for VR it is important to have in consideration an effect called Uncanny Valley where an object that looks like a human will give an eerie feeling in the user when this object presents irregularities or imperfections [6][7]. This effect can be reduced by either reducing its irregularities or making it less human such as reducing how realistic it looks.

1.2 Motivation, Objectives, and Research Questions

In recent years, conversational agents (CAs) have gained significant attention for their potential to enhance health-related communication, offer emotional support, and encourage users to open up about personal experiences [8][9][10][11]. As technology becomes more integrated into healthcare, researchers have begun exploring how CAs can be designed to foster trust and empathy, especially in sensitive contexts. One example is the case of the FCT funded project AViR (EXPL/CCI-INF/0298/2021), where a CA was used to encourage female users to share the experience of miscarriage [1]. In the context of this project, a study was done to determine whether the agent's appearance and speaking ability would impact user's empathy, affinity, likability, social presence and closeness, and which one was preferred for disclosure (Figure 1.1) [11].

It compared a non-human speaking agent, a mute female humanoid and a talking female humanoid agent controlled using the Wizard of Oz Technique. Their results showed that when interacting with these agents, the speaking humanoid (SH) agent presented a better emotional response in empathetic emotions such as Excited and Calm and when it came to Affinity with the user, both humanoid agents presented a higher Appeal and less Eeriness compared to the speaking non-humanoid (SNH). In terms of likability, the SH agent was the most preferred by over the majority of the participants, at 61.9% [11].

Participants appreciated various aspects of the SH agent, particularly her expressiveness (n=11), appearance (n=8), interactiveness (n=5), and voice (n=3) with the mute humanoid (MH) sharing the appreciation for her expressiveness (n=10) and appearance (n=8) while the SNH had her voice (n=5) as a positive feature while having an equal amount of likeness and dislikes for her appearance. Even taking the most liked agent (SH) not every aspect of the agent was liked.

Figure 1.1 - Speaking non-Humanoid (SNH), Mute Humanoid agent (MH) and Speaking Humanoid agent (SH), respectively [11].

Some of the agents' shortcomings were identified, including:

- Lacking facial expressions, with the humanoid agent's face barely moving while interacting with the user.
- Low responsiveness. While a conversation was possible, the speaking agents' responses eventually became repetitive.
- Repetitive body movements while talking were noticed after a while in humanoid agents.
- The agents were not fully able to act and respond empathically to the user.

These limitations bring up the importance of facial expressions, body language, and gaze in human interactions and comfort, how voice can help the user's perception, and the need for agents to respond to the user's emotional state. Upon analysis, it seems that more expressiveness is needed. This could be achieved by giving the agent more realistic facial and body movements, avoiding repetition in movement and agent oral responses alongside the reduction of the perceived lack of empathy.

Against this backdrop, this research aims to answer the following questions:

- RQ 1: Can an agent with improved body movements and more facial expressions reduce interaction discomfort and improve affinity and empathy?
- RQ 2: Can facial detection and eye tracking help the agent better detect user emotion?
- RQ 3: Can incorporating mirroring techniques improve users' sense of connection and empathy with an agent?
- RQ 4: Do users perceive non-hyper-realistic agents using synthesized voices favorably or negatively?

1.3 Document Structure

This thesis is organized into eight chapters to provide a comprehensive overview of the research:

- Chapter 1 - Introduction: This chapter talks about the study's background, why it is important to make virtual agents more empathetic, and the specific research questions that this work answers.
- Chapter 2 - State of the Art: Looks at the theoretical bases of human communication, human-computer interaction (HCI), and the psychological models of empathy and emotion that were used to make conversational agents.
- Chapter 3 - Software and Methodology: Talks about the iterative development process, the tools used (like the FAtiMA toolkit), and how the technical setup for the virtual world works.
- Chapter 4 - Development: Talks about how facial and eye tracking systems were put in place, how the Speaking Humanoid agent was changed, and how agent responses were improved through several rounds of testing.
- Chapter 5 - Pilot Study: This chapter talks about the experiment's steps, such as how to find participants, the tools used (ITC-SOPI and SUS), and the analysis of the data and results about the quality and usability of the agent.
- Chapter 6 - Conclusion: This chapter answers the research questions directly and suggests ways to make the AViR agent better in the future.

- Chapters 7 & 8 - References and Annexes: These chapters include a complete list of all the works that were cited, as well as extra materials like study posters, informed consent forms, and the questionnaires used in the pilot study.

(This page intentionally left blank.)

2 State of the Art

2.1 Introduction

Human communication has always influenced the development of human civilization, as it has made both relationships and cultural interactions easier, while also enabling various ways for people to share thoughts and information [12][13]. Empathy refers to identifying with, relating to, and responding to feelings - a particularly vital ability in this process [14]. It is one of the major hallmarks of genuine human interaction because it constitutes the building block of trust, cooperation, and cohesion among members of a society.

Increasingly, due to technological development, there has been a rising interest in bringing in these essential human aspects into digital interaction [7]. From simple text-based chatbots to today's most powerful virtual assistants, the development of conversational agents represents an effort toward building systems that take natural language inputs and provide natural language outputs to communicate with humans in meaningful, responsive, and even empathetic ways [14]. A number of challenges are involved in designing conversational agents that can actually mimic human-like behavior, in particular, in simulating empathy [5][7].

Virtual agents have the potential to change the concept of human-machine interaction in a number of ways, including better client service, personalized learning, mental health assistance, and the ability to perceive, interpret, and act upon human emotions. Research along the axis of emotion modeling and empathetic artificial intelligence (AI) has been heavily encouraged in an attempt to get these systems closer to human-like understanding [14]. Several models of emotion have been developed to mimic human-like reactions in virtual systems. In the sections that follow, we will examine these foundational models and their implications in creating empathetic virtual agents.

2.1.1 Human-Computer Interaction

Human-computer interaction (HCI) refers to the field of study concerning interaction by humans with computers and how to make systems more usable for humans [15]. Successful human-machine interaction relies on systems capable of communicating with the user in understandable, natural and easy ways [15] like Interactive Voice Responses (IVR) that can use automated phone systems to relay messages between voice or keypad inputs. Existing applications range from smartphones and smart speakers to automobiles. Intelligent personal assistants (IPAs) such as Siri, Mena, and Cortana have grown in recent years to let users perform tasks, access information, and interact with digital content using natural language. In fact, the high usage of IPAs across various platforms and devices justifies the level of utility and simplicity the assistants embody [15].

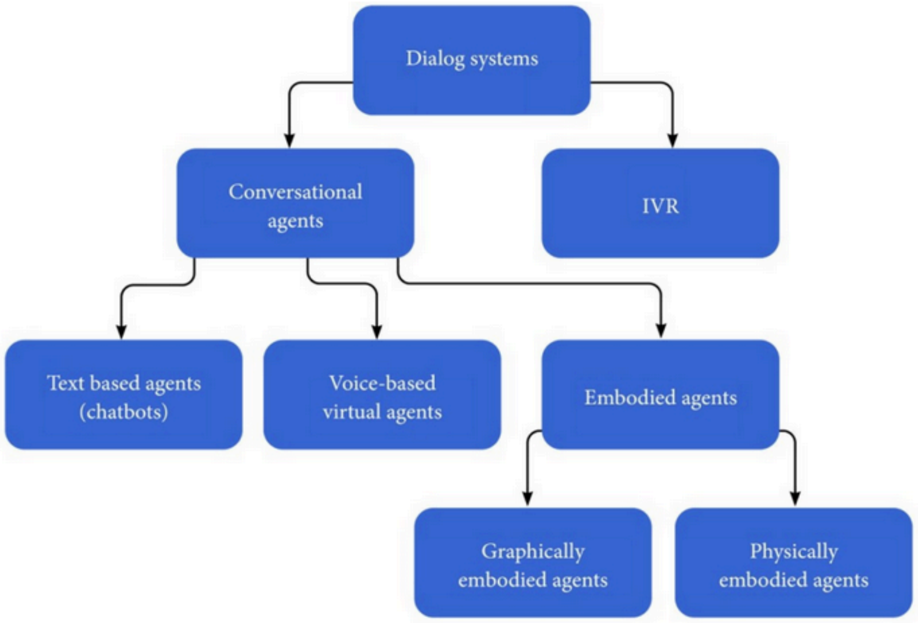


Figure 2.1 - Diagram visualization of types of conversational agents [14]

Virtual agents have had a key role in making systems easier to use, and reduced complexity in interpreting advanced commands or interfaces. CA applications are thus developing within the voice assistant and smart speaker market. IPAs have been introduced in several areas, such as information retrieval, healthcare, education, elderly care, and customer service [15]. This suggests that the development and use of conversational agents across a range of applications improve user's delight, accessibility, and service delivery as driven by HCI. Emotion identification and empathetic delivery, as relevant for such systems, are among the current points of interest in HCI research [7][16]. Improvements in this area are important for the development of interactions that can support applications requiring increased degrees of trust and emotional engagement.

2.1.2 Relevance for Mental Health Applications

Access to mental health care remains a major issue world-wide. Most countries face critical shortages of specialists in mental health [8]. There are about nine psychiatrists per 100,000 inhabitants in rich countries versus as few as 0.1 per 1,000,000 in poor countries [8]. Furthermore, major depression is the fourth leading cause of Disability-adjusted life year (DALYs) and the leading cause of years lived with disability globally [8]. This underlines the need to develop alternative means of meeting such an increasing demand for treatment, especially in areas and communities devoid of such resources.

One solution that has been proposed to address this limitation is the use of technology, especially conversational agents (CAs) [9][10]. Chatbots and virtual assistants have the potential to extend both reach and accessibility for mental health issues to help through emotional support, coping skills, and basic therapeutic interactions. Research suggests that chatbots may be particularly useful for patients who are anxious about discussing intimate or sensitive issues with a human therapist, as virtual agents may provide a confidential and non-judgmental space [8]. They are thus a potentially useful tool for groups of individuals who would otherwise not engage in conventional therapy. An example of such CA would be the Woebot, a chatbot powered by artificial intelligence that delivers Cognitive Behavioral Therapy (CBT) techniques through short, daily conversations [4]. As reported in a systematic review by Farzan et al., Woebot has demonstrated effectiveness in reducing symptoms of depression and anxiety, particularly among young adults [5]. The review highlights a randomized controlled trial in which users of Woebot reported a significant decline in depressive symptoms after only two weeks of use, compared to a control group. Additionally, Woebot has shown promise in building a strong therapeutic alliance, with users reporting a sense of connection and trust in the interaction despite the absence of a human therapist.

Another notable example is Wysa, an AI-driven mental health app that uses evidence-based therapeutic approaches such as CBT, Dialectical Behavior Therapy (DBT), and mindfulness [9][10] According to the same review, Wysa has been used by millions globally and has shown positive outcomes in reducing symptoms of anxiety and depression [5]. Studies included in the review found that Wysa users reported improvements in emotional regulation and psychological well-being over time, especially when using the app regularly. Wysa also offers optional human support, blending automated interaction with access to professional help when needed, thereby increasing its flexibility and appeal across different user groups.

For diagnosis of Major Depressive disorders (MDD), a research study at the University Hospital of Bordeaux tested an Embodied Conversational Agent (ECA) for diagnosing MDD through direct clinical face-to-face interviews [4]. The ECA worked on four interconnected modules: one interview manager, one 3D video display module, one speech synthesizer, and one speech recognizer. Interacting together, these modules enable interactive and realistic diagnostic interaction [4]. The patients' responses were scored for the presence of MDD according to the Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5) using a decisional algorithm tree [4]. Of the 35 cases diagnosed as suffering from MDD by the psychiatrist, the

ECA had 17 true positives and 18 false negatives. Thus, the accuracy was acceptable for a pilot study. Importantly, patients felt highly comfortable and trusted the ECA during their interactions. This reduced emotional barriers and the fear of judgment compared to traditional human interviews [4]. This again points to the capability of ECA in eliciting honest disclosures, which is very valuable in early diagnosis in mental health care [4].

2.2 Foundation of Human Emotion Models

While human-computer interaction (HCI) research has focused on improving user interaction, integrating emotional intelligence in the same systems requires further deliberation into human emotions [14][17]. Emotions form the basis of what makes us human, and are influential in our decisions, interactions with fellow humans, and relations with the world. Design for artificial agents must thus be based on a deep understanding of how emotions work. Over time, various models have been proposed to describe and categorize emotions [12][13][16][18]. Each has something different to say about the origin, dynamics, and effects of emotions.

This section goes through key frameworks that shaped research into emotions and their utilization in artificial systems. The Ortony, Clore, and Collins (OCC) model [12][16] presents a structured way to understand emotions grounded on the appraisal of events, actions, and objects. The Pleasure-Arousal-Dominance (PAD) model [13] takes a dimensional view, measuring emotions along axes such as pleasure and arousal. Meanwhile, the Behavioral Inhibition and Activation System (BIS/BAS) [18] focuses on the brain's motivational responses to stimuli.

2.2.1 Ortony, Clore and Collins Model

The OCC model defines emotions as arising from appraisals of three primary domains: events, actions, and objects. Each domain corresponds to a distinct source of value - goals, standards, and tastes, respectively - enabling the systematic classification of emotions [12]. Events are appraised based on their desirability relative to an individual's goals, actions are judged by their praiseworthiness with respect to social or moral standards, and objects are evaluated based on their appealing or unappealing nature in light of personal preferences [12][16].

Joy:	(pleased about) a desirable event
Distress:	(displeased about) an undesirable event
Happy-for:	(pleased about) an event presumed to be desirable for someone else
Pity:	(displeased about) an event presumed to be undesirable for someone else
Gloating:	(pleased about) an event presumed to be undesirable for someone else
Resentment:	(displeased about) an event presumed to be desirable for someone else
Hope:	(pleased about) the prospect of a desirable event
Fear:	(displeased about) the prospect of an undesirable event
Satisfaction:	(pleased about) the confirmation of the prospect of a desirable event
Fears-confirmed:	(displeased about) the confirmation of the prospect of an undesirable event
Relief:	(pleased about) the disconfirmation of the prospect of an undesirable event
Disappointment:	(displeased about) the disconfirmation of the prospect of a desirable event
Pride:	(approving of) one's own praiseworthy action
Shame:	(disapproving of) one's own blameworthy action
Admiration:	(approving of) someone else's praiseworthy action
Reproach:	(disapproving of) someone else's blameworthy action
Gratification:	(approving of) one's own praiseworthy action and (being pleased about) the related desirable event
Remorse:	(disapproving of) one's own blameworthy action and (being displeased about) the related undesirable event
Gratitude:	(approving of) someone else's praiseworthy action and (being pleased about) the related desirable event
Anger:	(disapproving of) someone else's blameworthy action and (being displeased about) the related undesirable event
Love:	(liking) an appealing object
Hate:	(disliking) an unappealing object

Figure 2.2 - 22 emotions identified in the OCC model [19]

Key Features of the OCC Model:

- **Appraisal Categories:**
  - *Event Appraisal:* Provides a judgment about the desirability of the event regarding an agent's goals. For example, candy being found is undesirable from health-related goals.
  - *Action Appraisal:* This focuses on whether the actions are in line with norms or standards. If the action is praiseworthy, it would be associated with pride, and if it is blameworthy, then with shame.
  - *Object Appraisal:* Centers on personal preferences, where emotions such as love or hate arise from the appealing or unappealing characteristics of an object.
- **Emotion Types:** The OCC model identifies 22 (later extended to 24) distinct emotion types, rejecting the notion of "basic emotions." These emotions are grouped into hierarchical categories, this would include:
  - *Event consequences:* emotions like joy and pity are determined by the desirability or undesirability of an event's outcomes.
  - *Agency of Actions:* Emotions like pride and re-approach arise from evaluations of actions in the form of praise worthy or blame worthy.

- *Aspects of Objects:* Feelings of love and hatred are based on a preference for objects.

- **Compound Emotions:** The model denotes complicated emotional states emerging from combinations of events, action and object appraisals. For example, gratitude arises from positive events caused by praiseworthy actions. Anger arises from negative events caused by blameworthy actions [12], [19].

2.2.2 Pleasure Arousal Dominance Model

The model by Mehrabian and Russell, the PAD, proposed in the 1970s [13], allows an approach to the study of the dimension of emotions for their mapping onto three axes: Pleasure, Arousal, and Dominance (i.e Figure 2.3). This would separate transient emotional states from enduring traits, yielding a dynamic representation of affective experiences [13].

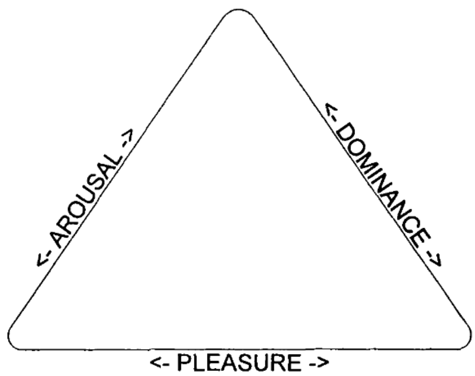


Figure 2.3 - Representation of states of the PAD model [13]

- **Dimensions of the PAD Model**
  - *Pleasure-Displeasure:* Represents the valence of a state, which can be understood as running the complete scope of emotions from positive feelings of joy, love, and relaxation to negative states like cruelty, boredom, and humiliation. Pleasure is closely tied to cognitive evaluations, with stimuli judged as more favorable inducing greater feelings of pleasure [13][20].
  - *Arousal-Non arousal:* Captures emotional intensity, with low arousal linked to relaxation and high arousal linked to mental alertness or physical activity, such as excitement or tension. Mehrabian's studies linked arousal levels to the

”information rate” of stimuli, with faster, more complex stimuli correlating with greater arousal [13].

- *Dominance-Submissiveness:* Measures a sense of control or influence over one’s environment, with emotions like power and anger linked to high dominance, while anxiety and loneliness correspond to low dominance [13][20].

- **Applications of the PAD Model**
  - *Event Appraisal:* Emotional relevance is appraised based on event pleasantness, modifying the agent’s PAD space position (e.g., positive events increase pleasure while stressful situations raise arousal and reduce dominance) [20].
  - *Emotion Dynamics:* PAD values dynamically evolve in response to emotionally relevant stimuli, slowly decaying toward neutral levels over time. This mechanism allows for realistic and context-sensitive emotional responses [20].
  - *Facial Expression Integration:* PAD is used alongside facial expression recognition modules to generate blended emotional states for lifelike agent behavior. For example, facial expressions might inform arousal and dominance levels, contributing to more lifelike agent behavior [21].

2.2.3 Behavioral Inhibition and Activation Systems

Biologically grounded, the Gray Behavioral Inhibition System-Behavioral Activation System represents a model that theorizes individuals process emotional and motivational signals based on neuron-bio-grounding principles. Within the realm of Reinforcement Sensitivity Theory (RST), this psycho-biologically analytical model theorizes an independent set of neural mechanisms regulating a basic behavior based upon environmental ambiance. In other terms, ”the behavior in response to various stimuli on the one side” conducted with BIS and a stimulation by appetitive stimuli (BAS), respectively [18][22].

The BIS is primarily a defensive system, meant to keep an organism away from harm by detecting and responding to signals of punishment, no reward, or threat. Sometimes referred to as a ”stop, look, and listen” system, it redirects attention to the environment in order to avoid negative outcomes [23]. By inhibiting risky behaviors and heightening anxiety, the BIS plays a critical role in ensuring survival [23]. Fear, anxiety, and caution are affective states linked with BIS activation, and thus avoidance behaviors are elicited in an attempt to reduce exposure to aversive events. In contrast, the BAS functions as a reward-driven, approach-oriented system that motivates individuals to move toward stimuli associated with potential rewards or positive outcomes. It promotes engagement in situations and activities expected to provide gratification or pleasure [24]. Reward likelihood is associated with the occurrence of hope, enthusiasm, or impulsive attitude following its activation, hence evokes behaviors that would increase reward amounts and maximize gains. This system is critical in motivating and goal-oriented behavior, especially in an environment where opportunities for gain are readily available [24].

One of the major ways in which the BIS/BAS model has been contributive is in relation to explaining human personality variation [18]. Individuals differ across the sensitivity of their biological BIS and BAS systems respectively, thereby influencing their dispositional emotional responses and behavioral directions [18]. People with a more sensitive BIS are more sensitive to punishment cues and are found to be anxious, to avoid risks, and cautious. Whereas individuals with more active BAS are more sensitive to reward cues and manifest traits like impulsivity, optimism, and strong goal achievement drive. It has been suggested that BIS and BAS interact to control behavior. Specifically, the activity of BAS may stimulate the BIS as an individual approaches its goal, because the possibility of obtaining a reward increases the chance of punishment [18]. The OCC, PAD, and BIS/BAS are some of the basic emotion models that are essential in understanding and generating human-like emotions in virtual agents.

However, these models cannot act as stand-alone but instead have to be adapted and then operated through practical tools and technologies. In the following section, we will see how such models are applied in systems such as GAMYGDALA and FAtiMA Toolkit, allowing empathetic and contextually-aware interactions. The linking of such theoretical mechanisms to real-life applications serves as a bridge between pure abstraction and functional empathetic AI systems. Together with the above models concerning emotions, one of the most relevant steps towards increasing user engagement has to do with the social interaction features of virtual agents. If integrated with non-verbal behaviors, natural language processing, and voice modulation, these can become much more human-like in the interactions that they provide and resonate with users. We will explore how these components will go about creating emotionally intelligent systems, providing a smooth connection between theoretical models and practical implementation.

2.3 Social Interaction and User Engagement

Social interaction is foremost in conversational systems. While up until now functionality has been the only interest, research is slowly turning these systems toward becoming emotionally intelligent and more appealing. In line with developing user-friendly interfaces, improvement of system cognition, and forging natural language dialogues which ring with users is in practice. Ultimately, the aim is to develop agents that would understand the user's feelings and react to them with empathy for trust and bonding [15].

2.3.1 Natural Language and Voice Interaction

Natural language processing (NLP) provides the backbone for conversational systems since it enables the system to understand the user and provide responses in a natural, contextually relevant way. Such systems work both text-based and voice-based to make integration seamless in everyday life, not complicating tasks and making technology further reachable [15].

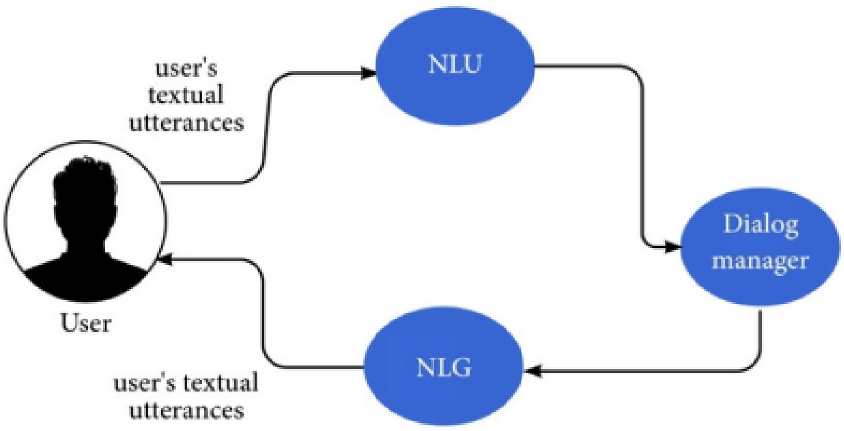


Figure 2.4 - Natural Language Understanding (NLU) and Natural Language Generation (NLG) between user and conversational agent (CA) [14]

Voice-based interaction gives depth to user engagement. Research has shown that voice characteristics, such as gender, pitch, and tone, play a huge role in user perception, trust, and likability [15]. This emotional dimension enables agents to capture user sentiments through acoustic cues and express their emotional states, hence making interactions more human-like [15].

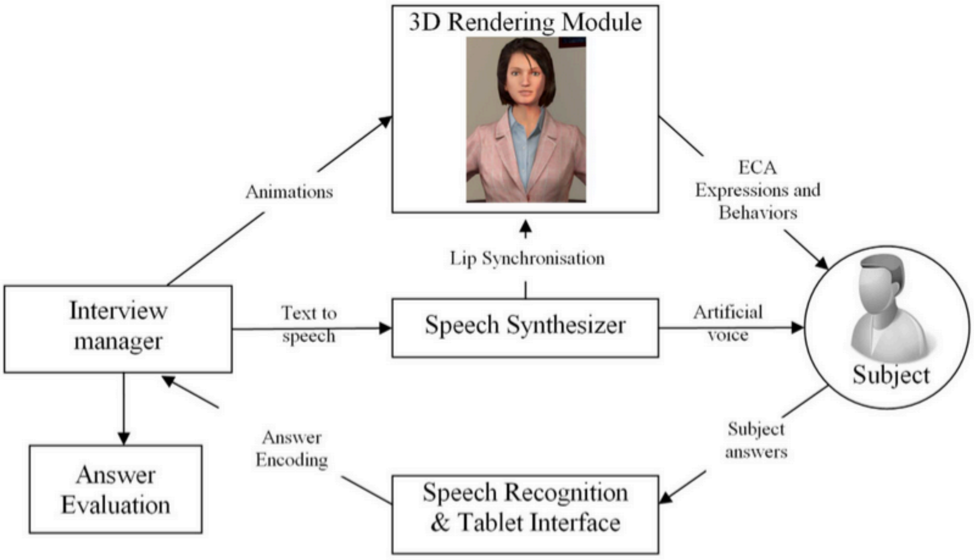


Figure 2.5 - Architectural necessity for a face-to-face clinical interview between user and ECA [4]

These include not only verbal but also non-verbal signals, such as facial expressions, head and hand movements, gaze, and posture. Cassell and Tartaro add that users respond to Embodied Conversational Agents (ECAs) in ways comparable to human interlocutors (Figure 2.5) [14]. Physiological and behavioral responses indicate that virtual agents can provoke emotional reactions comparable to real-life interactions [14]. This is an important ability in sensitive applications, such as healthcare, where empathy and trust are vital. Studies of ECAs have indeed shown immense promise in psychiatric applications, proposing that they provide much-needed soothing of patients and reduce certain barriers among those with emotional problems [4]. In the diagnosis of Major Depressive Disorders (MDDs), it is suggested that accuracy for ECAs were close to human psychiatrists, with specific improvements in more severe forms [4]. Better identification of facial expressions and voice prosody could enable greater improvements and extend their scope into mental health support [4].

These advancements signal a transformative shift brought about by natural language and voice interaction in the development of emotionally expressive and engaging agents. This kind of system will reduce the gap between human emotions and machine intelligence, thus completely revolutionizing user experiences for domains ranging from customer service to healthcare.

2.3.2 Non-verbal and Physical Behavior

Facial expressions, body language, and eye gaze are part of non-verbal communication and are deeply engraved in human interaction as a major source of emotions and intentions, mostly beyond words. The most important means of increasing an agent's believability and emotional expressiveness includes the simulation of such cues. They add trust, comfort, and user satisfaction, factors that constitute a key prerequisite for any form of effective interaction [6][7][15].

It has also been shown that agents able to depict human socio-emotional cues elicit more positive responses from users [6], [7]. The agent has to be both affectively empathetic, which means to reflect the user's emotions, and cognitively empathetic, considering the user's mental state before a response is given. This process works together to ensure interactions are easy and user engagement is greater [15]. Eye contact is perhaps one of the strongest indices of attention and empathy. The presence of agents simulating naturalistic gaze patterns enhances immersion and the quality of interaction [6][7].

Dynamic facial expressions, modeled with Facial Action Units (FAUs), as shown in Figure 2.6, based on Ekman's Six Basic Emotion Theory being those, happiness, sadness, anger, fear, surprise and disgust, also contribute significantly to enhancing nonverbal interactions. Wang et al. showed that avatars coupled with user affective states in the form of facial expressions or gestures increase emotional rapport and engagement by a significant degree [6]. Body language similarly conveys intent and emotion, and such physical cues would be employed by an embodied conversational agent to provide an intuitive and natural experience to a user, adding to the spoken language [7][15].

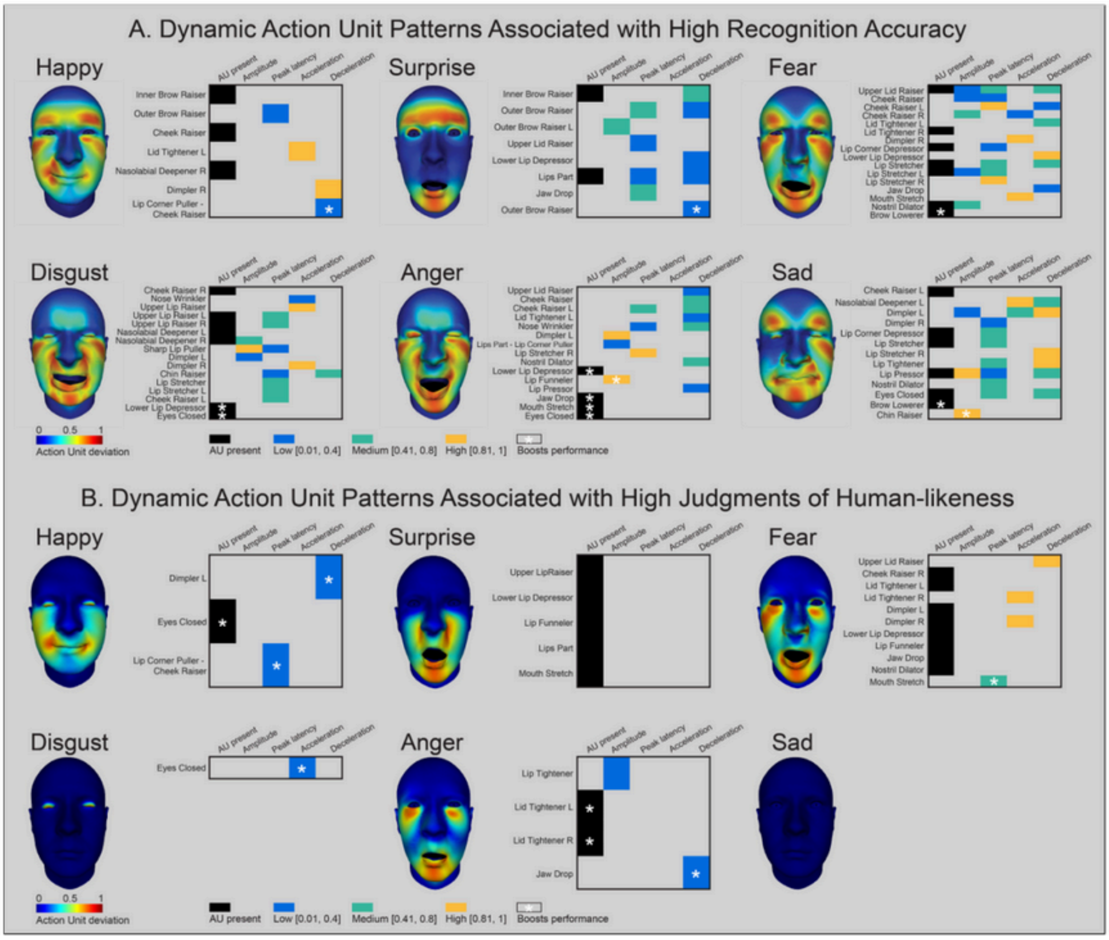


Figure 2.6 - The culturally-sensitive dynamic action units that are associated with high recognition accuracy in panel (A) and high judgments of human-likeness in panel (B) [6]

Though voice is not strictly physical interaction, it is closely connected to non-verbal interactions [16][21]. Some attributes of voice, such as pitch, tone, and loudness, may serve as carriers of emotional information which substantially influences user perception. The agents' ability to dynamically change these voice properties enables them to convey empathy and respond with adequacy to users' emotional states [15]. Other studies also indicate that users prefer conversational agents with control over voice parameters, such as gender and pitch, because such customization enhances trust and perceived relatability [15].

The key challenge with integrating non-verbal behaviors involves addressing the Uncanny Valley effect: near human-like features create discomfort because of small imperfections [6]. To minimize this, Scherer et al. pinpointed major combinations of facial action units most likely to be perceived as real emotions and, therefore, increase the believability of agents by reducing the perceptual discrepancies [6][7].

2.4 Emotion Regulation and Empathy Simulation

The virtual agents' ability to simulate empathy and regulate emotional responses constitute an urgent area of research in human-computer interaction. Virtual agents can show empathy by recognizing, understanding, and acting appropriately on the users' emotions - an activity that requires sophisticated emotion regulation mechanisms [7]. These mechanisms are often based on psychological theories trying to model human emotions and map them into computational frameworks. These agents would, in turn, develop trust and engagement through emulated human-like emotional processing for application areas ranging from healthcare to entertainment.

For effective emotion regulation, such agents need to assess and react to the user's emotions while constantly readjusting their behaviors. This emotional adaptiveness is about place-based constructs such as dimensional features of emotional states like valence and arousal [18]. These are the foundational tenets for most of the computational models of empathy.

2.4.1 Valence, Arousal and Cognitive Empathy

The development of virtual agents able to detect and respond to human emotions drew on fundamental theories, such as the Circumplex model of affect by Russell. This model characterizes emotions along the dimensions of valence-pleasure-displeasure and arousal-high-low activation [18]. These parameters are very important for the simulation of empathy and enable systems to estimate a user's emotional state in real-time and react accordingly. As a simple example, agents using bio potential signals and contextual information can provide personalized responses: empathic, encouraging, or congratulatory feedback-based on the user's apparent emotional state in job interview simulations [7].

This has been further extended to include natural user interfaces that effectively use facial expression recognition for gauging and responding to emotional cues. Ku et al. demonstrated how different intensities of emotional expressions in the avatars could influence the users' perceptions of valence and arousal, underlining such nuanced use of intensity-calibrated facial expressions to evoke affective responses [6].

2.4.2 Empathic Responses for User Satisfaction and Well-Being

These range from developing empathy in virtual agents to supporting mental health, teaching languages, and developing various skills for better user satisfaction [18]. The communication appropriateness achieved through better interaction timing used by the empathic agents results in greater satisfaction. In this context, Riek et al. reported that a robot's ability to imitate a user in head and mouth movements remarkably enhanced user satisfaction compared to a non-empathic robot [7].

It has been reportedly useful for embodied conversational agents in health-related applications to diagnose MDD, and this is also effective and specific to severity. ECAs can support clinicians to reduce their workload and make possible an accurate diagnosis [4]. Quality interaction with patients has been achieved in healthcare services [4]. In fact, chatbots and conversational agents have demonstrated potential in supporting engagement with therapeutic regimens, including CBT, and have contributed to improvements in psychological well-being by helping to reduce stress levels [8].

With the development of emotional intelligence being incorporated into virtual agents, consistency in empathic behaviors will become increasingly crucial. Behavioral inconsistency can hinder training and reduce learning effectiveness, highlighting the need for coherent design strategies that align with users' expectations and social norms [18].

2.5 Practical Uses of Empathic Agents

In recent years, companies, universities, and researchers have developed agents that apply empathic models discussed in Foundations of Human Emotion Models, utilizing both verbal and non-verbal social cues, such as speech, facial expressions, and body language, as outlined in Social Interaction and User Engagement.

- The GAMYGDALA emotion modeling framework extends the OCC model to provide a robust structure on integrating affective behavior into various games using agent emotion simulation [16]. The game would render internalized, emotive non-Playable Characters (NPCs) showing various social emotions, such as Figure 2.7, showing a guard feeling overwhelmed with his job. In another example, an NPC is angry toward the character but happy with his friend’s luck. Those emotional states would evolve in dynamism through continuous appraisal over what is happening in a game to show realistic changes in emotional states [16].

Figure 2.7 - Example of emotional interaction from a guard in GAMYGDALA [16]

In GAMYGDALA, emotions are associated with goals, which in turn are divided into active goals - a state that an NPC is trying to achieve - and maintenance goals - a state an NPC wants to maintain. An NPC will consider living a more significant priority than collecting resources, acting most expressively to danger than missed collections. This goal's organization will ensure that the NPCs respond in a relevant and meaningful manner to the happenings, thus allowing enhancement to take place in the gameplay experience. 16 of 24 OCC emotions are supported, creating subtle interactions between the players and NPCs [25]. By appraising events through the lens of goal relevance and likelihood, GAMYGDALA introduces complexity in character behaviors that add layers to make games feel more dynamic and personalized.

- The Comme il Faut (CiF) is a social interaction engine that develops believable and interactive social exchanges between characters in gaming environments [25]. Unlike typical NPC systems that depend on scripted interactions, CiF generalizes multi-character interactions into reusable "social exchanges.". CiF structures the social dynamics through four major representations: Social networks, Relationships, Cultural Knowledge Base, and Social Facts Knowledge Base [25]. These components collectively capture the social state of the game world, enabling NPCs to conduct meaningful, context-sensitive interactions. For example, the character's dialogue and actions can vary depending on their relationship to another NPC or with the game world's cultural expectations [25]. It has further been applied in games like Skyrim (Figure 2.8) by adding modified NPCs with all the capabilities of CiF-CK and testing them on two scenarios [25]. The Quest Scenario, in which the subjects engaged in structured narratives within characters, and the Open Scenario, where they allowed players to experiment in its sandbox with social NPCs for their impromptu interactions to be adapted.

Figure 2.8 - Dialogue from an NPC using CiF-CK in the game Skyrim [25]

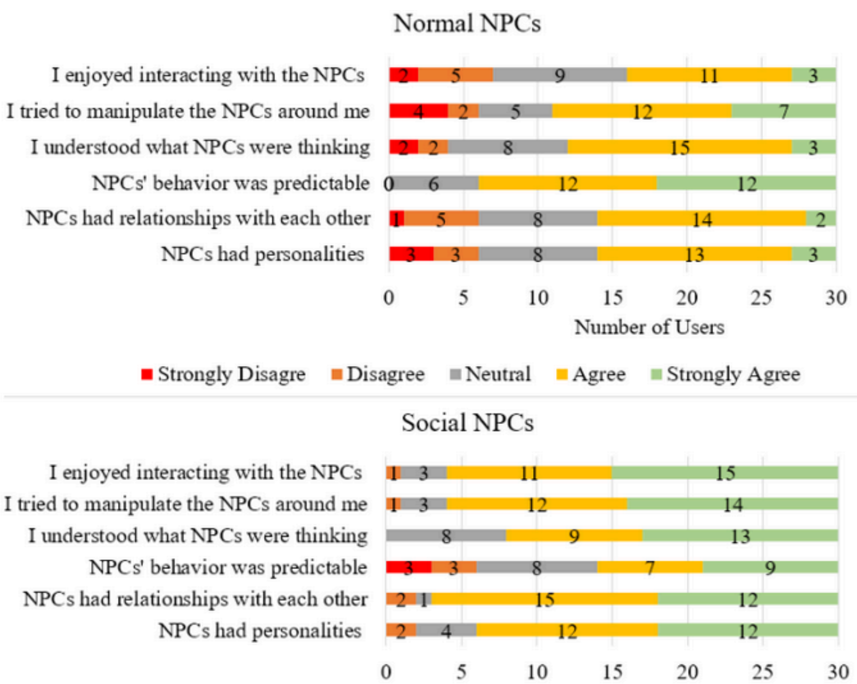


Figure 2.9 - Enjoyment from interacting with normal NPC vs. CiF-CK social NPC's. [25]

Players reported a much deeper sense of immersion and engagement with the social NPCs than with the traditional ones (Figure 2.9). This shows the potential of CiF for bringing a new dimension into storytelling and character interaction in games by making NPCs more relatable and believable [16].

- The FAtiMA Toolkit is an extended framework for developing virtual or robotic characters with socio-emotional intelligence [20]. Its architecture was first developed in the context of the FearNot. Project [20]. It integrates the Ortony, Clore, and Collins (OCC) model to simulate emotion-driven behaviors in 3D autonomous characters [20]. FATiMA has evolved over the years to include other constructs, such as empathy, psychological drives, and cultural dynamics that allow its application in different contexts like serious games and social skills training environments [20].

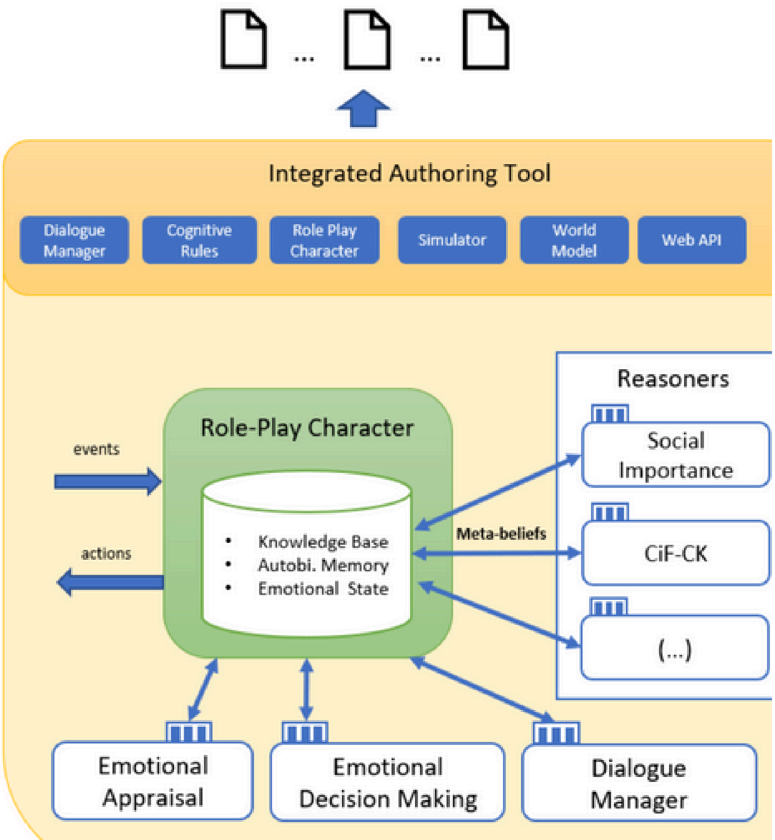


Figure 2.10 - FAtiMA Toolkit Components [20]

- Other strong features of this toolkit are the belief system representing the internal state of an agent and his/her understanding of the environment (Figure 2.10). The beliefs carry values, and they are laid out as “Composed Names.” It gives users freedom to define the scenarios, flexibility and no rigid conventions. It was never intended to model how humans would behave or carry out their functions, but to give agents an interactive response within their environments - be they in pedagogical uses or even in certain forms of therapy simulations. The modularity of the FAtiMA Toolkit and its strong basis in social and emotional theory makes it ideal for applications requiring nuanced human-agent interactions [20]. The features available within the FAtiMA Toolkit, along with its strong theoretical grounding in social and emotional modeling, make it particularly apt for applications requiring complex human-agent interactions, as found within games such as Space Modules Inc and Sports Team Manager [26]. In Space Modules Inc., for example, the player takes on the role of customer service and interacts with virtual customers, each demonstrating one of a number of unique states of emotions, including anger, stress, or uncooperativeness. The goal is increasing the customer's satisfaction due to the choice of available alternatives of dialogues. As provided in the literature, "The pedagogical goal of the game is to train players for a person's emotional state with verbal and non-verbal feedback. Acquire further experience on how to provide effective emotional responses" [26]. This shows the important function of the toolkit on keeping together emotional intelligence training on behalf of the player with funny play.

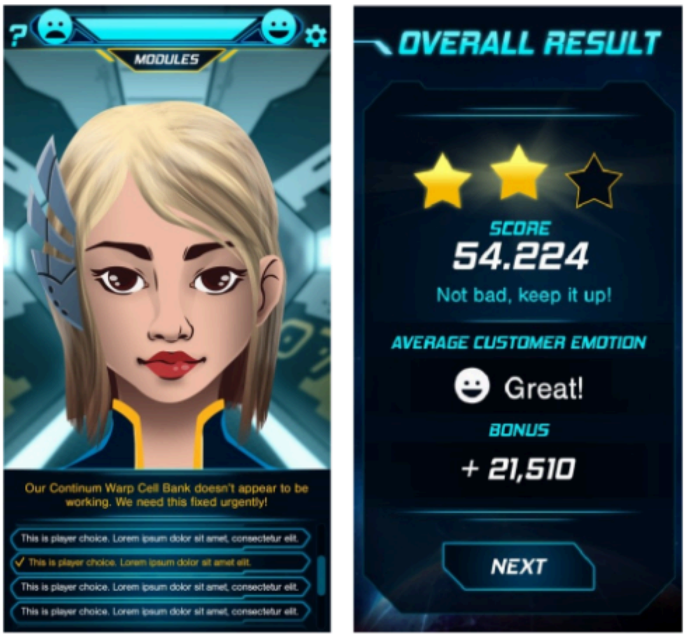


Figure 2.11 - Dialogue Screen (left image) and Result Screen (right image) Flow [26]

2.6 Ethical considerations

The rising popularity of conversational agents (CAs) and chatbots in critically sensitive domains such as medicine and mental health raises numerous ethical issues. While, on one hand, all these technologies offer the key benefits of accessibility, scalability, and higher user engagement, on the other they pose major challenges regarding security, privacy, transparency, and accountability.

2.6.1 Privacy Concerns and Data Security

Among the most important concerns is a total lack of comprehensive privacy protection for users of chatbots and conversational agents. Face-to-face interactions with clinicians need to follow severe privacy laws, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States [8]. However, most chatbots at present do not fall under these kinds of regulations. This makes users very exploited in terms of the various ways in which their data can be misplaced, sold, traded, bartered, or marketed by entities operating these systems. An innumerable amount of chatbots, operating specifically on the internet or generally on social media, may collect a wealth of personal information without the users' knowledge [8].

Besides, the architecture of such agents heavily relies on the storage and analysis of sensitive user data in contexts such as healthcare or education. In addition, the conversations regarding medical, mental health, or other personal issues are quite sensitive. Current technologies can hardly ensure anonymity and privacy in such dialogues. For example, virtual assistants normally record user audio and process it on remote servers, where unauthorized access and data breaches can occur [14].

2.6.2 Personalization vs. Privacy

The effectiveness of a conversational agent often relies on how personalizable interactions can be. Most personalization strategies involve implicit methods, including storing user data across sessions or explicit methods whereby, at the beginning, the agent requests user preferences [15]. This will contribute to user satisfaction because such responses will address the interests of the user. On the other hand, clearly doing so has major costs with respect to privacy. Users may reveal sensitive information, such as health, financial, or security data, in the process of natural, free-form conversations [15].

Researchers and developers highlight that to overcome these privacy challenges, users should be empowered with options on the amount of information shared. For instance, conversational agents should, at all times, inform users of the signals being collected from them, allow encryption options for the stored data, and allow them to practice the "right to be forgotten"

[15]. Even though this approach reduces the risk of privacy, it further makes the process of personalization cumbersome. Hence, this points to the challenge in balancing user satisfaction with ethical considerations [15].

Another major ethical issue is the lack of transparency shrouding the processes carried out by chatbots. Most people are unaware of how their information is actually managed [15]. This makes it difficult to obtain truly informed consent. Eventually, transparency is a needed quality if trust is to develop about their use, especially concerning critically valuable applications like human medical interventions. Also, chatbots need to explicitly show their limitations (e.g., not being able to ensure confidentiality) and give users easy and understandable terms of use [8].

In the absence of laws and regulations with respect to the usage of conversational agents, it is unclear who could be responsible in case of harm (e.g., if the chatbot gives wrong advice, fails to recognize a critical situation, or handles sensitive information inappropriately): creators, operators, or users themselves. This needs to be resolved through new legislation that will clearly outline the responsibilities of all stakeholders involved in the deployment of CAs [15].

2.7 Conclusion

This review has underlined how much has been achieved so far regarding the development of empathetic virtual agents. It also pointed to emotional intelligence as a key path toward further development of human-computer interactions. Virtual agents can show quite complex emotional responses using models such as the OCC [12][16], PAD [13][20][21], and the BIS/BAS [18][22]–[24] frameworks, thus enabling their use in applications for mental health, education, and entertainment. The development of systems like GAMYGDALA [16], CiF [25], and the FAtiMA Toolkit [20][26] represents practical integrations of these models into emotionally responsive and contextually appropriate agent behaviors. Such developments show the potential of these systems for trust creation, engagement, and increasing user satisfaction.

Taking into consideration the previous studies shown throughout the state of the art and their applications of emotions, voice and non-verbal behaviour, the AViR empathic agent [1][11] will be designed to confirm whether an empathic agent based on the Ekman's 6 basic emotion model, with 7 emotions (surprised, happy, sad, fearful, angry, confused, bored) can effectively simulate a conversation with a human and help users better cope with their daily-life problems. This while providing an empathic agent capable of satisfying the user expectations and bridging the gap between human empathy and artificial intelligence.

It is foreseen that improved facial and body movements will reduce the user discomfort by cutting the number of imperfections on the agent and using a more synthetic voice. Further, using a non-realistic agent is expected to help create affinity with the agent.

(This page intentionally left blank.)

3 Software and Methodology

3.1 Iterative Development Process

At the core of our methodology is the Iterative Development Process, characterized by cycles of planning, development, testing, and evaluation, facilitating constant refinement of the VR experiences being developed. Central to the iterative development process were weekly meetings with the supervisors and a multidisciplinary team, providing organized objectives to the project. Progress was consistently evaluated, and tasks related to testing, implementing, or fixing aspects of the project were assigned based on the discussions and presented work. The diverse expertise of the team allowed comprehensive evaluation and feedback, addressing multiple facets of the project. Criticisms and advice received were incorporated into weekly tasks, facilitating the project's evolution and refinement. This method allowed a swift response to emerging challenges and an effective incorporation of new findings.

3.2 Toolkits and software

Since the project worked on an existing project, a few technologies were already in use and their specific roles were:

- *Unity3D (https://unity.com/):* Unity3D was chosen as the primary game engine due to its wide-ranging capabilities in building complex 3D environments and compatibility with numerous VR platforms. The rich set of resources provided by Unity's active community, alongside its intuitive user interface, made it an ideal choice for implementing an agent that will receive multiple updates.

- *OpenXR for Unity:* Due to the diverse hardware compatibility OpenXR provides, it was the ideal platform to use alongside Unity. OpenXR provides a standard API that works across multiple VR hardware platforms, ensuring the project's accessibility and future scalability. The integration of OpenXR with Unity streamlined the previous development process and ensured smooth performance across diverse VR devices.

  (https://docs.unity3d.com/6000.1/Documentation/Manual/com.unity.xr.openxr.html)

- *GitHub (https://github.com/NeuroRehabilitation):* Selected for its comprehensive issue tracking, and intuitive user interface. It is also the version control standard at NeuroRehabLab, and, therefore, where the base project is located. GitHub's widespread adoption in the developer community also provides better opportunities for future collaborations and development.

- *Unity Assets (https://assetstore.unity.com/):* All the current assets used in the project come from various sources in the Unity Asset Store, used to expedite the previous development process.
- *Google Drive:* Since the AViR agent is only a small part of a larger structured project, all the assets are located in the NeuroRehabLab's Google Drive, agilizing the project's retrieval, with developers only needing to add the necessary assets in the project.

  (https://drive.google.com/drive/folders/1LTKLJ9LawiaV3eK4XcnFDSkxYC8FDAh2)
- *Meta Quest Pro:* The AViR project was originally developed using Meta Quest 2. Unfortunately, this headset is incompatible with our Research Questions on facial and eye tracking. As a result, the VR headset used for our study was the Meta Quest Pro. This headset is capable of providing the same six degrees of freedom namely, left/right, backward/forward and upward/downward movement alongside rotation movements like tilting, turning and nodding in a better capacity due to its upgraded depth adjustment, better lens, and use of Dual Liquid Crystal Display (LCD) compared to the single LCD use of Meta Quest 2. The controllers are also more responsive since Meta Quest Pro uses Self-Tracking controllers meaning they are not relying on the headset to provide movement.
- *Ready Player Me (https://readyplayer.me/) :* Ready Player Me is an easy to integrate 3D avatar creator that provides a full customizable avatar for games and apps. This tool not only is developer-friendly since it provides a Software Development Kit (SDK) for Unity, it is also web-based which means it doesn't need to have installations of software allowing the creation of this avatar for a multitude of platforms.

To track the user movements, we used the following software:

- *Meta Movement SDK:* It uses facial, eye and body tracking to bring its users' physical movements into the virtual space by using the abstracted signals that tracking provides to animate characters. It uses the tracking provided by the device to follow multiple parts of the face and body.

  (https://developers.meta.com/horizon/documentation/unity/move-overview/)
- *Visual Studio Code (https://code.visualstudio.com/):* To facilitate the development and use of the previous contents of the project and its updates, it was decided to use the Integrated Development Environment (IDE) Visual Studio Code. The primary reason

for its usage was its advanced debugging tools, which support multiple programming languages. It also offers a seamless Git Integration and a user-friendly interface.

3.3 Setup

First, the package was downloaded from the NeuroRehabLab GitHub repository (https://github.com/NeuroRehabilitation). The assets on the Drive linked directly on the same GitHub were also downloaded. To make sure everything was functional, the correct incorporation of all the assets in the project was confirmed before its first runtime.

After that, we installed the Meta Movement SDK package (

https://github.com/oculus-samples/Unity-Movement) and added it in the project using the git URL provided via Unity's.

When it was done, it was possible to confirm that the Meta Movement package was installed successfully when it appeared alongside the list of the other packages. After selecting it, we have access to its documentation, all the changes in versions and the licenses (Figure 3.2).

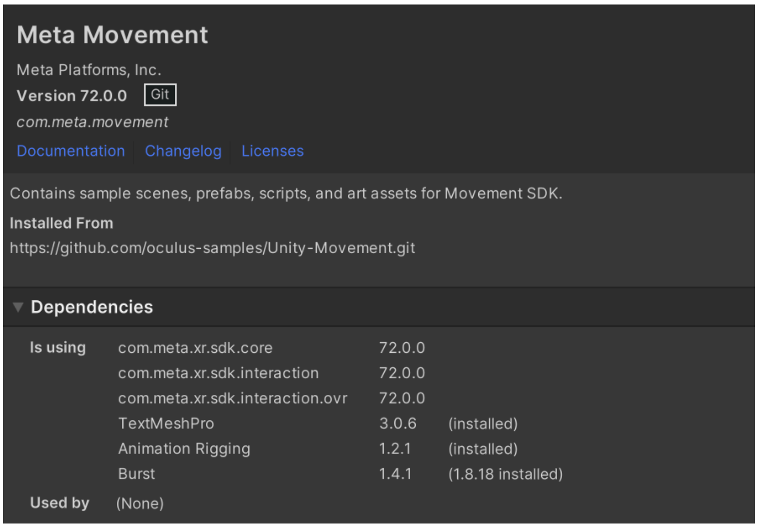


Figure 3.1 - Meta movement package page in Unity

After the package installation, changes were made considering that the version of Meta Movement (72.0.0) used was compatible with the Long-Term Service (LTS) version of Unity 2021.3, while the AViR project was developed in a previous version (2021.3.21f1).

Figure 3.2 - Previous AViR version (2021.3.21f1) and the new LTS version

With the update to the LTS version of Unity, the majority if not all packages were updated with it (example being the core hand and controller tracking xr.core in Figure 3.4. The entire Unity project was tested to guarantee its correct functioning.

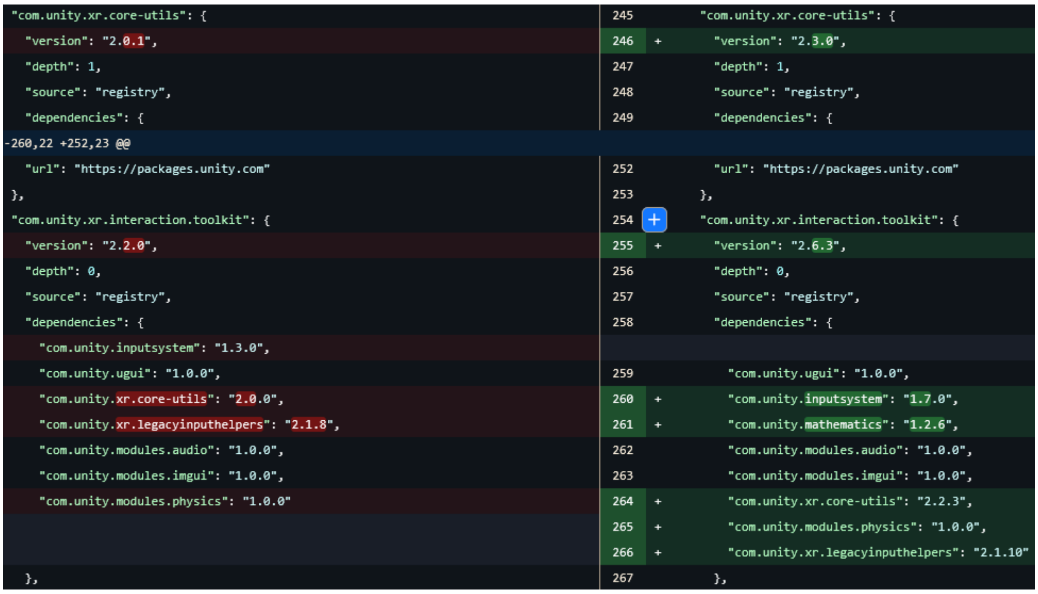


Figure 3.3 - Core functionality of project update versions

4 Development

4.1 Tracking

Early modifications included the addition of a new headset rig called OVRCameraRig that controls the new movement recorded by the user. This movement is tracked with the OVR Manager property. It contains various components for the quality provided to the player, such as a recommended Multi-Sample Anti-Aliasing (MSAA), a technique in computer graphics to smooth jagged edges producing a smoother color and the usage of local dimming to improve how the user sees the virtual world. In Figure 4.1 we can also see that it is possible to choose the tracking origin type, which we set up at the floor level, since it tracks the user's position and orientation relative to the floor, with floor height determined during boundary setup. The reason it was chosen compared to the other two is because it is capable of providing stable referencing for interpreting posture, height or movement.

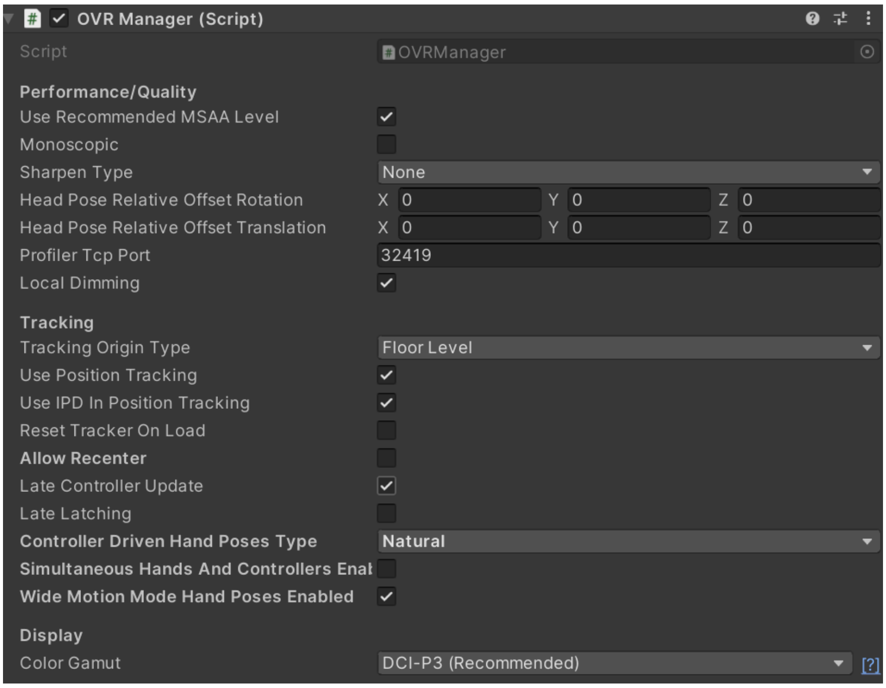


Figure 4.1 - Performance, Quality and start of the tracking definitions

The most important part of the Manager shows the Hand Tracking and the full body tracking permissions, seen in Figure 4.2. The hand Tracking Support indicates the type of hand tracking that is allowed. Since it could be used throughout the rest of the AViR Unity project, it was defined to support both controllers and direct hand tracking. The frequency was set to High, the minimum required for the avatar's hands to move properly or respond fast enough. At the bottom, we set up the body tracking support. All three options are required for our study (RQ3).

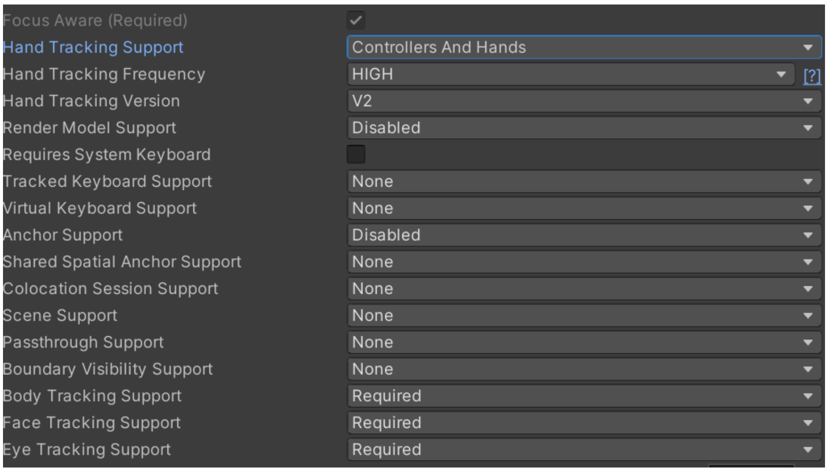


Figure 4.2 - Support for full body tracking

In the user's avatar itself, the Movement SDK adds a script to move the body, the retargeting Layer. The skeleton type was set to Full Body since we need almost full body tracking mostly due to the headset's inability to track the user's legs. Overall, our goal was to map the user's avatar animation data through pre-recording and the agent avatar's skeleton, so our agent could mimic human movement in a visually believable way. The Retargeting Layer is able to do so by identifying a set of common joints (shoulders, elbows, knees, hips) and adjusting the agent's pose to match the user's.

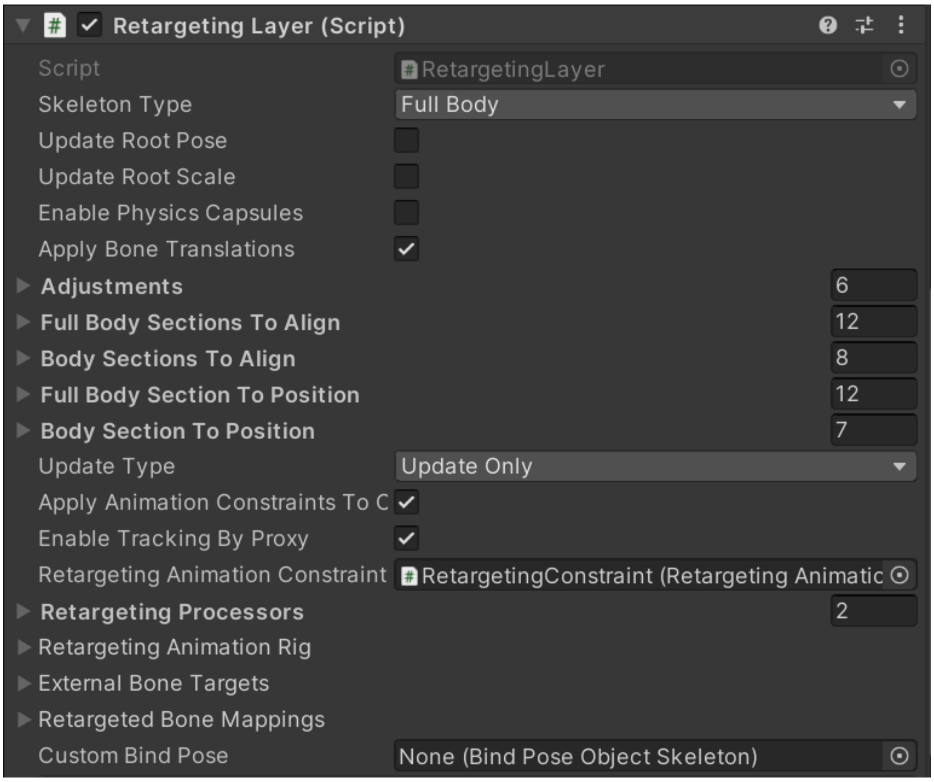


Figure 4.3 - Script that bind the avatar to the movement provided by the user

A challenge was found because of differences between the avatars used by Meta and those used in the early AViR agent. While Meta uses a Mixamo rig avatar (Figure 4.4), the AViR agent used Ready Player Me (Figure 1.1, Figure 4.22). While similar, they use different types of rigging.

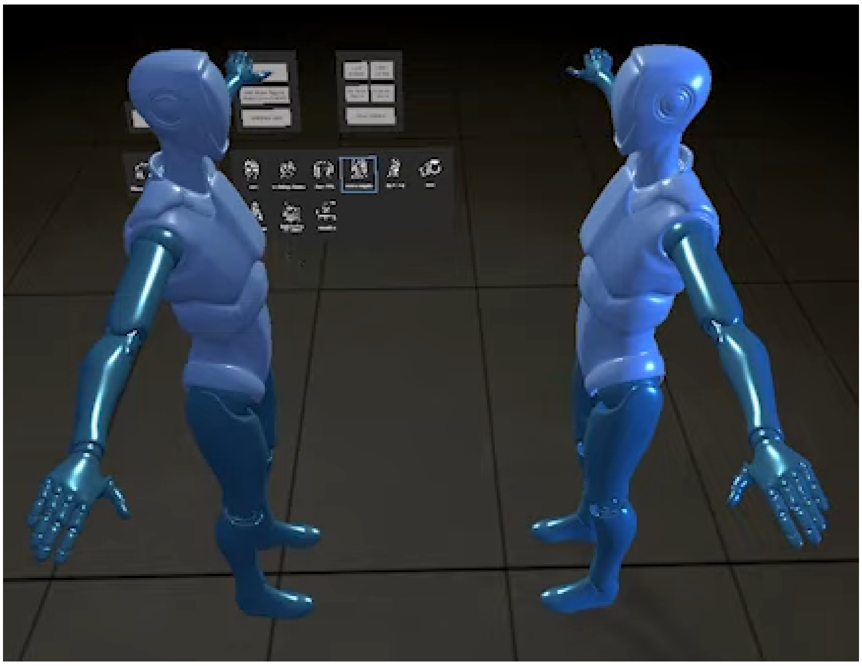


Figure 4.4 - Mixamo body rig

This problem was solved by using Ready Player Me's Dynamic Movement Loader and a URL Loader. The Dynamic Movement Loader provides a connection between the two avatars by providing a rest pose depending on the gender of our avatar and a configuration that is translated for each part of the Meta body to our Ready Player Me body. The latter works differently. Since Ready Player Me can provide and generate avatars from their own platform, we can make a call to generate the avatar on runtime instead of having it as an object in the game.

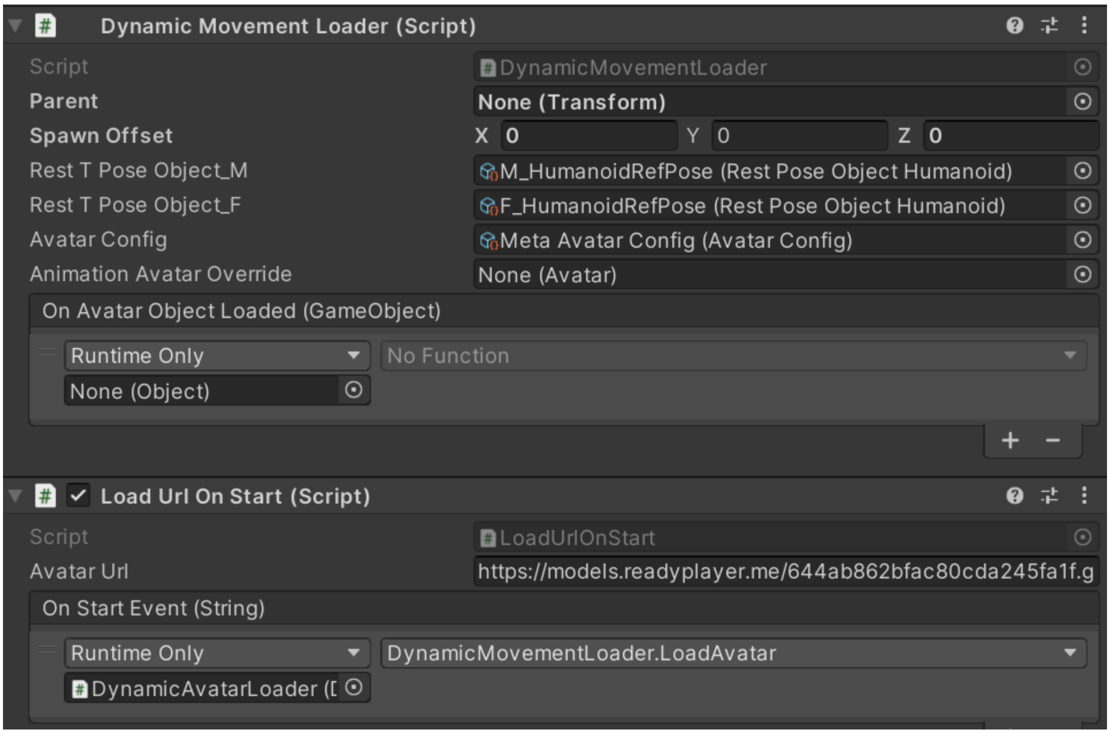


Figure 4.5 - Ready Player Me Script connector to the Meta Movement SDK

To start on the tracking of the user's avatar we created an Extract Model ID that would grab the avatar ID (set of numbers and letters after readyplayer.me in Figure 4.5). This solution not only erased any need for the developer to modify the avatar ID in multiple places, but also made it less prone to errors.

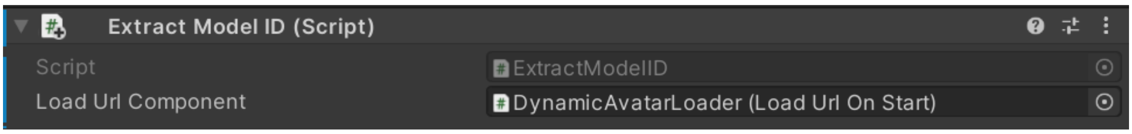


Figure 4.6 - Script that dynamically gets Ready Player Me ID

The ID from DynamicAvatarLoader extracted using Extract Model ID in Figure 4.6 is sent to 2 other scripts: the Face Tracking (Figure 4.7) and Body Tracking (Figure 4.11). First, it is sent to Face Tracking, which gets 2 data from other Meta Movement scripts: OVR Eye Gaze (tracks and makes the avatar eyes move and rotate) and OVR Expressions (Figure 4.8).

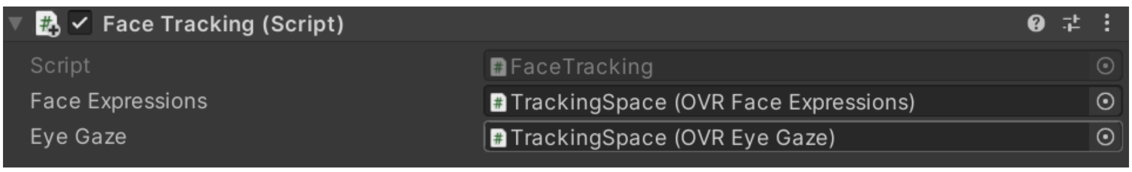


Figure 4.7 - Script that grabs eye movement and facial expressions.

As we can see in Figure 4.8, we have both the OVR Face Expressions and AR Kit Face; The AR Kit Face dictates which of the avatar's blendshapes are connected to the Meta Movement SDK Facial Action Coding System, and the OVR Face Expressions makes these blendshapes move.

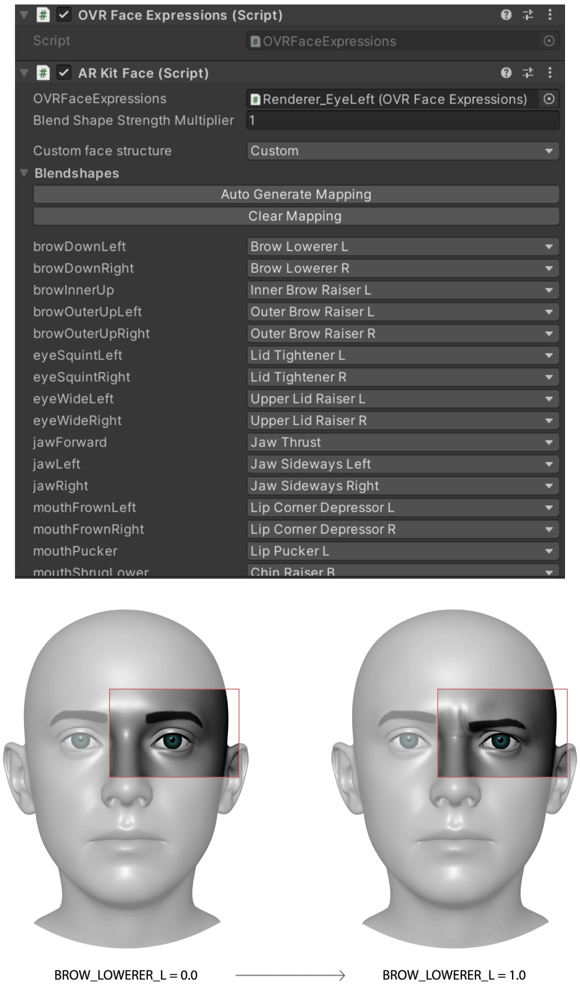


Figure 4.8 - Blendshapes connections and example of an FACS blendshape movement [27]

Facial Tracking works by detecting the 56 FACS from both the Eye Gaze and Facial Expressions, and identifying or matching them with the ones defined by Wang [6] (Figure 2.6) to create an emotion detector as seen in Figure 4.9 and 26. This data will be later registered by the AViR speaking avatar, which will consider the detected emotions when responding and moving. When an emotion has a weight value of 5 or higher the Current emotion will change to the registered emotion (Figure 4.9) and if none gets above 5 it will consider the emotions as neutral (Figure 4.10). The responses provided by the avatar takes in count what was the most dominant emotion in the past 10s before responding and making its own movements.

For the emotions chosen, Ekman's Six Basic Emotion Theory came first when picking the emotions for the avatar since those emotions are the most common emotions a person can demonstrate but, taking in consideration that the AViR agent primary goal is to interact empathically with users that experienced miscarriage, a heavy emotion such as disgust should not be included out of respect for the user and instead 2 emotions based on the Pleasure-Arousal-Dominance model were added, Confusion (Confuso, weak dominance) and Boredom (Aborrecido, weak Arousal) since this two emotions are also very common in conversation context.

```
[Face] Surpreso: 7
[Face] Feliz: 2
[Face] Triste: 4
[Face] Medroso: 8
[Face] Irritado: 0
[Face] Confuso: 0
[Face] Aborrecido: 0
[Eyes] Surpreso: 7
[Eyes] Feliz: 2
[Eyes] Triste: 4
[Eyes] Medroso: 8
[Eyes] Irritado: 0
[Eyes] Confuso: 0
[Eyes] Aborrecido: 0
Current Face: Medroso Current Eyes: Surpreso
```

Figure 4.9 - Multiple emotion detection

```
[Face] Surpreso: 0
[Face] Feliz: 8
[Face] Triste: 2
[Face] Medroso: 0
[Face] Irritado: 4
[Face] Confuso: 2
[Face] Aborrecido: 4
[Eyes] Surpreso: 0
[Eyes] Feliz: 2
[Eyes] Triste: 1
[Eyes] Medroso: 2
[Eyes] Irritado: 2
[Eyes] Confuso: 0
[Eyes] Aborrecido: 4
Current Face: Feliz Current Eyes: Neutro
```

Figure 4.10 - Detection of emotions according to the 5-point threshold

The Body Tracker is slightly different from Face Tracking, since it has 2 ways to detect an avatar: using Extract Model ID to grab the avatar as Face Tracking does; or by using an in-world avatar to detect body movement for testing purposes.

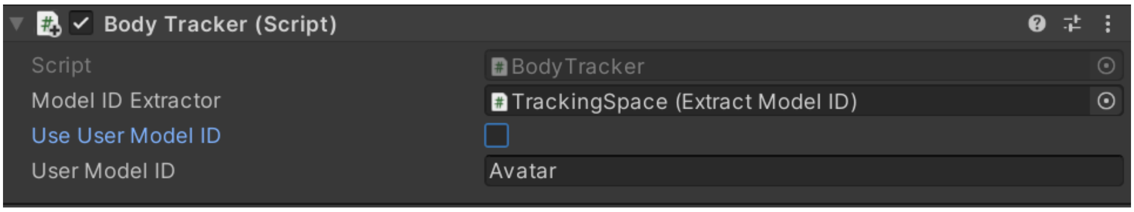


Figure 4.11 - Body Tracker capable of 2 methods of tracking

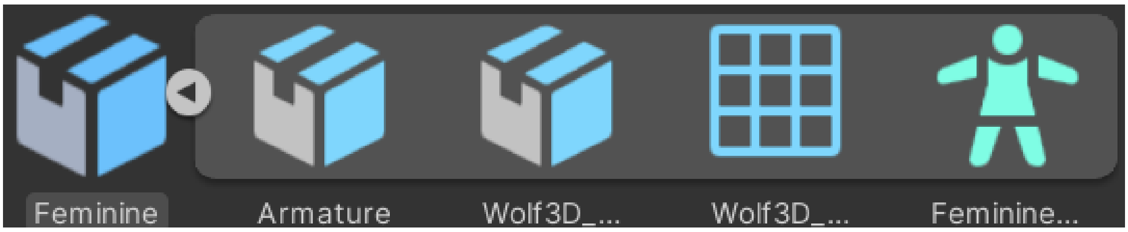


Figure 4.12 - Feminine humanoid avatar

The Body Tracker (Figure 4.11) tracks the avatar's bones while it moves, using the humanoid avatar provided by meta. It was set through an object that contained all the armature information of the feminine avatar (Figure 4.12), which was then added to the Animator (Figure 4.13).

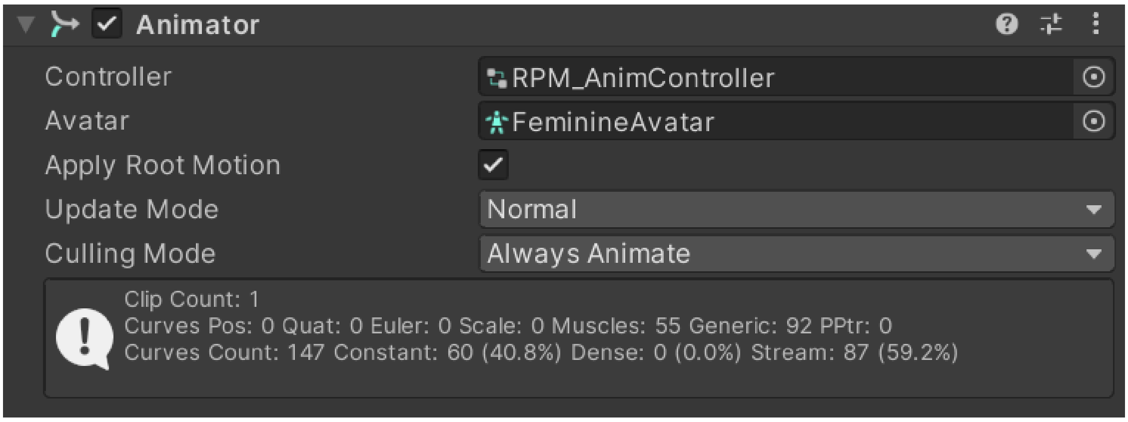


Figure 4.13 - Animator with the Avatar config of the Feminine object

This configuration must be set to Humanoid instead of Generic in its Rig Settings (Figure 4.14) for the avatar to move. Otherwise, bones will break and the avatar will be forced to make a T-Pose, a standard position indicating the object is not correctly set or is not working.

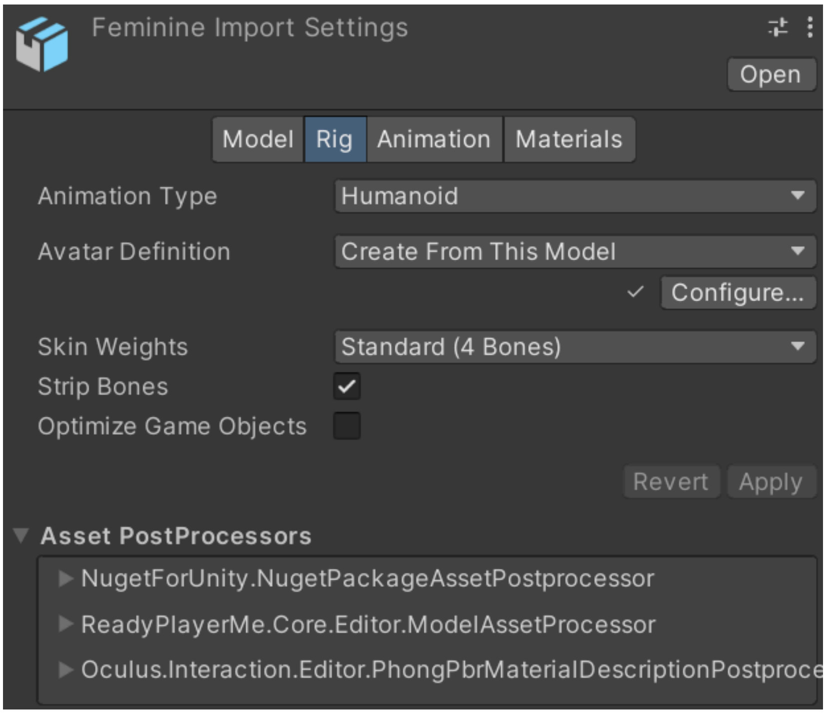


Figure 4.14 - Rig Settings

When it comes to mapping the avatar (Figure 4.15), it is visible on the left that the avatar's joints connect with one another, and the avatar's boundary to the floor. If the avatar is under or above the floor, its boundary is incorrect and needs to be fixed. On the right, it can be seen that the main joints of the body are divided in 4 sections: Body, Head, Left Hand and Right Hand.

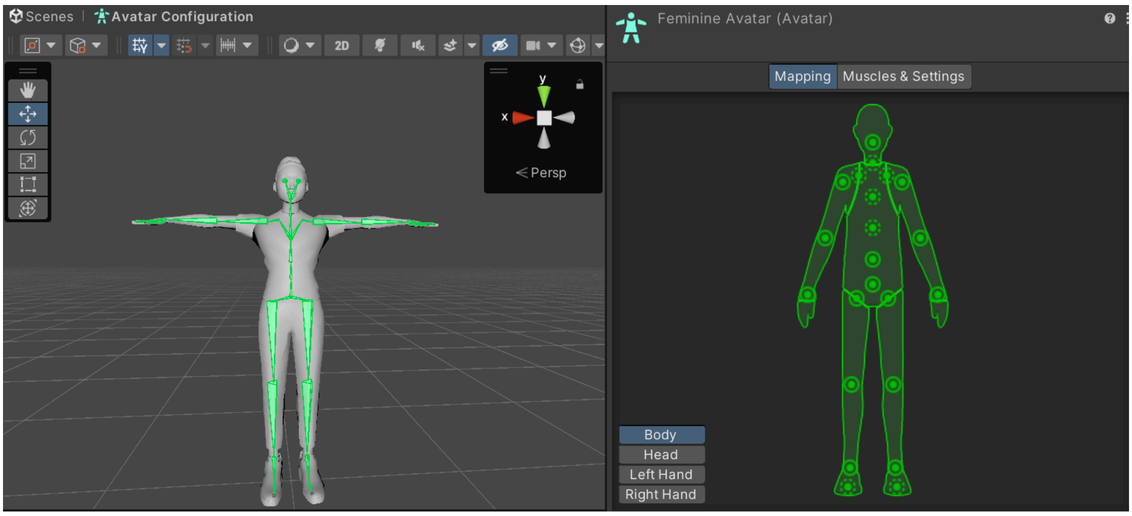


Figure 4.15 - Feminine Avatar Joint Mapping

In Figure 4.16, we can see the subgenre of joints are connected to the 4 main joints mentioned before and shown in Figure 4.15. For the body, it was 3 main bones: the Hip, Spine and Chest. The optional Bone Upper Chest, Head (not shown in Figure 4.16) controls the eyes, mouth and cheeks. The left and right arms have each one a set of Shoulder, Upper Arm and Lower Arm and Hands. It is to note that each Hand has control to each finger, with all the sets of finger joints being tracked when set to Hand tracking instead of controller.

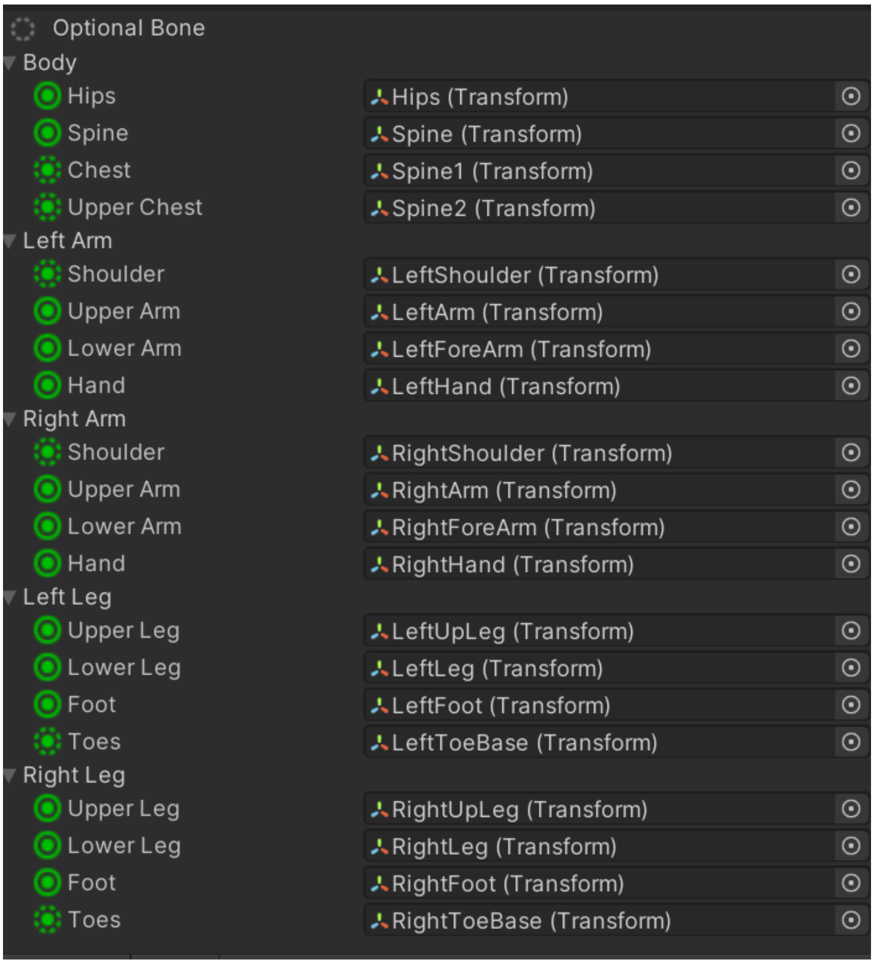


Figure 4.16 - Subset of Bones of the 4 main mapped bones

4.2 Speaking Humanoid Modifications

The agent's interaction with the user was modified. Previously, it used a limited set of pre-recorded audios prepared by a research assistant in Psychology and set to play one of the audios and animations picked by a psychologist using the Wizard-of-Oz method, a remote control using an Elgato Stream deck with 15 customizable LCD keys [1][11], depending on certain scenarios (user not talking, give reinforcement speech, etc,). The modified version is set to use 2 different services: Gemini and Azure Speech Services (Text-To-Speech and Speech-To-Text settings).

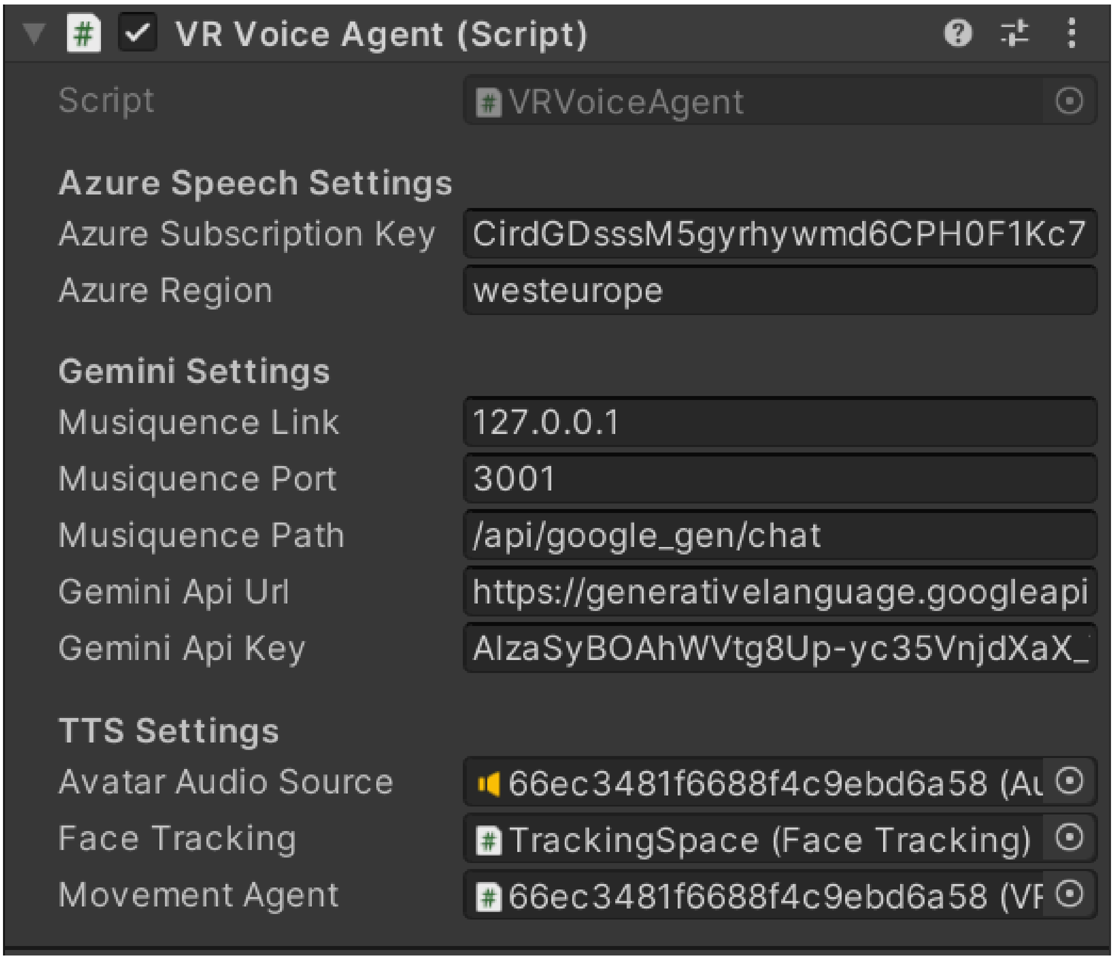


Figure 4.17 - Speaking Humanoid Voice Settings

To access these services, both need to have a working API key set up in their respective locations. For using Azure, alongside the key, it was needed to specify what region the key is from. For using Gemini, the version to be used needs to be specified with its URL (in our case, Gemini 2.5 flash). The algorithm is set to listen for User Speech and to catch short breaks until the user stops speaking, being the time set to 3.5 seconds. After this, it checks if any issues with the data were corrupted or not captured cleanly. If no issues are detected, the data is sent to Gemini. Then, Gemini checks for errors such as incorrect formatting, payment issues, among others. If no errors are found, it generates an appropriate response from the context provided by the user so the Agent can respond and maintain the flow of the conversation in the form of text, and sends it back to Azure so it can be transformed into audio so the agent can transmit the message back to the user (Figure 4.18).

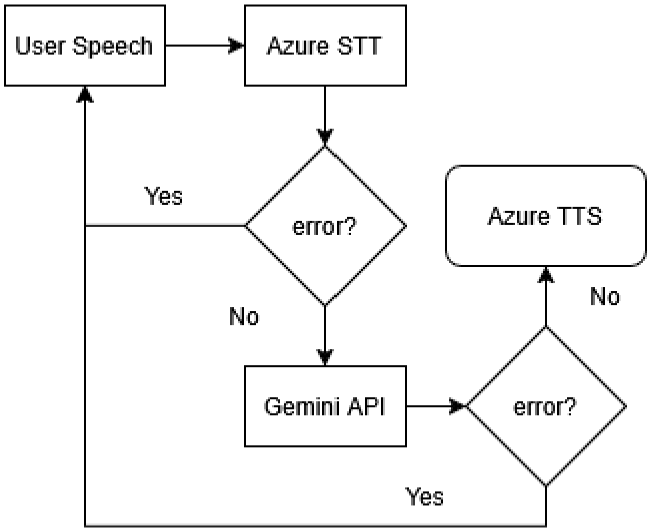


Figure 4.18 - Voice Agent Diagram from listening to the user to responding back

VR Voice Agent also deals with sending the information to the VR Movement Agent (Figure 4.21) so the responses and body movement of the agent match. But, before that, another step was needed, the recording of movements from the user (in this case, the research team) using the Movement Recorder.

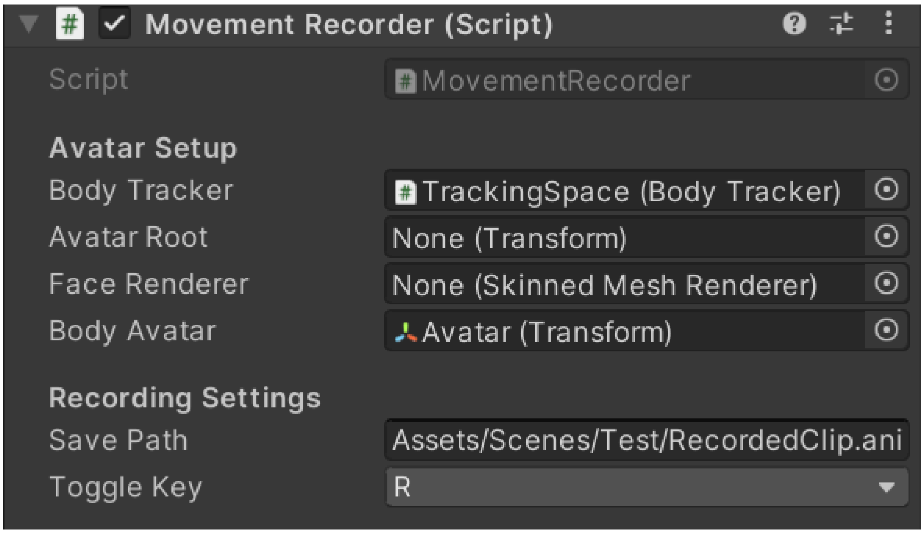


Figure 4.19 - Movement Recorder recording settings and double avatar setup

Movement Recorder grabs two avatars that are recording the movement of the user's avatar and, when pressing R, records all the movements (both facial and corporal) until the key is pressed again. Two avatars are being used due to a bizarre phenomenon caused by the Ready Player Me (Figure 4.5) compatibility framework with Meta Movement, which did not have a compatible connection between the Spine and Hips (Figure 4.16) when loading the avatar, creating a deformed lower body (Figure 4.20).

Figure 4.20 - Same animation when recorded using 2 avatars (left) vs 1 (right)

After testing every single recorded animation with the avatar in Figure 4.20, all the animations were then added to the previously mentioned VR Movement Agent seen in Figure 4.21. There are three important factors in this code:

1. There are distinct animations for when the avatar is speaking (e.g., hand movements (gestures), lip sync, trying to look at the user) compared to idle ones (e.g., low hand movement, no lip movement, looking away for a second or two and looking back at the user).

2. The animations are separated by emotions. For example, sad (triste) emotions will show the agent's avatar looking down more and smaller hand movements, while more active feelings like anger (irritado) show the agent moving her body more expressively.

3. To avoid the user's discomfort, VR Movement Agent component and VR Voice Agent component will coordinate the length of the recorded movements and the length of the audio to create a sense of realtime conversation using multiple types of movement instead of repeating the same movement.

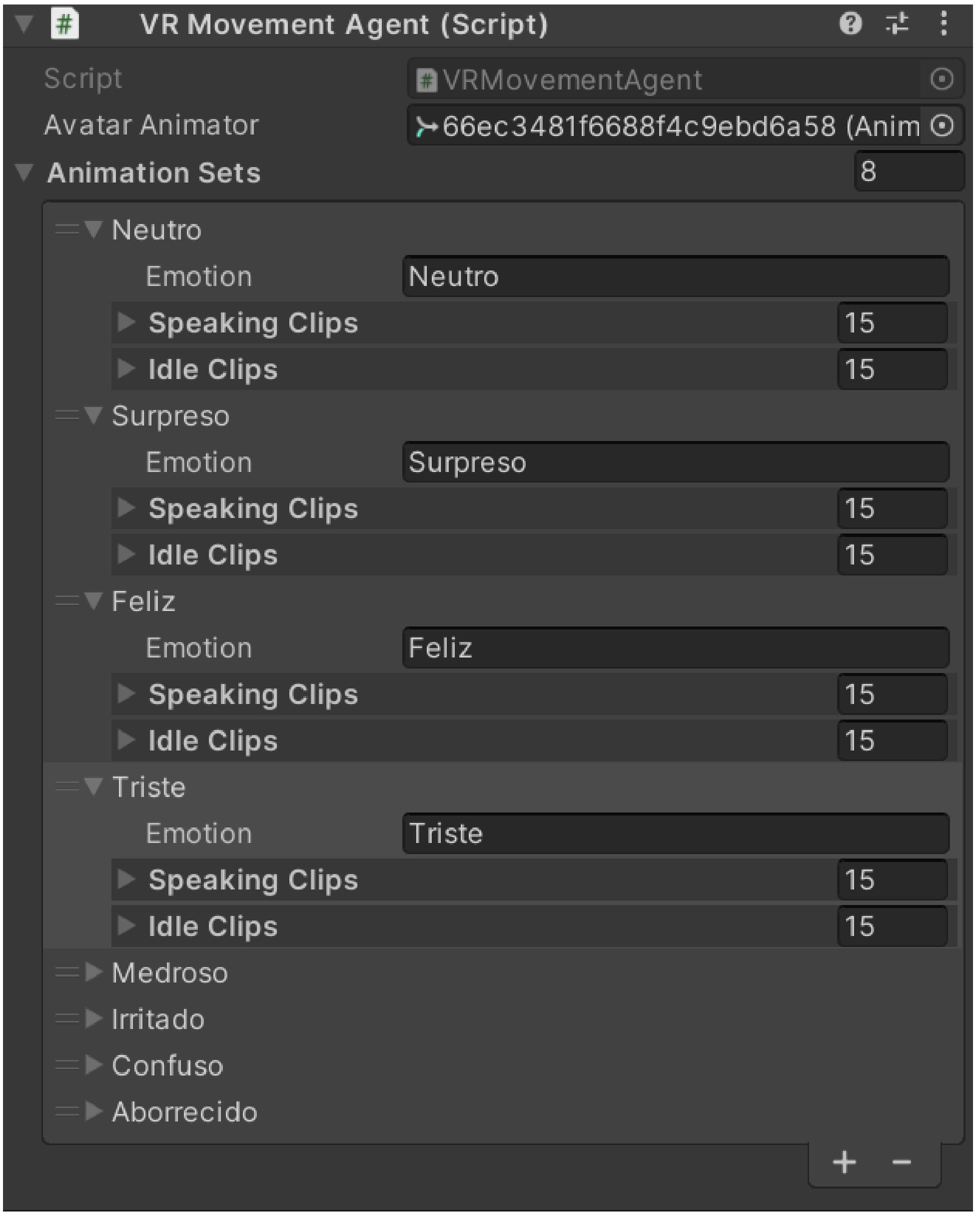


Figure 4.21 - List of emotions that the agent can express back to the user

In Figure 4.22, we can see a sample of the agent interacting with the user. On the left, the agent is being more expressive and trying to maintain eye contact with the user, while on the right the agent is being more reserved, not always trying to maintain eye contact with the user.

Figure 4.22 - Agent interactions with the user with active body movement (left) compared to more reserved movement (right)

4.3 Agent Responses

When interacting with a language model that was trained to be generalist and act friendly, it is necessary to verify how it replied and adapt to it. After making the first calls to the Gemini API, it was observed that everything parsed from its text to Azure to be voiced: the text, emoticons, lists, special characters, among others. To make sure no special characters are voiced by the avatar, 2 simple methods are implemented. First, the RemoveEmojis() method, which removes emojis by changing all Unicode categories Surrogate (\p{Cs}) symbols or other symbols (\p{So}) into (‘’). The second method, StripMarkdown(), converts the list markers (*, -, +) and numbered lists (1., 2.) into multiline mode, which changes special symbols into lines instead of input strings [28], removing bold and italics from the text and changing backticks, headers and blockquotes (@"[`#>]") to (‘’);

Then, specific behaviors were defined for the agent to introduce itself, and bid farewell. For greetings, the GetGreetingByTime() method was implemented. This method evaluates the time of the day and responds accordingly: ‘Bom dia’ is used between six o’clock and twelve o’clock, ‘Boa tarde’ between twelve o’clock and eight o’clock in the afternoon, and ‘Boa noite’ for the remaining hours of the day. For goodbye detection, the IsGoodbye() method analyzes the user’s input text and checks for common goodbye expressions, including ‘adeus’, ‘até logo’, ‘tchau’ and their variations.

To make sure the agent used the proper prompts, the ResetConversation() method creates a new system prompt every time it is called that prevails until it is called again, and clears the previous conversation history. Due to the way Gemini API operates, this history is stored in a List that contains the six latest messages between the agent and the user. In other words, it stores the three most recent messages from each participant to maintain conversational context while it avoids processing the entire conversation.

For the system prompt, 6 phrases were given that the agent (Gemini) should follow at all times:

1.	Respond in a calm way, responding always with empathic responses.
2.	Avoid language excessively enthusiastic or informal, responding with 30 words or less.
3.	Never refer to physical expressions, appearance, body or face.
4.	Always use natural and fluid language, like in a real conversation using always Portuguese from Portugal.
5.	Never use phrases with () and avoid numeric knowledge.
6.	Never end the conversation first and never repeat or use greetings like 'Bom dia' or 'Olá' in any response.

Table 1 - Rules the agent needs to follow while interacting with the user

While testing the agent, Gemini was notorious for not always replying in the most respectful way. Therefore, the first two rules were implemented. The first rule aims to prevent the agent from offering solutions or making statements that could be misinterpreted or harmful. The second one ensures that it responds in a more formal manner, avoiding overly pushy or excessively friendly behavior. The third rule aims to prevent judgmental remarks toward the user.

The fourth and fifth rules aim to guarantee the agent behaved and felt more 'human.' As indicated previously, Gemini tends to make markers or numbered lists, something that humans rarely do in a conversation. The fourth rule was not enough to prevent text generation from including explanations (as it often uses parentheses) or building lists of items when communicating back to the user.

The sixth and final rule was created because the agent's greeting and the main conversation are handled through different prompts. Since the agent first introduces itself and calls ResetConversation(), it is not aware that it has already greeted the user. Therefore, any greeting from the user could cause the agent to greet back a second time. All the rules suffered many alterations until they reached this final wording, especially due to the inclusion of strong words to make sure Gemini followed them.

Strong words such as 'Never' or 'Always' are needed to ensure Gemini does not have any ambiguity about what it can and cannot do. This is especially important when users try to force a conversation that could contradict the established rules. For example, if the fifth rule had been phrased ‘Do not use phrases with () and avoid numeric knowledge.’, and the user asked for a list of ingredients for a recipe, the agent could disregard this rule and respond ‘1. 2 eggs, 2. milk’ rather than in a more natural form, such as ‘you will need 2 eggs, milk,’.

4.4 System Refinement Testing

Prior to user testing the agent, five minor tests were conducted to evaluate its usability and identify potential issues. This allowed for the implementation of necessary corrections and enhancements before the usability study.

4.4.1 First Iteration

The first test, done with one participant knowledgeable in psychology, followed the same procedures (briefing the participant on the task, have any question clarified, signing the Informed consent and pre-experiment questionnaire) done to test the agent’s empathy. However, some major flaws that would lead to a perceived lack of empathy from the agent's part were identified midtest:

- The agent's excessive talk resulted in the users unawareness of when the agent concluded their response. This prompted the user to feel the agent was lecturing them rather than having a discussion.
- The agent’s gaze was not targeting the user’s eyes or face. Instead, a downward gaze made the agent feel little communicative, which made the user uncomfortable.
- Many of the animations showed the avatar moving its hands through its legs, resulting in feelings of oddness in the user.
- The agent was not detecting pauses made by the user, only reacting partially to what the user fully said.
- The agent spoke Brazilian Portuguese (words). While not a flaw per se, since the agent was created with Portuguese from Portugal in mind (voice), is envisioned to speak to Portuguese users in Portugal, using a different lexicon lowered the agent’s likeness to them.

Resolving these issues was essential, since they hindered the conversation flow and dynamics. Several changes were implemented accordingly:

- Reduce the number of words that Gemini produced to thirty or fewer; long enough to respond conversationally, but not long enough to feel like a lecture.

- To ensure that the agent maintained eye contact, and to avoid any pass-through issues, all animations were replaced.
- For the agent to detect fully what the user said, a pause detection was implemented.
- Ensured the agent speaks Portuguese from Portugal by changing the prompt, as seen in rule number 4.

4.4.2 Second Iteration

Once all the aforementioned issues were corrected, a new test conducted with two participants was performed following the same procedures. As a result of previous adjustments, the pause detection implementation caused the agent to interrupt the user. As shown in Figure 4.23, the agent is set to wait a specific amount of time (two and a half seconds) to check whether the user was speaking and then reset. During the test, the agent waited correctly the first time, but instead of resetting the wait cycle, it interpreted the second pause as one. As a result, the agent interrupted the user when the user paused again and proceeded to speak prematurely.

Additionally, the fact that the agent does not mention having a name reduces the user's connection with the agent.

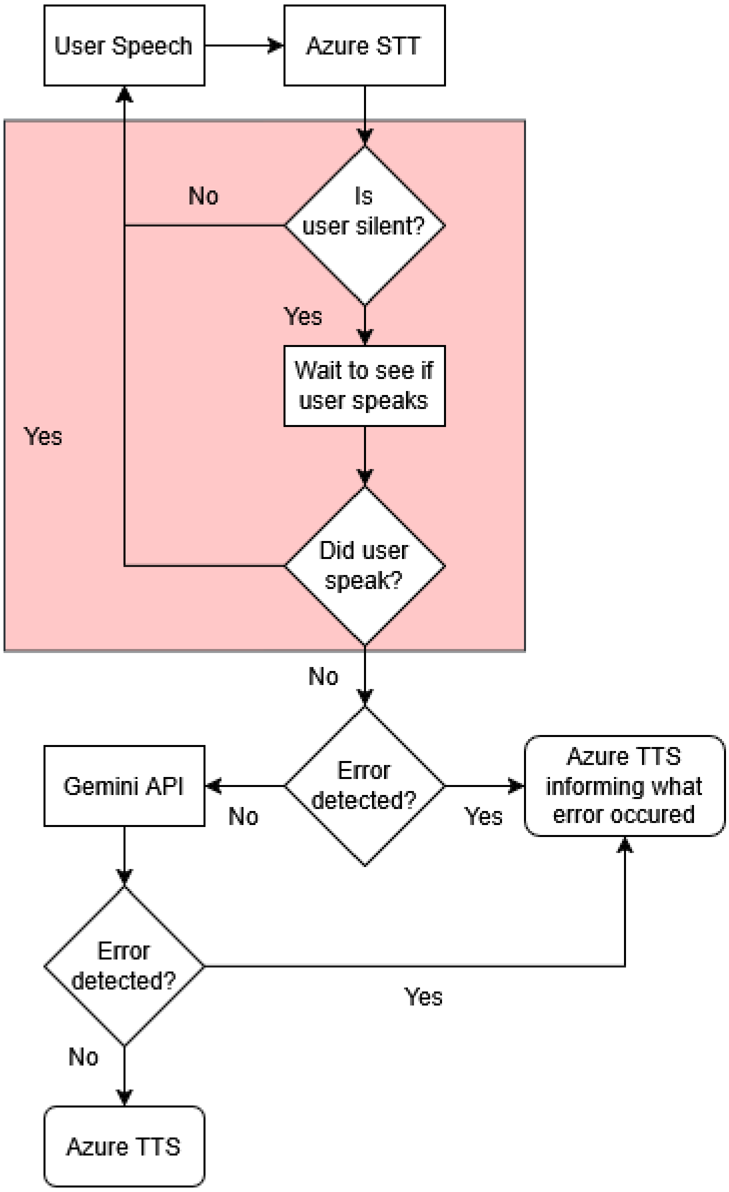


Figure 4.23 - New Voice Agent Diagram indicating the interruption issue location.

4.4.3 Third Iteration

For this study, done with the previous participants, the issue found in the previous test was promptly corrected. Previously, the user-silence check operated with a timer that started anytime the user stopped talking. While functional, it did not verify small utterances. Since the agent only considers two or more words as further speech after the first pause, fillers such as 'hm' or 'uh' are ignored. To avoid this problem, a boolean check verifying any spoken word was implemented alongside the timer. Furthermore, it was also considered that two and a half seconds of stillness could be insufficient. Therefore, the agent's silence interval was increased to five seconds.

No significant issues were identified after user testing, with this agent receiving the user's approval. However, the user mentioned that the agent's response time was low, which led to the fourth and fifth tests.

4.4.4 Fourth and Fifth Iterations

These tests, done with another tester, aimed to verify two features:

- The extent to which the agent's wait time can be reduced without interrupting the user.
- Refinement of the proposed emotion evaluation system.

The emotion system was thoroughly tested by the researcher prior to the fourth test, with a few changes implemented between the first and second test. Before the changes, all the blendshapes captured using the head-mounted display,translated to the OVR Face Expressions (Figure 4.8), and lastly converted into Ready Player Me blendshapes, were calculated separately by their weight in the emotion it was designated and added its value to it when detected. When the highest emotion (above five points) was detected, that emotion was validated. However, throughout development, every emotion was evaluated and given a score with non-regulated values. Further, its first implementation did not allow multiple inclusion (adding two or more Facial Expressions).

To rectify this, three major changes were made. First, the values for all Facial Expressions were regulated through a boolean method called CheckLevel(). This method returns a level switch after checking each Facial Expression based on four expression levels:

- Present. Any Facial Expression with this rule that registers an Expression Level value above 0.001f adds the designated score of the rule to the total emotion score.
- Low. Any Facial Expression with this rule that registers an Expression Level value between 0.01f and 0.40f adds the designated score of the rule to the total emotion score.
- Medium. Any Facial Expression with this rule that registers an Expression Level value between 0.41f and 0.80f adds the designated score of the rule to the total emotion score.
- High. Any Facial Expression with this rule that registers an Expression Level value above 0.81f adds the designated score of the rule to the total emotion score.
- When the Expression Level of the Facial Expression rule is under or above the designated level it returns false. For example, if a Facial Expression has the Low and Medium rules, it will return false if it registers a value below 0.01f or a value above 0.80f.

A new Struct Rule including 5 items on a list<> of facial expressions was created. This rule evaluates conditions that apply to one or multiple facial expressions, defines the required level for an expression, its corresponding emotion, the score it contributes to the final count, and whether it was an 'and' rule (multiple facial expressions required to meet the level), or an 'or' rule (only one emotion is required to meet criteria).

All rules created by the Struct are added to a read-only list, with separate lists for face and eye. When two rules containing the same expression, rule type and emotion have different expression levels, only the highest one is valid. For example:

- Face Expressions: NoseWrinklerL, NoseWrinklerR
- Expression Level: Low, Medium
- Emotion: 'Irritado' (Angry)
- Score: 2.5f (Low), 3.0f (Medium)
- Rule Type: 'Or' (only needs one of the facial expressions to match the expression level)

If the NoseWrinklerL was detected as Low and NoseWrinklerR was detected as Medium only the value of the NoseWrinklerR would be scored because it has the highest score between the two (3.0f for Medium compared to 2.5f for Low). Boredom was removed from the list of emotions for sharing too many facial expressions with the other emotions. The final set of emotions the agent is able to map is: surprised, angry, happy, sad, fearful and confused. Finally, 15 points is the maximum score of an emotion, being 5 the minimum required for an accurate detection of the user's emotion.

Regarding the agent's response time, upward scale testing was done. This test started with the user interacting with an agent set to respond after two and a half seconds of silence. Whenever the user was interrupted, the silence time threshold increased by quarter of a second. Each test took approximately 10 minutes each. The average silence time between the user pausing, thinking, and resuming speech was three seconds and seven hundred and fifty milliseconds. These results were implemented in the final logic, as seen in Figure 4.24.

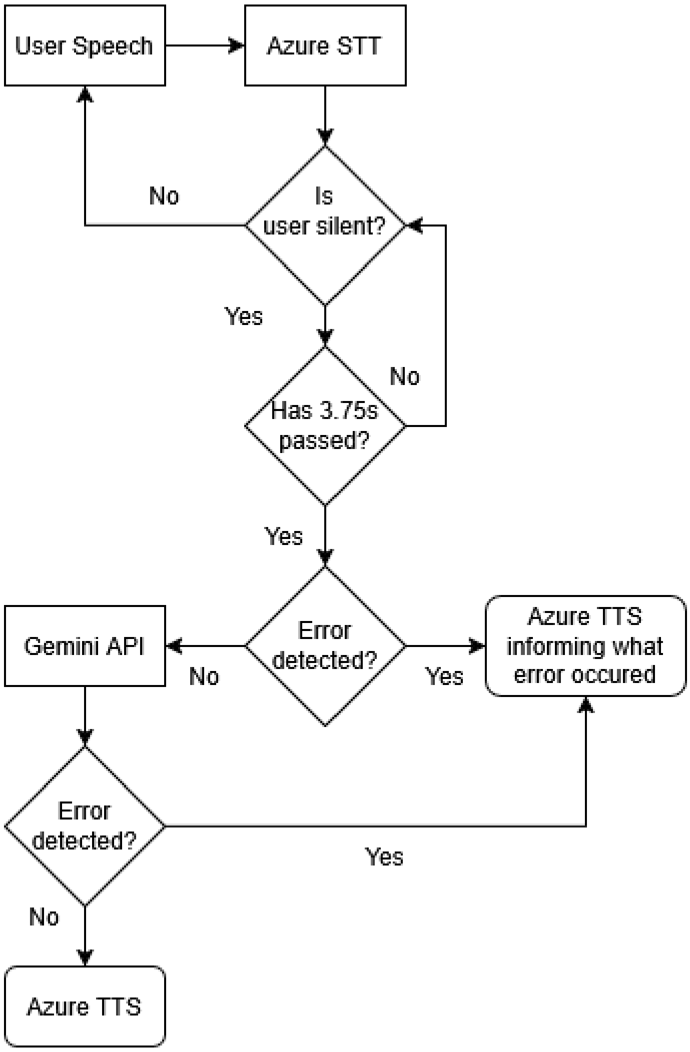


Figure 4.24 - Final Agent Diagram

(This page intentionally left blank.)

5 Pilot Study

5.1 Sample and Recruitment

The participants were recruited via posters, email, and word-of-mouth. In this study, adults (eighteen years old or above) with proficiency in Portuguese were eligible to participate. Participants were excluded if they had self-reported medical conditions or visual limitations that could affect the execution of the task or the use of the VR headset. All participants were informed of possible side-effects of headset usage, such as nausea, dizziness and ocular fatigue, and reminded that they could interrupt the session at any time. The recruitment poster (Annex 1), was placed in the information boards of the University of Madeira (UMa) and emailed to UMa and the Regional Agency for Development of Investigation, Technology and Innovation (ARDITI) for dissemination. All participants were encouraged to extend the invitation to anyone that could be interested in participating. All sessions were scheduled through the scheduling platform Calendly (https://calendly.com). This study has been approved by the Data Protection Committee (10993715) and the Ethics Committee (10084399) of the University of Madeira, and all participants signed an Informed Consent.

5.2 Instruments

For the study, two different instruments were used. The first one was a set of two digital questionnaires: Pre Questionnaire and Post Questionnaire. These were responded to before and after the interaction with the agent, respectively. The Pre Questionnaire consisted of eight questions designed by the researcher to collect demographic information, including age, gender, digital technologies knowledge, experience with virtual reality, among other factors. These questions were asked to characterize the sample and were not used to reach conclusions in the study. The Post-Questionnaire consists of three sections.

- The ITC-SOPI: it comprises the 2004 version of the Massive Virtual Reality Laboratory ITC Sense of Presence Inventory (ITC-SOPI) [29]. It consists of thirty-five questions, separated in two parts (A and B). Part A includes two questions assessing how the participants felt after the study, and part B consists of thirty-three questions on their overall experience during the study. This questionnaire uses a 5-Point Likert Scale to measure agreement with each statement, ranging from 1 (‘Strongly Disagree’) to 5 (‘Strongly Agree’). The questions are grouped into four factors: spatial presence (18 items), engagement (6 items), ecological validity (5 items), and negative effects (6 items).

- Quality and Likability: It consists of nine items (five multiple-choice and four open questions) that address how the users felt about the agent. Just like the ITC-SOPI, the multiple-choice items follow the Likert Scale, scaled from 1 ('Strongly Disagree') to 5 ('Strongly Agree'). The open questions are complementary for clarification whenever the participant disagrees with a statement. All open questions were optional.
- SUS-Based: It was adapted directly from the System Usability Scale (SUS) [30]. It consists of eight questions, and it uses Likert Scale (ranging from 1 'Strongly Disagree' to 5 'Strongly Agree') to assess usability and user-friendliness. Even though the original scale consists of 10 items, two questions were omitted in this adaptation:
    - 'I think that I would need the support of a technical person to be able to use this system.': it was omitted because the interaction with the agent will be supervised by a psychologist or researchers, in accordance with prior studies [1][11] and questions the participants might have will be clarified before, during and after use and any required assistance will be given, i.e help with putting the head mounted display (HMD), starting and restarting the system and other technical requirements.
    - 'I would imagine that most people would learn to use this system very quickly': it was omitted as the interaction relies on innate social intuition rather than the mastery of a complex technical interface. We assumed 'learning to use the system' is less about mastering an interface and more about applying the user's existing social intuition to a personal experience.

The final instrument is a per-session file saved into the hidden files of the headset, only accessible to the researcher. This file contains all messages sent by the agent to the participant, along with the agent's inferred emotion of the participant. This data was used to evaluate whether the agent's responses were appropriate for the detected emotions and whether any inappropriate response was generated. To protect the participant's privacy, no messages spoken by the participants were recorded.

5.3 Procedure

Upon signing the Informed Consent, the participants were briefed on their task and any enquiries they had were addressed. First, they filled in the pre-experiment questionnaire. Subsequently, participants were instructed to discuss any topic they would be willing to share in the presence of the researcher, but also advised not to discuss sensitive topics. The participants were reminded of their right to end the experiment at any moment. Then, they put on the HMD to interact with the agent, with the interaction being no longer than fifteen minutes before the agent bid farewell. Upon completion, all participants completed the post-experiment questionnaire, and were encouraged to inquire about or comment on any aspect of the study, such as the agent's behavior or the objective of the study. A participant reported feeling unwell after a few minutes interacting with the agent, which led to his/her participation being discontinued.

5.4 Data Analysis

All the data from the questionnaires in the study were saved directly from the Google Forms to Excel. Due to Excel saving the responses from the 5-Point Likert Scale into its named values (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree), a conversion to its numeral values was made.

To ensure that the Quality and Likability and the SUS-based sections correlate and can be administered in a real-world setting to evaluate the agent, a Cronbach's Alpha analysis was performed.

The Quality and Likability section contains five multiple-choice questions ($k = 5$) and four long-answer questions, and the SUS-Based section contains eight questions ($k = 8$). Because of sample variation, it was necessary to calculate the average ($\bar{x}$) of each question (Y_i) and the average score of each user test (X) in order to determine the deviation of each participant's response by subtracting the individual score (X_i) from the average ($X_i - \bar{x}$) and squaring the result $(X_i - \bar{x})^2$. The values of the squared deviations were then summed and divided by the number of participants minus one to obtain the sample variance; that is, for each part, $\sigma^2_{\gamma i} = (\sum(x_i - \bar{x})^2)/(N \text{ participants} - 1)$. The observed score variance was calculated using the total test score X for each participant, obtained by summing the scores across all k parts, as $\sigma^2_x = (\sum(X - \bar{X})^2)/(N \text{ participants} - 1)$. It is to note that, for negative questions, such as 'I found the system unnecessarily complex.', the values were reverse-coded using 6-n, i.e if the participant's response was 1n, it was converted to 5 in the Cronbach's Alpha calculation (6-1).

The Cronbach's Alpha coefficients were calculated for each question set. The Second set of questions (researcher's questions) resulted in $\alpha = 0.809$, and $\alpha = 0.755$ for the third set of questions (adapted SUS questions) (table). As seen in Arof et al. [31], both sets are at an acceptable level of reliability or above.

No	Coefficient of Cronbach’s Alpha	Reliability
1	More than 0.90	Excellent
2	0.80-0.89	Good
3	0.70-0.79	Acceptable
4	0.6-0.69	Questionable
5	0.5-0.59	Poor
6	Less than 0.59	Unacceptable

Table 2 - Reliability of Coefficient of Cronbach’s Alpha [31]

Besides the necessary values for the Cronbach's alpha, the Standard Deviation (SD) was calculated for all questions, alongside the Interquartile Ranges and the medians for each question. Also, to have a bigger understanding of where the data would follow on a bigger population, the margin of error was calculated in a ninety-five percentile confidence.

5.5 Results

5.5.1 Demographic Data

The study comprised fifteen participants, of whom sixty percent (n=9) were male and forty percent (n=6) were female. Most were aged between eighteen and twenty-five (n=12), with the last three participants between 26 and 33 (n=2), or older than 50 (n=1). Most participants (n = 13) were students, one was an employee, and one held an 'other' status.

Regarding their interaction with technologies, responses varied. Numerous participants rated their proficiency in digital technologies as rather high (n=11). However, most reported having minimal to no experience, with very few reporting considerable (n=2) and moderate (n=3) experience. Notably, when inquired about their engagement with other interactive technologies, such as simulators, interactive programs, or three-dimensional environments, eight indicated occasional usage, four stated they had utilized them only once and three participants reported frequent usage.

Another noteworthy aspect is the time dedicated to playing video games. While this question does not distinguish between games with NPCs and those without, numerous participants played videogames frequently, either once or many times a week or every single day (n=10), some occasionally several times a month (n=3) and the other individuals rarely engaged in videogames (n=2).

5.5.2 Post-Questionnaire

5.5.2.1 ITC-SOPI

As previously mentioned, the ITC-SOPI is divided into four factors: spatial presence, engagement, ecological validity, and negative effects.

According to the set of questions below, we can separate Spatial Presence items into four subsets:

- Self-Location, the feeling of being in the space (B7, B15, B21, B29)
- Interaction and agency, the ability to influence the environment (B3, B10, B11, B16, B28, B30, B1, B33)
- Sensory depth and realism, physiological sensations and 'tangibility' of the experience (B6, B19, B24, B25, B27)
- Social Presence, the presence of the agent (B20)

Nº	Question	Median (IQR)	Mean	Mode
B6	I felt that the characters and/or objects could almost touch me.	3(1.5)	2.7±0.6	3
B19	I could almost smell the different characteristics of the environment on display.	1(1)	1.6±0.8	1
B24	I felt I could reach out and touch things (in the displayed environment).	3(1.5)	3.1±0.4	4
B25	I felt that the temperature was altered to match the setting of the displayed environment.	2(2)	2.2±0.5	3
B27	I felt that all my senses were stimulated at the same time.	2(1.5)	2.5±0.5	2

Table 3 - Spatial Presence: Sensory depth and realism

Among the four subgroups of Spatial Presence, Sensory Depth and Realism received the lowest scores, with item B24 being the highest scoring item (M=3.1, SD=0.8) receiving below half the positive agreement (n=6). The two lowest of the group go to item B25 (M=2.2, SD=1.0) with mostly 'Strongly Disagree' (n=5) and a single 'Agree' response (n=1) and B19 being the item with the lowest score (M=1.6, SD=0.7) and receiving a high amount of 'Strongly Disagree' (n=8) and no agreement (n=0) responses. (table 3).

Nº	Question	Median (IQR)	Mean	Mode
B7	I felt like I was visiting the places in the displayed environment.	3(1)	3.3±0.4	3,4
B15	I felt like I was in the scenes being shown.	4(1)	3.5±0.2	4
B21	I felt surrounded by the displayed environment.	4(1)	3.7±0.4	4
B29	I felt as if I were in the same space as the characters and/or objects.	4(0)	3.9±0.3	4

Table 4 - Spatial Presence: Self-Location

On the other hand, Self-Location ended up with no negative Means, its highest being B29 resulted in a rather high Mean and low Standard Deviation (M=3.9, SD=0.6), with mostly 'Agree' (n=13) scores, and the remaining being 'Disagree' (n=1) and 'Strongly Agree' (n=1). Its lowest being item B7 (M=3.3, SD=0.8), scoring higher than the highest of the previously mentioned Sensory Depth and Realism B24, and receiving a lone negative feedback (n=1) and half the positive feedback (n=7) (table 4).

Nº	Question	Median (IQR)	Mean	Mode
B3	I felt that I was able to interact with the displayed environment.	4(1.5)	3.3±0.6	4
B10	I felt that I wasn't just observing something.	4(1)	3.4±0.3	4
B11	I had the feeling that I moved in response to parts of the displayed environment.	3(2)	2.7±0.7	3
B16	I felt that I could move objects (in the displayed environment).	3(2)	2.9±0.5	4
B28	I felt capable of altering the course of events in the displayed environment.	3(1)	3.3±0.6	3
B30	I had the feeling that parts of the displayed environment interacted with me.	3(2)	2.9±0.5	3,4
B31	The experience of moving things in the displayed environment was realistic.	3(1)	2.6±0.6	3
B33	I felt as if I were participating in the environment being displayed.	4(0.5)	3.6±0.5	4

Table 5 - Spatial Presence: Interaction and agency

For Interaction and agency, the highest scoring item was B33 (M=3.6, SD=1.0). This was followed by B3 and B28, which shared the same Mean (M=3.3, SD=1.2). The remaining items all fell below the Mean of 3: B16 (M=2.9, SD=1.0), B30 (M=2.9, SD=1.1), B11 (M=2.7, SD=1.3), and B31, which had the lowest Mean (M=2.6, SD=1.1) (table 5).

Nº	Question	Median (IQR)	Mean	Mode
B20	I had the feeling that the characters were aware of my presence.	4(1)	3.5±0.6	3,4

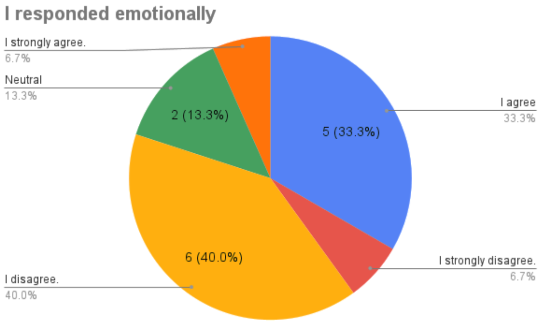
Table 6 - Spatial presence: Social Presence

Finally, Social Presence registered a rather high score among all questions, being the fourth highest scored mean (M=3.5) tied up with B15, but registering a higher Standard Deviation at SD=1.0, compared to B15's SD=0.8.

Nº	Question	Median (IQR)	Mean	Mode
A2	I felt like I had returned from a journey.	2(2)	2.8±0.7	2
B1	I felt immersed (in the environment displayed).	4(0)	4.0±0.2	4
B2	I lost track of time.	4(1.5)	3.6±0.6	4
B5	I felt that the content was true and genuine.	3(2)	2.8±0.7	4
B14	My experience was intense.	3(1)	2.6±0.6	3
B26	I responded emotionally.	3(2)	2.9±0.6	2

Table 7 - Engagement

When it comes to Engagement, it was almost unanimously positive (M=4.0, SD=0.4), with the majority (n=13) agreeing with the statement, and the remaining neutral (n=1) and agreeing strongly (n=1). On a similar note, B2 showed positive scores (M=3.6, SD=1.1), with the majority (n=10) agreeing with the statement, 1 scoring neutrally, and the remaining participants disagreeing (n=4). Despite A2 and B26 receiving the highest number of 'Disagree' (n=5 and n=6 respectively), B14 showed the lowest Mean (M=2.6, SD=1.2). The latter received the lowest amount of positive feedback (n=3) with only two 'Agree' and a 'Strongly Agree' responses. Similarly, B5 shows a lower Mean than B26, and the same Mean as A2 due to polarizing agreement (n=6) and disagreement (n=7), and few participants being neutral to the statement (n=2).



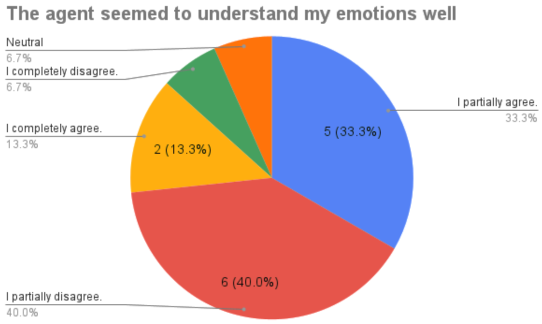


Figure 5.1 - Comparison between B26 and C3 questions

Nº	Question	Median (IQR)	Mean	Mode
B4	The environment displayed/presented seemed natural to me.	4(1)	3.3±0.4	4
B9	The content seemed credible to me.	3(2)	3.1±0.6	4
B13	I felt that the environment displayed was part of the real world.	2(1)	2.5±0.5	2
B17	The scenes depicted could truly occur in the real world.	4(2)	3.9±0.6	5
B23	I had a strong perception that the characters and objects were solid.	4(2)	3.1±0.5	4

Table 8 - Ecologic Validation and Naturality

In the Ecological Validity factor, B9 and B23 have the same Mean (M=3.1) but different Standard Deviation (SD=1.1 and SD=1.0, respectively) due to the same amount of participants disagreeing with the statement (n=5) but a different number agreeing with it: most (n=6) agreeing with item B9 against one strongly disagreeing; and a majority (n=8) also agreeing with item B23.

In this group, B17 shows the highest approval (M=3.9, SD=1.1), with most participants responding ‘Strongly Agree’ (n=6), and the remaining ‘Agree’ (n=3) and ‘Disagree’ (n=2) compared to the and . On the other hand B13 showed a majority in disagreement (n=8) with only half (n=4) agreeing with the statement, being the lowest scoring item of the Ecologic Validity group (M=2.5, SD=1.0).

Nº	Question	Median (IQR)	Mean	Mode
A1	I felt disoriented after the experience.	2(1.5)	2.0±0.5	2
B8	I felt tired.	1(1)	1.5±0.4	1
B12	I felt disoriented during the experience.	1(1.5)	1.7±0.5	1
B18	I felt eye fatigue.	1(0.5)	1.4±0.4	1
B22	I felt nauseous.	1(0)	1.3±0.3	1
B32	I felt like I had a headache.	1(0)	1.1±0.1	1

Table 9 - Negative Effects

Regarding Negative Effects, no items received a positive Mean, with the highest score being A1 (M=2.0, SD=0.9), sharing a similar number of agreement (n=1) with B18. Among the negative effects described in the questionnaire, headache received the highest amount of 'Strongly Disagree' (n=14), followed by nausea (n=12), eye fatigue (n=11) and tiredness (n=10). On the other hand, few felt neutral about being disoriented after the experience (n=3) or during it (n=4).

5.5.2.2 Agent's Quality and Likability Questionnaire

Nº	Question	Median (IQR)	Mode
C1	I felt comfortable interacting with the agent.	4(2)	2,4,5
C2	I felt that the agent's interactions were natural.	3(2)	4
C3	The agent seemed to understand my emotions well.	3(2)	2
C4	I thought the agent's emotional expressions were realistic.	3(2)	4
C5	The agent responded realistically during the conversation.	4(2)	4
C1.1	If you felt uncomfortable with the agent, please indicate why		
C2.1	If you felt that the interactions with the agent were not natural, please indicate why		
C6	If you could change something about the agent, what would it be? (It can be more than one thing)		
C7	Do you think empathic agents, like the one used in this study, are useful in contexts such as education and/or healthcare?		

Table 10 - Agent likeability

Regarding the agent's Likability, the responses had Means from neutral to good, as illustrated in Figure 5.2. C2 and C4 share the lowest Mean at M=2.9, with different Standard Deviation

(SD=1.1 and SD=1.2, respectively). Single differences between C2 and C4 were observed between the two, C2 received 3 'Disagree', 4 'Neutral' and 6 'Agree' and C4 receiving 4 'Disagree', 3 'Neutral' and 5 'Agree' while also getting a single 'Strongly Agree' (n=1). C3 was the biggest polarizing item, sharing the same amount of positive and negative feedback (n=7), only getting a single neutral response (n=1). C3's Mean (M=3.1, SD=1.3) is slightly positive due to receiving two 'Strongly Agree' responses compared to a single 'Strongly Disagree'.

Both C1 and C5 received the same amount of 'Strongly Agree' responses (n=4). While C5 received more 'Agree' responses (n=5) than C1 (n=4), it shows the same score (M=3.5) but different Standard Deviation (SD=1.4) than C1 (SD=1.2). This might result from C2 showing 'Strongly Disagree' responses (n=2), while C1 received none. C1 is the only question to share the same amount of responses (n=4) in three different levels, 'Disagree', 'Agree' and 'Strongly Agree.'

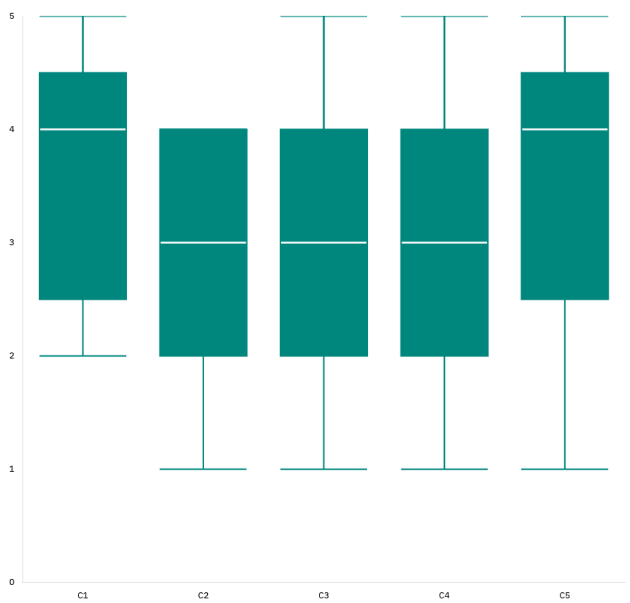


Figure 5.2 - Agent Likeability box and whisker chart

Regarding the open questions, when asked about the reasons why they felt uncomfortable with the agent (item C1.1), the participants responded that:

- The agent was artificial, not part of the participant's world.
- The static eye movement made them feel that the agent was a generic robot or something inanimate.
- The agent did not respond very well, answering things the participants did not ask or circling around a topic.

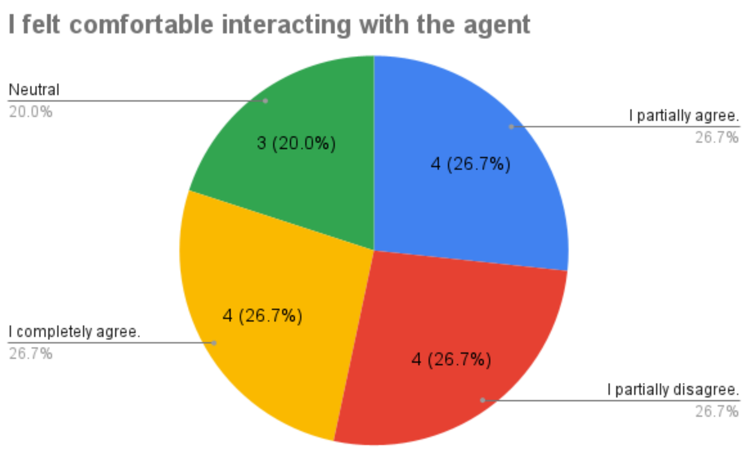


Figure 5.3 - Comfortability when interacting with the agent.

As for why the interactions did not feel natural (item C2.1), two reasons connected to the agent speech were given.

- Monotonous voice.
- The agent sometimes did not understand what the user said.

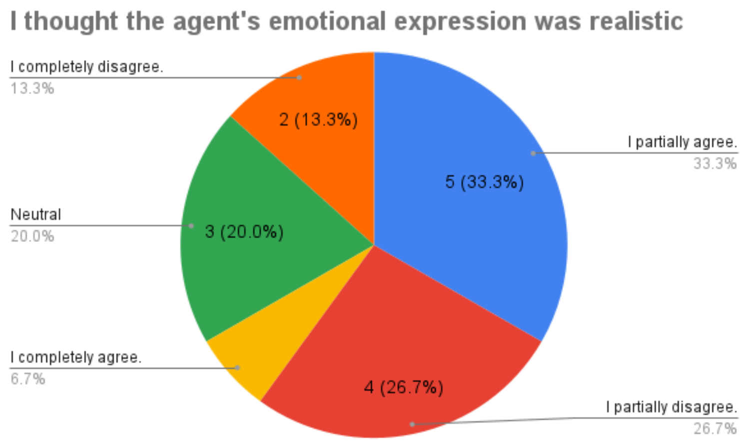


Figure 5.4 - Realism of agent emotional expressions responses

When it came to changes in the agent (C6), the requests were:

- Improve movement, especially the eyes.
- Improve the voice to no longer sound monotone.
- Introduce small talk so the conversation is not one sided.
- Different avatars.

When asked if agents like the one used in the study were useful in education or healthcare (C7), it was unanimous (n=14) that these types of agents are useful. However, one participant refrained from responding. Amidst the responses, a few (n=5) only answered ‘Yes’, while others (n=3) mentioned that real interactions should be followed alongside the use of virtual agents.

5.5.2.3 Usability

Nº	Question	Median (IQR)	Mode
D1	I think that I would like to use this system frequently.	3(1.5)	4
D2	I found the system unnecessarily complex.	1(1)	1
D3	I thought the system was easy to use.	5(1)	5
D4	I found the various functions in this system were well integrated.	4(2)	4,5
D5	I thought there was too much inconsistency in this system.	2(0.5)	2
D6	I found the system very cumbersome to use.	1(0)	1
D7	I felt very confident using the system.	4(2)	4
D8	I needed to learn a lot of things before I could get going with this system.	1(0)	1

Table 11 - Ecologic Validation and Naturality

For the SUS-based section, there are two sets of questions:

- Negative phrased questions: D2, D5, D6, D8
- Positive phrased questions: D1, D3, D4, D7

Among the positively phrased questions, ease of use (D3) received the highest score (M=4.7, SD=0.5) with a majority responding ‘Strongly Agree’ (n=10) and a few ‘Agree’ (n=5) It was followed by D4 and D7, both registering negative votes (n=1 and n=2 ‘Disagree’, respectively). D4 also got a few neutral responses (n=4) but got a higher mean (M=3.9, SD=1.0) than D7 (M=3.7, SD=1.3) because it received less negative responses (n=1) compared to D7 (n=4). The item on whether they would like to use the system (D1) got the lowest Mean (M=3.2, SD=1.1).

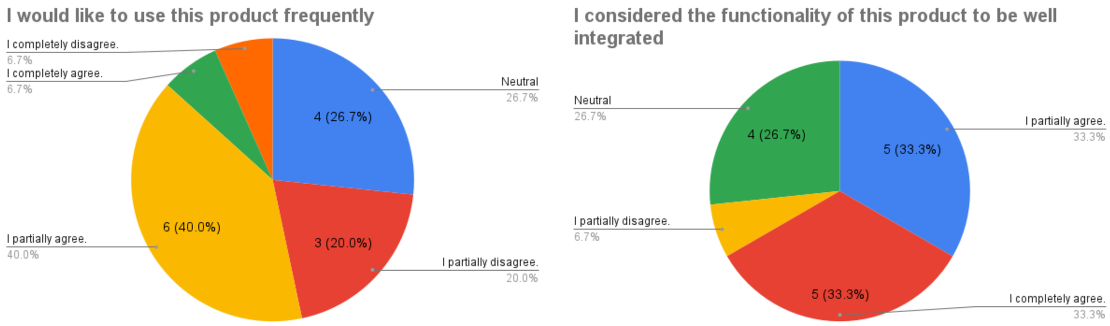


Figure 5.5 - Similarity between user willingness and integration

Among the negatively phrased items, D8 and D6 scored similarly, with D8 (M=1.1, SD=0.4) receiving one less ‘Disagree’ (n=2) than D6 (M=1.2, SD=0.4) and a majority of ‘Strongly Disagree’ (n=13 and n=12, respectively). Similarly, D2 had the same number of ‘Disagree’ responses as D6 (n=3), but a higher Mean (M=1.5, SD=0.7), as it got neutral responses (n=3). Lastly, D5 obtained the highest (M=2.1, SD=0.8), getting a few neutral responses (n=4) and a lone agreement (n=1).

5.6 Discussion

The pilot study had the purpose of testing human-machine interaction and verifying the participants’ sentiment towards the agent, checking if the agent responded empathically and did not make any participant uncomfortable. Also, not only it tested the interaction itself, but also the ambience to check if the space could help alleviate the interaction and allow the user to relax when speaking towards the agent.

In the Spatial Presence factor, the Sensory Depth and Realism items received the lowest scores (table 3). This was expected due to the minimal efforts made in the study to establish a sense of smell for the participants, explaining the discrepancy between ‘being able to almost smell the environment on display’ (B19) and the extent to which they ‘felt that all of their senses were stimulated at the same time’ (B27). Similar expectations were held for how much they ‘felt that the temperature was altered to match the setting of the displayed environment’ (B25), as no temperature fluctuations were indicated in the environment and no adjustment to real-world temperature was made.

Since the participants remained seated in the virtual office, any inquiries regarding movement would similarly be considered not pertinent to the study. Regardless, for ‘I had the feeling that I moved in response to parts of the displayed environment.’ (B11), moving according to the exhibited surroundings showed similar levels of acceptance and disagreement, while some participants expressed neutrality. Interestingly enough, movement (B11) and touch (‘I felt that the characters and/or objects could almost touch me. (B6), I felt I could reach out and touch things (in the displayed environment). (B24)) scored similarly, with B11 sharing the same mean as B6 with B24 scoring slightly higher. This could have led to the lack of movement and interaction with the scenario lowering the sense of touch.

Regarding the virtual room the agent was in, it was regarded rather positively. The fact that the Self-Location item B7, ‘I felt like I was visiting the places in the displayed environment.’, was the lowest was expected due to the outside view. On the right side of the room (seen on Figure 4.20, on the background), there’s an open window that shows an ocean view. Since there was no sound, smell or wind to be felt, the sense of ‘visiting the place’ was expectedly low. This is further reinforced by the highest scores of the Space Presence factor being obtained from the items on ‘being in the same space as characters and objects’ (B29) and to what extent they ‘felt surrounded by the displayed environment’ (B21).

Regarding the Engagement factor, our results suggest that contact with the agent and presence in the virtual environment resulted in most of the users losing track of time (table 7). This was evident from the number of participants who continued conversing with the agent until the agent said goodbye and concluded the session. Our results suggest that this might have been the result of the sense of immersion to the user (B1), similar to CiF-CK enhancement of NPCs and their interactions [25] and GAMYGDALA character complexity [16], and being able to participate in the environment through the conversation (B33).

In terms of content credibility (B9), we believe that the scenario’s naturalism (B4) and the agent’s realism (B23) led to the resulting mixed opinions, according to Tao et. al [32], which suggest that lower levels of appearance realism can negatively impact perceived trustworthiness and professional credibility of virtual entities. The non-photorealism of the virtual scenario can also explain the disagreement with them feeling like ‘being part of the real world’ (B13) (n=8 disagreed with the statement), as high-fidelity rendering is often a prerequisite for ‘place illusion’, the psychological sense that a virtual element truly inhabits the user’s physical reality [33]. However,the participants strong agreement with the possibility of it happening in real life (B17) (n=9 agreed with the statement) indicates that while the visual representation was non-photorealistic, the scenario maintained high ‘plausibility’, likely due to behavioral consistency or the logical flow of the social interaction [34].

Most participants (n=8) agreed that the environment was natural, while some (n=4) were neutral about the statement and only a few (n=3) thought it was unnatural. Zibrek et al. [33] explain this feeling of naturalness in a setting that isn't photorealistic by making a difference between "cognitive naturalness" and "physical realism." An environment can be considered realistic without high visual fidelity if it aligns with expected real-world spatial configurations, including accurate lighting and perspective, thereby improving processing fluency [33]. Kyrlitsias [34] also says that the idea of "congruence" is often more important than just photorealism. When the items and environmental behaviors are consistent and make sense, users can apply their real-world mental models to the virtual world.

Thankfully, negative effects of the virtual environment were very low, consistent with recent systematic reviews indicating that VR interventions are generally well-tolerated, with dropout rates due to cybersickness typically remaining below 13% [35][36].

The content being truthful and genuine (B5) was met with surprisingly polarizing opinions. It was considered neutral by very few (n=2), with some participants disagreeing completely (n=3) or agreeing completely (n=1). It is uncertain as to why the score is lower than expected, but verifying similar scored questions is plausible that the inability to move objects (B16), the participants felt like the agent or the objects were incapable of almost touching them (B6), the environment not feeling like part of the real world (B13) and the agent interaction not feeling natural (C2) led to this score.

While the ‘intensity of the experience’ (B14) should be low considering the majority of the participants casually interacted with the agent, it still received some agreement (n=3). Contrary to expectations, the number of participants who responded emotionally was relatively low (n=6), despite the fact that the conversation topics were self-selected. This finding is significant for analyzing the agent's likeability, particularly because item C3 ('the agent seems to understand my emotions well') scores correlate with item B26 (‘I responded emotionally.’).

As mentioned in our results, the ITC-SOPI item ‘I responded emotionally’ (B26) was associated with the item Quality and Likability item ‘The agent seemed to understand my emotions well’ (C3) due to similarities between negative (n=1 completely disagreed, n=6 disagreed in both) and positive responses (n=5 agreed in both, n=1 for completely agree in B26 and n=2 in C3), as seen in Figure 5.1. When analysing these results, participants who reported responding emotionally were more likely to agree that the agent identified their emotions, while those who reported the opposite were more likely to perceive the agent’s emotional identification as less accurate. Regarding the research question RQ2, ‘Can facial detection and eye tracking help the agent better detect emotion’, our results suggest that facial detection and eye tracking have the potential to support emotion detection when the participants respond emotionally.

The realism of the agent’s emotional expressions (C4) aligns with the findings for emotion detection (RQ3). Specifically, the positive (n=6) and negative (n=6) responses for realism closely match the feedback for "The agent seemed to understand my emotions well" (C3), which saw n=6 positive and n=7 negative responses. These results collectively address RQ3, suggesting that mirroring techniques can improve the user's sense of connection and empathy.

Continuing with RQ3, our findings indicate that mirroring participants' emotions may lead to improvements in the agent-user relationship when combined with perceived interaction comfort, as most participants reported feeling comfortable interacting with the agent (Figure 5.3).

When participants were asked about their reasons for feeling uncomfortable while interacting with the agent, some reported that the agent was overly artificial and disconnected from the real world. A participant mentioned, 'It was perfectly clear it wasn't a real person, I had the impression that I was talking to something inanimate', while another stated that 'The agent's eyes being static, meaning they weren't looking at me but only straight ahead while he moved his head, gave the impression of being a generic robot rather than an intelligent character.'. These responses suggest that eye-gaze behavior plays an important role in perceived interaction comfort, which is consistent with prior work on non-verbal behaviour [4][14][15]. The fact that head movement was implemented, but not eye-gaze dynamics, might explain the participants' discomfort while interacting with the agent. Research question RQ1 examined whether the proposed set of body movements and facial expressions could reduce interaction discomfort and improve affinity and empathy. Even though the findings suggest these features may contribute to an increase of affinity and empathy and reduce discomfort, the absence of eye-gaze behaviors may still elicit this effect.

Regarding research question RQ4, 'Do users perceive non-hyper-realistic agents using synthesized voices favorably or negatively?', participants reports of interaction discomfort and the agent's unnaturalness were often associated with reports of the agent being monotone, sounding robotic or artificial. These results may be influenced by:

- The limited availability of synthetic voices in European Portuguese.
- Reduced customization options and naturality in free-to-use voice solutions compared to paid alternatives.

Participants also referred to the agent's response time as an issue. As observed in our system refinement studies (section 4.4), despite lowering the agent's response delay to two seconds, its response latency was still noticeable. One possible explanation is that the agent did not provide cues indicating whether it was still listening or formulating a response. Issues with the responses also came in three ways:

- Speech recognition issues: the agent sometimes failed to understand the user's input, leading to incorrect responses. This was noticeable when the participant's primary language was not Portuguese.
- Limited conversational flexibility: the agent was unable to engage in small talk or provide more personalized responses.
- Repetitive confirmation requests: the agent often responded vaguely, and requested confirmation when attempting to explore topics more in depth.

When asked about recommended modifications to the agent, participants mentioned its customization. This is supported by studies that indicate that users prefer conversational agents with customizable voice parameters, such as gender and pitch, as it enhances trust and perceived relatability [15].

Despite the issues with the agent's responses, two thirds (n=10) of our participants agreed that the agent's functionalities (movement, responses and emotion detection) were well integrated (Figure 5.5). Half of the participants (n=7) indicated willingness to use the product, compared to four that were not. These findings appear to suggest that improvements to response quality and the implementation of eye-gaze behaviors could increase the participants' willingness to engage with the agent.

Based on the responses to the SUS-based section (table 11), participants considered this type of agent could be integrated into other sectors due to its perceived ease of use. When asked about its potential usefulness in contexts such as education and healthcare (C7), all participants agreed that their integration would be feasible, depending on implementation. Some participants noted that interacting with a non-human agent could help reduce fear of judgement:

- 'Yes, because some people may find it easier to interact with these empathic agents than real people.'
- 'Yes, because they are not a real person, they can alleviate the fear of being judged.'

Other participants mentioned that such agents could help learning and provide new perspectives:

- 'Yes, especially in the field of education. It would be a good incentive to have an agent talking to us and exploring new horizons.'
- 'Yes, I consider them useful. At first, it's a bit awkward, but later realized they have great relevance when it comes to helping people, giving advice or tips on how to solve problems, or understanding things from another point of view.'

- ‘I think they are useful in both education and healthcare because they convey credibility of talking to a real person, a character, and not just a chat or just sound. Having an agent that matches emotions and has expressions generates more empathy and, in a way, seems to understand us better.

Some participants also emphasized that, in sensitive areas such as health, these agents should be support or assistive tools, rather than replacements for human interactions or real interventions:

- ‘In an educational context they are excellent, however in a health context they are not ideal because a conversation with the agent is not comfortable.’
- ‘I believe that, in the future, these systems could play an important role in offering immediate support in times of need. However, I consider it essential that, after this initial intervention, they refer the person for appropriate human follow-up.’
- ‘Yes, even so, I don’t think they are capable of replacing a real human being.’

6. Conclusion

In this chapter, we synthesize our findings, highlighting our observations on empathic agents and evaluating their implications for future work. We also provide recommendations and guidelines based on our findings.

Addressing Research Questions

RQ1. Can an agent with improved body movements and more facial expressions reduce interaction discomfort and improve affinity and empathy? Using head mounted display sensors to capture facial movements and record body gestures (e.g., arm movements, chest rotation, hand and finger movements, and head rotation) were associated with higher ratings of presence, likability and usability. Although this could reduce perceived artificiality, more rigorous studies are needed to prove their effectiveness in reducing the perceived interaction discomfort. These findings support prior work supporting that body language can convey intent and emotion, providing an intuitive and natural interaction experience [6][7][15]. However, the absence of eye-gaze behaviors appears to have limited these improvements. Some participants reported feeling uneasy, which aligns with previous studies indicating that imperfections in a human-like artificial agent can trigger eeriness [6][7], needing to be addressed in future studies.

RQ2. Can facial detection and eye-tracking help the agent better detect user emotion? Our results suggest that the agent was more likely to detect emotions in participants who reported feeling emotional. This tendency can help support health-related communication by encouraging users' disclosure [8][9][10][11], and the introduction of conversational agents (CAs) in healthcare [4][7][9][10] just as Woebot [4][5]. It could also potentially alleviate the shortage of mental health specialists [8]. In contrast, participants who felt their responses were not emotional also tended to perceive the agent as unable to perceive their feelings. These findings indicate the need for further studies in both casual and emotional settings to improve the agent's user interaction and emotional intelligence [14][17]. Emotion identification and empathetic delivery are areas of interest in human-Computer interaction research [7][16]. Enhancing the agent's ability could increase user trust and emotional engagement, strengthening applications and interventions.

RQ3. Can incorporating mirroring techniques improve users' empathy with an agent? Using emotional detection based on models such as Ekman's Six Basic Emotion

[6], the Circumplex model of Russel [18], the Ortony, Clore and Collins model [12][16] and the Mehrabian and Russel PAD model [13], our results suggest that mirroring emotion techniques can improve the perceived empathic abilities and connection between the agent and the user during emotional interactions. However, these techniques might be less effective in technical conversations, where responses need to be objective.

RQ4. Do users perceive non-photorealistic agents using synthesized voices favorably or negatively? Our users generally perceived the synthesized voice of our agent negatively, and as not on par with intelligent personal assistants like Siri, Mena, or Cortana [15]. Limitations such as in tone and voices, particularly in less supported languages such as European Portuguese might have contributed for this perception. Our results highlight the importance of voice selection, especially when relying on free and limited voice services.

6.1 Improvements for the AViR agent

1. **Response times:** Participants reported the agent's response time as the main factor to the agent's unnaturalness and uncomfortable interaction. While introducing pause detection improved conversation flow and the agent's ability to respond without interrupting the user, the lack of visual cues indicating that the agent was switching from listening to speaking reduced responsiveness and affected the overall experience. Future improvements could introduce visual indicators (thinking gestures or spinning gears), or conditionals to reduce detection time, such as use of intonation (asking a question), detecting keywords to extend detection (so, and) or probability thresholds to know if the user has finished speaking to improve conversation flow.

2. **Small talk:** Incorporating small talk could enhance user experience by allowing the agent to respond to the user's non-verbal cues (e.g.,looking away, inspecting the room, looking through the window). The agent should also encourage the user to continue the conversation, and revisit previous topics or introduce related subjects to help the user's engagement during pauses.

3. **Agent Customization:** Allowing to customize some of the agent's features (e.g., gender,voice, and tone) or physical appearance (e.g., hair, clothing) could enhance trust and perceived relatability.

4. **Improve the agent's speech recognition and reasoning:** This study relied on the Oculus Virtual Audio Device for speech capture. This may have been less effective detecting speech from users with non-standard accents or non-native European

Portuguese. Future studies should consider dedicated microphones or allowing the participants to bring their own to improve recognition. Additionally, the reasoning engine used (Gemini API) is not ranked highly according to the Vellum Large Language Model (LLM) Leaderboard [39]. Therefore, updating to a newer version or model is recommended.

5. **Better privacy:** Updating the agent's reasoning model, although recommended, introduces privacy risks, as user data could be used to further train external models, breaching user confidentiality. Future work should consider updating to a more secure or locally controlled LLM, either trained in-house or restricted to a closed database.

6. **More animations and improvements in gestures:** Although participants did not explicitly report issues with the agent's gestures during the study, it is recommended to increase the range of hand, arm and facial movements for longer conversations. This study's sessions lasted a maximum of fifteen minutes, but longer conversations, i.e one hour or more, or repeated use, such limitation in gestures could reduce engagement.

(This page intentionally left blank.)

7 References

[1] M. S. Carneirão, J. G. Agrela, D. C. G. Mendes, R. Costa and S. Bermúdez i Badia, ”Providing Psychological Support After Mis-carriage Through a Customizable Virtual Reality Experience,” 2023 IEEE 11th International Conference on Serious Games and Applications for Health (SeGAH), Athens, Greece, 2023, pp. 1-7, doi: 10.1109/SeGAH57547.2023.10253763.

[2] T. Ellis, N. K. Latham, T. R. DeAngelis, C. A. Thomas, M. Saint-Hilaire and T. W. Bickmore, ”Feasibility of a virtual exercise coach to promote walking in community-dwelling persons with Parkinson disease,” Am J Phys Med Rehabil. 2013 Jun;92(6):472-81; quiz 482-5. doi: 10.1097/PHM.0b013e31828cd466.

[3] T. W. Bickmore, R. A. Silliman, K. Nelson, D. M. Cheng, M. Winter, L. Henault and M. K. Paasche-Orlow, ”A randomized controlled trial of an automated exercise coach for older adults,” J Am Geriatr Soc. 2013 Oct;61(10):1676-83. doi: 10.1111/jgs.12449.

[4] P. Philip, J. A. Micoulaud-Franchi, P. Sagaspe, et al., ”Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders,” Sci Rep 7, 42656 (2017). doi: 10.1038/srep42656.

[5] M. Farzan, H. Ebrahimi, M. Pourali and F. Sabeti, ”Artificial Intelligence-Powered Cognitive Behavioral Therapy Chatbots, a Systematic Review,” Iran J Psychiatry. 2025 Jan;20(1):102-110. doi: 10.18502/ijps.v20i1.17395.

[6] H. Wang, V. Gaddy, J. R. Beveridge and F. R. Ortega, ”Building an Emotionally Responsive Avatar with Dynamic Facial Expressions in Human—Computer Interactions,” Multimodal Technologies and Interaction. 2021; 5(3):13. doi: 10.3390/mti5030013.

[7] A. Paiva, I. Leite, H. Boukricha and I. Wachsmuth, ”Empathy in Virtual Agents and Robots: A Survey,” ACM Trans. Interact. Intell. Syst. 7, 3, Article 11 (September 2017), 40 pages. doi: 10.1145/2912150.

[8] A. N. Vaidyam, H. Wisniewski, J. D. Halamka, M. S. Kashavan and J. B. Torus, ”Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape,” The Canadian Journal of Psychiatry. 2019;64(7):456-464. doi: 10.1177/0706743719828977.

[9] P. M. Gardiner, K. D. McCue, L. M. Negash, T. Cheng, L. F. White, L. Yinusa-Nyahkoon, B. W. Jack and T. W. Bickmore, ”Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial,” Patient Educ Couns. 2017 Sep;100(9):1720-1729. doi: 10.1016/j.pec.2017.04.015.

[10] B. Jack, T. Bickmore, M. Hempstead, L. Yinusa-Nyahkoon, E. Sadikova, S. Mitchell, P. Gardiner, F. Adigun, B. Penti, D. Schulman and K. Damus, ”Reducing Preconception Risks Among African American Women with Conversational Agent Technology,” J Am Board Fam Med. 2015 Jul-Aug;28(4):441-51. doi: 10.3122/jabfm.2015.04.140327.

[11] D. C. G. Mendes, R. Costa, G. Agrela, S. B. i. Badia and M. S. Cameirão, ”Evaluating the Impact of Anthropomorphism and Verbal Interaction on Users' Perception of a Virtual Agent in Facilitating Self-Disclosure,” 2024 IEEE 12th International Conference on Serious Games and Applications for Health (SeGAH), Funchal, Portugal, 2024, pp. 1-8, doi: 10.1109/SeGAH61285.2024.10639572.

[12] G. L. Clore and A. Ortony, ”Psychological Construction in the OCC Model of Emotion,” Emotion Review, 5(4), 2013, pp. 335-343. doi: 10.1177/1754073913489751.

[13] A. Mehrabian, ”Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament,” Current Psychology 14, 1996, pp. 261–292. doi: 10.1007/BF02686918.

[14] M. Allouch, A. Azaria and R. Azoulay, ”Conversational Agents: Goals, Technologies, Vision and Challenges,” Sensors 2021, 21, 8448. doi: 10.3390/s21248448.

[15] S. Ghosh and S. Ghosh, ”Do Users Need Human-like Conversational Agents? – Exploring Conversational System Design Using Framework of Human Needs,” In Proceedings of the 3rd International Workshop on Conversational User Interfaces (CUI 2021), CEUR-WS.org, vol. 2950, 2021. https://ceur-ws.org/Vol-2950/paper-07.pdf.

[16] A. Popescu, J. Broekens and M. van Someren, ”GAMYGDALA: An Emotion Engine for Games,” IEEE Transactions on Affective Computing, vol. 5, no. 1, 2014, pp. 32-44. doi: 10.1109/T-AFFC.2013.24.

[17] T. Shinozaki, Y. Yamamoto and S. Tsuruta, ”Context-based counselor agent for software development ecosystem,” Computing, 2013. doi: 10.1007/s00607-013-0352-y.

[18] M. Saberi, S. DiPaola and U. Bernardet, ”Expressing Personality Through Non-verbal Behaviour in Real-Time Interaction,” Frontiers in Psychology, 12, 2021. doi: 10.3389/fpsyg.2021.660895.

[19] B. D. Steunebrink, M. Dastani and J. J. Meyer, ”The OCC model revisited,” In Proceedings of the 4th Workshop on Emotion and Computing, 2009, pp. 1-8, doi: 10.13140/RG.2.1.2514.8005.

[20] S. Mascarenhas, M. Guimarães, R. Prada, P. A. Santos, J. Dias and A. Paiva, ”FAtiMA Toolkit: Toward an Accessible Tool for the Development of Socio-emotional Agents,” ACM Trans. Interact. Intell. Syst. 12, 1, Article 8 (March 2022), 30 pages. doi: 10.1145/3510822.

[21] H. Boukricha, I. Wachsmuth, A. Hofstätter and K. Grammer, ”Pleasure-arousal-dominance driven facial expression simulation,” 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, Amsterdam, Netherlands, 2009, pp. 1-7, doi: 10.1109/ACII.2009.5349579.

[22] S. A. Murphy and M. J. Hine, ”The Role of Motivational Systems and Emotions in a Virtual Task,” 2009 42nd Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 2009, pp. 1-10, doi: 10.1109/HICSS.2009.452.

[23] D. Watson, D. Wiese, J. Vaidya and A. Tellegen, ”The Two General Activation Systems of Affect: Structural Findings, Evolutionary Considerations, and Psychobiological Evidence,” Journal of Personality and Social Psychology, 76, 1999, pp. 820-838. doi: 10.1037/0022-3514.76.5.820.

[24] S. L. Gable, H. T. Reis and A. J. Elliot, ”Behavioral activation and inhibition in everyday life,” Journal of Personality and Social Psychology, (78), 2000, pp. 1135-1149. doi: 10.1037/0022-3514.78.6.1135.

[25] M. Guimarães, P. Santos and A. Jhala, ”CiF-CK: An architecture for social NPCS in commercial games,” 2017 IEEE Conference on Computational Intelligence and Games (CIG), IEEE Press, 2017, pp. 126–133. doi: 10.1109/CIG.2017.8080425.

[26] S. Mascarenhas, M. Guimarães, R. Prada, J. Dias, J. Santos, P. Star, K. Hirsh, B. Spice, E. Kommeren and R. Rob, ”A Virtual Agent Toolkit for Serious Games Developers,” 2018 IEEE Conference on Computational Intelligence and Games (CIG), 2018, pp. 1-7. doi: 10.1109/CIG.2018.8490399.

[27] Meta. (2024). Movement SDK for Unity: Face Tracking. Meta Horizon Documentation.

https://developers.meta.com/horizon/documentation/unity/move-face-tracking/

(Accessed 12 January 2026)

[28] Microsoft. (2024). Regular expression options. Microsoft Learn.

https://learn.microsoft.com/en-us/dotnet/standard/base-types/regular-expression-options

(Accessed 17 January 2026)

[29] i2 media research. (n.d.). ITC-Sense of Presence Inventory (ITC-SOPI).

https://www.i2mediaresearch.com/sopi (Accessed 16 January 2026)

[30] J. Martins, C. Gonçalves, F. Branco, M. Castroucho, R. Goncalves and J. J. P. C. Rodrigues, ”European Portuguese Validation of the System Usability Scale (SUS),” Procedia Computer Science, vol. 64, 2015, pp. 293-300, doi: 10.1016/j.procs.2015.09.273.

[31] K. Mohd Arof, S. Ismail and A. L. Saleh, ”Contractor's Performance Appraisal System in the Malaysian Construction Industry: Current Practice, Perception and Understanding,” International Journal of Engineering & Technology, vol. 7, no. 3.9, 2018, p. 46, doi: 10.14419/ijet.v7i3.9.15272.

[32] Z. Tao, Y. Liu, J. Qiu and S. Li, ”Impact of virtual avatar appearance realism on perceptual interaction experience: a network meta-analysis,” Front Psychol. vol. 16, 2025, doi: 10.3389/fpsyg.2025.1624975.

[33] K. Zibrek, S. Martin and R. McDonnell, ”Is photorealism important for perception of expressive virtual humans in virtual reality?” ACM Transactions on Applied Perception, vol. 16, no. 1, 2019, pp. 1–19, doi: 10.1145/3349609.

[34] C. Kyrlitsias and D. Michael-Grigoriou, ”Social interaction with agents and avatars in immersive virtual environments: A survey,” Frontiers in Virtual Reality, vol. 2, 2022, doi: 10.3389/frvir.2021.786665.

[35] A. Sakhare, J. Stradford, R. Ravichandran and J. Pa, ”Conducting a VR clinical trial in the era of COVID-19,” Frontiers in Virtual Reality, vol. 2, 2021, doi: 10.3389/frvir.2021.639478.

[36] Y. Li, I. Shiyanov and B. Muschalla, ”Assessing cybersickness in a virtual reality intervention among older adults living in nursing homes: a feasibility study,” Disability and Rehabilitation: Assistive Technology, 2025, pp. 1–12, doi: 10.1080/17483107.2025.2503909.

[37] S. Martín, V. Álvarez, B. García-López, V. M. González and J. R. Villar, ”VR-Photosense: A virtual reality photic stimulation interface for the study of photosensitivity,” Advances in Intelligent Systems and Computing, 2021, pp. 178–186, doi: 10.1007/978-3-030-87869-6_17.

[38] A. Voinescu, K. Petrini and D. Stanton Fraser, ”Presence and simulator sickness predict the usability of a virtual reality attention task,” Virtual Reality, vol. 27, 2023, pp. 1967–1983, doi: 10.1007/s10055-023-00782-3.

[39] Vellum. (2024). ”LLM Leaderboard,” https://www.vellum.ai/llm-leaderboard (Accessed 17 January 2026).

8 Annexes

8.1 Study Poster

8.2 Informed Consent

Consentimento Informado, Esclarecido e Livre

Título do estudo: Estudo de usabilidade de um agente artificial empático

Investigador Principal:
Paulo Afonso Pinto Belim, Universidade da Madeira, 962878817, 2095219@student.uma.pt

Outros Investigadores:
Mónica da Silva Cameirão, Universidade da Madeira, monica.cameirao@staff.uma.pt
Diana C. Gonçalves Mendes, Universidade da Madeira, diana.mendes@staff.uma.pt

PORQUE ESTÁ ESTE ESTUDO A SER REALIZADO? Este estudo faz parte de um projeto de investigação da Universidade da Madeira que pretende avaliar a eficácia de tecnologias de Realidade Virtual (VR) na promoção da recuperação cognitiva e do bem-estar. No âmbito desta investigação, surge o interesse específico pela usabilidade de agentes virtuais no contexto da saúde mental, com o objetivo de analisar a sua capacidade de reconhecer emoções expressas pelos utilizadores e responder de forma empática, contribuindo, assim, para o desenvolvimento de intervenções terapêuticas mais eficazes.

EM QUE CONSISTE O ESTUDO? Cada participante será convidado a participar numa sessão individual em laboratório, conduzida pelo investigador principal. A sessão inicia-se com uma breve explicação e assinatura do termo de consentimento, seguida da experiência em Realidade Virtual, durante a qual o participante interage com um agente virtual programado para detectar expressões emocionais (movimentos oculares, expressões faciais, gestos e movimentos corporais) e oferecer respostas empáticas em tempo real. Após a interação, será solicitado ao participante que preencha questionários sobre a perceção de empatia do agente, sugestões de melhoria e preferências de interação. Não serão realizadas gravações de áudio ou vídeo em nenhuma fase.

QUANTAS SESSÕES INCLUI O ESTUDO? O estudo consiste numa única sessão individual para cada participante, realizada num laboratório da Universidade da Madeira ou no ARDITI. Cada sessão tem a duração aproximada de 20 a 40 minutos e inclui três fases principais: (1)

receção e esclarecimento de dúvidas, assinatura do termo de consentimento e breve questionário demográfico (cerca de 10 minutos); (2) experiência de Realidade Virtual com um agente empático, onde serão monitorados movimentos oculares, expressões faciais, gestos e postura corporal (até no máximo 15 minutos); e (3) preenchimento de questionários de avaliação de usabilidade, perceção de empatia e sugestões de melhoria (cerca de 10 minutos).

QUEM PODE PARTICIPAR NO ESTUDO? Podem participar voluntários adultos (com 18 anos ou mais) proficientes em Língua Portuguesa, sem critérios aleatórios ou divisão em grupos de controle. A participação pressupõe ausência de condições que impeçam o uso de VR (ex.: epilepsia não controlada).

EXISTEM RISCOS PARA QUEM PARTICIPA NESTE ESTUDO? O principal desconforto potencial está associado ao uso de VR, podendo surgir náuseas, vertigens ou fadiga ocular. Para minimizar estes riscos, poderão ser feitas pausas imediatas a pedido do participante.

EXISTEM BENEFÍCIOS PARA QUEM PARTICIPA NESTE ESTUDO? Embora não existam benefícios diretos garantidos, os participantes terão a oportunidade de conhecer tecnologias avançadas de Realidade Virtual e agentes empáticos, contribuindo para futuras aplicações em saúde mental.

EXISTEM CUSTOS PARA QUEM PARTICIPA NESTE ESTUDO? A participação neste estudo não implica qualquer custo financeiro direto. Os participantes não serão cobrados por qualquer parte do protocolo experimental ou pelo uso dos equipamentos de Realidade Virtual. Contudo, cada participante é responsável pelas suas despesas de deslocação até ao local do estudo.

QUEM TERÁ ACESSO AOS DADOS? COMO SERÁ ASSEGURADA A CONFIDENCIALIDADE E O ANONIMATO? QUANDO SERÃO DESTRUÍDOS OS DADOS? Os dados recolhidos serão codificados com identificadores anónimos e armazenados em servidores seguros da Universidade da Madeira, acessíveis apenas à equipa de investigação autorizada. Será garantida a confidencialidade e o anonimato, e os dados serão eliminados 2 anos após a publicação dos resultados do estudo.

HÁ RETORNO DE RESULTADOS AO PARTICIPANTE? Os participantes poderão solicitar, por email, um resumo geral dos resultados do estudo, sem identificação individual, num prazo de até seis meses após conclusão da análise de dados.

QUAIS SÃO OS DIREITOS DE QUEM PARTICIPA NESTE ESTUDO? A sua participação é totalmente voluntária. Pode recusar participar ou interromper a qualquer momento, sem penalização ou perda de direitos. O investigador poderá interromper a sua participação se houver justificação fundamentada, sem implicar prejuízos para o participante.

QUEM POSSO CONTACTAR PARA QUAISQUER DÚVIDAS OU QUESTÕES SOBRE ESTE ESTUDO? Para questões ou esclarecimentos, contacte:

- Paulo Afonso Pinto Belim (Tel.: 962 878 817, Email: 2095219@student.uma.pt)
- Mónica da Silva Cameirão (Email: monica.cameirao@staff.uma.pt)
- Diana C. Gonçalves Mendes (Email: diana.mendes@staff.uma.pt)

UNIVERSIDADE da MADEIRA

TERMO DE CONSENTIMENTO INFORMADO

Eu, abaixo-assinado, fui informado sobre o propósito e procedimentos deste estudo. Foi-me garantido que todos os dados relativos à identificação dos participantes neste estudo são confidenciais e que será mantido o anonimato. Sei que posso recusar-me a participar ou interromper a qualquer momento a participação no estudo, sem nenhum tipo de penalização por este facto. Compreendi a informação que me foi dada, tive oportunidade de fazer perguntas e as minhas dúvidas foram esclarecidas.
Aceito participar de livre vontade no estudo acima mencionado. Também autorizo a divulgação dos resultados obtidos no meio científico, desde que seja garantido o anonimato.

_________________________________ _______________

ASSINATURA DO PARTICIPANTE DATA

_________________________________ _______________

ASSINATURA DO INVESTIGADOR DATA

8.3 ITC-SOPI

Independent Television Commission/ i2 media research ltd.

Sense of Presence Inventory©

Instruções para Aplicação e Pontuação

1. Aplicação

- Aplique o questionário a participantes não-especialistas imediatamente após estes terem experenciado os conteúdos que estão a ser alvo de avaliação experimental
- Os participantes não devem ter conhecimento do objetivo do estudo nem que o questionário se destina a medir presença.
- O questionário deve ser preenchido de forma rápida, com respostas imediatas aos items pela ordem que estes são apresentados.
- Informações acerca do conhecimento prévio dos participantes acerca de uso de conteúdos multimedia são potencialmente importantes e devem ser sempre recolhidos.
- Como valor de referência, o preenchimento completo do questionário deve demorar entre 5 a 7 minutos.

2. Pontuação

- Cada questionário completado resulta em pontuações de 4 factores (cada um gerado através do cálculo da média de todos os items que contribuem para cada um dos factores) por experiência e por participante.
- Os factores, e os items que constituem para os mesmos são:
    - **Presença Espacial**
        - *média dos items B3, B6, B7, B10, B11, B15, B16, B19, B20, B21, B24, B25, B27, B28, B29, B30, B31, B33*
    - **Engagement**
        - *média dos items A2, B1, B2, B5, B14, B26*
    - **Validade Ecológica / Naturalidade**
        - *média dos items* B4, B9, B13, B17, B23
    - **Efeitos Negativos**
        - *média dos items A1, B8, B12, B18, B22, B32*
- As pontuações de cada um dos fatores não pode ser combinada num fator geral de "experiência global" – os resultados para cada fator devem ser analisados individualmente.
- Dados em falta devem ser calculados usando medias – mas deve ser minimisado através de análise cuidada aquando da aplicação do questionário.
- Quando os ambientes avaliados com recurso ao questionário ITC-SOPI não contêm personagens, recomenda-se que o item B20 seja ignorado no cálculo do Presença Espacial – o B20 refere-se especificamente a personagens no ambiente.

Preâmbulo

Por favor leia as instruções seguintes antes de continuar

Instruções: Estamos interessados em descobrir o que você sente acerca da experiência que acabou de ter no 'AMBIENTE EXIBIDO'. Usamos o termo "ambiente exibido" aqui, e ao longo do questionário, como referência ao filme, vídeo, jogo ou ambiente virtual com o qual acabaste de ser confrontado. Algumas questões referem-se ao "CONTEÚDO" do ambiente exibido. Com isto queremos dizer a história, cena ou eventos ou qualquer outro aspecto que possas ter visto, ouvido ou sentido acontecer no ambiente exibido. O ambiente exibido e o seu conteúdo (incluindo a representação de pessoas, animais ou cartoons são referidos com o"PERSONAGENS" são diferentes do "MUNDO REAL": o mundo onde vives diariamente. Por favor tem esta página como referência se tiveres dúvidas sobre o significado de alguma questão.

Este questionário está dividido em duas partes, Parte A e Parte B. A Parte A questiona-te acerca dos teus pensamentos e sentimentos assim que o ambiente exibido acabou. A Parte B refere-se aos teus pensamentos e sentimentos enquanto estavas a experienciar o ambiente exibido. Por favor não leve muito tempo a responder cada uma das questões. Geralmente a sua primeira resposta é a melhor. Para cada questão, escolha a resposta mais próxima ao teu sentimento.

Por favor lembre-se que não existem respostas certas ou erradas – estamos apenas interessados nos TEUS pensamentos e sentimentos acerca do ambiente exibido. Por favor não comente aspectos deste questionário com algum potencial participante nesta experiência pois pode afetar as suas respostas.

Todas as respostas serão tratadas com confidencialidade.

PARTE A

Por favor indique o QUANTO CONCORDA OU DISCORDA com cada uma das seguintes afirmações fazendo um círculo em apenas UM dos números utilizando a escala de 5 pontos

(Discordo fortemente)	(Discordo)	(Não concordo nem discordo)	(Concordo)	(Concordo fortemente)
1	2	3	4	5

DEPOIS DA MINHA EXPERIÊNCIA NO AMBIENTE EXIBIDO...

1. Senti-me desorientado..1 2 3 4 5

2. Tive a sensação de ter regressado de uma viagem...1 2 3 4 5

PARTE B

Por favor indique o QUANTO CONCORDA OU DISCORDA com cada uma das seguintes afirmações fazendo um círculo em apenas UM dos números utilizando a escala de 5 pontos

(Discordo fortemente)	(Discordo)	(Não concordo nem discordo)	(Concordo)	(Concordo fortemente)
1	2	3	4	5

DURANTE A MINHA EXPERIÊNCIA NO AMBIENTE VIRTUAL...

1. Eu senti-me envolvido (no ambiente exibido)……...1 2 3 4 5
2. Eu perdi a noção do tempo...1 2 3 4 5
3. Eu senti que consegui interagir com o ambiente exibido.................................1 2 3 4 5
4. O ambiente exibido/apresentado pareceu-me natural.......................................1 2 3 4 5
5. Eu senti como que o conteúdo era verdadeiro e genuíno.................................1 2 3 4 5

MASSIVE
VIRTUAL REALITY LABORATORY
INESCTEC utad

6. Eu senti que, as personagens e/ou os objetos, quase me podiam tocar.............1 2 3 4 5

(Discordo fortemente)	(Discordo)	(Não concordo nem discordo)	(Concordo)	(Concordo fortemente)
1	2	3	4	5

<u>DURANTE</u> A MINHA EXPERIÊNCIA NO AMBIENTE VIRTUAL...

7. Eu senti que estava a visitar os lugares no ambiente exibido…........................1 2 3 4 5

8. Eu senti-me cansado...1 2 3 4 5

9. O conteúdo pareceu-me credível..1 2 3 4 5

10. Eu senti que não estava apenas a observar algo...1 2 3 4 5

11. Eu tive a sensação que me movimentei em resposta a partes do ambiente exibido.…………………………………………………………………….…1 2 3 4 5

12. Eu senti-me desorientado..1 2 3 4 5

13. Eu senti que o ambiente exibido era parte do mundo real………….............1 2 3 4 5

14. A minha experiência foi intensa...1 2 3 4 5

15. Eu tive a sensação de estar nos cenários exibidos..1 2 3 4 5

16. Eu senti que podia mover objetos (no ambiente exibido)………..................1 2 3 4 5

17. As cenas representadas poderiam verdadeiramente ocorrer no mundo real....1 2 3 4 5

18. Eu senti fadiga ocular...1 2 3 4 5

19. Eu quase podia sentir o cheiro de diferentes características do ambiente exibido…………………………………………………………...…………....1 2 3 4 5

20. Eu tive a sensação que os personagens estavam conscientes da minha presença……………………………………………………………………….1 2 3 4 5

21. Eu senti-me rodeado pelo ambiente exibido………......................................1 2 3 4 5

22. Eu senti náuseas..1 2 3 4 5

23. Eu tive uma forte percepção de que os personagens e objetos eram sólidos...1 2 3 4 5

24. Eu senti que podia alcançar e tocar em coisas (no ambiente exibido)………1 2 3 4 5

25. Eu senti que a temperatura foi alterada para coincidir com o cenário do ambiente exibido…………………………………………………………................1 2 3 4 5

(Discordo fortemente)	(Discordo)	(Não concordo nem discordo)	(Concordo)	(Concordo fortemente)
1	2	3	4	5

DURANTE A MINHA EXPERIÊNCIA NO AMBIENTE VIRTUAL...

26. Eu respondi emocionalmente..1 2 3 4 5

27. Eu senti que todos os meus sentidos foram estimulados ao mesmo tempo.....1 2 3 4 5

28. Eu senti-me capaz de alterar o curso dos eventos no ambiente exibido…......1 2 3 4 5

29. Eu senti como se estivesse estado no mesmo espaço que as personagens e/ou objetos………………………………………………………………...…….1 2 3 4 5

30. Eu tive a sensação que partes do ambiente exibido interagiam comigo..........1 2 3 4 5

31. A experiência de mover coisas no ambiente exibido foi realista...................1 2 3 4 5

32. Eu senti que tinha uma dor de cabeça...1 2 3 4 5

33. Eu senti-me como se estivesse a participar no ambiente exibido…...............1 2 3 4 5

Se existe algo que queira referir por favor utilize este espaço:

____
____
____
____
____
____
____
____
____
____
____
____
____

8.4 System Usability Scale (SUS)

System Usability Scale (SUS)

	1 Discordo fortemente	2	3	4	5 Concordo plenamente
Acho que gostaria de utilizar este produto com frequência.					
Considerei o produto mais complexo do que necessário.					
Achei o produto fácil de utilizar.					
Acho que necessitaria de ajuda de um técnico para conseguir utilizar este produto.					
Considerei que as várias funcionalidades deste produto estavam bem integradas.					
Achei que este produto tinha muitas inconsistências.					
Suponho que a maioria das pessoas aprenderia a utilizar rapidamente este produto.					
Considerei o produto muito complicado de utilizar.					
Senti-me muito confiante a utilizar este produto.					
Tive que aprender muito antes de conseguir lidar com este produto.					

8.5 Pre-Questionnaire

Estudo de Usabilidade para o desenvolvimento de um agente artificial empático - Pré-Questionário

* Indicates required question

1. Qual a sua idade: *

Mark only one oval per row.

	18-25	26-33	34-41	42-50	51-59	60+
Idade:	⬭	⬭	⬭	⬭	⬭	⬭

2. Qual o seu género: *

Mark only one oval per row.

	Feminino	Masculino	Outro	Prefiro não dizer
Género:	⬭	⬭	⬭	⬭

3. Qual o seu estado civil: *

Mark only one oval per row.

	Solteiro	Num relacionamento	Casado	Divorciado	Viuvo	Prefiro não dizer
Estado civil:	⬭	⬭	⬭	⬭	⬭	⬭

4. Qual o seu estado profissional: *

Mark only one oval per row.

	Estudante	Empregado	Desempregado	Trabalhador Independente	Reformado	Ou
Estado profissional:	⬭	⬭	⬭	⬭	⬭	⬭

5. Como classificaria o seu nível de conhecimentos técnicos em tecnologias digitais? *

Mark only one oval per row.

	Nenhum	Baixo	Moderado	Elevado	Muito Elevado
Conhecimento:	⬭	⬭	⬭	⬭	⬭

6. Qual a sua experiência com realidade virtual (VR): *

Mark only one oval per row.

	Nenhuma	Pouca Experiência	Alguma Experiência	Muita Experiência
Experiência:	⬭	⬭	⬭	⬭

7. Com que frequência joga videojogos? *

Mark only one oval per row.

	Nunca	Raramente	Algumas vezes por Mês	Uma vez por semana	Algumas vezes por semana	Todos os dias
Frequência:	⬭	⬭	⬭	⬭	⬭	⬭

8. Já utilizou outras tecnologias interativas (ex.: simuladores, ambientes 3D, aplicações interativas)? *

Mark only one oval per row.

	Não	Sim, apenas uma vez	Sim, ocasionalmente	Sim, com frequência
Utilizou outras tecnologias interativas:	◯	◯	◯	◯

8.6 Post-Questionnaire

Estudo de Usabilidade para o desenvolvimento de um agente artificial empático - Pós-Questionário

Nesta primeira parte, responda em termos da sua experiência que acabou de ter 'ambiente exibido'. Usamos o termo 'ambiente exibido' aqui e ao longo do questionário, como referência ao ambiente virtual com o qual acabou de ser confrontado.

Algumas das perguntas foram retiradas do ITC-SOPI i2 media research ltd. 2004 e todos os direitos são reservados para eles.

* Indicates required question

1. **Depois** da minha experiência no ambiente exibido... *

Mark only one oval per row.

	Discordo Fortemente	Discordo	Neutro	Concordo	Concordo fortemente
Senti-me desorientado	⬭	⬭	⬭	⬭	⬭
Tive a sensação de ter regressado de uma viagem	⬭	⬭	⬭	⬭	⬭

2. **<u>Durante</u>** a minha experiência no ambiente virtual... *

Mark only one oval per row.

	Discordo Fortemente	Discordo	Neutro	Concordo	Concordo fortemente
Eu senti-me envolvido (no ambiente exibido)	⬭	⬭	⬭	⬭	⬭
Eu perdi a noção do tempo	⬭	⬭	⬭	⬭	⬭
Eu senti que consegui interagir com o ambiente exibido	⬭	⬭	⬭	⬭	⬭
O ambiente exibido pareceu-me natural	⬭	⬭	⬭	⬭	⬭
Eu senti que o conteúdo era verdadeiro e genuíno	⬭	⬭	⬭	⬭	⬭
Eu senti que, as personagens e/ ou objetos, quase me podiam tocar	⬭	⬭	⬭	⬭	⬭
Eu senti que estava a visitar os lugares no ambiente exibido	⬭	⬭	⬭	⬭	⬭
Eu senti-me cansado	⬭	⬭	⬭	⬭	⬭
O conteúdo pareceu-me credivel	⬭	⬭	⬭	⬭	⬭
Eu senti que não estava apenas a observar algo	⬭	⬭	⬭	⬭	⬭

Eu tive a sensação que me movimentei em resposta a partes do ambiente exibido	◯	◯	◯	◯	◯
Eu senti-me desorientado	◯	◯	◯	◯	◯
Eu senti que o ambiente exibido era parte do mundo real	◯	◯	◯	◯	◯
A minha experiência foi intensa	◯	◯	◯	◯	◯
Eu tive a sensação de estar no cenário exibido	◯	◯	◯	◯	◯
Eu senti que podia mover objetos (no ambiente exibido)	◯	◯	◯	◯	◯
A cena representada poderia verdadeiramente ocorrer no mundo real	◯	◯	◯	◯	◯
Eu senti fadiga ocular	◯	◯	◯	◯	◯
Eu quase podia sentir o cheiro de diferentes caracteristicas do ambiente exibido	◯	◯	◯	◯	◯
Eu tive a sensação que a personagem estava	◯	◯	◯	◯	◯

consciente da minha presença					
Eu senti-me rodeado pelo ambiente exibido	○	○	○	○	○
Eu senti náuseas	○	○	○	○	○
Eu tive uma forte percepção de que a personagem e objetos eram sólidos	○	○	○	○	○
Eu senti que podia alcançar e tocar em coisas (no ambiente exibido)	○	○	○	○	○
Eu senti que a temperatura foi alterada para coincidir com o cenário do ambiente exibido	○	○	○	○	○
Eu respondi emocionalmente	○	○	○	○	○
Eu senti que todos os meus sentidos foram estimulados ao mesmo tempo	○	○	○	○	○
Eu senti-me capaz de alterar o curso dos eventos no ambiente exibido	○	○	○	○	○
Eu senti como se estivesse estado no mesmo espaço que a	○	○	○	○	○

personagem					
Eu senti que partes do ambiente exibido interagiam comigo	⬭	⬭	⬭	⬭	⬭
A experiência de mover coisas no ambiente exibido foi realista	⬭	⬭	⬭	⬭	⬭
Eu senti que tinha uma dor de cabeça	⬭	⬭	⬭	⬭	⬭
Eu senti como se estivese a participar no ambiente exibido	⬭	⬭	⬭	⬭	⬭

Nesta segunda parte, as questões são em relação à sua experiência com o agente e como se sentiu ao interagir com ela.

3. Senti-me confortável ao interagir com o agente *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Interação:	⬭	⬭	⬭	⬭	⬭

4. Sentiu-se desconfortável com o agente? Indique porque:

____
____
____
____
____

5. Senti que as interações do agente eram naturais. *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Interações Naturais:	⬭	⬭	⬭	⬭	⬭

6. Se sentiu que as interações com o agente não eram naturais, indique porque:

____
____
____
____
____

7. O agente pareceu perceber bem as minhas emoções. *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Emoções:	⬭	⬭	⬭	⬭	⬭

8. Achei que a expressão emocional do agente era realista. *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Emoções do agente:	⬭	⬭	⬭	⬭	⬭

9. O agente respondeu de forma realista durante a conversa. *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Respostas realistas:	⬭	⬭	⬭	⬭	⬭

10. Se pudesse mudar algo no agente, o que seria? (Pode ser mais do que uma coisa):

11. Acha que os agentes empáticos, como o que foi utilizado neste estudo, são úteis em contextos como educação e/ou saúde?

12. Acho que gostaria de utilizar este produto com frequência *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Utilizar Produto:	⬭	⬭	⬭	⬭	⬭

13. Eu acho este sistema desnecessariamente complexo. *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completament
Complexidade:	⬭	⬭	⬭	⬭	⬭

14. Achei o produto fácil de utilizar *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Facilidade:	⬭	⬭	⬭	⬭	⬭

15. Considerei que a funcionalidade deste produto estavam bem integradas. *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Boa Integridade:	⬭	⬭	⬭	⬭	⬭

16. Achei que este produto tinha muitas incosistências. *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamer
Inconsistências:	⬭	⬭	⬭	⬭	⬭

17. Considerei o produto muito complicado de utilizar *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completament
Complexidade:	⬭	⬭	⬭	⬭	⬭

18. Senti-me muito confiante a utilizar este produto *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Confiante:	⬭	⬭	⬭	⬭	⬭

19. Tive que aprender muito antes de conseguir lidar com este produto *

Mark only one oval per row.

	Discordo Completamente	Discordo Parcialmente	Neutro	Concordo Parcialmente	Concordo Completamente
Aprender:	⬭	⬭	⬭	⬭	⬭