

# POLITECNICO DI TORINO

Master in Ingegneria Informatica  
(Computer Engineering)



**Politecnico  
di Torino**

Master's Degree Thesis

## Virtual reality body swapping to improve the training of soft skills and self-assessment

Supervisors

Prof. Fabrizio LAMBERTI

Prof. Sofia SEINFELD TARAFÀ

Co-supervisors

Dr. Davide CALANDRA

Dr. Gabriele PRATTICÒ

Candidate

**Chiara DE GIORGI**

December 2022

## Abstract

Immersive Virtual Reality applications have been proven to be effective in training soft skills and in managing anxiety. Many studies on this technology have focused on public speaking and on the treatment of phobias. Apparently, there is an interesting domain represented by job interviews that has not been fully considered yet. In fact, most of the work done in this context leveraged desktop-based Virtual Reality, a configuration which does not enable to take full advantage of the potential of immersive technologies.

Training before having a job interview, however, could be a precious asset for students who are about to finish their studies and need to apply for job positions, especially if they have never experienced one. This training could indeed be done with traditional methods; by means of Virtual Reality, though, an individual can see him or herself from the outside and better assess his or her performance. Compared with, e.g., practicing while watching in the mirror, there is so much more that can be done in immersive Virtual Reality. This technology offer the possibility to observe the whole scene where the job interview takes place, both from an external viewpoint as an observer without a direct involvement, as well as from another person's perspective as if his or her virtual body was that of the individual being interviewed.

Based on the above premises, this thesis work explores the two configurations above, aiming at disclosing which one could be more useful to self-assess an individual performance while answering questions in a job interview. Rather than actually developing a tool to support practicing in this scenario, the ultimate goal is to identify how such a tool could be most effective, i.e., whether observing the interview from an "objective" or "neutral" viewpoint or from the "subjective" viewpoint of the interviewer whose job is to carry out the evaluation.



# Summary

This thesis presents a study on the change of perspective, body swap and embodiment in Virtual Reality. In particular, a Virtual Reality application has been developed and tested: it consists of a job interview simulation to train soft skills and trigger self-assessment through the change of perspective. The users were immersed in a scene where a typical office was represented: each participant embodied an avatar representing herself or himself while an interviewer in front of him or her asked some typical questions of a job interview. The participants tested three configurations: one without perspective swap, one in which they had the opportunity to see their replies to the answers from an external point of view and another one that enabled them to see themselves from the perspective of the interviewer. Six questions have been selected after a confrontation with experts in the field of Human Resources, who suggested the main topics that come up in these situations. Considering that the participants of the experiments were all Italians, during the simulation the questions have been asked in their mother tongue, in order not to provoke additional anxiety related to speaking English.

The main purpose of the study is to leverage visual perspective changes to improve self-assessment during soft skills training. In fact, the idea is to exploit the potential of Virtual Reality to research the advantages of this technology in order to find out some of the configurations that could be more useful in the training of soft skills specifically in the context of a job interview and to manage anxiety.

The context of a job interview has been chosen for this study because training before having a job interview could be a precious asset for students who are about to finish their studies and need to apply for job positions, especially if they have never experienced one. However, the experiment was open to other students and people in general who expressed the need to practice and enhance their skills for a job interview. This training could indeed be done with traditional methods; by means of Virtual Reality, though, an individual can see him or herself from the outside and better assess his or her performance.

The idea of this thesis was inspired by a previous study about self-counseling. In this experiment the participants sat in a room in front of Sigmund Freud, who asked them to talk about personal problems. Then, the participant had the possibility



to reply, as it happens in a therapy session in the real world. In this phase the user embodied a self-avatar and could see Freud talking. Then, in a second phase, the user swapped bodies with Sigmund Freud and could give himself or herself advice to solve the issue. Such a study demonstrated that the participants had a strong illusion of presence and body ownership and the sensation that Freud was talking with them. What is more interesting is that the participants who swapped bodies with Freud reported that, if their responses changed over the sessions of the experiment, this was due to the fact that they were able to see themselves from the outside, as if they were another person with another perspective talking to themselves. The present research, with respect to such study on self-counseling, focuses on another topic, i.e., job interview, and, in addition to the configuration with body swap, examines a configuration with Third-Person Perspective.

Many studies on change of perspective, body swap and embodiment in Virtual Reality have been taken into account during the design phase of the present work. In particular, it is evident that sense of presence and embodiment can be triggered by both First Person Perspective and Third Person Perspective, even though results are generally better with the first one. Moreover, it has been proven that training soft skills in Virtual Environments is effective and that this kind of simulation in Immersive Virtual Reality can induce anxiety in the participants and, at the same time, help them to handle it. The research carried out until now has focused mainly on public speaking: in these studies, the participants could see themselves from the First Person Perspective (1PP) of the avatar giving the speech and from a Third-Person Perspective (3PP). Here a third configuration, similar to one of the conditions taken into account in this work, is absent: the possibility to experience the speech from the 1PP of someone in the audience, who would represent a person with the role to assess the performance.

Apparently, there is an interesting domain represented by job interviews that has not been fully considered yet. In fact, most of the work done in this context leveraged desktop-based Virtual Reality, a configuration which does not enable to take full advantage of the potential of immersive technologies. There are few job interview simulations requiring to use a Headset and two of these are “Be Confident” and Bodyswaps “Job interview simulator”. The first one requires that the user is asked three questions, which are always the same, and has the possibility to answer them. On the other hand, Bodyswaps application allows users to choose between three responses to every question: they have to pronounce one of them and an AI analyze the social cues to give advices to improve. In addition to this, the users have the possibility to see their virtual avatar answering the question from the point of view of the interviewer. However, there is no empirical evidence about the effectiveness of this change of visual perspective and avatars, in comparison to more traditional methods only based on having a first person perspective of the interviewee avatar.

The experiment presented in this thesis consists of three conditions, one of which represents a “Baseline” since the user experiences only the perspective of the interviewed. The main focus of this work is on the comparison between the other two, which are the “*External POV*” and the “*Interviewer’s POV*” conditions: they enable the participants to switch respectively to an external point of view and to the one of the interviewer. Each condition starts with the interviewer asking a question to which the participant has to answer through voice and gestures: this answer is recorded. Then, in the conditions which involve a change of perspective, the users watch the recording of the interviewer asking the question and of themselves answering it. To make clearer how the experiment takes place, it is possible to call *A* the phase when the user listens to the question and answers, and *B* the repetition of this phase (for the “Baseline” condition) or the phase involving switch of perspective (*External POV* and *Interviewer’s POV*): with this naming convention the structure of the whole experiment is *ABAB ABAB ABAB*.

During each experiment, the results on the effectiveness of the application for the training of soft skills, embodiment and emotional reactions have been measured through five questionnaires in total. The first one was administered before starting the experience. Then, since it is important to keep track of how users perceive each condition, one questionnaire was administered after each of them. Finally, the last questionnaire was administered at the end of the experiment. Even though some questions are personalized to fit with the experiment, they originate from validated questionnaires.

Building on the knowledge of the studies mentioned above, the hypotheses of this study can be summarized as follows.

- *H1*: the embodiment is expected to be high when participants have a First Person Perspective either of the interviewee or of the interviewer, with no significant differences between them, and lower in the external Point Of View (POV).
- *H2*: no significant differences are expected in presence and plausibility among the three conditions.
- *H3*: participants should experience less emotional distress and anxiety scores in the conditions where they have either the interviewer’s Point of View or an external one when compared to only practicing from the interviewee’s perspective;
- *H4*: more positive and less activating emotions should be evoked by the *External POV* condition compared with the other two, while more positive and less activating emotions are expected in the interviewer’s perspective in comparison with the interviewee’s perspective.

- *H5*: participants should find more useful the conditions where they can swap perspectives (either interviewer's or external) to improve and assess their performance in the job interview, with respect to the condition where they can only practice from the First Person Perspective of the interviewee (*Baseline* condition).

The application created for this study has been developed in Unity. Some Unity assets were necessary for the implementation, such as RootMotion's *FinalIK* for the setting of Inverse Kinematics on the avatars, Crazy Minnow Studio's *SALSA* for the lip sync and Eli Davis' *Record and Play* for the recording of the participants' movements. The hardware necessary for the experiment to take place, apart from a computer with a Virtual Reality-ready graphics card, are an HTC Vive headset, two Valve Index controllers, and two Vive trackers to track the movement of elbows.

The sample for this experiment was composed of 19 participants, many of whom (68,42%) were bachelor's or master's students. The 73,68% of the participants have studied or still study either Computer Engineering or Cinema Engineering. The participants have rated positively the experience: the PSSUQ questionnaire on the quality of the system concerning both hardware and software produced an average score of 4,48 out of 5, while the questionnaire on the quality of the Virtual Reality environment produced a score of 4,49 out of five. However, most of the results did not bring statistically significant information. The ones with statistical significance are related to the usefulness of each condition concerning the improvement of job interview skills and the assessment of verbal and non-verbal language during the interview. In fact, the results partially confirmed hypothesis *H5*, since the *Interviewer's POV* condition was considered the best one to improve job interview skills and to assess the effectiveness of both verbal and non-verbal language. The second place went to the *External POV* condition and the last one to the *Baseline* condition. However, this podium changes for the question concerning verbal responses. In fact, in this case, *External POV* and *Baseline* conditions are tied.

# Table of Contents

|   |      |
|---|------|
| <b>List of Tables</b>                                     | IX   |
| <b>List of Figures</b>                                    | X    |
| <b>Acronyms</b>   | XIII |
| <b>1 Introduction</b>                                     | 1    |
| 1.1 Applications of Virtual Reality . . . . .             | 1    |
| 1.2 Conception and aim of the study . . . . .             | 2    |
| 1.3 Brief description of the thesis . . . . .             | 3    |
| 1.4 Research problems and hypotheses . . . . .            | 3    |
| 1.5 Overview of the thesis . . . . .                      | 4    |
| <b>2 State of the Art</b>                                 | 6    |
| 2.1 Background . . . . .                                  | 6    |
| 2.2 Perspective swap in Virtual Reality . . . . .         | 7    |
| 2.3 Embodiment in Virtual Reality . . . . .               | 8    |
| 2.3.1 Influence of perspective . . . . .                  | 8    |
| 2.3.2 Influence of virtual body realism . . . . .         | 8    |
| 2.3.3 Body swap and embodiment . . . . .                  | 9    |
| 2.4 Learning in Virtual Environments . . . . .            | 9    |
| 2.5 Virtual Reality for soft skills training . . . . .    | 10   |
| 2.6 Summary . . . . .                                     | 13   |
| <b>3 Materials and methods</b>                            | 16   |
| 3.1 Experimental design . . . . .                         | 16   |
| 3.1.1 Condition A: "Baseline" . . . . .                   | 17   |
| 3.1.2 Condition B: "External POV" . . . . .               | 17   |
| 3.1.3 Condition C: "Interviewer's POV" . . . . .          | 18   |
| 3.1.4 Questions of the interview . . . . .                | 19   |
| 3.1.5 Configuration of questions and conditions . . . . . | 20   |

|          |  |           |
|----------|--|-----------|
| 3.2      | Sample . . . . .                                   | 20        |
| 3.3      | Measures . . . . .                                 | 21        |
| 3.3.1    | Initial questionnaire . . . . .                    | 21        |
| 3.3.2    | Post-condition questionnaires . . . . .            | 23        |
| 3.3.3    | Final questionnaire . . . . .                      | 25        |
| 3.4      | Hardware and software . . . . .                    | 26        |
| 3.4.1    | Hardware . . . . .                                 | 27        |
| 3.4.2    | Software . . . . .                                 | 29        |
| 3.5      | Experimental procedures . . . . .                  | 31        |
| 3.5.1    | Initial settings . . . . .                         | 32        |
| 3.5.2    | Familiarization with the environment . . . . .     | 33        |
| 3.5.3    | First phase: simulation and recording . . . . .    | 34        |
| 3.5.4    | Second phase . . . . .                             | 35        |
| 3.5.5    | End of a condition and of the experiment . . . . . | 36        |
| <b>4</b> | <b>Software implementation</b>                     | <b>37</b> |
| 4.1      | Structure of the Unity project . . . . .           | 37        |
| 4.2      | Preparation of main scene . . . . .                | 38        |
| 4.3      | Implementation of the avatars . . . . .            | 39        |
| 4.3.1    | Realization of the 3D models . . . . .             | 39        |
| 4.3.2    | Animations of the interviewer . . . . .            | 40        |
| 4.3.3    | Setting of IK . . . . .                            | 42        |
| 4.3.4    | Setting of lip sync . . . . .                      | 43        |
| 4.4      | Calibration and initial settings . . . . .         | 44        |
| 4.4.1    | Selection of avatar . . . . .                      | 44        |
| 4.4.2    | Trackers setting and calibration . . . . .         | 44        |
| 4.4.3    | Height calibration . . . . .                       | 45        |
| 4.4.4    | Questions and conditions settings . . . . .        | 45        |
| 4.5      | Finite State Machine . . . . .                     | 46        |
| 4.6      | Game Manager and Scene Manager . . . . .           | 48        |
| 4.6.1    | Game Manager . . . . .                             | 48        |
| 4.6.2    | Scene Manager . . . . .                            | 49        |
| 4.7      | Users' interaction with the interviewer . . . . .  | 50        |
| 4.8      | Recording and replay . . . . .                     | 50        |
| 4.8.1    | Transforms recording . . . . .                     | 51        |
| 4.8.2    | Audio recording . . . . .                          | 51        |
| 4.8.3    | Transforms replay . . . . .                        | 52        |
| 4.8.4    | Audio replay . . . . .                             | 53        |

|          |   |    |
|----------|---|----|
| <b>5</b> | <b>Results</b>  | 54 |
| 5.1      | Statistics of the results . . . . .                           | 54 |
| 5.1.1    | Characteristics of the sample . . . . .                       | 54 |
| 5.1.2    | Evaluation of the experience . . . . .                        | 55 |
| 5.1.3    | Embodiment during the replay . . . . .                        | 55 |
| 5.1.4    | Comparison of the conditions . . . . .                        | 55 |
| 5.2      | Discussion on the results . . . . .                           | 56 |
| 5.2.1    | Hypothesis <i>H1</i> : embodiment . . . . .                   | 56 |
| 5.2.2    | Hypothesis <i>H2</i> : presence and plausibility . . . . .    | 57 |
| 5.2.3    | Hypothesis <i>H3</i> : emotional distress . . . . .           | 57 |
| 5.2.4    | Hypothesis <i>H4</i> : emotional reactions . . . . .          | 57 |
| 5.2.5    | Hypothesis <i>H5</i> : comparison of the conditions . . . . . | 58 |
| <b>6</b> | <b>Conclusions</b>  | 59 |
| 6.1      | Future work . . . . .   | 60 |
| <b>A</b> | <b>Listings</b>   | 61 |
| A.1      | Animations of the interviewer . . . . .                       | 61 |
| A.2      | Finite State Machine . . . . .                                | 61 |
| A.3      | Record and Play . . . . .                                     | 62 |
|          | <b>Bibliography</b>   | 65 |

# List of Tables

|     |  |    |
|-----|--|----|
| 3.1 | Order in which conditions and questions were administered during<br>the experiment . . . . . | 20 |
|-----|--|----|

# List of Figures

|      |  |    |
|------|--|----|
| 2.1  | Prospective of the participant in [7]  | 9  |
| 2.2  | Prospective of Freud in [7]  | 10 |
| 2.3  | View of auditorium from the view of presenter in [5]                                   | 11 |
| 2.4  | Male (left) and female (right) doppelgangers during the speech in [41]                 | 12 |
| 2.5  | An example of job interview simulation in a non-immersive virtual environment [42]     | 13 |
| 2.6  | BeConfident job interview simulation [8]   | 14 |
| 2.7  | Bodyswaps job interview simulation [9]   | 14 |
| 3.1  | View of the user while the interviewer is asking a question                            | 17 |
| 3.2  | Replay from an external point of view  | 18 |
| 3.3  | Replay from the interviewer’s point of view  | 19 |
| 3.4  | Self-Assessment Manikin questionnaire administered during the experiment               | 22 |
| 3.5  | Shortened version of the STAI questionnaire administered during the experiment         | 23 |
| 3.6  | Questions on embodiment (FPP: baseline - EPOV: external POV - IPOV: interviewer’s POV) | 24 |
| 3.7  | Questions on presence and plausibility   | 25 |
| 3.8  | SPIN questionnaire   | 25 |
| 3.9  | Questions on the quality of the system   | 26 |
| 3.10 | HTC Vive Pro   | 28 |
| 3.11 | Valve Index controllers  | 28 |
| 3.12 | Moment in which the participant familiarizes with the environment                      | 34 |
| 3.13 | Panel signaling the current question   | 35 |
| 3.14 | Panel signaling the end of a condition   | 36 |
| 4.1  | Calibration scene  | 38 |
| 4.2  | Overview of the main objects in the office   | 38 |
| 4.3  | Avatar of the interviewer  | 39 |
| 4.4  | Female user’s avatar   | 40 |



|      |  |    |
|------|--|----|
| 4.5  | Male user's avatar . . . . .   | 40 |
| 4.6  | Animator of the interviewer avatar . . . . .   | 41 |
| 4.7  | Main Blend Tree of interviewer's animator . . . . .  | 41 |
| 4.8  | Position of elbow <i>goals</i> (represented by colored blue and yellow circles) for the female avatar . . . . .                        | 42 |
| 4.9  | Hierarchy of the <i>Player</i> . . . . .   | 43 |
| 4.10 | Setting of SALSA for the interviewer's avatar . . . . .  | 44 |
| 4.11 | Representation of what the experimenter sees during the calibration (bigger window) and what the participant sees (smaller window) . . | 46 |
| 4.12 | Schema of the FSM implemented for the application . . . . .  | 47 |
| 4.13 | The <i>Game Manager</i> Game Object with the scripts attached to it . .  | 49 |
| 4.14 | The <i>Scene Manager</i> Game Object with the scripts attached to it . .   | 50 |
| 5.1  | Results on the comparison of conditions . . . . .  | 56 |



# Acronyms

**VR**

Virtual Reality

**HMD**

Head Mounted Display

**VE**

Virtual Environment

**1PP**

First-person perspective

**3PP**

Third-person perspective

**POV**

Point-of-view

**UI**

User Interface

**IK**

Inverse Kinematics

**SAM**

Self-Assessment Manikin

**STAI**

State-Trait Anxiety Inventory

**SPIN**

Social Phobia Inventory

**FSM**

Finite State Machine

# Chapter 1

## Introduction

### 1.1 Applications of Virtual Reality

Virtual Reality (VR) can make a positive contribution in different fields such as education, cultural heritage, healthcare, and psychology. For example, it can be used in schools in order to give students the possibility to closely analyze things they can not see, such as cells [1]. In addition to this, Virtual Reality can enable people to visit places that can no longer be visited, such as the *Domus Aurea* in Rome [2], or even explore from the inside the paintings of Van Gogh [3]. Another example is represented by immersive simulations for medical workers to practice clinical reasoning, decision-making and critical communication skills in risk-free environments [4].

The present thesis has a multidisciplinary nature and covers knowledge from fields such as psychology, computer science, human-computer interaction, and computer graphics. The analysis of the state of the art highlights various types of Virtual Reality applications which can help people to overcome their phobias or to practice soft and hard skills. These studies demonstrated the efficacy of this tool to improve soft skills and to overcome fears and anxiety. In fact, situations that trigger a phobia or a fear can be reproduced in Virtual Reality giving users the possibility to practice and get habituated to the situation [5][6]: vast amount of evidence has shown that VR can lead to a strong sense of immersion and presence, making the users feel as if they were really inside the virtual environment. However, despite the perceived realism, in VR users are still aware that the virtual scenes are only simulations, which helps them to feel safer and more in control of the feared situations.

## 1.2 Conception and aim of the study

The idea of this thesis was inspired by a previous study about self-counseling [7]. In this experiment the participants sat in a room in front of Sigmund Freud, who asked them to talk about personal problems. Then, the participant had the possibility to reply, as it happens in a therapy session in the real world. In this phase the user embodied a self-avatar and could see Freud talking. Then, in a second phase, the user swapped body with Sigmund Freud and could give himself or herself advice to solve the issue. This study demonstrated that the participants had a strong illusion of presence and body ownership and the sensation that Freud was talking with them. What is more interesting is that the participants who swapped bodies with Freud reported that, if their responses changed over the sessions of the experiment, this was due to the fact that they were able to see themselves from the outside, as if they were another person with another perspective talking to themselves.

Building on the knowledge of this previous study, the present thesis aims to leverage visual perspective changes to improve self-assessment during soft skills training. In particular, the study case is a job interview. In fact, the literature shows that job interview simulations have been developed mostly in non-immersive Virtual Reality. Therefore, users are not fully immersed in the environment. At the moment, there are two released applications simulating a job interview by means of a headset [8][9], but they do not fully exploit the potential of Virtual Reality. In fact, thanks to this technology it is possible to enable people to see themselves from the outside in order to assess their verbal and non-verbal language. However, considering how stressful a job interview can be, in particular for students who have never experienced one and need to apply for job positions, this topic is important and needs attention.

For this reason, the aim of this work is to research the advantages of this technology in order to find out some of the configurations that could be more useful in the training of soft skills specifically in the context of a job interview and anxiety management. In fact, while traditional job simulations do not enable participants to watch their performance from the outside, this is possible in Virtual Reality. In the real world, the most similar configuration to this one is practicing in front of the mirror. However, in this case, one would see himself or herself being still in his or her own body, aware of the fact that he or she is facing a reflection of himself or herself, while Virtual Reality can enable people to see themselves from an external point of view without embodying any avatar or even from another person's virtual body. Past research suggests that these configurations can influence how the participants evaluate themselves: this is what has been investigated in this work. In fact, the proposed application has not been designed to be just a tool to practice for a job interview, considering also that some improvements should be done in order to make it complete. It is, instead, a study on the most effective

configuration to develop such tools.

Considering the social exigence which led to the conception of this work, the main target are students in their last year of studies. However, the experiment is open to other students and people in general who express the need to practice and enhance their skills for a job interview.

### 1.3 Brief description of the thesis

In order to conduct the present study, an application has been developed in Unity, as it is the game engine mostly used in the research field and is taught in this university's courses. On the other hand, the hardware necessary for the experiment to take place, apart from a computer with a VR-ready graphics card, are an HTC Vive headset, two Valve Index controllers, and two Vive trackers to track the movement of elbows.

The application consists of a scene where a typical office is represented: each participant embodies an avatar representing herself or himself while an interviewer in front of him or her asks some typical questions of a job interview. There are three types of conditions: the first is more traditional and the user can just answer the question while facing the interviewer, another gives the opportunity to see the avatar of the user from a third-person perspective and the last one enables the participant to swap body with the interviewer. In this way it is possible to overcome the limits of reality and explore new ways to strengthen soft skills and self-assessment.

### 1.4 Research problems and hypotheses

Each condition has been designed taking into account past studies, which have been useful to formulate some hypotheses.

- *H1*. As far as embodiment is concerned, the expected results are high scores when participants have a First-Person Perspective either of the interviewee or of the interviewer, with no significant differences. On the other hand, embodiment should be lower and the feeling of being outside of one's own body should be stronger in the external Point Of View (POV) compared to the rest of the conditions.
- *H2*. No significant differences are expected in presence and plausibility among the three conditions.
- *H3*. As far as emotional reactions and impact are concerned, participants should experience less emotional distress in the conditions where they have

either the interviewer's Point of View or an external one when compared to only practicing from the interviewee's perspective: this is due to the fact that *Interviewer's POV* and *External POV* conditions implicate self-distancing and having a different perspective. For the same reason, anxiety scores are expected to be lower after going through the external perspective and the interviewer's perspective with respect to the interviewee's perspective.

- *H4*. In order to evaluate the emotional reactions, the Self-Assessment Manikin questionnaire [10] has been used: more positive values for *valence* and lower for *arousal* should be evoked by the *External POV* condition compared with the other two. Moreover, *valence* is expected to be more positive and *arousal* lower in the interviewer's perspective in comparison with the interviewee's perspective.
- *H5*. Finally, the hypotheses on the usefulness of the experience are related to the aim of this study: participants should find more useful the conditions where they can swap perspectives (either interviewer's or external) to improve and assess their performance in the job interview, with respect to the condition where they can only practice from the First-Person Perspective of the interviewee (*Baseline* condition).

## 1.5 Overview of the thesis

This thesis is composed of six chapters, which illustrate different aspects.

1. *Introduction*. The current chapter is an introduction to the context, origin and content of this thesis.
2. *State of the Art*. In this chapter, after a brief description of important factors of VR related to this work, there is an overview of previous studies on topics that are important in order to understand the content of the thesis.
3. *Materials and methods*. In this part there is the description of the experimental design, the experimental protocol and the measures taken into account in order to elaborate the results. In addition to this, the hardware and the software necessary to develop the application are illustrated.
4. *Software implementation*. The chapter on implementation focuses on the application developed in Unity to conduct this study. In particular, different aspects are taken into account here, from the 3D modeling and animation to the programming and setting of the project.
5. *Results*. Here the results of the experiments are shown.



6. *Conclusions.* Lastly, there are the conclusions concerning this study and the possible future work to improve it.

# Chapter 2

## State of the Art

### 2.1 Background

In Virtual Reality applications three factors are particularly important: *presence*, *immersion*, and *interaction*. Presence is the sensation to be part of the Virtual Environment, the illusion of being there despite the fact that you know it is not real [11]. Thus, it is strictly correlated with the plausibility of the events and the objects, for both their appearance and their properties. In fact, the sense of presence includes the *illusion of place* on the one hand and the *illusion of plausibility* on the other, where the first one is the sensation to be in a place even knowing it is not true and the second one is the sensation that something that is happening in the virtual world is happening in the real one, too [12]. High levels of immersion and interaction lead to a high sense of presence, which can trigger embodiment.

As far as the *sense of embodiment* is concerned, it could be considered as composed of three sub-components: the sense of self-location, the sense of agency, and the sense of body ownership [13].

The *sense of self-location* consists of a precise volume in space when you feel located: it refers to the experience of being inside a body, not inside a precise world. As a matter of fact, this sensation can be broken by out-of-body experiences, where one sees his or her physical body from the outside. Unlike *presence*, which is the result of the relationship between one's self and the environment, *self-location* refers to the relationship between one's self and one's body.

The *sense of agency*, instead, is closely connected with active movements. In fact, it is enhanced when the predicted consequences of actions are congruent with the actual consequences of them.

Finally, the sense of *body ownership* requires that one feels to possess a body and perceives it as the source of the experienced sensations. In conclusion, we

can use the term *sense of embodiment* in order to refer to the union of sensations correlated with being inside, having, and controlling a body [14].

## 2.2 Perspective swap in Virtual Reality

Many studies have focused on the effects that the change of perspective in immersive Virtual Reality applications can have on users, demonstrating that perspective has an important role.

Firstly, it is evident that the First Person Perspective, which implies that the user sees the Virtual Environment from the point of view of a virtual avatar, leads to better immersion [15]. As a matter of fact, when a user finds herself or himself exposed to a threat (for example height) and has the possibility to choose the point of view to have, it is more feasible that she or he will choose an external point of view in comparison to the one of a virtual avatar: the subjects tend to perceive the Third Person Perspective as safer than First Person Perspective [16]. Moreover, this perspective enables the user to gain a better performance while executing a task [17][18][19].

On the other hand, according to the review of the state of the art, in the majority of cases space awareness is higher in Third Person Perspective than in First Person Perspective [15][16][17]. However, if the dynamism of a task is taken into account, it results show that the First Person Perspective is better for static to dynamic tasks, while the Allocentric Third Person Perspective is better than both First Person Perspective and Egocentric Third Person Perspective for moderately dynamic tasks, where *allocentric* refers to the fact that the virtual camera is independent of the user's avatar and *egocentric* means that it follows the avatar. Anyway, it is demonstrated that the difference between the type of perspectives is less pronounced when the task is more dynamic [18].

As far as the sense of presence and the sense of agency are concerned, the results are very similar between First Person and Third Person Perspective and they can reach high levels in both cases [15][17].

There are some other considerations that could be done about perspective swap. Firstly, a study led at the University of Duisburg-Essen in Germany demonstrates that users of Virtual Reality games prefer First Person Perspective over the Third Person Perspective since it helps to better focus on in-game actions and generates higher involvement, while the Third Person Perspective induces to focus more on the player than on the game [20].

Another interesting research is about how visual perspective influences the formation of memories for events: it demonstrated that Third Person Perspective experiences create Third Person Perspective memories and that, even though Third Person Perspective memories are as vivid and visually accurate as First Person

Perspective memories, the first ones may contain more accurate spatial information [21].

## 2.3 Embodiment in Virtual Reality

### 2.3.1 Influence of perspective

As anticipated, the *sense of embodiment* refers to the union of sensations correlated with being inside, having, and controlling a body. It has been proved that perspective has a great influence on it. However, according to the state of the art on the subject, there are different deductions depending on the experiment.

On the one hand, some results led to the conclusion that embodiment is also possible in the Third Person Perspective: in fact, this point of view is able to trigger the illusion to own a virtual body if complemented with a rich multi-sensory congruence, which plays a minor role in First Person Perspective. Moreover, according to other research, ownership toward a virtual body can be experienced only when the latter and the physical body partially or totally overlap, and this can happen also from a Third Person Perspective [22].

On the other hand, there are studies proving that First Person Perspective is fundamental to elicit full body ownership illusion [23][24] and that the sole effect of seeing a realistic virtual body in the same position of the physical one is sufficient, regardless of congruent multi-sensory and/or head-based sensorimotor cues.

### 2.3.2 Influence of virtual body realism

Besides perspective, another important factor that influences the sense of embodiment is the realism of the virtual body. In fact, it has been demonstrated that full body ownership illusion can be strengthened by the appearance of the avatar, in particular skin texture and clothes: when the users are represented by a realistic human body even congruent multi-sensory and/or sensorimotor cues are superfluous, while in the case of a fake humanoid their importance is much greater [23]. It seems that in every study in which a user is led to embody different characters with different levels of realism, for example from a stick person to a 3D scan of the user or from a robot to a doppelganger, the more the avatar resembles the user the stronger is the embodiment [25][26][27][28]. However, it is possible to arouse embodiment also when the user wears the shoes of an avatar of a different gender: this is what has been proved by a research where male subjects experienced body transfer illusion of a life-sized virtual human female seen from a First Person Perspective [29].

### 2.3.3 Body swap and embodiment

Embodiment can be achieved also when the user swaps her or his body with another one. For example, the subjects who experienced a Virtual Reality self-counseling where they alternatively assume the role of themselves expressing their problems to Freud (see Figure 2.1) and of Freud himself giving them advice (see Figure 2.2) found the body swap effective: they were able to identify themselves with Freud to such an extent that they improved their responses to their personal problems [7]. Another noteworthy study demonstrated that embodiment could be so strong even in the case of body swap with another person that people should be able to literally shake hands with themselves while experiencing ownership of another person's body: here the experimenter had cameras mounted on the head in order to capture the participants' body and the latter wore an headset showing what cameras captured, thus they were able to see their physical bodies from the shoulder to slightly above the knees [30].



Figure 2.1: Prospective of the participant in [7]

## 2.4 Learning in Virtual Environments

Several studies related to training in Virtual Environments have proved their efficiency: more specifically, performance tends to be superior to traditional methods in the education field [31], there is a positive effect on the generation of ideas and the sense of presence is facilitated [32].

However, if this already happens with “traditional” Virtual Environments, it is enhanced in Immersive Virtual Reality applications. In fact, in this case, the



**Figure 2.2:** Prospective of Freud in [7]

engagement is much higher and there is not the frustration of desktop applications due to the slow and tedious movements of the mouse and keyboard, but the movement is more immersive and fluid [33]. In particular, it has been demonstrated that Virtual Reality is very attractive and effective in teaching both hard and soft skills and that hard skills can be even better taught than with traditional methods [34]. Another study focused on the use of Virtual Reality in order to teach abstract concepts or decision-making methodologies: in this case the results are positive, too. The main disadvantage detected in this study were the user interface and the ways of interaction, since some participants found difficult even to move around and look at objects [35]. Finally, a recent study about the effects of interactivity and immersion on cognitive and affective variables of a framework named CAMIL (Cognitive Affective Model of Immersive Learning) is worthy to mention. This study has shown that the process of learning involves situational interest and embodied learning and that, when immersion is low, interactivity gains more importance for agency and embodied learning, while high immersion leads to larger interest. However, a negative effect on declarative memory has been observed, maybe due to incongruency between actions of the body and the learning content [36].

## 2.5 Virtual Reality for soft skills training

As anticipated, Virtual Reality applications may be useful to train both hard and soft skills. Some studies focused on the latter, proving that VR is an effective tool [6][5][37][38][39] and that users experience less anxiety in comparison with

traditional methods [32][34][35].

It is also for this reason that most of them lead to the conclusion that Immersive Virtual Reality is a better tool for training soft skills than other methods [37][38]. However, it seems that this effect depends on the complexity of the social skills to be improved: it appears to be marginally significant in programs aimed at developing basic skills, not statistically significant in those for moderate complexity outcomes, and statistically significant in programs to develop complex outcomes. The social skills referred to as basic are those requiring a single social competency to perform (e.g. recognizing emotions), while complex social skills involve most types of social competencies (e.g. holding an entire conversation) [38].

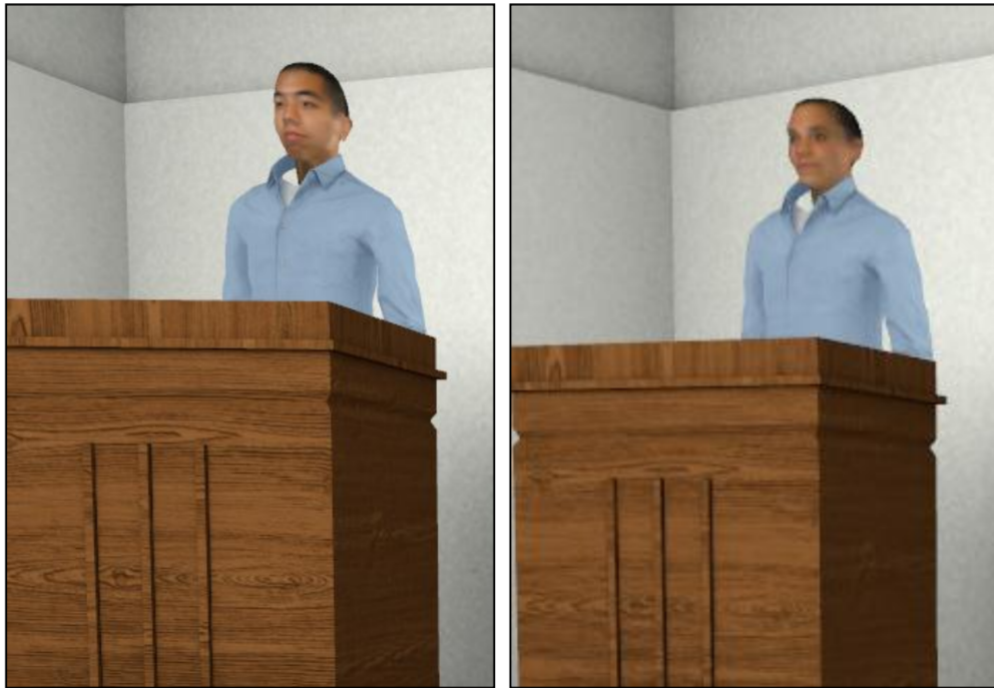


**Figure 2.3:** View of auditorium from the view of presenter in [5]

In particular, it seems that Immersive Virtual Reality has a great impact on perspective-taking and empathy thanks to higher levels of presence and embodiment, while “pen-and-paper” methods seem to be more effective in facilitating decision-making [39]. In addition to this, this kind of application enables the users to overcome the fear of public speaking [6] and elicits anxiety, which can be used to manage it in a proper way, thus resulting in a significant decrease in anxiety rates [5] (scenario visible in Figure 2.3). As far as anxiety is concerned, it is important that in these applications the trainees experience it because this demonstrates that immersion is strong, but at the same time they should be able to realize that it is not real: a combination of realistic elements in a way that is not feasible in reality could help in this intent [40]. Another possible use of Virtual Reality is in the field of treatment of phobias, by reproducing exposure therapy, which is a step-by-step exposure to anxiety, in a Virtual Environment: research shows that it

is not more effective, but comparable to traditional methods, i.e., in vivo exposure. The main reasons are probably higher credibility and expectancy compared to in vivo exposure and quicker progress of patients due to a perception of increased control and safety [37].

One important feature that Virtual Reality has over traditional methods is the possibility to trigger out-of-body experiences, which means that the user is able to see him or herself from the outside. For example, a study focused on the training for public speaking compared the situation in which users are invited to just imagine themselves giving a successful speech with one where they watch animated virtual humans with their faces performing it. According to the results, male users found the second situation better because it provoked less anxiety. On the other hand, female users did not agree, due to the lack of similarity between them and the virtual avatars, and of the realism of the doppelgangers' gestures [41] (see Figure 2.4).



**Figure 2.4:** Male (left) and female (right) doppelgangers during the speech in [41]

As far as Virtual Reality job simulations are concerned, until now it seems that most of them are not developed in Immersive VR (an example is visible in Figure 2.5). However, some of these applications are interesting because of the use of social cues recognition techniques in order to react and adapt to the user's behaviour [43][44][42], where social cues are nonverbal aspects of our communication with one





**Figure 2.5:** An example of job interview simulation in a non-immersive virtual environment [42]

another (e.g. eye contact, posture, tone of voice).

There are few job interview simulations requiring to use a Headset and two of these are “Be Confident” [8] (in Figure 2.6) and Bodyswaps “Job interview simulator” [9]. The first one requires that the user is asked three questions, which are always the same, and has the possibility to answer them. On the other hand, Bodyswaps application allows to choose between three responses to every question: the user has to pronounce one of them and an AI analyze the social cues to give advices to improve, as it is possible to see in Figure 2.7. In addition to this, the users have the possibility to see their virtual avatar answering the question from the point of view of the interviewer. However, there is no empirical evidence about the effectiveness of this change of visual perspective and avatars, in comparison to more traditional methods only based on having a first-person perspective of the interviewee avatar.

## 2.6 Summary

To summarise, there are some important topics of this work that have been profusely examined by previous studies, but the union of some of these has not been considered yet.



Figure 2.6: BeConfident job interview simulation [8]

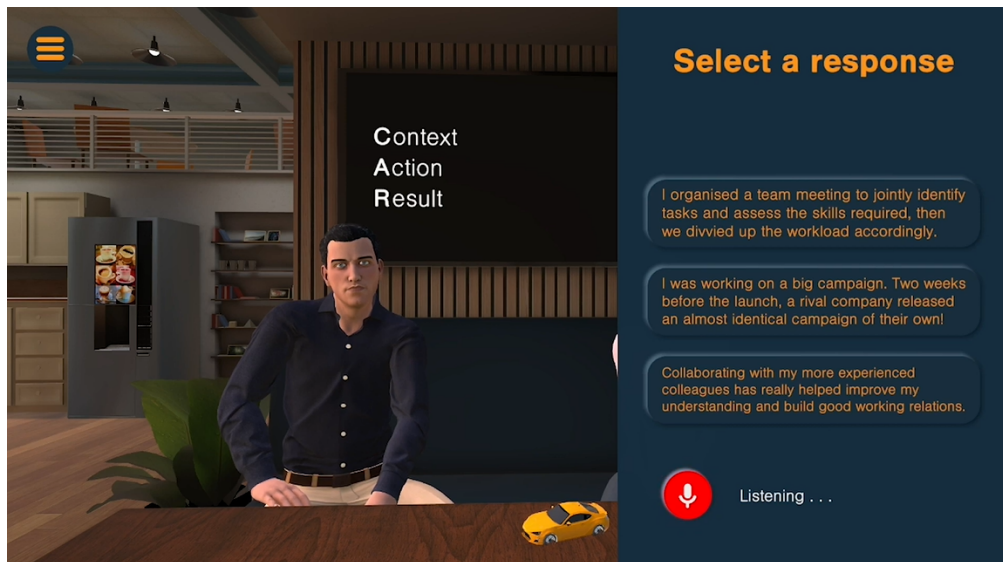


Figure 2.7: Bodyswaps job interview simulation [9]

In particular, it is evident that sense of presence and embodiment can be triggered by both First Person Perspective (1PP) and Third Person Perspective (3PP), even though results are generally better with the first one.

Moreover, it has been proven that training soft skills in Virtual Environments is effective and that this kind of simulation in Immersive Virtual Reality can induce anxiety in the participants and, at the same time, help them to handle

it. The research carried out until now has focused mainly on public speaking: in these studies the participants could see themselves from the 1PP of the avatar giving the speech and from a 3PP. Here a third configuration, similar to one of the conditions taken into account in this work, is absent: the possibility to experience the speech from the 1PP of someone in the audience, who would represent a person with the role to assess the performance. On the other hand, the study related to self-counselling with Freud tested the equivalent of this configuration, since the participants could give themselves advice from the 1PP of Freud: this proved that participants were able to distance themselves from their subjective point of view and to change perspective. The aim of this work is also to verify if this effect can be triggered also in the case of a job interview. However, in the mentioned study, participants could experience only 1PP of their avatar and 1PP of Freud, but not an external point of view, too. The comparison between these two ways to change perspective (3PP and 1PP of the person who should assess you or give you advice from an objective point of view) is missing in the state of the art and has been explored in this work.

In conclusion, this work does not only aim to implement a tool to practise for a job interview, since there are already some softwares in Immersive Virtual Reality focused on this type of applications and past studies have proven the effectiveness of such applications. Rather, the purpose is to explore two configurations that lead to an objective point of view and that have not been compared yet, in order to find out which one could be more useful to train soft skills for a job interview.

## Chapter 3

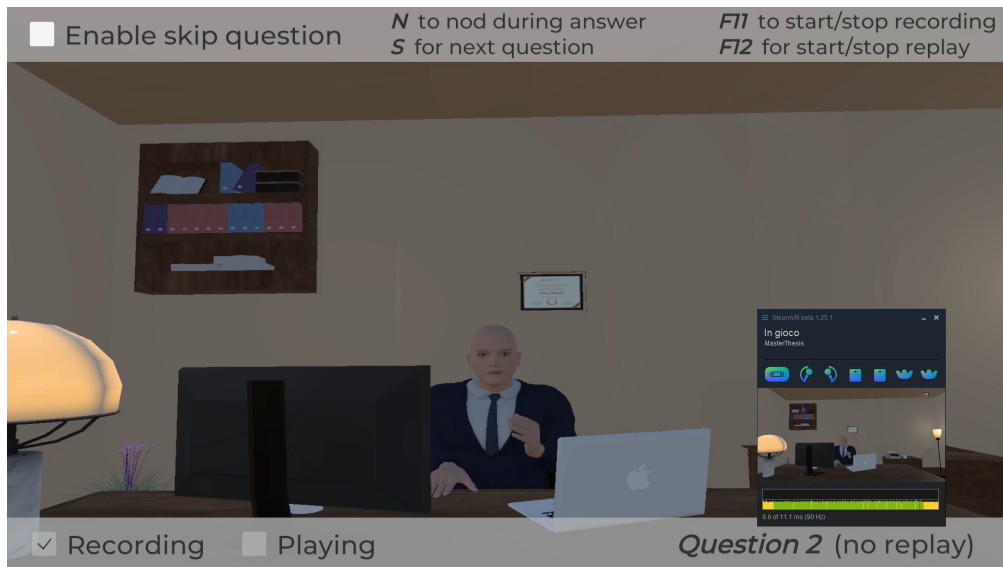
# Materials and methods

### 3.1 Experimental design

The experiment consists of three conditions, one of which represents a "Baseline" since the user experiences only the perspective of the interviewed. The main focus of this work is on the comparison between the other two, which are the "*External POV*" and the "*Interviewer's POV*" conditions: they enable the participants to switch respectively to an external point of view and to the one of the interviewer.

Each condition starts with the interviewer asking a question to which the participant has to answer through voice and gestures, as it is possible to see in Figure 3.1: this answer is recorded. Then, in the conditions which involve a change of perspective, the users watch the recording of the interviewer asking the question and of themselves answering it. For each condition there are two questions, thus at the end of the experiment, the participants will have answered six questions in total. To make clearer how the experiment takes place, it is possible to call *A* the phase when the user listens to the question and answers, and *B* the repetition of this phase (for the "Baseline" condition) or the phase involving switch of perspective ("*External POV*" and "*Interviewer's POV*"): with this naming convention the structure of the whole experiment is *ABAB ABAB ABAB*.

The order of the conditions and of the questions is random to the participants, but the experimenter is able to change it in the calibration scene of the application. The reason for the random order is that it is important to assure that the results do not depend on a precise sequence of questions and conditions or pair question-condition. For example, if the participants find the "*Interviewer's POV*" easier, it has to be clear that it is not because the questions of this condition were easier to answer or because this condition has been experienced as the last one.



**Figure 3.1:** View of the user while the interviewer is asking a question

### 3.1.1 Condition A: "Baseline"

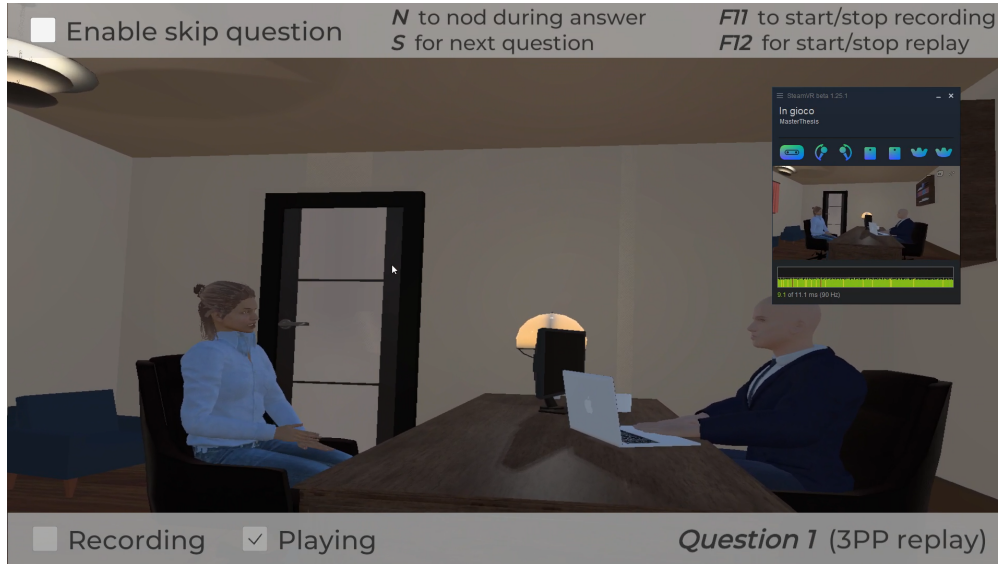
This condition consists of a single phase where the participants listen to the interviewer and answer the question from the perspective of the interviewee, but this phase is repeated immediately afterward. In fact, contrary to the other two conditions, the “Baseline” has been designed without a replay because it represents the current state of the art of job interview simulations in VR, where there is no perspective swap. In addition to this, it represents the equivalent of a job interview in the real world.

However, in order to have the same exposure to all the questions, there is a repetition. In fact, in the other two conditions, every question and answer are listened to twice: the first one during the practice and the second one during the replay. Thus, since in the "Baseline" there is no replay, the questions are practiced twice: this assures a fully counterbalanced design. At the same time, the experiment tests the effectiveness of repeating twice the same question in order to improve the soft skills for a job interview.

### 3.1.2 Condition B: “External POV”

The first phase of this condition is the same as the "Baseline": the interviewer asks a question and the participant answers. Then, the user experiences a change of perspective and sees the same scene from a Third-Person Perspective. In particular, a camera is positioned at eye level next to the desk between the avatars of the

interviewee and the interviewer. Thus, the user will be exactly in the middle between the avatars, as it is possible to see in Figure 3.2. The fact that the Point-Of-View is external implies that the user does not have any virtual body. From this perspective, the user watches the recording of the interviewer asking him or her the current question and of his or her answer to it.



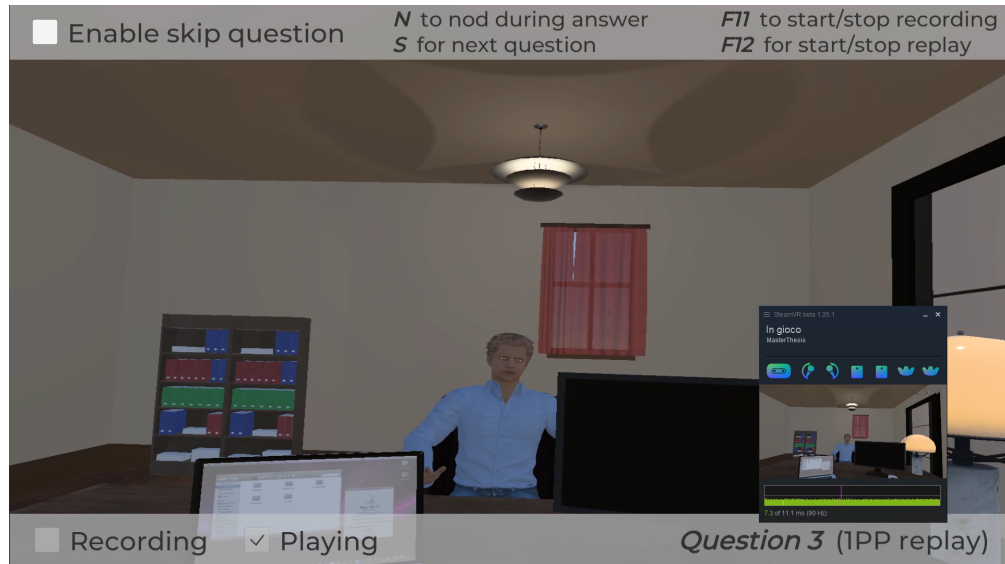
**Figure 3.2:** Replay from an external point of view

According to the state of the art, while experiencing the external Point-Of-View, the participants should feel "safer" [16] and be able to self-distance from themselves in order to manage anxiety and decrease it [6][5]. What is expected from this condition is that the participants will learn to assess their performance from an objective point of view.

### 3.1.3 Condition C: “Interviewer’s POV”

This condition, too, implies that during the first phase the participants practice the answer to a question from the interviewer, and in the second one they watch themselves being asked and answering the same question. However, in this case, they see their own avatars from the perspective of the interviewer, as it appears in Figure ???. In particular, they have full control of the interviewer’s body, in order to better trigger the embodiment towards his avatar. The aim is to test if this leads to self-distancing and to a change of perspective in self-assessment of the performance during the interview. More in detail, this condition should demonstrate if the perception that the participants have of themselves is altered or not by the fact that they see themselves from the point of view of the person

with the role of judging the outcome of the interview and the suitability for a job position.



**Figure 3.3:** Replay from the interviewer's point of view

A similar configuration to this one has been already implied before in the state of the art. For example, a study on "Embodied Perspective as a Victim of Sexual Harassment" [45] involved a phase where a male participant embodied an avatar among other men harassing a woman in front of them. Then, a second phase enabled the participant to experience the same situation from the body of the woman, having full control of her body.

### 3.1.4 Questions of the interview

The questions selected for the job interview simulation are the result of discussions with experts in the field of Human Resources, who suggested the main topics that come up in these situations. These questions, which are six in total as mentioned before, are the following.

1. Could you tell me about yourself and briefly describe your studies and professional past experience?
2. Do you prefer working independently or on a team? Why?
3. What did you do in the last year to improve your knowledge?
4. When you're balancing multiple projects, how do you keep yourself organized?



5. How do you deal with pressure or stressful situations?
6. What type of work environment do you prefer?

Considering that the participants of the experiments were all Italians, during the simulation the questions have been asked in their mother tongue, in order not to provoke additional anxiety related to speaking English and alter the results.

### 3.1.5 Configuration of questions and conditions

It is important that during the experiment the occurrence of each question and each condition is balanced in order not to generate bias in the results. For this reason, their order has been designed before starting the experiments and is represented in Table 3.1. In this table *FPP* refers to the "*baseline*" condition, while *EPOV* is the "*External POV*" condition and *IPOV* the "*Interviewer's POV*" condition. Each row of the table is related to a participant: each entry of the row contains the name of a condition and the questions asked during that condition, exactly in the order they had during the experiment. The pattern represented by Table 3.1 has been repeated every six participants. The number of the questions (e.g., Q1, Q2 etc.) refers to the numbered list of the previous paragraph.

| Id | 1st condition and questions | 2nd condition and questions | 3rd condition and questions |
|----|-----------------------------|-----------------------------|-----------------------------|
| 1  | FPP (Q1-Q6)                 | EPOV (Q2-Q5)                | IPOV (Q3-Q4)                |
| 2  | IPOV (Q2-Q5)                | EPOV (Q3-Q4)                | FPP (Q1-Q6)                 |
| 3  | FPP (Q3-Q4)                 | IPOV (Q1-Q6)                | EPOV (Q2-Q5)                |
| 4  | IPOV (Q4-Q3)                | FPP (Q6-Q1)                 | EPOV (Q5-Q2)                |
| 5  | EPOV (Q5-Q2)                | IPOV (Q4-Q3)                | FPP (Q6-Q1)                 |
| 6  | EPOV (Q6-Q1)                | IPOV (Q5-Q2)                | FPP (Q4-Q3)                 |

**Table 3.1:** Order in which conditions and questions were administered during the experiment

## 3.2 Sample

Since the topic of the application is a job interview and it has been developed in a university context, the main target of this experiment is students in their last year, eventually searching for a job at the moment. However, practicing for a job interview could be useful also to other students or to people who have already completed their studies or even work. For this reason, during the recruitment for the tests, no restrictions have been applied to the requirements to participate. As



a consequence, the age range of the participants is wide, from 21 to 31. Since this gap in age could implicate a different level of experience in job interviews, the results are examined while taking into account how many job interviews each participant has already done in a real context, too.

### 3.3 Measures

During each experiment, the results on the effectiveness of the application for the training of soft skills and on embodiment have been measured through five questionnaires in total. The first one was administered before starting the experience. Then, since it is important to keep track of how users perceive each condition, one questionnaire was administered after each of them. Finally, the last questionnaire was administered at the end of the experiment.

Even though some questions are personalized to fit with the experiment, they originate from validated questionnaires used in some studies mentioned before.

#### 3.3.1 Initial questionnaire

The initial questionnaire contains some demographic information about the participant and other questions about her or his current emotional state.

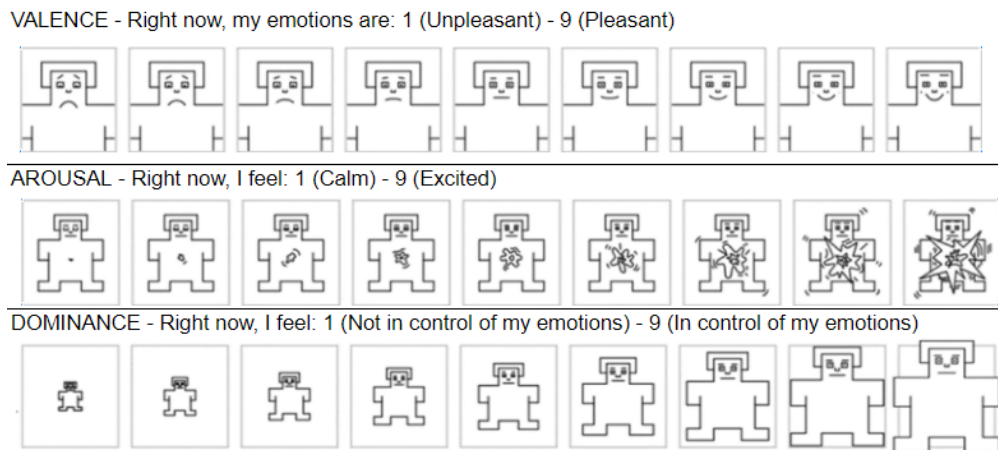
More in detail, the demographic information required are:

- age
- gender
- professional status (Bachelor Student, Master student, PhD Student, Post-doc Researcher, Employee, other)
- Italian level (because the questions during the interview are in Italian)
- English level (because the questionnaires are in English)
- experience with programming
- how much she or he uses tools for Virtual Reality
- how much she or he plays video-games
- her or his field of expertise

The questions about experience with programming, VR, videogames, and the field of expertise can be useful to understand the results of all the questionnaires, particularly concerning the difficulty to use the system and its quality.

After these general questions, the participants were asked about their emotions while they were filling in this questionnaire: the results of this part are used as a baseline in comparison to the same questions asked after each condition. This part helps to understand, for example, if the experiment evoked negative or positive emotions in the participant because she or he already was in this mood or not. The choice of questionnaires to be used for the assessment of the participants' emotional states relied on two validated questionnaires widely used in psychology and human-computer interaction: the *Self-Assessment Manikin* and the *State-Trait Anxiety Questionnaire*.

- The *Self-Assessment Manikin* (SAM) [10] is a picture-oriented assessment technique that enables to directly measure the valence, arousal, and dominance of a person's emotional reaction to a high range of stimuli. *Valence* indicates whether the elicited emotion is positive (high values) or negative (low values), while *arousal* measures how much the elicited emotion is "activating", thus how much the participant is excited (high values) or calm (low values), and through *dominance* the participant conveys if he or she feels "in control" of the emotion (high values) or controlled by it (low values). The SAM can be administered using different scales: a 5-point scale, a 7-point scale, or a 9-point scale. The preferred scale for this experiment was the 9-point scale, to give the participants as many options as possible to express their emotions. This technique has been chosen because it represents a quick and easy method to register the emotional state of the participants. The questions are visible in Figure 3.4.



**Figure 3.4:** Self-Assessment Manikin questionnaire administered during the experiment

- The *State-Trait Anxiety Inventory* questionnaire (STAI) [46] consists of a

validated instrument for measuring anxiety in adults. It differentiates the "state anxiety", which is a temporary condition, from the "trait anxiety", which, on the other hand, is more general and long-standing anxiety. This questionnaire is used worldwide and it includes forty questions with a range of 4 possible responses to each. However, for this experiment, we used a validated version of the STAI that only includes six items [47]. The purpose of these questions is to evaluate how much the subject is anxious before starting the experiment in comparison with the results of the STAI administered after each condition, in order to understand how much the VR simulation affected the rate of the subject's anxiety. This questionnaire can be seen in Figure 3.5.

**STAI (pre) | 1 (Not at all), 2 (Somewhat), 3 (Moderately), 4 (Very Much)**

---

**Right now I feel calm:**

---

**Right now I am tense:**

---

**Right now I feel upset:**

---

**Right now I feel content:**

---

**Right now I feel relaxed:**

---

**Right now I feel worried:**

**Figure 3.5:** Shortened version of the STAI questionnaire administered during the experiment

### 3.3.2 Post-condition questionnaires

After each condition, the participants had to complete a questionnaire concerning different topics: embodiment, presence, plausibility and usefulness of the experience, and emotional reactions and impact.

- The questions on embodiment (in Figure 3.6) are adapted from the *ad-hoc* questionnaire realized for a study on counseling in VR through embodied self-dialogue [7] and a standardized questionnaire on Avatar Embodiment [48]. Both questionnaires use a 7-point Likert scale. These questions measure ownership, agency, and self-location towards the virtual bodies of both the interviewee and the interviewer. They are divided into two parts, one referring to the moment in which the participants answer the questions of the interviewer while being in the body of the interviewee and the other to the moment in which the participants watch the replay of their response. The questions referring to the embodiment of the interviewer's avatar in the first part (while answering) are expected to always produce low values.

- The questions on presence and plausibility (in Figure 3.7), too, are adapted from Slater et al. [7] and use a 7-point Likert scale.
- The usefulness of the experience was measured with *ad-hoc* questions focused on the effectiveness of the perspectives in improving job interview skills and eliciting objective self-assessment of verbal and non-verbal responses to the questions of the interview.
- Finally, emotional reactions and impact were measured respectively with SAM and STAI.

According to the condition just tested, the questions of the post-condition questionnaire slightly change concerning embodiment and the usefulness of the experience. In fact, after the "baseline" condition, the questions on embodiment do not include the part on the replay. Then, in this case, the questions on the usefulness of the experience do not include the part referring to the change of perspective. As far as

| Embodiment Answering  |  |
|---|--|
| <b>INTERVIEWEE Embodiment   1 (Strongly disagree) - 4 (neutral) - 7 (Strongly Agree)</b>  |  |
| Although the virtual body of the <b>interviewee</b> did not look like me physically, I felt that the virtual body of the <b>interviewee</b> was my own body.                                |  |
| I felt that the movements of the virtual body of the <b>interviewee</b> were caused by my own movements.  |  |
| I felt as if my <i>body</i> was located where I saw the virtual body of the <b>interviewee</b> .  |  |
| I felt "out of my body".  |  |
| <b>INTERVIEWER Embodiment   1 (Strongly disagree) - 4 (neutral) - 7 (Strongly Agree)</b>  |  |
| Although the virtual body of the <b>interviewer</b> did not look like me physically, I felt that the virtual body of the <b>interviewer</b> was my own body.                                |  |
| I felt that the movements of the virtual body of the <b>interviewer</b> were caused by my own movements.  |  |
| I felt as if my <i>body</i> was located where I saw the virtual body of the <b>interviewer</b> .  |  |
| I felt "out of my body".  |  |
| Embodiment Replay   |  |
| <b>INTERVIEWEE Embodiment   1 (Strongly disagree) - 4 (neutral) - 7 (Strongly Agree)</b>  |  |
| Although the virtual body of the <b>interviewee</b> did not look like me physically, I felt that the virtual body of the <b>interviewee</b> was my own body.                                |  |
| I felt that the movements of the virtual body of the <b>interviewee</b> were caused by my own movements.  |  |
| I felt as if my <i>body</i> was located where I saw the virtual body of the <b>interviewee</b> .  |  |
| I felt "out of my body".  |  |
| I had the illusion that when I saw the avatar of the interviewee answering the questions from an external point of view/interviewer perspective I was seeing my virtual body.               |  |
| <b>INTERVIEWER Embodiment   1 (Strongly disagree) - 4 (neutral) - 7 (Strongly Agree)</b>  |  |
| Although the virtual body of the <b>interviewer</b> did not look like me physically, I felt that the virtual body of the <b>interviewer</b> was my own body.                                |  |
| I felt that the movements of the virtual body of the <b>interviewer</b> were caused by my own movements.  |  |
| I felt as if my <i>body</i> was located where I saw the virtual body of the <b>interviewer</b> .  |  |
| I felt "out of my body".  |  |
| I had the illusion that when I saw the avatar of the interviewer asking and listening to the questions from an external point of view/interviewer perspective I was seeing my virtual body. |  |

**Figure 3.6:** Questions on embodiment (FPP: baseline - EPOV: external POV - IPOV: interviewer’s POV)

SAM and STAI are concerned, they do not change from one condition to another and are identical to the ones in the initial questionnaire, since they are supposed to be compared. In fact, the aim is to evaluate the difference between the beginning of the experiment and a precise condition. The only difference is in the moment the questions refer to: while in the initial questionnaire the questions are “*Right now* my emotions are / I feel ...”, in this case, they are “*During the VR condition* my emotions are / I feel ...”. Whereas, questions on presence and plausibility are the same for every condition.

| Presence and plausibility   1 (Strongly disagree), 7 (Strongly agree)  |
|--|
| I had the sensation of being in a job interview situation. (+7 corresponds to the normal sensation of being in a job interview)  |
| When I think about the VR experience, I remember the virtual office as a place where I have been.                                |
| I felt that the virtual experience was realistic.  |
| I had the feeling of sharing the virtual office with the other person (the interviewer), as if we were really in the same place. |
| I felt that the virtual interviewer was realistic.   |

Figure 3.7: Questions on presence and plausibility

### 3.3.3 Final questionnaire

At the end of the experience, the participants answer some general *ad-hoc* questions on job interviews in a real context. In particular, they are asked how many job interviews they have performed, how good they consider themselves in answering job interview questions, and how anxious they are in these situations.

Then, in order to better evaluate the results of the questionnaires related to emotions, there are some questions taken from the *Social Phobia Inventory* (SPIN) [49], as it is possible to see in Figure 3.8. This is a 17-item questionnaire that has been developed by the Psychiatry and Behavioral Sciences Department at Duke University to measure the severity of social anxiety disorder. In this experiment this questionnaire has great importance, considering that eventual high values in the anxiety of participants could be provoked either by, for example, the intimidating nature of the interviewer or by the inclination of the subjects. As a consequence, taking into account these parameters helps us in evaluating the realism of the experience and the usefulness of the experience in controlling the anxiety rates, which should happen according to the State of the Art.

| SPIN   1 (Not at all), 5 (Extremely)                                 |
|--|
| I am afraid of people in authority                                   |
| I am bothered by blushing in front of people                         |
| Parties and social events scare me                                   |
| I avoid talking to people I don't know                               |
| Being criticized scares me a lot                                     |
| I avoid doing things or speaking to people for fear of embarrassment |
| Sweating in front of people causes me distress                       |
| I avoid going to parties   |
| I avoid activities in which I am the center of attention             |
| Talking to strangers scares me                                       |
| I avoid having to give speeches                                      |
| I would do anything to avoid being criticized                        |
| Heart palpitations bother me when I am around people                 |
| I am afraid of doing things when people might be watching            |
| Being embarrassed or looking stupid are among my worst fears         |
| I avoid speaking to anyone in authority                              |
| Trembling or shaking in front of others is distressing to me         |

Figure 3.8: SPIN questionnaire

Then, the final questionnaire includes some questions to evaluate the usability of the whole system (both hardware and software) through some questions from the *Post-Study System Usability Questionnaire* (PSSUQ) [50] and the questionnaire

on the *Quality of Environment of an immersive Virtual Reality environment* [51]. These questions are visible in Figure 3.9.

| Quality of the System [PSSUQ]   1 (Not at all), 5 (Extremely)            |
|--|
| I was able to complete the tasks and scenarios quickly using this system |
| I felt comfortable using this system                                     |
| The interface of this system was pleasant                                |
| I liked using the interface of this system                               |
| Overall, I am satisfied with this system                                 |

| Satisfaction [QoE]   1 (Strongly Disagree), 5 (Strongly Agree)              |
|---|
| I enjoyed experiencing the virtual environment                              |
| The system was easy to use  |
| I needed to learn a lot of things before I could get going with this system |
| I would have liked more time in the virtual environment                     |
| My experience did not meet my expectations                                  |

**Figure 3.9:** Questions on the quality of the system

Finally, the last questions invite the participants to draw up four rankings of the three conditions considering respectively which one helped them to better improve their job interview skills, which one helped them to assess their verbal responses more objectively, which one helped them to assess their non-verbal responses more objectively and which one they preferred overall. To conclude, they are asked to give two positive comments and two negative comments on the experiment.

### 3.4 Hardware and software

The application has been initially conceived for Oculus Quest 2, since it is a standalone device: this means that it does not need to position base stations in the room for head and hand tracking and, as a consequence, it allows the user to be more comfortable, without having to deal with wires connected to the headset. In addition to this, the choice of this device was due to its major diffusion in comparison with other headsets, thus this would have been a key feature for an application of this type. Last but not least, the use of Oculus would have allowed the user to take advantage of the hand and finger tracking without needing the controllers: in fact, it is possible to enable this feature on Oculus Quest 2 and use the inside-out cameras of the headset to detect the position and orientation of hands and the disposal of fingers without needing more hardware. Therefore, in this configuration, the user would have needed to wear only a headset, and so she or he would have experienced interactions that are as natural as possible.

However, during the first phase of the implementation, a few problems appeared. In fact, the first trials to test the accuracy of the Inverse Kinematics, which is necessary to provide the user with control of the virtual character's body, led to

poor results. In particular, the elbow movements of the real person and the ones of the virtual body appeared incongruous: this undermined the embodiment, which is fundamental in this work. For this reason, other devices have been added to the system: two VIVE Trackers 2018 and two straps in order to put them on the elbows of the user. As a consequence, two base stations for tracking were needed, too.

The use of Vive Trackers required Steam's tool for Virtual Reality as software, that is SteamVR. However, no evidence of compatibility between Oculus Hand Tracking and SteamVR has been found. For this reason, the Oculus headset was no more necessary. In fact, in order to use the VIVE devices and Oculus together, it was only possible to use VIVE's tracking system: this means that another VIVE tracker was required on the Oculus headset, resulting in complications that are not worth the effort. This led to the current configuration of the system.

### 3.4.1 Hardware

As far as the hardware is concerned, the configuration involves a headset HTC Vive Pro, two SteamVR base stations 2.0, two VIVE Trackers 2018, two Valve Index controllers, and two wrist straps used as elbow straps.

#### HTC Vive Pro

HTC Vive Pro headset (in Figure 3.10) is a tethered device produced by HTC. As a tethered device, in comparison with a standalone one, it allows high-end virtual experiences.

More in detail, the Vive Pro's screen is a dual AMOLED 3.5" diagonal and has a resolution of 1440x1600 pixels per eye, with a refresh rate of 90Hz and a field of view of 110 degrees. The headset includes Hi-Res certificate headphones and integrated microphones. As far as the sensors are concerned, this HMD is provided with SteamVR Tracking, G-sensor, gyroscope, proximity, and Eye Comfort Setting (IPD).

#### Valve Index controllers

In order to have an interaction as natural as possible, Valve Index Controllers (in Figure 3.11), also known as "Knuckles", have been chosen for this work: a "floating" strap enables the user to hold the controllers with open hands. In addition to this, thanks to sensors for thumb and finger positions and pressure sensitivity, these controllers have advanced finger-tracking capability.

The tracking of these controllers is done by SteamVR 2.0 sensors compatible with SteamVR base stations 1.0 and 2.0. Their haptic feedback is due to a High Definition Resonance Linear Actuator (HD LRA) and each of them includes a



**Figure 3.10:** HTC Vive Pro

Track Button with Force Sensor, a Grip Force Sensor, Finger Tracking, and an Inertial Measurement Unit.



**Figure 3.11:** Valve Index controllers



### 3.4.2 Software

The game engine used to develop the VR application is Unity, together with the OpenVR SDK, and the programming language is C as a consequence. The pairing of the devices with the PC and settings of the controllers' interactions required, as mentioned before, SteamVR software. Some Unity assets were necessary.

#### Unity

*Unity* is a cross-platform game engine developed by Unity Technologies and is considered easy to use for beginner developers. It can be used to develop three-dimensional and two-dimensional games, as well as interactive simulations and the like. The version used for this project is 2020.3.36f1.

#### OpenVR

*OpenVR* is a Software Development Kit and Application Programming Interface developed by Valve for supporting SteamVR and other Virtual Reality headsets devices. It is used as the default API and runtime by the SteamVR platform and it represents the interface between the VR hardware and software.

#### SteamVR

*SteamVR* is a runtime embedded in the Steam client to power Virtual Reality experiences. It allows to experience VR content by means of different types of devices, since it supports Valve Index, HTC Vive, Oculus Rift, Windows Mixed Reality, and others. SteamVR enables users to manage several areas of their VR experience, such as the room setup to define the play area, and the device status and management to update firmware, pair new devices, change audio settings, mirroring, and customize some features.

#### FinalIK

RootMotion's *FinalIK* is a Unity asset that has been used in order to set the *Inverse Kinematics* of the avatars the user has control of. It is written in C, it is optimized for great performance, and is provided with a warning system to safeguard from null references and invalid setups. Two *FinalIK* components were used: *VRIK* and *FullBodyBipedIK*. *VRIK* is a high-speed full-body solver dedicated to Virtual Reality avatars. However, since it was not satisfying for the movements of wrists, the rotation of arms has been managed using *FullBodyBipedIK*.

## Lipsync SALSA

In this work the avatars' movement of lips while talking was generated with Crazy Minnow Studio's *SALSA LipSync Suite* and, during the recording phase of the experiment, with its plugin *SalsaMicInput*. *SALSA* (Simple Automated LipSync Approximation) is a real-time system to create lip synchronization from an audio input without needing for phoneme mapping or baking. This allows to implement lipsync into a project with minimal effort and time.

Another feature of Salsa is the *Eyes* component. *Eyes* is an eye and head tracking and random generation system which creates realistic eye movements and supports a cross-eyed look when targets are close.

A strong advantage of Salsa is the presence of presets for different supported model types called "*OneClicks*", which simplify the settings of all the parameters necessary for a realistic result. In fact, by applying a *OneClick* on a supported model, a *SALSA* module with visemes configuration, an *EmoteR* instance for emphasis emotes, and a *Eye* module with head, eyes, and eyelids configuration are automatically configured.

As mentioned before, also a specific plugin of *SALSA* was necessary in this project: *SalsaMicInput*. This plugin adds the real-time Unity microphone as input for the settings of the *SALSA* component and only requires a Unity Audio Source for microphone input operation.

## Record and Play

Ali Davis' *Record and Play* is a Unity asset that makes it possible to record what happens in the scene and then watch it. In particular, it can capture transforms of Game Objects and custom events. However, it does not allow saving the recording in the build of the Unity project, so some features have been added to it.

This tool has been conceived as a per-object recording tool, thus it allows to optimize recordings for a specific use case. Custom events and metadata can be added to the recordings, too. Finally, developers can provide custom event handlers called during a playback that correspond to the times they were logged in the recording.

## MBLab and Realillusion Character Creator

For the generation of the avatars, two for the user (male and female) and one for the interviewer, two different tools to generate custom avatars have been used. Some initial trials involved *MakeHuman*, but it has been discarded since the blendshapes that it generates are more related to emotions than to speech. This led to using the *MBLab* for the avatar of the interviewer, an add-on for Blender which enables the generation of characters with a high level of customization: one can choose

the gender, height, age, proportions, the color of skin and eyes, and many other attributes. The model generated by means of this tool is already provided with a rig and blendshapes simulating the phonemes. On the other hand, *MBLab* does not provide a set of clothes to put on the 3D model and not even hair: the only way to do it is to create or find some 3D models of clothes, adapt them to the *MBLab*'s template character and makes use of a proxy embedded in the add-on in order to adapt and attach the clothes to the body. The result of this process is good, even if it is not perfect, but the process itself can take a lot of time.

For this reason, in the last phase of this project, another tool was employed to generate the avatars of female and male users: *Realillusion Character Creator*. The free version of the software has been sufficient to create a personalized character with clothes on already. It has been selected among the softwares that are compatible with *SALSA*'s "*OneClicks*". As a consequence, the process to both create the avatar and set its configuration for *SALSA* became much shorter.

## **Blender**

Blender is a free and open-source 3D computer graphics software. More in detail, it supports the entire 3D pipeline: modelling, rigging, animation, simulation, rendering, compositing, and motion tracking. Blender's API for Python scripting enables advanced users to customize the application by developing specialized tools. Its interface uses OpenGL to provide a consistent experience.

## **Mixamo and Rokoko**

Mixamo is a 3D rigging and animation tool. It provides a large amount of high-quality 3D characters already rigged and animations. In addition to this, it allows also to upload and rig custom characters, and to animate them. This tool has been used to animate the avatar of the interviewer in the fastest and most realistic way.

However, since Mixamo was not able to rig the avatar of the interviewer, a tool for re-targeting has been used to solve this issue: *Rokoko* add-on for Blender. This free add-on assigns existing motion capture data onto the custom-rigged character.

## **3.5 Experimental procedures**

For each experiment, the first step after the participants have arrived is to illustrate what they are going to do. In particular, they apprehend that their task is to practice a job interview simulation in Virtual Reality, where they will answer six questions. The text used during the experiments is the following.

*You are going to experience a simulated job interview in Virtual Reality: the person in front of you in the virtual environment will ask you 6*

*questions typical of a job interview. The purpose is to test 3 different conditions through which you will be able to review the answers you gave to the interviewer. In the "Baseline" version you should rely only on your perspective, possibly going over your answers in your mind. In the "External POV", after each answer you will be able to watch yourself again from an external perspective. Finally, in the "Interviewer's POV" you will be able to watch yourself again from the point-of-view of the interviewer. For each condition you will be asked 2 of the 6 total questions.*

*Before starting the experiment, you will have to fill in a questionnaire. After each step, the application will signal you to leave the virtual environment to answer a questionnaire. The last questionnaire will be administered to you at the end of the experiment.*

*When the application starts, there will be a calibration moment, where I will give you instructions on what to do. After that, you will be able to begin the simulation.*

*If you are unable to continue with the experiment, you can ask me to stop it at any time and withdraw from the experiment without giving any explanation to the experimenter.*

The last sentence is important because, if a participant perceives the situation in the virtual environment as real and experiences too much anxiety, continuing with the experiment could be too stressful.

After the participant is informed about the content of the experiment, he or she is invited to fulfil a consent form and a module about generic information, such as age, gender and field of study. Some questions ask the participants about their previous experience with applications in Virtual Reality, with video games and with programming languages, in order to better assess the results of the questionnaires about the usability of the system and the quality of the experience. Then, the participant will compile a questionnaire to evaluate the emotions and the level of anxiety of the subject at that moment.

Once the preparation is ended, the user is ready to start the experiment. The tester helps him or her to wear the headset, the trackers and the controllers and explains which buttons of the controllers will be necessary for the application: the trigger and A and B buttons. Then, when the participant is ready, the experimenter starts the application.

### **3.5.1 Initial settings**

When the Virtual Reality application is started, the participant will enter a simple environment where the experimenter records some settings useful for the main part

of the experiment. In particular, a UI visible only to the experimenter allows her or him to give the trackers their role in the application (right or left elbow) and to choose the order of the questions and of the conditions. In addition to this, the user can choose an avatar (female or male). After this, the experimenter asks the participant to double-check in order to be sure to stay in the yellow square, which will become green when the correct position is reached, then to reach and hold the T-pose, so that the straight legs and arms of the person form a capital letter T. At this moment, the participant is asked to press the right trigger of the controller: this is to calibrate the height of the virtual character according to his or hers and to calibrate the so-called *goals* of the virtual character's elbows according to the current position of the trackers, i.e., the components of an IK chain used to control the direction that the chain bends in. Both the choice of the avatar's gender and the calibration of height can be done by the experimenter, too, by pressing respectively *C* and *U* keys on the keyboard. After this setting, the experimenter presses a toggle in order to prevent the participant from making other changes, then by pressing the *A* key of the keyboard she or he can give the participant the possibility to try the Inverse Kinematics settings and verify that everything is working fine. It is also possible for the experimenter to test if the audio is working properly in input and output: the *Enter* key of keyboard allows to play an audio and test output, while a toggle enables Salsa for the virtual character in order to verify that the microphone is working well. After everything is set, the experiment itself can begin: the user is invited to sit down, close his or her eyes, then the experimenter switch to the main scene by pressing *S* and asks the participant to re-open the eyes.

### 3.5.2 Familiarization with the environment

At the beginning of the experiment, it is important that the users, especially those who have never experienced Virtual Reality applications before, get used to interacting with the virtual environment, feel comfortable and perceive the virtual body as their own. For this reason, as soon as the main scene is loaded, the users are invited to familiarize themselves with their virtual body (avatar) from 1PP and with the environment. When they enter the scene, the participants see themselves in the Virtual Environment seated on a chair in front of the interviewer.

Firstly, they are invited to move their heads in order to observe the environment. The experimenter asks them where they are and to describe what they see.

There is also a virtual mirror to their right (visible in Figure 3.12), that allows them to see their avatar moving in accordance with their real-time movements. The experimenter invites them to look right and asks what they see. When they mention the mirror, the experimenter asks what happens if they move and in particular if their movements are coherent with what they see in the mirror. In



**Figure 3.12:** Moment in which the participant familiarizes with the environment

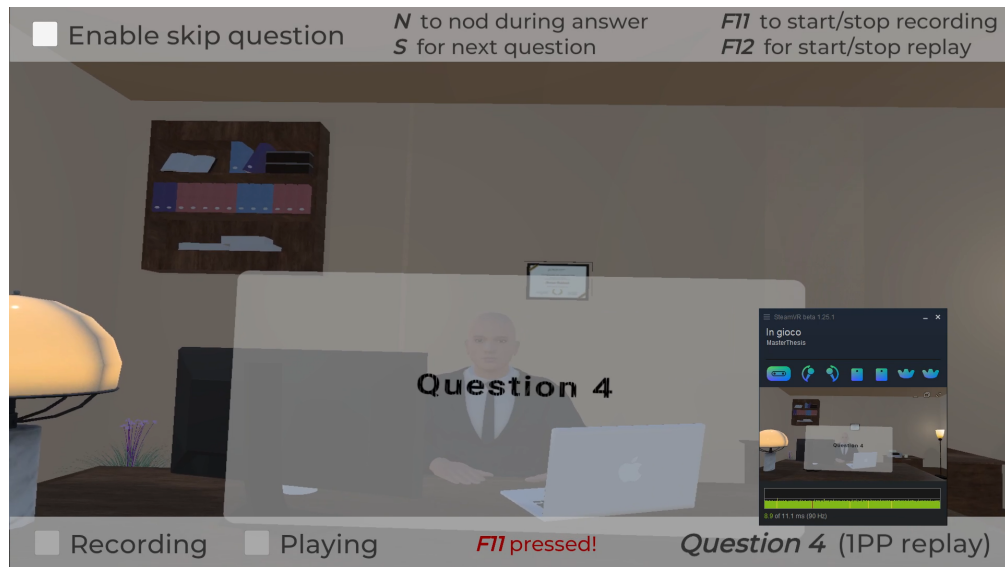
fact, the mirror is used to enhance embodiment through visuomotor correlations, as it happens in the study of self-counselling mentioned in Chapter 2 [7]. After this moment, the participants should perceive the virtual body as if it was their own.

### 3.5.3 First phase: simulation and recording

During the first phase of each condition, the interviewer asks a question to the participant and the answer is recorded. In two conditions out of three, the recording will be necessary for the second phase, while in the “Baseline” condition, it will be only used by the experimenter to better analyse each experiment.

The control of this phase is in the experimenter’s hands. After she or he presses the *F11* key, a panel appears in front of the user to say which question will be asked (e.g., “Question 1”, “Question 2” etc.), as it is possible to see in Figure 3.13. Then, when the panel has disappeared, the interviewer asks a question, and the recording of users’ movements and voices starts. Each of the six questions has pre-recorded audio and animations and the order of the questions changes in each experiment, as mentioned before. The fact that the recording starts while the interviewer is still talking, enables us to keep track of the reactions of the participants to the question. When the participants have finished answering, the experimenter presses the *F11* key again to stop the recording. Even though this phase is the same for every condition, the instructions that the experimenter gives to the testers before entering it are different.

- Before the first phase of the “*Baseline*” condition, the experimenter reminds



**Figure 3.13:** Panel signaling the current question

the participant that he or she is going to listen and answer one question twice and that this will happen for two questions in total. The first time in which the participant listens and answers the question represents the first phase of the condition.

- On the other hand, at the beginning of the first phase of the “*External POV*” and the “*Interviewer’s POV*” conditions, the experimenter explains to the participant that he or she is going to listen to one question and answer and invites him or her to observe what will happen next. After the participant has finished answering and the experimenter has stopped the recording, a UI panel appears in front of the participant to signal that the recording is ready and that he or she can press any button of the controllers to start the replay. Then, the second phase can start.

### 3.5.4 Second phase

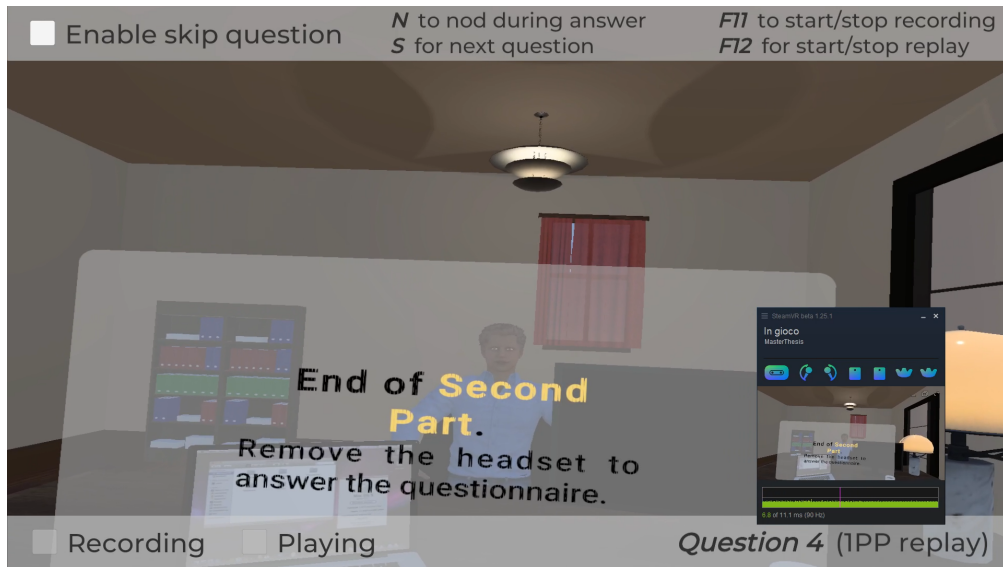
This phase of the conditions consists of a second exposure to the same question the participants have just answered to. However, it has a different configuration according to the condition.

- In the “*Baseline*” condition, the second phase starts when the experimenter presses *F11* again, enabling the participant to listen and answer the question again, and ends with the experimenter pressing the same key. Thus, in this condition first and second phases coincide.

- The second phase of “*External POV*” and “*emphInterviewer’s POV*” conditions begins when the participant presses any button of the controllers or the experimenter presses *F12*. This action starts the replay of the user answering the current question. When the replay has ended, the experimenter presses *S* to switch to the next question.

### 3.5.5 End of a condition and of the experiment

At the end of each condition, a UI panel (in Figure 3.14) appears in front of the user to signal that the condition is ended and to invite him or her to exit VR in order to fulfil a questionnaire. These questionnaires are related to embodiment, presence and plausibility, the usefulness of the experience, emotional reactions, and emotional impact during the condition. However, some questions change according to the condition, so the experimenter needs to administer the right questionnaire.



**Figure 3.14:** Panel signaling the end of a condition

When the last condition finishes, the UI panel signals the end of the experiment, too. In this case, the experimenter helps the participant to remove all the hardware, then she or he administers the questionnaire related to that condition and, in the end, the “Post-experiment questionnaire”. The latter consists of questions on social anxiety and the quality of the system.



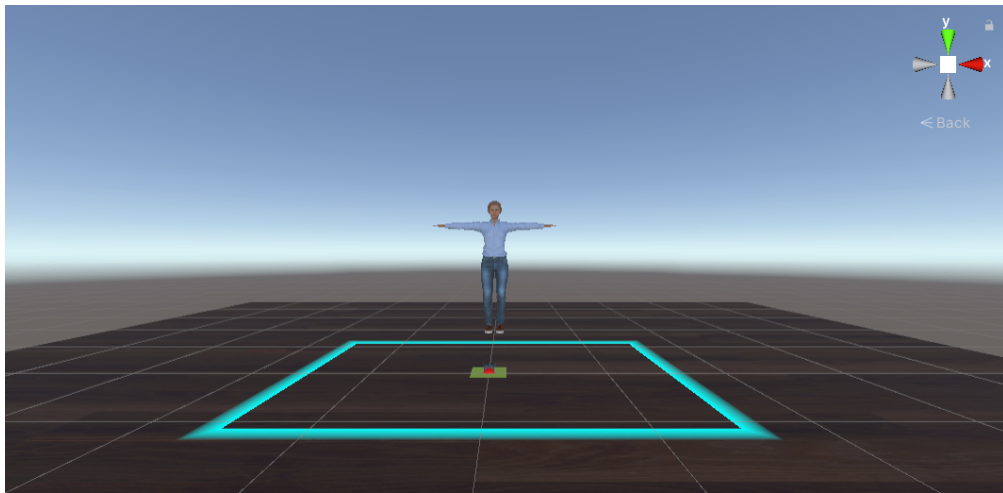
# Chapter 4

## Software implementation

### 4.1 Structure of the Unity project

The Unity project consists of four scenes in total.

- The *Calibration scene* (visible in Figure 4.1) is the starting scene of the application, where some important configurations are set.
- The *User scene* is a scene representing a virtual office, in which the *Camera rig* and the IK are set to the avatar of the user. Here the answering to the questions, thus the recording, takes place.
- The *3PP scene* is equivalent to the *User scene* as far as the 3D assets are concerned. However, in this case, the *Camera rig* is the child of an empty and no IK is set. It is dedicated to the replay phase of the *External POV* condition.
- The *1PP scene*, on the other hand, remains identical to the *User scene* in the aspect, but the *Camera rig* and the IK are set to the avatar of the interviewer. This scene is used for the replay of the *Interviewer's POV* condition.



**Figure 4.1:** Calibration scene

## 4.2 Preparation of main scene

The virtual experience takes place in the recruiter's office of an unknown company, as it is possible to see in 4.2. The elements that suggest this are mainly a desk with a computer on it, some shelves with documents, a printer and a graduation diploma. A mirror has been added to help the participants in embodying the virtual avatar representing themselves. All the 3D objects in the scene are free models found on



**Figure 4.2:** Overview of the main objects in the office

the website *Sketchfab* or free assets from the Unity Asset store. Some changes have

been applied to a few of them using *Blender* and some textures have been realized with *Substance Painter* to make the scene more realistic.

### 4.3 Implementation of the avatars

The implementation of the avatars required some steps, from the realization of the 3D models to the setting in Unity for the Inverse Kinematics and the lip sync.

#### 4.3.1 Realization of the 3D models

The avatar of the interviewer was realized using Blender's add-on *MBLab*, as it is possible to see in Figure 4.3. This tool allowed to choose age, gender and other attributes like the color of eyes and skin. However, it was necessary to manually add some clothes to the character. The 3D models of the clothes have been adapted to the template character of this add-on using Blender's tool *proportional editing*. Once this process has been completed, it was possible to apply the clothes to the character thanks to *MBLab's Proxy fitting* tool, which automatically adapted them to the customized character. However, before exporting the 3D model of the interviewer into Unity, it was necessary to apply all Blender's modifiers, since Unity does not recognize them. This phase took some days because the *Proxy fitting* tool is not always precise enough, so some adjustments with *proportional editing* and *weight paint* were necessary.

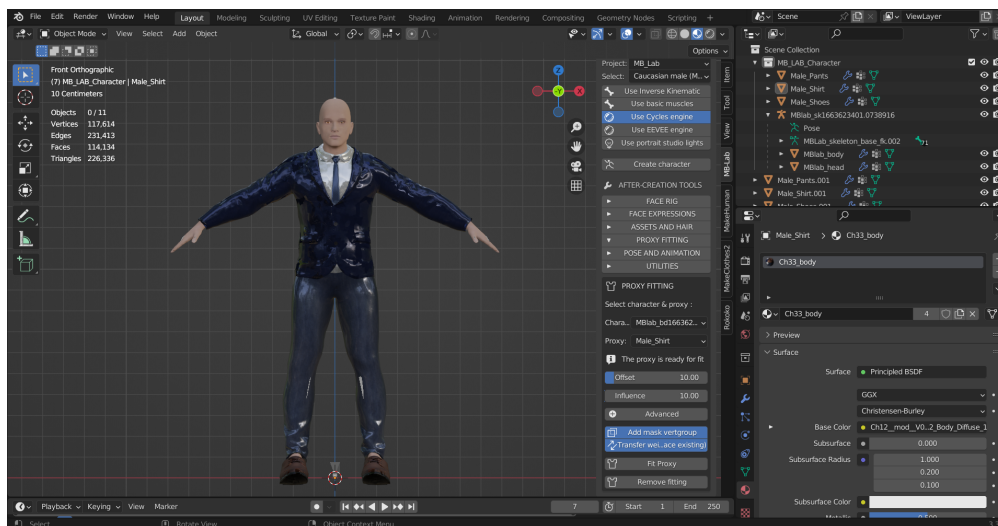


Figure 4.3: Avatar of the interviewer

For this reason, in order to speed up the realization of the characters, the two avatars for the users have been realized using the free version of Realillusion's

*Character Creator*. Personalized versions of the two template characters (male and female) have been used for the realization of this project. The results can be seen in Figure 4.4 and Figure 4.5.



Figure 4.4: Female user's avatar



Figure 4.5: Male user's avatar

### 4.3.2 Animations of the interviewer

The animations of the interviewer were realized using Mixamo, Blender, and Unity. The idle, the nod and one of the talking animations have been realized specifically

for this project, while the other two talking animations were downloaded from Mixamo and mapped on the rig of the character using Rokoko. Then, some work was necessary in Unity in order to mix all the animations and adapt them to each question. In fact, a combination of animations has been realized for each question. This result was achieved using Unity’s *blend tree* and some scripting.

In particular, the *default state* of the Animator is represented by the idle animation without *loop*. Then, the Animator (in Figure 4.6) automatically activates the Blend Tree (in Figure 4.7) when the idle animation ends. It is possible to exit the Blend Tree by pressing the *N* key, which activates the nod animation. As far as the Blend Tree is concerned, it is composed of a 1D Blend Tree between the idle animation and another Blend Tree: the transition from one to the other is managed by the float “isTalking”, which triggers the idle when its value is 0 and the Blend Tree when its value is 1. Using a float instead of a bool enables to make a smooth transition between the animations, so values between 0 and 1 will produce in-between animations. The inner Blend Tree is a 2D Freeform Cartesian using the float “Blend1” and “Blend2” as parameters: these values allow to make transitions among three types of talking animations.

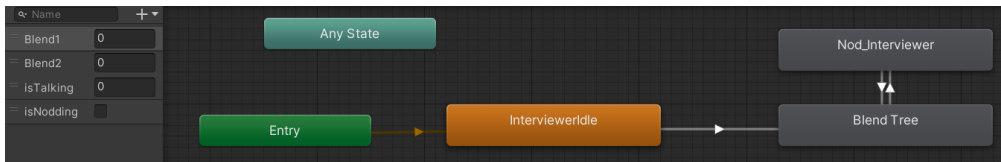


Figure 4.6: Animator of the interviewer avatar

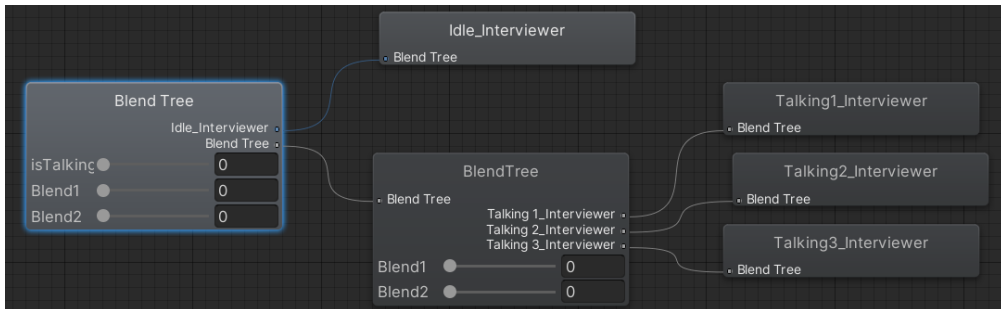


Figure 4.7: Main Blend Tree of interviewer’s animator

The Blend Tree is managed in the `Question.cs` class of the application. Here there is the definition of a `struct` called `QuestionItem` containing the number of a question (`int`), the audio clip related to that question (`AudioClip`), the condition of this question (`enum`), and a `Dictionary` called `animationValues`. This `Dictionary` is used to set the “isTalking”, “Blend1” and “Blend2” parameters of the inner Blend

Tree to the desired values in precise moments: the values and the instants in which they have to change have been identified taking into account which animation best suits the tone of the voice and the content of what is being said in a precise moment. An example on the setting of Dictionary value can be seen in Listing A.1.

When it is necessary to start the interviewer’s animation, a function in the `InterviewerPlay.cs` class calls a `Coroutine` which allows to switch from a value to the other smoothly. Then, the end of the audio clip triggers an `Event` that calls a `Coroutine` to interrupt all talking animations and go back to the idle.

### 4.3.3 Setting of IK

The setting of Inverse Kinematics (IK) required a lot of attention because it is fundamental for embodiment.

The main component that has been used is the `VR IK` component of the `FinalIK` Unity asset. It consists of a part called “References” where each bone of the rig is given the correct role in the IK chain. Then, in the “Solver” section it is possible to set *targets* and *goals*. On the one hand, a *target* represents an object which specifies the desired transform for the bone assigned to it. On the other one, a *goal*, as mentioned in Chapter 3, is the component of an IK chain used to control the direction that the chain bends in. In this work, it was necessary to create a target for the head, the pelvis, the hands and the feet, while goals were created for the elbows and the knees. Targets of head and hands were created by duplicating the corresponding bones of the rig’s hierarchy. Some adjustments for positions and rotations in *play mode* were needed. Then, the head target has been set as *child* of the `CameraRig`’s `Camera`, while the right and the left hand targets as *children* of the right controller manager and left controller manager (i.e. `SteamVR` prefabs `vr glove right` and `vr glove left`): this setting is due to the fact that head and hands have to follow the movements of the HMD and the controllers accordingly. Similarly, the goals for the elbows have been realized by duplicating the bones of the upper arms and moving them a bit further behind, as shown in Figure 4.8.



**Figure 4.8:** Position of elbow *goals* (represented by colored blue and yellow circles) for the female avatar

Then, they have been set as “children” of the Game Objects that manage the trackers. In order to have better results for the movements of wrists, hand targets of the VRIK component have been used to influence only the position of the avatars’ mesh and the elbow goals have not been set here. In fact, the influences of hand targets for rotation and the influence of elbow goals have been set only in the FullBodyBipedIK in the “Left Arm” and “Right Arm” sections. As far as the feet and pelvis are concerned, since the users are seated during the experiment, they have fixed positions that have been changed in play mode until the results became satisfactory. It is possible to see the hierarchy of targets and goals in Figure 4.9.

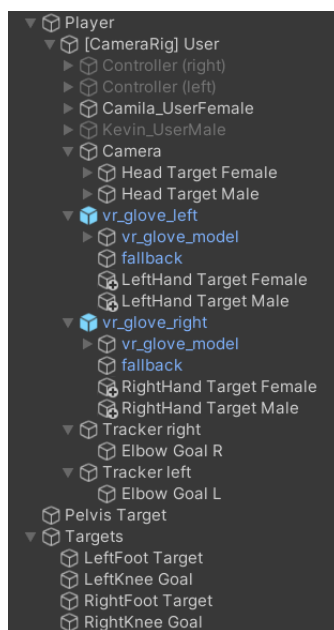


Figure 4.9: Hierarchy of the *Player*

#### 4.3.4 Setting of lip sync

The lip sync for the avatars was managed with the Unity asset *SALSA*. As mentioned in 3, for the avatars of users the setting was automatic thanks to *SALSA*’s OneClick for characters generated with Realillusion Character Creator. However, for the interviewer’s avatar it was necessary to realize the visemes manually. This operation has been done following the guidelines on *SALSA*’s documentation [52] and part of the result can be seen in Figure 4.10.

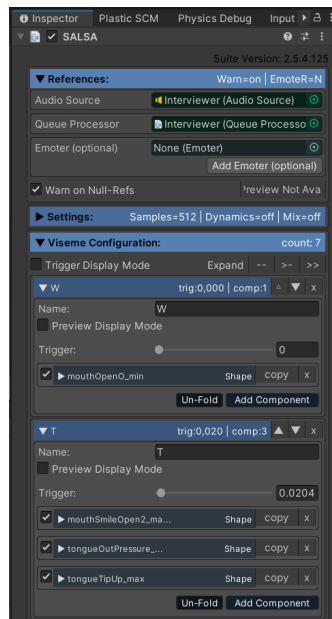


Figure 4.10: Setting of SALSAS for the interviewer’s avatar

## 4.4 Calibration and initial settings

The first scene the users enter when the application is started is dedicated to the settings and the calibration necessary for the experience. In particular, in this phase the participants can choose their avatar and calibrate their height, while the experimenter sets the trackers and the order of questions and conditions.

### 4.4.1 Selection of avatar

When the application is started, the participant sees in front of her or him a male or a female avatar. By pressing the trigger of the left controller, he or she is able to change the avatar. In fact, the press of this button de-activates a mesh and activates the other one.

### 4.4.2 Trackers setting and calibration

The SteamVR’s component `Steam VR Tracked Object`, used to reflect the movements of a Vive tracker on a Game Object, requires setting the index of the device it refers to. For this reason, in the calibration scene a User Interface (UI) was added to enable the experimenter to check which is the correct index for each tracker. In fact, in the scene there are two Game Objects with the `Steam VR Tracked Object` component, one for the left tracker and one for the right tracker. Both of them



have one *empty* and a 3D object as children: the *empty* represents the *goal* for the elbow, while the 3D object is a sphere, red for the right tracker and green for the left tracker. The spheres are used to visualize the positions of the trackers. In fact, when the movements of the red sphere match the ones of the right tracker and the movements of the green sphere match the ones of the left tracker, then it means that the device indexes of the `Steam VR Tracked Object` components are correctly set. Thanks to the UI, the experimenter can press “ - ” or “ + ” buttons to scroll all the devices until the index is the right one for each component.

In addition to the setting of the correct index for the trackers, another feature was added to have good tracking of the elbow regardless of how trackers are positioned on the arms of the participants. This ensures that the movements of the avatars’ elbows are as much accurate as possible. Two *empties* which can be referred to as “dummy goals” were positioned in the scene as they were the *goals* of an avatar (see 4.2.3), which was then deleted. So, every time one enters the calibration scene, the script `FollowY.cs` copies the Y coordinate of a tracker position into the position of the corresponding *goal*. However, since the “dummy goals” have been positioned with respect to the position of the avatar, the user should be in the same place in order to have the best result: for this reason, in the place of the avatar a yellow square has been added. When the user is on the yellow square it becomes green: this is done using a collider. In this way, it is possible to copy the correct transform for the elbow goal according to the current position that the tracker has in the real world.

### 4.4.3 Height calibration

It is necessary to correctly map the movements of participants on the avatars: in order to do this, the dimensions of the avatars should be similar to the participants’, so the height calibration should be done. For this reason, when the participant presses the trigger of the right controller, the Y position of the `CameraRig`’s camera is saved. Then, the scale to apply to the avatars of both the interviewer and the user is calculated as  $height/r.Height$ , where *height* refers to the height of the participant, thus to the Y coordinate of the camera’s position, and *r.Height* refers to the height of the 3D character. The value of the participant’s height is saved into the Game Manager in order to apply it also in the other scenes.

### 4.4.4 Questions and conditions settings

The order of questions and conditions is managed with a drag-and-drop UI, consisting of a canvas with a `vertical layout` and an image per entry, as it is possible to see in Figure 4.11. At the press of the *S* button, when the change of scene from the *Calibration scene* to the *User scene* is triggered, the current order of both

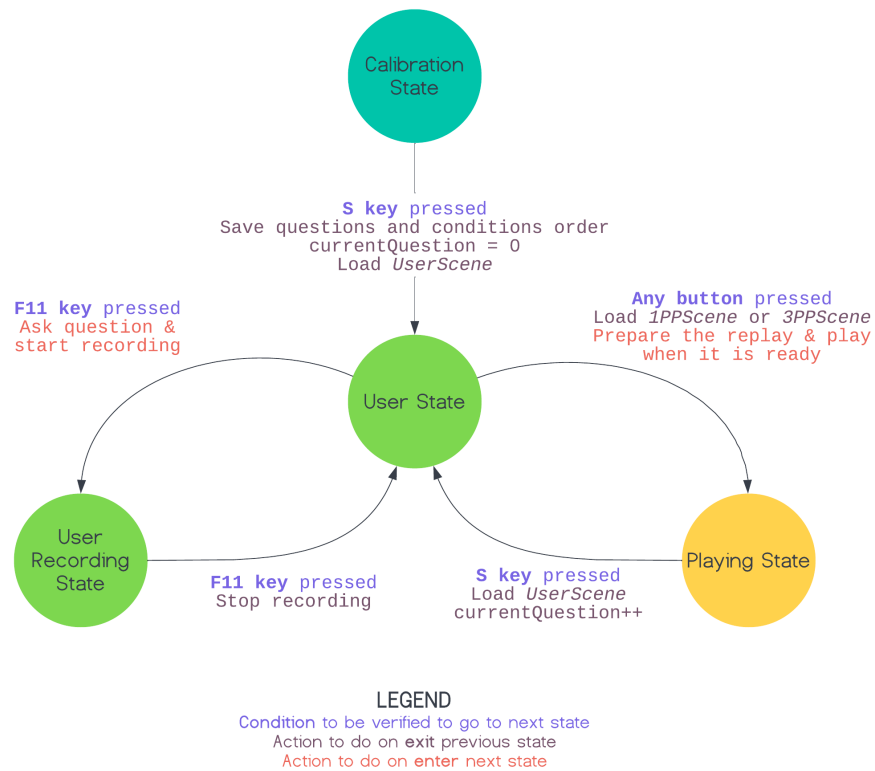
conditions and questions is saved into the Game Manager.



**Figure 4.11:** Representation of what the experimenter sees during the calibration (bigger window) and what the participant sees (smaller window)

## 4.5 Finite State Machine

Considering the phases that characterize the application, a Finite State Machine (FSM) has been realized in order to manage them in a proper way. An FSM is an abstract machine that can be in exactly one state at a given time. This is exactly what happens in this application: there is an initial moment dedicated to calibration, a moment for recording, one for replay and the other one when there is no calibration neither recording nor replay.



**Figure 4.12:** Schema of the FSM implemented for the application

In particular, in this application the states are:

- an initial state represented by the “*Calibration State*”, dedicated to all the settings and calibrations that take place in the *Calibration Scene*;
- “*User State*”, which represents the moment when the user is in the *User Scene* and the recording has not started yet or has already finished;
- “*User Recording State*”, which represents the moment when the user is in the *User Scene* and the recording is in progress, thus she or he is listening to the question or has already listened to it and is answering;
- “*Playing State*”, during which the user can be either in the *3PP Scene* or in the *1PP Scene* and the replay is in progress or has already finished.

As it is possible to see in Figure 4.12, the transition from the *Calibration State* to the *User State* occurs when the *S* key is pressed: this triggers the saving of the current order of questions and conditions (according to the order set in the

*Calibration Scene*), the initialization of the current index of the questions to 0 and the loading of the *User Scene*. This index is used to know which is the current question and to find it in the list of questions.

The transition from the *User State* to the *User Recording State* is triggered by pressing the *F11* key. At the `OnEnterState` function (i.e. the function called when the application enters a state) of the *User Recording State* the functions used for playing the question of the interviewer (both audio and animation) and start recording the audio and the movements of the user are invoked. In order to go back to the *User State* when the user has finished answering, the *F11* key has to be pressed again. Thus, on the `OnExitState` of the *User Recording State*, the function that stops recording is called.

Finally, the transition from the *User State* to the *Playing State* is triggered by the user when she or he presses any button of either the right or the left controller. More specifically, in `UserStateASMB.cs`, which manages the *User State*, the condition related to the current question is checked in order to call the right function of the Game Manager: if the condition is the *External POV* the function will be `Load3PPScene`, while if the condition is *Interviewer's POV* the function called is `Load1PPScene`), as it is possible to see in Listing A.2. These functions start the asynchronous load of the correct scene and change the state when the scene is loaded, as shown in Listing A.3 for the case of the *External POV* condition.

## 4.6 Game Manager and Scene Manager

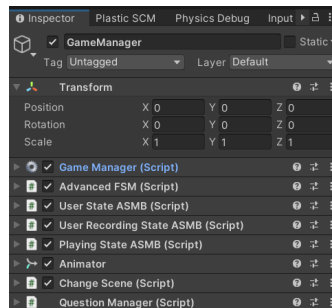
Two important parts of the Unity project are the *Game Manager* and the *Scene Manager*: all the main functions of this application are related to these two elements.

### 4.6.1 Game Manager

The Game Manager is the only script whose instance is not destroyed at the change of scene. For this reason, everything that has no references in the scenes and that should be used in more than one scene is put into the *Game Manager*. The Game Manager consists of a Game Object (i.e. an *empty*) created in Editor in the *Calibration scene*. Some scripts are attached to this Game Object, as it is possible to see in Figure 4.13:

- `GameManager.cs` contains the function to initialize the Game Manager, the ones to capture the events related to the press of the controllers' buttons and the functions to trigger the change of state and the change of scene;
- `AdvancedFSM.cs` initializes the *Advanced Final State Machine* (i.e. the FSM implemented in this application), by iterating on all the states and adding the *listeners* to the `OnStateEnter`, `OnStateExit` and `OnStateUpdate` functions;

- `UserStateASMB.cs`, `UserRecordingStateASMB.cs` and `PlayingStateASMB.cs` define the behaviour of the *Advanced Final State Machine*'s states;
- an Animator represents the FSM;
- `ChangeScene.cs` and `QuestionManager.cs` contain functions and variables that are useful for the whole duration of the application. In particular, the first script is used to call the asynchronous load of the scene and the fade to black from one scene to another. The other, instead, keeps the list of all the questions and initializes it with the correct values while leaving the *Calibration scene* (i.e. when the *S* key is pressed).



**Figure 4.13:** The *Game Manager* Game Object with the scripts attached to it

## 4.6.2 Scene Manager

Similarly to the Game Manager, the Scene manager consists of a Game Object with some scripts attached to it. However, in this case, its instance is related to the scene in which it is created. For this reason, it is used to collect all the functionalities that are useful in more than one scene but have references in each scene. The script representing the Scene Manager itself changes according to the scene: in the *User Scene* it is `UserSceneManagement.cs`, while for both *3PP Scene* and *1PP Scene* it is `PlayingSceneManagement.cs`. The *Calibration Scene* does not have a Scene Manager because it was not necessary. Both `UserSceneManagement.cs` and `PlayingSceneManagement.cs` extend the class `SceneManagement.cs`, so that the functions that do not have to change according to the scene, e.g. the `Start` function, are called from the base class (i.e. `SceneManagement.cs`) and the functions that have to produce different results are called by the children of the base class (i.e. `UserSceneManagement.cs` and `PlayingSceneManagement.cs`).

As it is evident from Figure 4.14, the scripts attached to the Scene Manager are:

- `UserSceneManagement.cs` or `PlayingSceneManagement.cs`, according to the scene;

- `ApplyInitialSettings.cs`, which applies all the settings derived from the *Calibration Scene* and contains the references to the *Player* Game Object (see Figure 4.9), the Game Objects related to the trackers and the rig that should be calibrated;
- `RecordAndPlay.cs`, which is the script with the aim of managing the recording and the replay and has, as a consequence, the references to the Game Objects that have to be recorded and to the Game Object of the interviewer;
- `UIManagement.cs`, the script that manages the UI visible to the user.

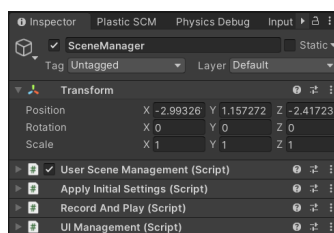


Figure 4.14: The *Scene Manager* Game Object with the scripts attached to it

## 4.7 Users' interaction with the interviewer

In the current work, the interactions of users with the interviewer are managed by the experimenter. In fact, the experimenter decides when to start asking a question by pressing the *F11* key, which starts the recording at the same time. When the user has finished talking, the experimenter presses *F11* again to stop the recording. The questions do not change according to the answers of the participants, so there is no interviewer's feedback from this point of view. However, the experimenter, by pressing the *N* key when the user is answering, can trigger a nod animation in order to make the dialogue more realistic.

## 4.8 Recording and replay

The recording of users' movements and their voices has been managed independently and, as a consequence, the replay, too. In fact, the Unity asset Record and Play has been used to record the *Transforms* (i.e. the component used to store and manipulate the position, rotation and scale of an object) related to the user's avatar bones, while the audio of the microphone has been recorded using Unity's *Microphone* API. In the application developed for this work, the management of recording and replay of both *Transforms* and audio is done in `RecordAndPlay.cs`.

The recording of both transforms and audio is done in the *User Scene*, while the replay in *3PP Scene* and *1PP Scene*.

### 4.8.1 Transforms recording

Thanks to Inverse Kinematics, the movements of participants correspond to the ones of their avatars. Thus, in order to keep track of participants' non-verbal language it is sufficient to record the avatars' movements, i.e. the Transforms of their bones. According to the documentation of the *Record and Play* asset, the steps to generate a recording are the following:

1. creation of a **Recorder**, which is a class of the asset “used for building recordings and keeping up with all the subject recorders” [53];
2. attachment of the **SubjectBehavior** script to any Game Object that should be recorded;
3. start of the recording by calling the function **Start**;
4. stop of the recording by calling the function **Stop**;

These steps have been performed in the current application, too. In this case, the Game Objects to be recorded are represented by the bones of the user's avatar rig, which have been added to a list in order to attach a **SubjectBehavior** script to each of them by iterating on the list. The Game Objects are assigned to the list in Editor. The **Start** function of **SceneManagement.cs** calls the **StartRecordAndPlay** function of the **RecordAndPlay.cs** script to initiate the recorder and set the Game Objects that need to be tracked, as is evident from Listing A.4.

While the setting of the recorder is managed in the Scene Manager, the start of the recording is invoked in the **OnEnterState** of the *User Recording State*. As mentioned in the documentation of *Record and Play*, this is done by calling **recorder.Start()**. Then, the recording is ended with **recorder.Finish()**, which creates a **Recording** object. This is saved in the Game Manager in order to use it in the scene where the replay is shown. In addition to this, the recording is saved in the application data for any further analysis by the experimenter: in order to do it, a function to convert the **Recording** object into a JSON and save it into the application data path is called.

### 4.8.2 Audio recording

*Record and Replay* provides management of custom events that should have been useful to capture the audio into the recording created by this asset. However,

contrary to expectations, this has generated problems in the synchronization between audio and animations. Thus, the audio has been managed separately.

In order to record the correct audio input, the `RecordAndPlay.cs` script keeps track of the microphone used as input by SALSA's `SalsaMicInput` plugin. This component saves the microphone's data samples in an Audio Clip of `loopLenght` seconds (5 is the default value). However, this value has been changed to 300, considering 5 minutes as the maximum duration for an answer to a question of the interview. In fact, this Audio Clip is the one that is saved when the recording finishes.

The start of audio recording does not require a lot: it is only necessary to call the `StopMicrophone` function of `SalsaMicInput` to set the Audio Clip to null and call the `End` function of Unity's `Microphone` API, which stops microphone data acquisition. Then, `SalsaMicInput`'s `StartMicrophone` function is called to start capturing the microphone input again. This procedure assures that the position in samples of the microphone is always 0 when the recording starts.

In case 5 minutes are not sufficient, a `Coroutine` keeps track of the current position in samples of the microphone with respect to the Audio Clip, so that, if the samples fill the Audio Clip, this one can be saved and another one is created. This is shown in Listing A.5. Each audio created is saved into the Game Manager.

The end of the audio recording requires more steps than the start, as is evident in A.6. In particular:

- it is necessary to save the current position in samples of the microphone (i.e. `micSamples`) to understand how many samples have to be saved: this ensures that the Audio Clip saved will be as long as necessary, thus avoiding to store 5 minutes of audio if the participant replies in less time;
- an array of `micSamples` floats is created;
- the data of the microphone input are saved into the array;
- a new Audio Clip is created and the array values are set as its data with the `SetData` function of Unity's `AudioClip` API.

Similarly to the `Recording`, the audio clip has to be saved into application data, too. This is done thanks to the `SavWav.cs` class of a GitHub repository called "Unity-Microphone-Recorder"[54] and specifically by calling the function `SaveWav`.

### 4.8.3 Transforms replay

The documentation of *Record and Play* asset summarizes the steps necessary to play the `Recording` previously created.



1. The first step is to build a `PlaybackBehavior` instance using its `Build` function: it is called when the FSM enters the *Playing State*. Since by default the playback instantiates cubes to represent the objects tracked in the recording, in the current application it was necessary to implement the `IActorBuilder` interface. This enabled setting the bones of the avatar used for the replay as “actors” of the playback, as shown in Listing A.7.
2. The `Play` function of `PlaybackBehavior` is called to start the replay.

#### 4.8.4 Audio replay

The functions related to the replay of the audio are contained in the `AudioClipsPlay` class. In particular, the `SetAudioClips` function get the audio clips previously recorded from the Game Manager and the `PlayAudio` calls a `Coroutine` used to play all the audio clips sequentially. Most times the audio clip recorded should be only one, but, in case a participant in the experiment talks for more than five minutes, this function assures that all the clips are played.

# Chapter 5

## Results

### 5.1 Statistics of the results

#### 5.1.1 Characteristics of the sample

The sample for the experiment was composed of 19 participants: 11 males and 8 females. Most of them (68,42%) are bachelor's or master's students and 15,79% are Ph.D. Students in the field of Computer Science and Virtual Reality. The average age of participants is 25.

As far as the experience with VR is concerned:

- 13,33% have never experienced Immersive Virtual Reality tools;
- 20,00% claim that they might have happened to use Immersive Virtual Reality tools;
- 10,00% sometimes use Immersive Virtual Reality tools;
- 6,67% quite often use Immersive Virtual Reality tools;
- 13,33% use Immersive Virtual Reality tools every day.

The average value for previous experience with VR is 2,79 out of 5 and the standard deviation is 1,76.

Considering that the questionnaire questions have been asked on a scale from 1 to 5, the average value concerning experience with programming is 3,74, with a standard deviation of 1,58. In fact, 73,68% of the participants have studied or study Computer Engineering or Cinema Engineering.

### 5.1.2 Evaluation of the experience

The overall evaluation of the users' experience was satisfying.

More in detail, the average value of the responses to the PSSUQ questionnaire, on both hardware and software, was 4,68 on a scale from 1 to 5. In addition to this, the average evaluation of the Quality of the Virtual Reality Environment was 4,49 out of 5.

The majority of participants (73,68%) did not suffer ill effects due to the Virtual Reality application. In addition to this, the amount of ill effects suffered by the rest of the sample has low relevance, since the maximum value was 22,4 out of 64.

### 5.1.3 Embodiment during the replay

It is clear that, during the phase of the replay, the embodiment of the interviewer's avatar was triggered only in the *Interviewer's POV* condition. In fact, the p-value considering the results of the *External POV* condition as the first group and the *Interviewer's POV* condition as the second one, is 0 for all the questions concerning the interviewer's embodiment.

On the other hand, as far as the embodiment of the interviewee is concerned, there is no statistical evidence in the results.

### 5.1.4 Comparison of the conditions

All the questions on the preferred condition produced statistically significant results, as it is possible to see in Figure 5.1. In order to report the results, we refer to each question with a name.

- *Rank overall* refers to the question "Which variant you preferred overall?" (item 1 in Figure 5.1).
- *Rank help improving* refers to the question "Which variant can help you better in improving your job interview skills?" (item 2 in Figure 5.1).
- *Rank verbal* refers to the question "Which variant enabled you better to assess the effectiveness of your verbal responses to the job interview more objectively?" (item 3 in Figure 5.1).
- *Rank not verbal* refers to the question "Which variant enabled you better to assess the effectiveness of your non-verbal language during the interview more objectively?" (item 4 in Figure 5.1).
- *Computed overall* represents the aggregate metric on *Rank overall* (item 5 in Figure 5.1).

| Item | Rank |     |     | p-value         |                 |                 |             |
|------|------|-----|-----|-----------------|-----------------|-----------------|-------------|
|      | FPP  | EPP | IPP | Friedman        | FPP/EPP         | FPP/IPP         | EPP/IPP     |
| 1    | 3    | 2   | 1   | <b>,003</b>     | <b>,049</b>     | <b>,002</b>     | <b>,126</b> |
| 2    | 3    | 2   | 1   | <b>&lt;.001</b> | <b>,049</b>     | <b>&lt;.001</b> | <b>,020</b> |
| 3    | 2    | 2   | 1   | <b>,071</b>     | <b>,754</b>     | <b>,136</b>     | <b>,016</b> |
| 4    | 3    | 2   | 1   | <b>&lt;.001</b> | <b>&lt;.001</b> | <b>&lt;.001</b> | <b>,506</b> |
| 5    | 3    | 2   | 1   | <b>&lt;.001</b> | <b>,003</b>     | <b>&lt;.001</b> | <b>,017</b> |

**Figure 5.1:** Results on the comparison of conditions

Taking into account Friedman’s Test and the p-value for each of these variables, the results can be summarized as follows. According to the naming convention we will use to refer to the condition, *FPP* represents the *Baseline* condition, *EPP* represents the *External POV* condition and *IPP* the *Interviewer’s POV* condition.

- Friedman’s Test for *Rank overall* is 0. In particular, the p-value between *FPP* and *EPP* is 0,049, while it is 0 between *FPP* and *IPP*, and 0,02 between *EPP* and *IPP*.
- Friedman’s Test for *Rank help improving* is 0,071, so it has no statistic relevance. However, the p-value between *EPP* and *IPP* is 0,016.
- Friedman’s Test for *Rank verbal* is 0. More in detail, both the p-value between *FPP* and *EPP* and the p-value between *FPP* and *IPP* are 0.
- *Computed overall* is consistent with the explicit metrics and in this case the Friedman’s Test result is 0, too. However, in comparison with *Rank overall*, the aggregate metric acquires statistical relevance in the comparison between *EPP* and *IPP*, too. In fact, the p-value between *EPP* and *IPP* is 0,017. In addition to this, the p-value between *FPP* and *EPP* is 0,003 and between *FPP* and *IPP* it is 0.

## 5.2 Discussion on the results

Considering that most of the results do not have statistical significance, no conclusions can be drawn from the hypotheses of this experiment, except for *H5*, as it is possible to see in Subsection 5.2.5. However, some considerations can be done taking into account the overall responses to the questionnaires.

### 5.2.1 Hypothesis *H1*: embodiment

According to this hypothesis, the embodiment was expected to be high when participants had a First-Person Perspective either of the interviewee or of the

interviewer. However, this has not always occurred. In fact, considering the responses to the questionnaires on embodiment during both the answering and the replay phase, the participants' sense of embodiment has been heterogeneous. The following considerations have been done considering the answers to the question "Although the virtual body of the interviewer did not look like me physically, I felt that the virtual body of the interviewer was my own body"

- Some subjects (31,58%) confirmed the hypothesis, since they felt embodied in the interviewer and not in their avatar, even though they could recognize their past movements in the movements of the interviewee. In particular, they answered the question on the interviewer's virtual body with a value  $>4$ .
- Another sample of participants (31,58%) did not realise they had the control of the interviewer's avatar, thus their sense of embodiment towards him was low. This was due to the fact that they did not move a lot while watching the replay or they were too focused on their avatars to notice it.
- The interesting result is that other participants (31,58%) felt embodied in both the interviewee's and the interviewer's avatars at the same time.

### 5.2.2 Hypothesis *H2*: presence and plausibility

The hypothesis on the absence of significant differences in presence and plausibility among the three conditions, is confirmed. Only some participants increased values of plausibility after the first condition that they experienced, because they felt more immersed and accustomed to the virtual environment.

### 5.2.3 Hypothesis *H3*: emotional distress

As far as emotional distress and anxiety are concerned, the hypothesis is not confirmed. In fact, no evidence of their decrease has been found during either the external or the interviewer's perspective.

### 5.2.4 Hypothesis *H4*: emotional reactions

Similarly to *H3*, there is no evidence about the fact that more positive and less activating emotions have been elicited in the external perspective in comparison with the other two and in the interviewer's perspective in comparison with the interviewee's.

### 5.2.5 Hypothesis *H5*: comparison of the conditions

According to the results of the ranks that the participants compiled at the end of the experiment, the preferred condition was the *Interviewer's POV* condition regardless of the parameters taken into account. In fact, it was considered the best condition to improve job interview skills and to assess the effectiveness of both verbal and non-verbal language. The second place went to the *External POV* condition and the last one to the *Baseline* condition. However, this podium changes for the *Rank verbal* question, i.e. the one concerning verbal responses. In fact, in this case, *External POV* and *Baseline* conditions are tied.

The results reported above partially confirm hypothesis *H5*, according to which participants should find more useful the conditions where they can swap perspectives (either *Interviewer's POV* or *External POV* conditions) to improve and assess their performance in the job interview simulation, with respect to the *Baseline* condition. In fact, it is true that participants have found it more useful to swap perspectives, *Interviewer's POV* has been preferred over *External POV* condition. In addition to this, considering the assessment of verbal responses, *External POV* and *Baseline* conditions have the same rating.

## Chapter 6

# Conclusions

The present thesis is the result of a study on embodiment, perspective swap and body swap which led to the development of an application in Virtual Reality aiming at helping the improvement and self-assessment of soft skills for a job interview.

As mentioned in Chapter 5, most of the results obtained through the questionnaires administered during the application testing phase do not have statistical significance and only hypothesis *H5* can be verified. Despite this, the participants have found the application to be useful for both training soft skills and self-assessment. In particular, some of the comments the participants provided after the experience are related to the fact that it is a “useful tool, especially for anxious people”, that “only through such a tool it is possible to experience an external or the interviewer’s point of view” or even that it is “more realistic than just practicing in front of a mirror in the real world”. Other participants claimed that they managed to identify with the interviewer’s role. As a consequence, they were able to understand the sensation they would have given in a real context and self-assess themselves as the person who should evaluate them, which was one of the goals of this work. As far as the *External POV* condition is concerned, it has been deemed more useful than the *Baseline* condition but the overall results see the *Interviewer’s POV* condition in first place. This is exactly what was expected: the conditions with a change of perspective were preferred over the one without it. Some participants made considerations from an emotional standpoint as well. In fact, they claimed that they could feel the pressure. Moreover, the estrangement from one’s own body enabled one of them to be more relaxed and decrease anxiety. One participant specified that simply the fact of being in a virtual environment helped him or her to be less anxious, while another attributed the decrease in anxiety to the increasing practice.

## 6.1 Future work

The work described in this thesis does not aim at giving definitive responses on the effects of perspective and body swap for training soft skills. It should represent, instead, a starting point. In fact, future work needs to be done in order to make the study more relevant. Firstly, it is necessary to expand the sample of participants and verify if it is sufficient to have statistically significant results. Otherwise, the measures and the experimental design should be examined and eventually modified.

As far as the VR application is concerned, some improvements can be done, too. For example, an enhancement of the graphic aspect of the scene and of the avatars, especially the interviewer's, could improve the realism of the simulation. In addition to this, the use of trackers for the feet could help embodiment, especially in the case when participants cross their legs, and more reactions from the interviewer could be added. It could be interesting to insert some metrics used by Human Resources experts to evaluate the candidates for a job position so that the user could receive an evaluation based on objective parameters, such as too much hesitation in the answers or too many gestures while talking. For a complete application, it could be useful to give the users the possibility to practice every question as many times as necessary and insert an AI to change the next question according to the answer the interviewee gave to the previous one.

To summarize, this work introduced the comparison between the First-Person perspective of an avatar who should not represent the user and Third-Person perspective in a context of a job interview simulation to train soft skills and manage anxiety. The application has been appreciated by the users and the perspective which has been proven to be most effective was the First-Person perspective of the interviewer.



# Appendix A

## Listings

### A.1 Animations of the interviewer

---

```
1 animations.Add(new Tuple<float, string>(-1, "Blend1"), new
   Tuple<float, float>(0, 0.71f));
2 animations.Add(new Tuple<float, string>(0, "isTalking"),
   new Tuple<float, float>(0, 1));
3 animations.Add(new Tuple<float, string>(2.9f, "Blend1"),
   new Tuple<float, float>(0.71f, 0));
4 animations.Add(new Tuple<float, string>(2.9f, "Blend2"),
   new Tuple<float, float>(0, 1));
```

---

**Listing A.1:** Example of setting of Blend Tree values according to the question in the *CreateAnimationForQuestion* function (*Question.cs*)

### A.2 Finite State Machine

---

```
1 if (GameManager.Instance.recording != null)
2 {
3     if (GameManager.Instance.currentCondition ==
   ExperimentCondition.ThirdPPReplay)
4         GameManager.Instance.Load3PPScene();
5
6     else if (GameManager.Instance.currentCondition ==
   ExperimentCondition.FirstPPReplay)
7         GameManager.Instance.Load1PPScene();
8 }
```

---

**Listing A.2:** *UserStateASMB.cs*: check on the current condition and call of the function to change state and scene

---

```

1 internal void Load3PPScene()
2 {
3     _changeScene._onSceneLoaded.AddListener(Playing);
4     _changeScene.LoadScene(_3PPScene);
5 }

```

---

**Listing A.3:** *GameManager.cs*: function to change to *Playing State* and load *3PP Scene*

## A.3 Record and Play

---

```

1 // Create a recorder
2 recorder = ScriptableObject.CreateInstance<Recorder>();
3 for (int i = 0; i < objectsToTrack.Count; i++)
4     SubjectBehavior.Build(objectsToTrack[i], recorder);

```

---

**Listing A.4:** *RecordAndPlay.cs*: initial setting of the recorder

---

```

1 IEnumerator ClipCount()
2 {
3     while (recorder.CurrentState() ==
4           RecordingState.Recording &&
5           Microphone.IsRecording(micInput.selectedMic))
6     {
7         var micSeconds =
8           Microphone.GetPosition(micInput.selectedMic) /
9           micInput.sampleRate; // becomes 0 after loopLength
10          seconds
11         if (micSeconds < micSecondsPrev)
12         {
13             // Create an array
14             var samples = new
15             float[micInput.audioSrc.clip.samples];
16             // Fill the array with the audio data
17             micInput.audioSrc.clip.GetData(samples, 0);
18             string recName =
19             DateTime.Now.ToShortDateString().Replace('/', '-') + "_" +
20             DateTime.Now.ToLongTimeString().Replace(':', '-');
21             // Creation of the audioClip

```

---

```

16         AudioClip audioClip = AudioClip.Create(recName,
micInput.audioSrc.clip.samples,
micInput.audioSrc.clip.channels,
micInput.audioSrc.clip.frequency, false);
17         // Copy of the audio data in the audioclip
18         audioClip.SetData(samples, 0);
19
20         // Cliplengtht reached: save it and start with
another
21
RecPlayUtils.SaveAudioToGameManager(micInput.audioSrc.clip);|
22     }
23
24     // Previous value of seconds becomes current (for
next check)
25     micSecondsPrev = micSeconds;
26     yield return null;
27 }
28 }

```

---

**Listing A.5:** *RecordAndPlay.cs*: coroutine to keep track of the microphone's current position in samples

---

```

1 void StopRecordingAudio()
2 {
3     // Get current position (in samples) of the microphone
4     var micSamples =
Microphone.GetPosition(micInput.selectedMic);
5     // Create an array as big as the samples we need
6     var samples = new float[micSamples];
7     // Fill the array with the audio data of the Microphone
clip
8     micInput.audioSrc.clip.GetData(samples, 0);
9
10    string recName =
DateTime.Now.ToShortDateString().Replace('/', '-') + "_"
+ DateTime.Now.ToLongTimeString().Replace(':', '-');
11
12    // Creation of the AudioClip
13    AudioClip audioClip = AudioClip.Create(recName,
micSamples, micInput.audioSrc.clip.channels,
micInput.audioSrc.clip.frequency, false);
14    // Copy of the audio data in the audioclip
15    audioClip.SetData(samples, 0);
16

```

```
17 // Firstly, save the recording & audio to the
    GameManager to be ready for replay
18 RecPlayUtils.SaveAudioToGameManager(audioClip);
19
20 // Then start async save of recording & audio
21 // GameManager.Instance.recordingTask = Task.Run(async
    () => await RecPlayUtils.SaveAudioToWavAsync(audioClip,
    audioClip.name)); // NON FUNZIONA
22 StartCoroutine(SaveAudio());
23 }
```

---

**Listing A.6:** *RecordAndPlay.cs*: function to stop audio recording and save the captured audio into an Audio Clip

---

```
1 public Actor Build(int subjectId, string subjectName,
    Dictionary<string, string> metadata)
2 {
3     // Find the Game Object in the list, in order to
    represent the movements with itself
4     GameObject actor = objectsToTrack.Find(x =>
    x.gameObject.name == subjectName);
5     return new Actor(actor);
6 }
```

---

**Listing A.7:** *RecordAndPlay.cs*: method to set the bones of the avatar's rig as actors of the replay

# Bibliography

- [1] Avantis Systems. *ClassVR - Investigating cells*. URL: <https://www.classvr.com/case-studies/science-investigating-cells-age-14-16/> (cit. on p. 1).
- [2] Coop Culture. *Domus Aurea Experience*. URL: <https://parcocolosseo.it/en/guided-tours/domus-aurea-experience/> (cit. on p. 1).
- [3] Exhibithion Hub. *Van Gogh: the immersive experience*. URL: <https://www.exhibitionhub.com/exhibitions/van-gogh-immersive-experience/> (cit. on p. 1).
- [4] Virtro. *Immersive Healthcare*. URL: <https://www.youtube.com/watch?v=H7dDZUpD624> (cit. on p. 1).
- [5] Merve Nazligul, Murat Yilmaz, Ulas Gulec, Mert Gözcü, Rory O'Connor, and Paul Clarke. «Overcoming Public Speaking Anxiety of Software Engineers Using Virtual Reality Exposure Therapy». In: Sept. 2017, pp. 191–202. ISBN: 978-3-319-64217-8. DOI: 10.1007/978-3-319-64218-5\_15 (cit. on pp. 1, 10, 11, 18).
- [6] Fabrizio Palmas, Jakub Cichor, David Plecher, and Gudrun Klinker. «Acceptance and Effectiveness of a Virtual Reality Public Speaking Training». In: Oct. 2019. DOI: 10.1109/ISMAR.2019.00034 (cit. on pp. 1, 10, 11, 18).
- [7] Mel Slater, Solène Neyret, Tania Johnston, Guillermo Iruetagoiena, Mercè Álvarez de la Campa, Miquel Alabernia, Bernhard Spanlang, and Guillem Feixas. «An experimental study of a virtual reality counselling paradigm using embodied self-dialogue». In: *Scientific Reports* 9 (July 2019), p. 10903. DOI: 10.1038/s41598-019-46877-3 (cit. on pp. 2, 9, 10, 23, 24, 34).
- [8] Virtro. *Be Confident VR*. URL: <https://www.youtube.com/watch?v=SkCQi1SwH8Y> (cit. on pp. 2, 13, 14).
- [9] Bodyswaps. *Bodyswaps - Job Interview Simulator*. URL: <https://www.youtube.com/watch?v=vq30aQkwgWA> (cit. on pp. 2, 13, 14).

- [10] Margaret M. Bradley and Peter J. Lang. «Measuring emotion: The self-assessment manikin and the semantic differential». In: *Journal of Behavior Therapy and Experimental Psychiatry* 25.1 (1994), pp. 49–59. ISSN: 0005-7916. DOI: [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9). URL: <https://www.sciencedirect.com/science/article/pii/0005791694900639> (cit. on pp. 4, 22).
- [11] Helene Wallach, Marilyn Safir, Roy Samana, Idan Almog, and Reut Horef. «How Can Presence in Psychotherapy Employing VR Be Increased? Chapter for Inclusion in: Systems in Health Care Using Agents and Virtual Reality». In: vol. 337. Mar. 2011, pp. 129–147. ISBN: 978-3-642-17823-8. DOI: 10.1007/978-3-642-17824-5\_7 (cit. on p. 6).
- [12] Mel Slater. «Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments». In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364 (Dec. 2009), pp. 3549–57. DOI: 10.1098/rstb.2009.0138 (cit. on p. 6).
- [13] N Braun, S Debener, N Spsychala, and et al. «The Senses of Agency and Ownership: A Review.» In: *Frontiers in psychology* 9 (Apr. 2018), p. 535. DOI: 10.3389/fpsyg.2018.00535 (cit. on p. 6).
- [14] Konstantina Kilteni, Raphaela Groten, and Mel Slater. «The Sense of Embodiment in Virtual Reality». In: *Presence Teleoperators amp Virtual Environments* 21 (Nov. 2012). DOI: 10.1162/PRES\_a\_00124 (cit. on p. 7).
- [15] Sofia Seinfeld, Tiare Feuchtner, Antonella Maselli, and Jörg Müller. «User Representations in Human-Computer Interaction». In: *Human-Computer Interaction* 36 (Feb. 2020), pp. 1–39. DOI: 10.1080/07370024.2020.1724790 (cit. on p. 7).
- [16] Henrique Debarba, Sidney Bovet, Roy Salomon, Olaf Blanke, Bruno Herbelin, and Ronan Boulic. «Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality». In: *PLOS ONE* 12 (Dec. 2017), e0190109. DOI: 10.1371/journal.pone.0190109 (cit. on pp. 7, 18).
- [17] Geoffrey Gorisse, Olivier Christmann, Etienne Armand Amato, and Simon Richir. «First- and Third-Person Perspectives in Immersive Virtual Environments: Presence and Performance Analysis of Embodied Users». In: *Frontiers in Robotics and AI* 4 (2017). ISSN: 2296-9144. DOI: 10.3389/frobt.2017.00033 (cit. on p. 7).
- [18] Naval Bhandari and Eamonn O’Neill. «Influence of Perspective on Dynamic Tasks in Virtual Reality». In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 939–948. DOI: 10.1109/VR46266.2020.00114 (cit. on p. 7).

- [19] Rajiv Khadka and Amy Banic. «Effects of Egocentric Versus Exocentric Virtual Object Storage Technique on Cognition in Virtual Environments». In: Mar. 2020, pp. 205–209. DOI: 10.1109/VRW50115.2020.00044 (cit. on p. 7).
- [20] Katharina Emmerich, Andrey Krekhov, Sebastian Cmentowski, and Jens Krüger. «Streaming VR Games to the Broad Audience: A Comparison of the First-Person and Third-Person Perspectives». In: (Jan. 2021) (cit. on p. 7).
- [21] Heather Iriye and Peggy Jacques. «Memories for third-person experiences in immersive virtual reality». In: *Scientific Reports* 11 (Feb. 2021), p. 4667. DOI: 10.1038/s41598-021-84047-6 (cit. on p. 8).
- [22] Antonella Maselli and Mel Slater. «Sliding perspectives: Dissociating ownership from self-location during full body illusions in virtual reality». In: *Frontiers in Human Neuroscience* 8 (Sept. 2014), pp. 1–19. DOI: 10.3389/fnhum.2014.00693 (cit. on p. 8).
- [23] Antonella Maselli and Mel Slater. «The building blocks of the full body ownership illusion». In: *Frontiers in human neuroscience* 7 (Mar. 2013), p. 83. DOI: 10.3389/fnhum.2013.00083 (cit. on p. 8).
- [24] Valeria Petkova, Mehrnoush Khoshnevis, and H. Henrik Ehrsson. «The Perspective Matters! Multisensory Integration in Ego-Centric Reference Frames Determines Full-Body Ownership». In: *Frontiers in Psychology* 2 (2011). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2011.00035. URL: <https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00035> (cit. on p. 8).
- [25] Kornelius Kammler-Sücker, Annette Löffler, Dieter Kleinböhl, and Herta Flor. «Exploring Virtual Doppelgängers As Movement Models To Enhance Voluntary Imitation». In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* PP (Oct. 2021), pp. 1–1. DOI: 10.1109/TNSRE.2021.3120795 (cit. on p. 8).
- [26] Geoffrey Gorisse, Olivier Christmann, Samory Houzangbe, and Simon Richir. «From Robot to Virtual Doppelgänger: Impact of Avatar Visual Fidelity and Self-esteem on Perceived Attractiveness». In: May 2018, pp. 1–5. DOI: 10.1145/3206505.3206525 (cit. on p. 8).
- [27] Geoffrey Gorisse, Olivier Christmann, Samory Houzangbe, and Simon Richir. «From Robot to Virtual Doppelgänger: Impact of Visual Fidelity of Avatars Controlled in Third-Person Perspective on Embodiment and Behavior in Immersive Virtual Environments». In: *Frontiers in Robotics and AI* 6 (Feb. 2019). DOI: 10.3389/frobt.2019.00008 (cit. on p. 8).
- [28] Marc Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. «The effect of avatar realism in immersive social virtual realities». In: Nov. 2017, pp. 1–10. DOI: 10.1145/3139131.3139156 (cit. on p. 8).

- [29] Mel Slater, Bernhard Spanlang, Maria Sanchez-Vives, and Olaf Blanke. «First Person Experience of Body Transfer in Virtual Reality». In: *PloS one* 5 (May 2010), e10564. DOI: 10.1371/journal.pone.0010564 (cit. on p. 8).
- [30] Valeria Petkova and H.Henrik Ehrsson. «If I Were You: Perceptual Illusion of Body Swapping». In: *PloS one* 3 (Feb. 2008), e3832. DOI: 10.1371/journal.pone.0003832 (cit. on p. 9).
- [31] María Murcia López. «The effectiveness of training in virtual environments». PhD thesis. University College London, 2018. URL: <https://discovery.ucl.ac.uk/id/eprint/10063411> (cit. on p. 9).
- [32] Ioannis Doumanis, Daphne Economou, Gavin Robert Sim, and Stuart Porter. «The impact of multimodal collaborative virtual environments on learning: A gamified online debate». In: *Computers Education* 130 (2019), pp. 121–138. ISSN: 0360-1315. DOI: <https://doi.org/10.1016/j.compedu.2018.09.017>. URL: <https://www.sciencedirect.com/science/article/pii/S0360131518302537> (cit. on pp. 9, 11).
- [33] Chiara Zizza, Adam Starr, Devin Hudson, Sai Shreya Nuguri, Prasad Calyam, and Zhihai He. «Towards a social virtual reality learning environment in high fidelity». In: Jan. 2018, pp. 1–4. DOI: 10.1109/CCNC.2018.8319187 (cit. on p. 10).
- [34] Filip Górski, Przemysław Zawadzki, Paweł Buń, and Beata Starzyńska. «Virtual Reality Training of Hard and Soft Skills in Production». In: *Proceedings of the 23rd International ACM Conference on 3D Web Technology*. Web3D '18. Poznań, Poland: Association for Computing Machinery, 2018. ISBN: 9781450358002. DOI: 10.1145/3208806.3219787. URL: <https://doi.org/10.1145/3208806.3219787> (cit. on pp. 10, 11).
- [35] Beata Starzyńska, Filip Górski, and Paweł Buń. «Virtual Reality System for Learning and Selection of Quality Management Tools». In: Mar. 2018, pp. 264–274. ISBN: 978-3-319-77699-6. DOI: 10.1007/978-3-319-77700-9\_27 (cit. on pp. 10, 11).
- [36] Gustav Petersen, Giorgos Petkakis, and Guido Makransky. «A study of how immersion and interactivity drive VR learning». In: *Computers Education* 179 (Jan. 2022), p. 104429. DOI: 10.1016/j.compedu.2021.104429 (cit. on p. 10).
- [37] Mark Powers and Paul Emmelkamp. «Virtual reality exposure therapy for anxiety disorders: A meta-analysis». In: *Journal of anxiety disorders* 22 (Feb. 2008), pp. 561–9. DOI: 10.1016/j.janxdis.2007.04.006 (cit. on pp. 10–12).



- [38] Matt C. Howard and Melissa B. Gutworth. «A meta-analysis of virtual reality training programs for social skill development». In: *Computers Education* 144 (2020), p. 103707. ISSN: 0360-1315. DOI: <https://doi.org/10.1016/j.compedu.2019.103707>. URL: <https://www.sciencedirect.com/science/article/pii/S036013151930260X> (cit. on pp. 10, 11).
- [39] Marcus Tan, Stefanie Chye, and Kylin Teng. «“In the shoes of another”: immersive technology for social and emotional learning». In: *Education and Information Technologies* 27 (Mar. 2022), pp. 1–24. DOI: 10.1007/s10639-022-10938-4 (cit. on pp. 10, 11).
- [40] Marianne Mast, Emmanuelle Kleinogel, Benjamin Tur, and Manuel Bachmann. «The future of interpersonal skills development: Immersive virtual reality training with virtual humans». In: *Human Resource Development Quarterly* 29 (Jan. 2018), pp. 125–141. DOI: 10.1002/hrdq.21307 (cit. on p. 11).
- [41] Laura Aymerich-Franch and Jeremy Baileson. «The use of doppelgangers in virtual reality to treat public speaking anxiety: a gender comparison». In: Mar. 2014 (cit. on p. 12).
- [42] Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, Rosalind Picard, and Mohammed Hoque. «MACH: My automated conversation coach». In: Dec. 2014. DOI: 10.1145/2493432.2493502 (cit. on pp. 12, 13).
- [43] Tobias Baur, Ionut Damian, Patrick Gebhard, Kaska Porayska-Pomsta, and Elisabeth Andre. «A Job Interview Simulation: Social Cue-Based Interaction with a Virtual Character». In: Sept. 2013. DOI: 10.1109/SocialCom.2013.39 (cit. on p. 12).
- [44] Patrick Gebhard, Tobias Baur, Ionut Damian, Gregor Mehlmann, Johannes Wagner, and Elisabeth Andre. «Exploring Interaction Strategies for Virtual Characters to Induce Stress in Simulated Job Interviews». In: vol. 1. May 2014 (cit. on p. 12).
- [45] S. Neyret, X. Navarro, and A. et al. Beacco. «An Embodied Perspective as a Victim of Sexual Harassment in Virtual Reality Reduces Action Conformity in a Later Milgram Obedience Scenario.» In: *Sci Rep* 10, 6207 (2020). DOI: <https://doi.org/10.1038/s41598-020-62932-w> (cit. on p. 19).
- [46] C.D. Spielberger and R.L. Gorsuch. *State-trait Anxiety Inventory for Adults: Sampler set*. Mind Garden, 1983. URL: <https://books.google.it/books?id=K8BatwAACAAJ> (cit. on p. 22).
- [47] TM Marteau and H Bekker. «The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI)». In: *The British journal of clinical psychology* 31 (Sept. 1992), pp. 301–6. DOI: 10.1111/j.2044-8260.1992.tb00997.x (cit. on p. 23).

- [48] Tabitha Peck and Mar Gonzalez-Franco. «Avatar Embodiment. A Standardized Questionnaire». In: *Frontiers in Virtual Reality* 1 (Feb. 2021). DOI: 10.3389/frvir.2020.575943 (cit. on p. 23).
- [49] Kathryn Connor, Jonathan Davidson, L. Churchill, Andrew Sherwood, Edna Foa, and Richard Weisler. «Psychometric properties of the Social Phobia Inventory (SPIN)». In: *The British journal of psychiatry : the journal of mental science* 176 (May 2000), pp. 379–86. DOI: 10.1192/bjp.176.4.379 (cit. on p. 25).
- [50] James R. Lewis. «Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies». In: *International Journal of Human-Computer Interaction* 14.3-4 (2002), pp. 463–488. DOI: 10.1080/10447318.2002.9669130 (cit. on p. 25).
- [51] Darragh Egan, Sean Brennan, John Barrett, Yuansong Qiao, Christian Timmerer, and Niall Murray. «An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments». In: June 2016, pp. 1–6. DOI: 10.1109/QoMEX.2016.7498964 (cit. on p. 26).
- [52] Crazy Minnow Studio. *SALSA LipSync Suite - Online Documentation*. URL: <https://crazyminnowstudio.com/docs/salsa-lip-sync/> (cit. on p. 43).
- [53] Eli Davis. *Record and Play: API reference*. URL: <https://recolude.gitlab.io/RecordAndPlay3D/> (cit. on p. 51).
- [54] Schaggo. *Unity-Microphone-Recorder*. <https://github.com/Schaggo/Unity-Microphone-Recorder.git>. 2020 (cit. on p. 52).