UNIVERSITÀ DEGLI STUDI DI UDINE
hic sunt futura

www.uniud.it

PhD course in:

Computer Science, Mathematics and Physics

Cycle XXXVI

Thesis title

# Virtual reality systems for anxiety mitigation: exploring desktop, immersive and augmented reality

Candidate
Marta Serafini

Supervisor
Luca Chittaro

Year 2024

Institute contacts
Department of Mathematics, Computer Science and Physics
University of Udine
Via delle Scienze, 206
33100 Udine — Italy
+39 0432 558400
https://www.dmif.uniud.it/

*A Diletta*

# Abstract

In recent years, virtual reality (VR) has emerged as a technological tool that could be used to mitigate anxiety, a condition that significantly impacts the well-being and health of a large population. In the literature, VR systems for anxiety mitigation can be classified into two main categories: VR exposure (VRE) therapy systems and VR systems for relaxation training. However, limited research has been conducted to compare different display types. This thesis focuses on three different types of anxiety, exploring and comparing different types of display: desktop VR, immersive VR, and augmented reality (AR) in both categories of VR systems for anxiety mitigation. In particular, we propose and evaluate VR systems designed to mitigate social anxiety-related disorders, i.e., school exam anxiety and public speaking anxiety, comparing the above mentioned display types. Then, we propose and assess a VR system with biofeedback for relaxation training, targeting anxiety mitigation more broadly, comparing desktop and immersive VR. We conclude the thesis by evaluating the therapeutic effects of our VR system with biofeedback for relaxation training on a clinical trial that involved patients with fibromyalgia.

# Contents

# List of Figures

# List of Tables

# Introduction

Anxiety is a condition that affects many individuals and has debilitating effects. According to the World Health Organization, anxiety and depression collectively cost the global economy 1$ trillion annually.

A valuable technique for individuals who wish to reduce the anxiety levels they experience is exposure therapy. This technique involves repeated exposure to the feared situation, leading to desensitization, progressively reducing the elicited anxiety, and making the situation less fearful [2]. There are three main approaches to exposure therapy: (1) in vivo exposure (IVE) that exposes individuals to the feared situation in real-world environments, (2) imaginal exposure that involves mentally imaging the feared situation, and (3) virtual reality exposure (VRE) that uses virtual environments (VEs), i.e., simulated environments created through computer generated images, to repeatedly expose individuals to the feared situation.

IVE and imaginal exposure approaches have traditionally been used in psychology to treat various types of anxiety. Over the past two decades, VRE has been increasingly used as an alternative to IVE and imaginal exposure in the treatment of anxiety (e.g., [3–6]).

In recent years, there has been a growing interest in augmented reality exposure (ARE) as an exposure approach. ARE uses interactive digital devices, such as computers, smartphones, tablets, or head mounted displays (HMDs), to virtually represent the feared situation within the individual's real-world environment. Unlike VRE, in ARE individuals are not immersed in the virtual world but are instead exposed to the feared situation virtually represented inside the real world.

Previous studies have shown that the fear response elicited in ARE is similar to that elicited in VRE [7,8]. Moreover, the effectiveness of ARE as a treatment tool for specific anxiety disorders has been found to be comparable to both VRE [9,10] and IVE [10,11]. Furthermore, outcomes achieved with ARE in treating specific phobias are maintained during follow-up assessments [11,12].

An alternative to exposure therapy for reducing anxiety levels involves the use of relaxation techniques. They could improve personal well-being [13] by lowering stress-related symptoms [14,15], relieving anxiety [16–18], and controlling postoperative pain [19,20]. Furthermore, relaxation-based interventions for medical conditions can be beneficial as an adjunct to standard medical care [21].

In particular, a specific relaxation technique that has been shown to reduce stress, anxiety, and depressive symptoms is slow and deep diaphragmatic breathing [15,17,22]. In addition to conventional instructor-led classes, breathing exercises can be learned through computer-based tools such as mobile breathing training apps [23]. More recently, immersive VR solutions for learning slow and deep diaphragmatic breathing have also been proposed [24–27].

In addition to these advancements, some authors have recently included biofeedback in VR systems for breathing and relaxation training. Biofeedback detects users' affective state by measuring their physiological activity and "feeds back" the detected information to users in real-time. In this way, it aims to enable users to learn over time how to change their physiological activity to enhance health and performance [28], reduce stress-related symptoms [29], and increase their feeling of well-being [13]. Existing biofeedback systems for relaxation provide users with biofeedback on breathing (e.g., [27]), cardiac (e.g., [30]), electrodermal (e.g., [31]), and brain activity (e.g., [32]).

This thesis explores the use of three different types of displays (i.e., desktop VR, immersive VR, and AR) to mitigate anxiety using both exposure therapy and relaxation techniques. In particular, this thesis:

- Proposes and assesses exposure systems in desktop VR, immersive VR or AR to mitigate specific social anxiety-related disorders, namely, oral exam anxiety and public speaking anxiety.

- Proposes and assesses a system with biofeedback for relaxation training comparing the immersive VR and desktop VR to mitigate anxiety in general.
- Assesses the system with biofeedback for relaxation training in a clinical trial.

The thesis is organized as follows. Chapter 1 provides an overview of the existing literature on VRE systems and VR systems with biofeedback for relaxation training. The chapter first compares different types of display used in exposure therapy. It then provides an overview of related work on VRE and ARE systems for social anxiety-related disorders, i.e., exam anxiety and public speaking anxiety. Subsequently, it presents and compares related research on VR systems with biofeedback for relaxation training. Chapter 2 focuses on VRE systems for exam anxiety. The chapter first illustrates the VRE system for oral exam anxiety we propose in desktop VR. Subsequently, it presents the feasibility and the home trial studies we conducted to evaluate the effectiveness of the system in eliciting anxiety in individuals. Chapter 3 focuses on VRE systems for public speaking anxiety. The chapter first introduces the exposure system for public speaking anxiety we propose in immersive VR and AR. It then presents the feasibility study we conducted to assess the effectiveness of our system in eliciting anxiety in individuals comparing two different types of display, namely immersive VR and AR. Chapter 4 focuses on VR systems with biofeedback for relaxation training. It first describes the immersive VR system with biofeedback for relaxation training we propose. Then, the chapter illustrates the two evaluations we conducted to assess its effectiveness in helping individuals to relax, specifically (i) comparing real biofeedback with sham biofeedback and (ii) comparing between immersive VR with desktop VR. Chapter 5 focuses on a real clinical setting where we assessed the effectiveness of our immersive VR system with biofeedback for relaxation training on a sample of patients with fibromyalgia. The chapter first explains fibromyalgia, highlighting its comorbidities with anxiety, and provides an overview of the existing literature on immersive and non-immersive VR systems for fibromyalgia. Then, the chapter describes the evaluation we conducted to assess the effectiveness of our system in helping patients reduce pain and anxiety levels. Finally, the thesis draws conclusion and outlines future work.

# 1

# VRE and ARE systems, and VR systems with biofeedback for relaxation training

In this chapter we introduce the main topics of this thesis: (i) VRE and ARE systems for social anxiety-related disorders, and (ii) VR systems with biofeedback for relaxation training. In particular, in Section 1.1, we provide an overview of the different approaches to exposure therapy, using three different types of displays, i.e., desktop VR, immersive VR, and AR. In Section 1.2, we present an overview of related work on VRE and ARE systems for exam anxiety and public speaking anxiety. In Section 1.3, we categorize the types of biofeedback mappings with users' physiological parameters within VR systems with biofeedback for relaxation training. Then, we present and compare related work on VR systems with biofeedback for relaxation training.

## 1.1 Different types of display in exposure therapy systems for anxiety mitigation

Over the past two decades, VRE has been increasingly used as an alternative to IVE and imaginal exposure in the treatment of anxiety [3–6]. VRE offers several advantages over IVE and imaginal exposure:

- It is more practical than IVE since it allows for exposures that might be challenging to implement in vivo (e.g., airplane flight to treat aviophobia), while maintaining comparable effectiveness in mitigating anxiety (e.g., [3]).
- It shows equal effectiveness in mitigating different types of anxiety disorders [3]. Therefore, VRE is a viable alternative for individuals unwilling or too scared to engage in IVE or have difficulty imagining the feared situation.
- It allows easier access to feared stimuli than IVE and allows for full control over the exposure process. Manipulations that would be impractical in IVE scenarios, such as repeating the takeoff of a virtual flight multiple times, become feasible in VRE [33].
- It allows for the use of a variety of specifically developed scenarios, helping in the prevention of the renewal of fear [34].
- It exhibits a low refusal rate as individuals generally prefer it over IVE [35].
- It ensures the absence of real danger for individuals since the feared situation is virtually represented.
- It guarantees confidentiality to individuals, a feature that may not always be feasible in IVE. For example, privacy can be compromised during IVE therapy conducted in an airplane flight [33].
- Its exposures are less expensive than IVE exposures, particularly in specific phobias. For instance, it offers a more economical solution for treating aviophobia compared to IVE therapy.

While the current literature primarily explores immersive VR systems for exposure therapy, the use of desktop VR systems remains unexplored. Although HMDs can increase users' sense of immersion, they restricts the use of the system to laboratory settings and to people who own an

HMD. Conversely, the use of widely available hardware would extend the accessibility of VRE systems to a broader audience, also at home. Moreover, although modern HMDs have made significant progress in limiting symptoms of motion sickness, they still carry the risk of negative side effects, such as nausea, headaches, and dizziness [36, 37]. On the other hand, the use of a VRE system using commonly available hardware would help minimize health risks. Furthermore, existing studies on the effects of immersive VR on users suggest that VR can lead to more visual fatigue than traditional screens [38]. Hence, this thesis also considers the use of desktop VR VRE systems.

Throughout the last decade, several studies have assessed the efficacy of VRE as a therapy for anxiety disorders. Some of these studies have compared results obtained on different types of anxiety disorders [3, 39, 40], while others have focused on a specific anxiety disorder, such as specific phobias [5, 6, 41, 42], social anxiety [43–47], or public speaking anxiety [43, 48, 49].

Multiple studies have shown that VRE reduces anxiety and phobia symptoms (e.g., [5]) and can lead to significant behavioral changes in real-life situations (e.g., [41]). Moreover, meta-analyses have indicated that the efficacy of VRE and IVE are comparable for different anxiety disorders [3, 39, 40], or specifically for social anxiety [46] and public speaking anxiety [43, 49].

In recent years, ARE has gained increasing interest. Research has shown that ARE can be equally effective as a treatment tool for specific anxiety disorders when compared to both VRE [7,8] and IVE [8,9]. Besides sharing the advantages of VRE, ARE offers the following further benefits:

- It facilitates exposure therapy in multiple real-world settings, allowing individuals to experience the virtual representation of the feared situation while remaining physically present in their real surroundings. In this way, individuals can experience their feared situation in the actual places they encounter difficulties, facilitating the transfer of the acquired skills to the real world.

- It exhibits higher ecological validity than VRE because individuals remain present in their real surroundings while engaging with the virtual feared situation.
- It reduces the time and costs of system development. Unlike VRE, which requires the creation of an entire VE, ARE only needs modeling the feared situation.
- It allows individuals to see their own body instead of a virtual representation of their body, as seen in VRE. This enables them to interact with the virtual feared situation using their own body, thereby enhancing the realism of their experience.

Despite the advantages ARE has over VRE, research on the feasibility and effectiveness of ARE systems are still limited. Existing ARE systems almost exclusively focus on the treatment of small animal phobias, primarily spiders (e.g., [9, 11, 50–52]) and cockroaches (e.g., [9, 11, 52–54]), with two exceptions addressing claustrophobia [8] and acrophobia [7], respectively.

Given the presented advantages that an ARE system can offer over a VRE system, it would be valuable to explore further the use of ARE systems in different types of anxiety disorders like social anxiety and public speaking anxiety.

In the following subsection, we provide an overview of the existing literature concerning VRE and ARE systems for social anxiety-related disorders, specifically exam anxiety and public speaking anxiety, that are the central topics of this thesis. The subsection focuses on systems that use computer-generated 3D VEs and virtual agents.

## 1.2 VRE and ARE systems for social anxiety-related disorders

### 1.2.1 Exam anxiety

In the following, we provide an overview of the existing literature concerning feasibility studies on VRE systems for exam anxiety. Since no trials in the literature have explored systems that simulate exams using virtual agents, we also focus on related studies. Specifically, we concentrate on trials of VRE systems that simulate one-on-one conversations in which the user is asked questions by a person assessing his/her skills in job interview contexts.

#### Feasibility studies

There are only three proposals of VRE systems for exam anxiety in the literature [55–57], while no ARE systems for exam anxiety have been proposed. Table 1.1 synthetically highlights the main differences between the existing works.

Table 1.1: Comparison of the main differences of studies conducted on VRE systems for exam anxiety.

| Reference | Type of exam | Type of display | Studied VR effect | Participants |
|---|---|---|---|---|
| [55] | Written | HMD | Elicited anxiety | 21 university students |
| [56] | Written | HMD | Elicited anxiety | 22 adolescents |
| [57] | Written | HMD | NS[a] | NS |

[a] NS=Not specified information

TAVE (Test Anxiety Virtual Environments) [55] is an immersive VRE system that provides users with three VEs, to be experienced in chronological order: student's home, representing the day before and then the morning of the exam; followed by subway journey to the exam location;

and finally school corridor and classroom where the exam takes place. A feasibility study was conducted to assess whether the different VEs of TAVE could elicit significantly different responses in students. The system was tested on 11 students with high levels of exam anxiety and ten students with low levels of exam anxiety. After experiencing each VE, students' level of anxiety was measured with two self-report questionnaires. Results showed that students with high exam anxiety experienced higher levels of anxiety than students with low exam anxiety in all the VEs. However, the level of anxiety in the two groups did not progressively increase with the sequence of VEs but peaked during the subway journey.

Kwon et al. [56] proposed an immersive VRE system that includes two VEs: house on the day before the exam and school on the day of the exam. They assessed the feasibility of the VRE system in inducing different levels of exam anxiety in adolescents. However, in addition to experiencing the VEs in an exposure session, participants performed also a meditation session to regulate anxiety immediately after the exposure session. The system was tested on 21 adolescents whose general anxiety and exam anxiety were measured with two self-report questionnaires at the beginning of the study, before they were exposed to the first VE. The study also measured participants' HRV with a physiological sensor during VEs exposure. After each session, adolescents' level of anxiety was measured. The main results of the study showed a significant difference in anxiety only between exposure and meditation sessions. The study does not instead specify if there were significant differences in perceived anxiety between the exposure sessions.

Luo et al. [57] developed a VRE system for exam anxiety containing three VEs aimed at eliciting increasing anxiety levels: home on the day before the exam, school entrance on the day of the exam, and classroom where the exam takes place. The system was assessed on middle school adolescents: unfortunately, while the paper states that participants experienced varying degrees of anxiety during VE exposures, it does not illustrate a study that might provide evidence about the system's ability to elicit increasing levels of anxiety.

The VRE systems for exam anxiety described above have three aspects in common:

1. They concern a written test. In TAVE [55], students are seated in a classroom and take a written test with multiple-choice, general knowledge questions. In [56], adolescents are seated in a classroom where a written test is about to start, but the test itself is not simulated. In [57], students are seated in a classroom and, after selecting their preferred subject from a list of five options, proceed to take a written test with questions from a college entrance test. In some educational systems, students take oral tests as an adjunct or an alternative to written tests, and several studies have highlighted the benefits of oral tests as assessment method [58–60]. The reliance on oral tests changes from country to country, for example they are central to the Italian educational system [61, 62], while the US educational system prefers written tests [59, 60, 63, 64]. As already mentioned, oral tests elicit higher levels of anxiety than written tests [65], and the availability of VRE systems for oral tests could thus benefit a large population of students worldwide. Moreover, the need for engaging interactions with the professor in oral exams introduces an additional challenge for students who also suffer from social anxiety. Therefore, it is crucial to provide students with scenarios in which they are not only exposed to an exam situation but also to the interaction with the professor.

2. Current VRE systems for exam anxiety require using an HMD. As previously described in Section 1.1, HMDs restrict the use of the system to laboratory settings and to individuals who own such devices. Conversely, the use of widely available hardware, such as computer monitors, would allow a much larger number of students to benefit from the VRE system, also at home, while also reducing health risks associated with HMD usage.

3. The studies on VRE systems for exam anxiety have exposed participants to the different VEs in a fixed order. In [55], both the group with high exam anxiety and the group with low exam anxiety were exposed to the three VEs in chronological order: first the home, then the subway, then the classroom VE. In [56], all participants were exposed to the two VEs in the same order: first the house, then the school VE. Exposing all participants to the VEs in a fixed order could lead to order effects, i.e., the participants' responses to the VEs could be affected by the participants' responses elicited in the previously experienced VEs.

**VRE systems for one-on-one conversations in job interview contexts**

Existing trials of VRE systems that simulate one-on-one conversations involve first-person interactions, where users engage in conversations with a virtual agent that asks work-related questions [66–71]. In particular, one study investigated the impact of the virtual agent realism on participants' anxiety levels during interview simulations [68]. Findings indicated that participants' anxiety was more affected by the virtual interviewer behavior than its degree of realism. Four studies were conducted using the same VRE system [66,67,69]. Specifically, Hartanto et al. [66] and Morina et al. [67] showed the feasibility of eliciting anxiety through virtual social interactions, while Kampmann et al. [69] compared the VRE therapy with IVE therapy and a waitlist control group. Results showed that both treatment groups improved social anxiety levels compared with the waitlist group, with IVE obtaining better results than VRE. Bouchard et al. [70] compared the same groups using a VRE system that offered participants various scenarios for training: public speaking, job interviews, conversations with supposed relatives, performing under the scrutiny of strangers, handling criticism, or managing insistence situations. Results showed that the VRE system was more effective than IVE in treating social anxiety. Furthermore, Zainal et al. [71] compared a VRE system that simulated job interviews and informal dinner party scenarios with a waitlist control group. Results showed a greater reduction in social anxiety symptoms, job interview fear, and trait worry among participants who used the VRE system, in comparison to the waitlist control group.

While the studies of Hartanto et al. [66] and Kwon et al.[68] focused exclusively on job interview scenarios, other authors explored different scenarios but none of them addressed oral exams. Among these studies, three involved participants experiencing all scenarios of the VRE system [66, 67, 69], while in the remaining studies, participants chose the scenarios for their training [70, 71].

The studies described have the following aspects in common:

- None of them concerns trials of VRE systems that simulate oral exams. The availability of such systems might benefit a large population of students worldwide who are affected by exam anxiety at different levels of intensity. Indeed, the Organization for Economic Cooperation and Development found in its report on students' well-being that 56% of students worldwide experienced high levels of exam anxiety, even when adequately prepared for tests [72].

- Current VRE systems have never been trialed by participants at home or in the actual place they are intended to be used. Instead, they have always been tested within laboratory settings.

- Current studies are carried out on systems that require an additional person using the Wizard of Oz technique to control the virtual agent behavior [68, 69, 73], script [67, 69, 70, 74], or response timing [66]. Only one trial involves a system that does not require this technique, but it consists of 360° videos with actors following a predefined script [71].

- No existing trial considered systems that allow users to customize their conversations with the virtual agent. A greater level of customization would provide users with a personalized experience that closely aligns with their specific exam peculiarities.

### 1.2.2 Public speaking anxiety

Literature on VRE systems for public speaking anxiety can be categorized into two types of studies:

1. Feasibility studies, which aim to assess the effectiveness of the system in eliciting anxiety levels.

2. Treatment studies, which aim to assess the efficacy of the system in mitigating anxiety.

#### Feasibility studies on VRE systems

Among feasibility studies on VRE systems [75–79], two studies compared anxiety levels and physiological responses of participants when exposed to real or virtual audiences [75,79].

In [79], authors found that participants experienced significantly higher anxiety levels while giving a speech in front of a real audience, a virtual audience, and an empty virtual auditorium than before starting their speech. Participants' heart rate (HR) and anxiety levels elicited by the real and the virtual audiences were similar, while HR and heart rate variability (HRV) measures revealed that the real audience elicited significantly higher arousal than the empty virtual auditorium.

A within-subjects study [75] exposed participants to a real audience and three virtual audiences of different sizes. Results showed similar anxiety levels across the audiences, but participants' HR was higher during the speech task than after the speech task in all conditions. Moreover, HR was significantly higher in the small virtual audience condition than the medium and the big audience conditions, but only when the small audience was the first virtual condition experienced by participants.

Other studies on VRE systems for public speaking anxiety have used different audience sizes [80–83], but only in [75] the effects of this aspect on participants were specifically analyzed. It is worth mentioning that another study investigated the effect of audience size on preservice teachers [84], finding that facing a big class elicited higher stress and HR levels compared to a small class.

**Treatment studies on VRE systems**

Regarding treatment studies, previous research has shown the effectiveness of VRE systems in reducing public speaking anxiety [80–83, 85–88]. Positive outcomes were maintained in follow-ups at one-month [81] and three-month [80, 87] intervals. One study has shown that VRE treatment, compared to a no-treatment condition, reduced participants' public speaking anxiety levels and lowered their HR [85]. Other studies have shown that VRE treatments for public speaking anxiety succeeded in lowering levels of distress [81], anxiety, arousal, and HR [83]. Furthermore, the comparison between IVE and VRE treatment for public speaking anxiety in [88] showed that both approaches were effective in reducing anxiety levels, despite the real audience eliciting higher levels of anxiety than the virtual audience.

**Features of VRE studies**

In general, both types of studies on VRE systems for public speaking anxiety involved participants giving a speech either in front of an empty room [78, 79, 85], or a virtual audience. The virtual audience could have a predefined behavior that could be neutral [75, 77–79, 87], or positive or negative [77, 81, 85, 86, 88]. The VEs used in these studies can be classified into three categories:

1. A room with an audience arranged around a single table consisting from three to 18 virtual agents [75, 77, 78, 80–82].
2. A classroom with an audience seated behind tables arranged in multiple rows, with a number of virtual agents from six to 46 [75, 81–83, 86–88].

3. An auditorium with an audience formed by a number of at least 20 virtual agents [79, 80, 82, 85].

In two studies, participants could personalize the VE and the audience characteristics [82,83]. The duration of the speech presented to the virtual audience varied across the studies. Some speeches lasted from one to three minutes [77, 82, 87, 88], while others were five or six minutes long [75,77, 79,81]. In [83], the speech duration was 20 minutes, divided into 5-minute blocks. Regarding the preparation time given to participants, most studies allowed two to five minutes to prepare the speech [75,78,79,81,83,86,88]. In contrast, two studies provided participants with longer preparation periods of two days [77] and two weeks [85], respectively.

Regarding the topic of the speech, it could be predefined for all participants [75, 79, 81–83, 85, 88], freely chosen by participants [77], or selected from a limited list of options [78, 86]. Eight of these systems require the intervention of an outside operator, typically the therapist or experimenter, to trigger audience reactions at the appropriate time, such as asking the speaker a question [75,77,80,82,85,86]. Table 1.2 summarizes the main features of VRE systems for public speaking anxiety.

Table 1.2: Features of studies on VRE systems for public speaking anxiety.

| **Ref.** | **Virtual audience** | **VE** | **OO**[1] | **Speech time** | **Speech prep. time** | **Speech topic** |
|---|---|---|---|---|---|---|
| [85] | None, positive, negative | Auditorium with NS[2] number of agents | Yes | 2 min | 2 weeks | NS |
| [78] | None, neutral | 5 agents seated around a table | No | NS | 5 min | Selected from a NS list of topics |
| [79] | None, neutral | Auditorium with 20 agents | No | 5 min | 5 min | Bhutan's culture, form of government, and geography |
| [75] | Neutral | 3 agents seated around a table, 6 or 15 agents seated behind tables arranged in multiple rows | Yes | 5 min | 5 min | Job interview as police officer, teacher, travel tour guide, and salesperson |
| [77] | Neutral, positive, negative | 8 agents seated around a table | Yes | 5 min | 2 days | Free topic |
| [81] | Positive, negative | 10 or 18 agents seated around a table, 48 agents seated behind tables arranged in multiple rows | No | Max 6 min | 3 min | Selected between: (i) attractions and character of Melbourne, (ii) pros and cons of public transport, (iii) importance of managing our environment |
| [86] | Positive, negative | Max 32 agents seated behind tables arranged in multiple rows | Yes | NS | 3 min | 3 topics randomly selected from a list of 5 controversial NS topics |
| [88] | Positive, negative | 25 agents seated behind tables arranged in multiple rows | No | 2 min | 2 min | Should teenagers spend more time in nature?, what should the dream house look like?, is graffiti an art form?, can happiness be bought? |

[1]OO=Outside operator
[2]NS=Not specified information

| Ref. | Virtual audience | VE | OO | Speech time | Speech prep. time | Speech topic |
|---|---|---|---|---|---|---|
| [80] | Positive, negative | 5 agents seated around a table, auditorium with 22 agents | Yes | NS | NS | NS |
| [82] | Positive, negative | Customizable number of agents seated around a table | Yes | 1–3 min | NS | Present yourself, exercises i.e., name words starting from a specific letter |
| [83] | Custom. (positive, neutral, negative) | Customizable number of agents (6, 12, 20) seated behind tables arranged in multiple rows | No | 3 min | 20 min divided in 5-min blocks | Experience of being a college student |
| [87] | NS | 10 agents seated behind tables arranged in multiple rows | No | NS | 1–2 min | NS |

### ARE studies and VRE vs. ARE studies

There is a research gap regarding the feasibility of ARE systems for public speaking anxiety and their effectiveness in eliciting levels of anxiety and distress. Furthermore, the comparison of immersive VR and AR in eliciting anxiety in the same feared situation has been explored only in two previous studies for specific anxiety disorders, showing mixed results [7,8]. One study [8] found that the exposure to an environment with claustrophobic characteristics displayed in immersive VR resulted in higher anxiety levels than when displayed in AR. However, the measurement of participants' HRV showed the opposite result. In [7], an immersive VR environment and an AR environment with acrophobic characteristics elicited similar anxiety levels in participants.

Given the current research gap in ARE systems for public speaking anxiety, there is an opportunity to explore their potential in eliciting anxiety and distress to provide an exposure environment where individuals can train to give presentations in front of a virtual audience. Additionally, further investigation is needed to assess the impact of different audience sizes in VR and extend research to ARE systems.

### Audience gaze behavior studies

Understanding individuals' gaze behavior during a speech in front of a virtual audience may enhance the design of audience behaviors in systems for public speaking training. Previous studies have analyzed individuals' gaze behavior during speech presentations to a virtual audience [75,76, 89–92]. Three studies compared healthy individuals with patients with social anxiety disorder [76, 91] or with high social anxiety levels [90]. Specifically, in [76], participants with social anxiety disorder looked less at the virtual audience than healthy participants. In [91], participants with social anxiety disorder looked less at the pre-recorded audience than healthy participants. Moreover, both groups looked more at audience agents with negative or neutral behaviors than areas unrelated to the audience. In [90], healthy participants looked significantly less at pre-recorded audience agents with negative behavior than those with neutral or positive behavior. In contrast, these differences were not observed in participants with high social interaction anxiety levels. Moreover, healthy participants looked significantly more at audience agents with positive behavior and less at agents with negative behavior than participants with high social interaction anxiety levels.

In three additional studies [75,89,92], researchers investigated individuals' gaze behavior while giving a speech in front of a virtual audience. In [92], female participants who had previously attended a course on presentation skills looked more at the virtual audience than female participants who had not attended any course on presentation skills. In [89], participants with high public speaking anxiety levels looked less at uninterested pre-recorded audience agents compared to interested ones.

A further study [93] assessed the participants' gaze behavior towards a virtual audience of 11 members who listened to a virtual speaker's speech. Among the results, it emerged that participants looked more at agents who always looked at them than agents who never looked at them. It is important to note that in this study, participants had to observe the virtual audience and did not give a speech in front of them. This might potentially yield different results.

In summary, the studies in the literature that have investigated gaze behavior towards the audience during a public speech showed that participants with high social anxiety levels looked less at the audience than healthy participants [76,90,91], and looked more at interested agents than uninterested agents [89]. However, it is worth noting that the presented studies did not investigate whether the behavior of virtual audience agents influenced participants' gaze behavior [76, 92], and the studies that did explore this aspect utilized desktop VR systems [90, 91], used pre-recorded audiences [89–91], or did not employ eye-tracking systems to collect eye gaze data [75], which may have limited the accuracy of the collected data. Therefore, there is a need for further investigation into individuals' gaze behavior towards the virtual audience during a public speech.

## 1.3 VR systems with biofeedback for relaxation training

After presenting VRE and ARE systems for social anxiety-related disorders, in this subsection we focus on VR systems with biofeedback that use relaxation training to reduce anxiety levels. Specifically, we first provide an overview of the types of biofeedback mappings with users' physiological parameters within VR systems with biofeedback for relaxation training. Subsequently, we present and compare existing studies on VR systems with biofeedback for relaxation training.

The Association for Applied Psychophysiology and Biofeedback defines biofeedback as a process that enables individuals to learn how to change their physiological activity to improve their health and performance [28]. Biofeedback employs sensors to measure the user's physiological parameters and then provides the collected information to the user in real-time. Immediate feedback helps the user gain voluntary control over the physiological process and induces favorable changes. Over time, these changes can persist without continuously using instrumentation [28]. Biofeedback can be employed to increase awareness of breathing activity, as well as other physiological parameters to alleviate stress [94]. Indeed, increased awareness of breathing leads to anxiety reduction (e.g., [95]) and improved relaxation, (e.g., [96]). For these reasons, slow and deep breathing is commonly used in anxiety and stress reduction approaches, (e.g., [15,97]).

VR systems with biofeedback for relaxation training can use either non-immersive displays (such as PC monitors) or immersive displays (such as VR HMDs). Table 1.3 synthetically shows the main features of such systems. As shown in the table, nine systems reproduce natural environments in land, maritime, or underwater settings [26,27,30,32,98–102]. Three other systems use the VE of an office, a maze, or a scary mansion to train relaxation during stressful situations [103–105]. Two systems use an empty VE containing only a single 3D object whose movements match users' breathing or skin conductance (SC) [31,106]. Almost all VR systems with biofeedback for relaxation training employ a single physiological sensor whose value is mapped into one or more VE elements. As shown in Table 1.3, seven systems use physiological measurements of breathing activity, four systems use HR or HRV, while other four systems use SC, muscle or brain activity. Only two systems use more than one physiological measurement: in [32], alpha and theta brain waves are mapped into two different VE elements; in [103], data from facial muscle activity, SC and HR are used to derive a single stress value. It is worth noting that four systems do not use traditional breathing sensor to record users' breathing data. They respectively use a hand controller placed on user's abdomen [26,98], a handmade spirometer-like device [101], and a microphone [102].

To provide the user with feedback on his/her current physiological activity, existing VR systems with biofeedback for relaxation training map user's physiological parameters into one or more of the following five categories:

1. *Attributes of VE elements*: the system changes one or more attributes of elements within the VE following user's physiological activity. These elements are naturally embedded in the VE, e.g., clouds, plants, or fog. Ten systems map user's physiological activity into attributes of VE elements such as color, brightness, position, size, or quantity, e.g., amount of fog/clouds, or flames of a campfire. Incorporating feedback into the elements of the VE increases the salience and attractiveness of feedback, fostering motivation and focus [24,30], and improving users engagement.

2. *2D data visualizations*: the system employs two-dimensional data

visualizations to display user's physiological activity. Unlike the previous category, these visualizations are overlaid on the VE. They can take different forms, such as icons, graphs, or circles that change in response to the user's physiological activity. Four systems use 2D visualizations. In [105], a heart-shaped 2D icon shows to the user his/her current HR by increasing the color filling of the icon as HR increases. In [100], a line graph displays the user's current breaths per minute (bpm) with a line, comparing it to another line that shows the breathing rate of 5.5 bpm. In [27, 102], the user can observe his/her breathing through a circle that grows with each inhalation and shrinks with each exhalation.

3. *Locomotion*: user's breathing activity controls a locomotion technique to allow the user navigate the VE. Three systems use breath-based locomotion. In [26, 27], the user must maintain slow and deep breathing to smoothly and continuously navigate the VE. In a non-immersive platform game described in [101], the user controls the vertical position of a virtual fish through his/her inhalations and exhalations and must collect as many starfish as possible. The starfish are arranged following a sinusoidal path so the easiest and most comfortable way to collect them is by maintaining a slow and continuous deep breathing. It should be noted that when using immersive displays, a mapping on locomotion may increase the risk of motion sickness because the user's point of view moves continuously while his/her head remains still, causing a sensory conflict between the visual and vestibular systems [107].

4. *Task difficulty*: the system adapts the task difficulty to a physiological parameter of the user. Following the operant conditioning paradigm (i.e., the method of learning that encourages behavior change by using rewards and punishments [108]), the system trains the user to self-regulate his/her physiological activity in stressful situations. To facilitate success in the task, the user needs to maintain his/her

physiological activity under (or above) a predefined threshold value. Exceeding the threshold indicates increased arousal of the user, who is penalized by increasing the task difficulty. One system maps user's HR on task difficulty [105]. Since increasing task difficulty may elicit negative emotions and stress in users, this mapping allows users to train themselves to regulate such emotions in the presence of stressors [109].

5. *Virtual character*: the system changes the appearance or behavior of a virtual character based on the physiological activity of the user. One system maps user's physiological parameters on a virtual character [103]. The user's data from facial muscle activity, SC, and HR are used to derive the user's stress level that is embodied into the affective state and behavior of a virtual character. Higher user's stress level leads to worse character's behavior, e.g., displaying anger and struggling in the completion of its tasks.

Table 1.3: Categorization of VR systems with biofeedback for relaxation training.

| Ref. | Display | VE | Physiological data used for biofeedback | Mapping of physiological activity | Type of mapping |
|---|---|---|---|---|---|
| [103] | Non-immersive (PC monitor) | Office setting | Cardiac activity, electrodermal activity, muscle activity of zygomaticus major and corrugator supercilii | Data from facial muscle activity, SC, and HR are used to derive a single stress value that is mapped on the behavior of a virtual character | Virtual character |
| [105] | Non-immersive (PC monitor) | Three VEs: maze with a spirit, maze with a ball, table seen from above with two pairs of hands | Cardiac activity | Greater HR increases the color filling of a heart-shaped 2D icon; maintaining HR low allows to outrun the spirit from the maze; greater HR increases the ball size; greater HR increases opponent speed in a hand-slapping competition | Task difficulty; 2D data visualizations |
| [104] | Non-immersive (PC monitor) | Scary mansion | Brain activity | Data from brain activity are used to derive a single relaxation value which is mapped on the gradation of light glowing from the avatar's helmet | Visual attributes of VE elements |
| [101] | Non-immersive (PC monitor) | Underwater setting | Breathing activity | Inhalations move the avatar upwards and exhalations move it downwards, resulting in forward locomotion | Locomotion |
| [26] | Immersive (VR HMD) | Two VEs: hilly setting, maritime setting | Breathing activity | Slow breathing moves the user forward in the VE; inhalations and exhalations change the color of elements in the VE, the grass growth, and the emission of particles from blossoms | Locomotion; visual attributes of VE elements |
| [98] | Immersive (VR HMD) | Hilly setting | Breathing activity | Inhalations and exhalations change the color of flowers, rocks and tree fruits | Visual attributes of VE elements |

| Ref. | Display | VE | Physiological data used for biofeedback | Mapping of physiological activity | Type of mapping |
|---|---|---|---|---|---|
| [99] | Immersive (VR HMD) | Three VEs: glade with a campfire, waterfall, maritime setting | Cardiac activity | Greater HR increases fire intensity, sea waves movement, waterfall movement, and size of pre-selected words or images related to personal stressful events | Visual attributes of VE elements |
| [27] | Immersive (VR HMD) | Underwater setting | Breathing activity | Slow and deep breathing moves the user forward in the VE; shallow breathing applies gravity which lies the user to the ground; inhalations and exhalations change the color and illumination of plants; inhalations and exhalations grow and shrink a 2D circle | Locomotion; 2D data visualizations; visual attributes of VE elements |
| [100] | Immersive (VR HMD) | Maritime setting | Breathing activity | Breath activity is shown with a line graph; maintaining optimal breathing rate makes fog disappear from the VE | 2D data visualizations; visual attributes of VE elements |
| [102] | Immersive (VR HMD) | Glade with a tree | Breathing activity | Inhalations and exhalations grow and shrink a 2D circle | 2D data visualizations |
| [31] | Immersive (VR HMD) | Empty (only a single 3D object) | Electrodermal activity | Lower SC moves the sun position until below the horizon, then it rises the moon | Visual attributes of VE elements |
| [32] | Immersive (VR HMD) | Maritime setting | Brain activity | Greater theta wave lifts the user's floating position in the VE; greater alpha wave increases the opacity of an energy bubble surrounding the user | Visual attributes of VE elements |
| [106] | Immersive (VR HMD) | Empty (only a single 3D object) | Breathing activity | Inhalations and exhalations move a 3D cloud-shaped object closer and farther away from the user | Visual attributes of VE elements |

| Ref. | Display | VE | Physiological data used for biofeedback | Mapping of physiological activity | Type of mapping |
|---|---|---|---|---|---|
| [30] | Immersive (VR HMD) | Maritime setting | Cardiac activity | Greater HRV turns on lights, clears sky from clouds, moves a sailing boat, increases wind and wave sound volume, activates up to 7 lamps on a landing stage, lights campfire and flashlights with crackling sound | Visual and auditory attributes of VE elements |

### 1.3.1 Design and results of studies

The studies of VR systems with biofeedback for relaxation training can be categorized in non-comparative studies, within-subjects studies, or between-subjects studies, as shown in Table 1.4. Eight systems were evaluated with a longitudinal study using within-subjects (three studies) or between-subjects design (five studies). The most frequent sample size involved from 8 to 25 participants (seven studies). Larger sample sizes involved 35 to 45 (three studies), 60 to 72 (four studies), 86 to 138 (three studies) participants. One study involved 411 participants. Table 1.4 summarizes previous studies about the effects of VR systems with biofeedback for relaxation training.

Measures used in the studies can be categorized into self-reports by participants (subjective measures) and derived from user's physiological parameters (objective measures). As shown in Table 1.4, level of anxiety is the most used subjective measure, followed by relaxation, while the most used objective measures are HR and HRV, followed by breathing activity and SC.

In five studies, participants were asked to relax, while in ten studies they were asked to maintain slow and diaphragmatic breathing following the rhythm of a pacer or an audio guided meditation. In two studies, participants were asked to perform meditation exercises.

Regarding objective measures, five studies showed that cardiac (HRV or HR) parameters improved during or after the use of the system. However, no study found significantly different values of cardiac parameters between VR with biofeedback and VR without biofeedback conditions. Tinga et al. [106] evaluated instead the immersive VR placebo condition against the immersive VR biofeedback condition finding a lower HR in the immersive VR placebo condition than the immersive VR biofeedback condition. However, it should be kept in mind that the system provides a primitive and limited immersive VR experience that only displays a cloud in an empty VE. Regarding subjective measures, twelve studies found an improvement in anxiety, relaxation, stress, or pain scores after using the system. In three studies, no significant differences were found on subjective measures between biofeedback and non-biofeedback conditions, but two studies found differences between an immersive and non-immersive VR biofeedback system [24,30]. The immersive VR biofeedback system used a rich biofeedback that mapped users' physiological activity on attributes of multiple VE elements, whereas the non-immersive VR biofeedback system used a simple biofeedback that mapped users' physiological activity on the color of a 2D circle. Results showed that the rich biofeedback led to less mind wandering and both greater relaxation self-efficacy and focus on the present moment than the simple biofeedback [24]. Moreover, participants who tried the rich biofeedback perceived a faster passing of time, expressed greater intention to use the system, and were more likely to recommend it than participants who tried the simple biofeedback [30]. Finally, both studies found that the rich biofeedback resulted in a more enjoyable experience than the simple biofeedback.

Considering the cost and complexity of adding biofeedback to a VR relaxation system, such choice should be adequately supported by evidence. In particular, the biofeedback system should provide more accurate feedback, and make relaxation easier to achieve than the same system without biofeedback. However, the assessment of the effectiveness of biofeedback in improving relaxation is rarely addressed in proposals of such VR systems. A placebo condition should instead be added to the study when evaluating biofeedback systems. The placebo condition is a control condition in which unaware users are given a sham treatment instead of the real one to assess the actual effectiveness of the real treatment. It is commonly used in medical studies because factors such as user suggestibility can lead to measuring positive effects, and improvements in well-being, even with sham treatments. Therefore, the evaluation of VR systems with biofeedback should carefully consider how traditional, non-VR systems with biofeedback have been evaluated in medical studies, e.g., [110–113], and follow those research methods to assess if their proposed treatment is actually better than a sham treatment. However, among the available studies of VR systems with biofeedback for relaxation training, only two considered sham biofeedback [103, 106]. In [103], desktop VR was used, but neither the relaxation effects nor physiological measurements were analyzed for statistical analysis. In [106], the system provided a very primitive and limited immersive VR experience (i.e., an empty VE that only displays a cloud moving towards and away from the user). The low number of previous studies leads to the need to investigate further whether the effects of VR systems with biofeedback for relaxation training are actually due to the use of real biofeedback by comparing it with placebo biofeedback.

Table 1.4: Categorization of previous studies of VR systems with biofeedback for relaxation training. Letters in the column "Task" refer to the conditions in the column "Experimental conditions". The column "P" provides information about the number of participants.

| Ref. | Design | P | Experimental conditions | Measures | Task | Main statistically significant findings |
|---|---|---|---|---|---|---|
| [103] | Within-subjects | 35 | (A) Desktop VR biofeedback with a single-sensor stress detection algorithm (B) Desktop VR biofeedback with a multi-sensor stress detection algorithm (C) control condition (desktop VR placebo biofeedback) | *Subjective measures*: perceived biofeedback quality, difficulty of relaxation training *Objective measures*: none | Stay calm and relaxed to allow the virtual character to remain focused on progressing in its task | Perceived biofeedback quality higher in desktop VR biofeedback with the single-sensor stress detection algorithm rather than the desktop VR placebo biofeedback |
| [99] | Longitudinal, between-subjects | 20 | (A) Immersive VR biofeedback (8 sessions over an unspecified period) (B) control condition (no treatment) (C) Mobile VR without biofeedback (8 sessions over an unspecified period) | *Subjective measures*: anxiety, worry *Objective measures*: HR, SC | (A) Stay relaxed by observing the flickering campfire (Sessions 1,2), the waves lapping on a shore (Sessions 3,4), the waterfall (Sessions 5,6), stressful images (Sessions 7,8) (C) as (A) | Anxiety lower after treatment than before treatment in VR conditions; worry lower after treatment than before treatment in mobile VR without biofeedback and no treatment conditions |

| Ref. | Design | P | Experimental conditions | Measures | Task | Main statistically significant findings |
|---|---|---|---|---|---|---|
| [101] | Within-subjects | 16 | (A) control condition (casual conversation with the experimenter) (B) Desktop VR biofeedback (first time) (C) control condition (action maze chase video game) (D) Desktop VR biofeedback (second time) (E) control condition (casual conversation with the experimenter) (F) traditional relaxation activity (G) control condition (casual conversation with the experimenter) | *Subjective measures*: none *Objective measures*: HRV (RMSSD) | (A) converse with the experimenter about the introduction of the study (B) play the desktop VR biofeedback game (i.e., maintain slow breathing to move the fish up and down following a sine wave, collecting as many stars as possible) (C) play the action maze chase video game (D) play the immersive VR biofeedback game (E) converse with the experimenter about the VR and game experiences (F) relax as much as possible (G) converse with the experimenter about all the previous tasks | RMSSD lower during action maze chase video game than all other conditions; RMSSD higher during second desktop VR biofeedback game than first casual conversation with the experimenter |

| Ref. | Design | P | Experimental conditions | Measures | Task | Main statistically significant findings |
|---|---|---|---|---|---|---|
| [26] | Longitudinal, within-subjects | 45 | (A) control condition (no treatment)<br>(B) Immersive VR biofeedback (6 sessions during one week) | *Subjective measures*: experience satisfaction, usefulness of diaphragmatic breathing and relaxation, simulator sickness, ease of use, ease of performing diaphragmatic breathing, breath awareness, relaxation, stress, burnout, relaxation self-efficacy<br>*Objective measures*: breathing activity, share (i.e., percentage of the total training session duration) of inhalations and exhalations | (B) Maintain diaphragmatic breathing to move along a predefined path and advance in the game | Share of inhalations and share of exhalations for each immersive VR biofeedback session higher than the average of all previous sessions (except for share of inhalations in Session 5); ease of performing diaphragmatic breathing higher in Sessions 3, 5, 6 than the average of all previous sessions; breath awareness and relaxation self-efficacy higher after immersive VR biofeedback treatment than no treatment; stress and burnout lower after immersive VR biofeedback treatment than no treatment |
| [27] | Non-comparative | 86 | Immersive VR biofeedback | *Subjective measures*: anxiety, positive and negative affect, performance pressure, experience satisfaction<br>*Objective measures*: breathing activity | Maintain diaphragmatic breathing to move, and explore freely the VE | Anxiety lower after immersive VR biofeedback session than before immersive VR biofeedback session |

| Ref. | Design | P | Experimental conditions | Measures | Task | Main statistically significant findings |
|---|---|---|---|---|---|---|
| [100] | Longitudinal, between-subjects | 12 | (A) Immersive VR biofeedback (B) control condition (no treatment) | *Subjective measures*: anxiety, pain, sense of presence, system usability *Objective measures*: none | (A) Maintain slow breathing following a sine wave | No analysis of statistical significance |
| [104] | Longitudinal, between-subjects | 136 | (A) Desktop VR biofeedback (B) control condition (puzzle platform video game) | *Subjective measures*: experience satisfaction, anxiety *Objective measures*: none | (A) stay relaxed to light up the game scene and advance in the game (B) play the puzzle platform video game | No statistically significant results |
| [102] | Within-subjects | 20 | (A) Immersive VR biofeedback (B) audio guided meditation | *Subjective measures*: global health, emotional distress/anxiety, emotional distress/anger, pain *Objective measures*: none | (A) breathe following the rhythm of a pacer (B) maintain awareness of breath while doing a meditation exercise (body scan) | Pain lower after session than before session in both conditions; distress/anxiety lower after session than before session in audio guided meditation |
| [31] | Non-comparative | 411 | Immersive VR biofeedback | *Subjective measures*: relaxation *Objective measures*: HR, SC, breathing activity | Stay relaxed | Relaxation higher after immersive VR biofeedback session than before VR biofeedback session |

| Ref. | Design | P | Experimental conditions | Measures | Task | Main statistically significant findings |
|---|---|---|---|---|---|---|
| [114] | Longitudinal, between-subjects | 138 | (A) Desktop VR biofeedback (6 times during three weeks)<br>(B) control condition (platform video game, 6 times during three weeks) | *Subjective measures*: anxiety, anxiety-reducing expectations about the experience<br>*Objective measures*: none | (A) stay relaxed to evade a spirit, exit from a maze without hitting the walls with a ball, and win a hand slapping contest<br>(B) play the platform video game | Anxiety lower after treatment than before treatment in both conditions; linear decrease in anxiety higher in desktop VR biofeedback than platform video game |
| [106] | Between-subjects | 60 | (A) Immersive VR biofeedback<br>(B) control condition (Immersive VR placebo biofeedback)<br>(C) control condition (empty VE without 3D objects) | *Subjective measures*: tension, calmness, experience satisfaction<br>*Objective measures*: HR, HRV (RMSSD), alpha wave and theta wave (theta to alpha ratio) | Stress arithmetic task; breathe following an audio guided meditation | Tension and HR lower in breathing task than stress arithmetic task; calmness, theta to alpha ratio, and RMSSD higher in breathing task than stress arithmetic task; HR better in immersive VR placebo biofeedback than immersive VR biofeedback |
| [115] | Longitudinal, within-subjects | 8 | (A) Immersive VR biofeedback (6 sessions)<br>(B) control condition (no treatment)<br>The two conditions are alternated during four weeks | *Subjective measures*: anxiety, disruptive classroom behavior<br>*Objective measures*: none | (A) Maintain diaphragmatic breathing to move and explore freely the VE | Anxiety lower in overall immersive VR biofeedback sessions than overall no treatment sessions; anxiety lower immediately after and two hours after session than before session in immersive VR biofeedback |

| Ref. | Design | P | Experimental conditions | Measures | Task | Main statistically significant findings |
|---|---|---|---|---|---|---|
| [105] | Longitudinal, within-subjects | 8 | Desktop VR biofeedback (8 sessions during four weeks) | *Subjective measures*: experience satisfaction, difficulty of training, behavioral problems in everyday life, anxiety *Objective measures*: none | Stay relaxed to evade a spirit, exit from a maze without hitting the walls with a ball, and win a hand-slapping contest | No analysis of statistical significance |
| [30] | Between-subjects | 68 | (A) Immersive VR biofeedback (B) 2D colored circle biofeedback (C) control condition (video of a natural environment) | *Subjective measures*: mood (i.e., good-bad mood, alertness-tiredness, rest-unrest subscales), experience enjoyment, intention to use, recommendation, perception of passing of time, attentional focus *Objective measures*: HR, HRV (RMSSD, SDNN, LF/HF ratio, coherence ratio) | (A) maintain deep diaphragmatic breathing following the rhythm of a pacer (B) as (A) (C) stay relaxed while watching a video of a natural environment | Rest-unrest mood higher in immersive VR biofeedback than video of a natural environment; HR lower during session than after session in all conditions; RMSSD higher during session than before and after session in all conditions; SDNN, coherence ratio, LF/HR ratio higher in biofeedback conditions than video of a natural environment; experience enjoyment and attentional focus higher in VR biofeedback than other conditions; intention to use and recommendation higher in immersive VR biofeedback than 2D colored circle biofeedback; perception of passing of time quicker in VR biofeedback than other conditions |

| Ref. | Design | P | Experimental conditions | Measures | Task | Main statistically significant findings |
|---|---|---|---|---|---|---|
| [24] | Between-subjects | 60 | (A) Immersive VR biofeedback<br>(B) 2D colored circle biofeedback | *Subjective measures*: relaxation, relaxation self-efficacy, mind wandering, focus on the present moment, attentional resources, experience enjoyment<br>*Objective measures*: HR, HRV (RMSSD, coherence ratio) | Attentional resource task, breathe following the rhythm of a pacer; attentional resource task | HR lower after session than before session in both conditions; relaxation higher after session than before session in both conditions; relaxation self-efficacy and focus on the present moment higher in VR biofeedback than 2D colored circle biofeedback; mind wandering lower in immersive VR biofeedback than 2D colored circle biofeedback; coherence ratio and RMSSD higher during session in both conditions; experience enjoyment higher in immersive VR biofeedback than 2D colored circle biofeedback |
| [116] | Longitudinal, between-subjects | 25 | (A) Immersive VR biofeedback (8 sessions over an unspecified period)<br>(B) control condition (no treatment)<br>(C) Mobile VR without biofeedback (8 sessions over an unspecified period) | *Subjective measures*: anxiety<br>*Objective measures*: HR, SC | (A) Stay relaxed while observing the flickering campfire (Sessions 1,2), the waves lapping on a shore (Sessions 3,4), the waterfall (Sessions 5,6), stressful images (Sessions 7,8)<br>(C) as (A) | HR and anxiety lower after session than before session in VR conditions; anxiety lower after treatment than before treatment in VR conditions |

| Ref. | Design | P | Experimental conditions | Measures | Task | Main statistically significant findings |
|---|---|---|---|---|---|---|
| [32] | Within-subjects | 43 | (A) Immersive VR biofeedback (B) Immersive VR without biofeedback (C) Desktop VR without biofeedback | *Subjective measures*: meditation depth (i.e., negative feelings, relaxation, self-reflection, emotions arisen, sense of presence, feeling of non-duality) *Objective measures*: none | perform meditation exercises (body scan and focused attention); stress memory task after each meditation exercise | Negative feelings lower in immersive VR conditions than desktop VR without biofeedback; relaxation, self-reflection, emotions arisen, and feeling of nonduality higher in immersive VR conditions than desktop VR without biofeedback |
| [98] | Between-subjects | 72 | (A) Immersive VR biofeedback (B) Immersive VR without biofeedback | *Subjective measures*: user experience, focus on breath *Objective measures*: breathing activity, share (i.e., percentage of the total training session duration) of inhalations and exhalations, HRV (RMSSD, LF, HF) | Focus on breath and maintain slow diaphragmatic breathing | Focus on breath higher in immersive VR biofeedback than immersive VR without biofeedback; share of inhalations and share of exhalations higher in immersive VR biofeedback than immersive VR without biofeedback |

# 2

# Desktop VR as an exposure method for oral exam anxiety

This chapter proposes a system that exposes individuals to oral exam scenarios in desktop VR, where a virtual agent acts as an examiner to induce anxiety. As mentioned in Section 1.2, the use of an HMD can enhance users' sense of immersion but restricts the use of the system to laboratory settings and to individuals who own an HMD. In contrast, the use of widely available hardware should allow a broader range of individuals to access and benefit from the VRE system, also at home.

To be effective, systems for exposure therapy must be able to elicit anxiety in individuals. Therefore, before using the system as a tool for anxiety mitigation, it is essential to validate its capacity to elicit anxiety. In this regard, we have conducted a feasibility study to assess that the three difficulty levels of our VRE system elicit significantly different levels of anxiety in individuals. Then, we conducted a qualitative study of a home trial to gather participants' insights regarding their subjective experience with our system.

In the following sections, we first introduce our VRE system for oral exam anxiety. Subsequently, we provide detailed descriptions of the feasibility and trial studies we have carried out. The results of these studies have been respectively published in [117] and [118].

## 2.1 The proposed VRE system

The VRE system was developed for Windows and MacOS operating systems, using Unity version 2021.3.3f1. It simulates oral exams scenarios in which students customize the pool of questions they might be asked by the virtual examiner (VX)[1]. The system is organized in two main sections: exam customization and exam simulation, both of which will be explained in detail in the following two subsections.

### 2.1.1 Exam customization

The student customizes the exam simulation by defining the pool of questions through a specific interface that is displayed when the system is launched (Figure 2.1). For each question, the student can also specify the available time to answer the question, in a range between one and four minutes. By clicking the “Add” button, the question is added to the pool of questions.

Before an exam simulation, the system requires:

- A microphone test (described in the following).
- At least ten questions entered.
- The answer times associated by students to their entered questions had to exhibit a variation, with the maximum difference in the number of questions sharing the same answer time being no more than one. This balancing was necessary for experimental purposes to ensure that the VX asked the student about the same number of questions (five) and each session had the same length (12 minutes).

By clicking the “Confirm” button, the VRE system checks compliance with the first two rules described above. By clicking the “Start” button, if all requirements above are met, the exam simulation starts. If the

[1]The virtual examiner (VX) will be referred to with the pronoun “it” for the rest of the thesis.

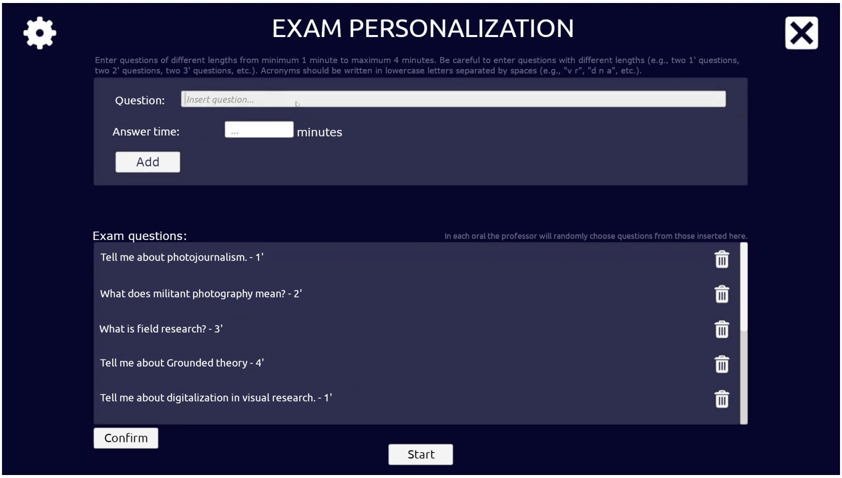



Figure 2.1: The exam customization section of the proposed VRE system. The interface was in the participants' language (Italian). All textual parts have been translated here into English for the reader's convenience.

microphone test has not been performed, the VRE system asks the student to read aloud a neutral sentence ("The square is a geometric figure with four equal sides") while the VRE system analyzes the students' voice volume. The system detects the maximum sound volume and considers 75% of that value as the minimum sound level threshold to determine whether the student is speaking or not.

### 2.1.2 Exam simulation

In the exam simulation, the student is seated in a virtual office, facing a VX that sits behind a desk, as shown in Figure 2.2. The 3D model used for the VX is the "Business_Female_03" character from Microsoft's Rocketbox library [1]. The VX speaks with the voice called "Elsa (Neural)" in the Azure Cognitive Services text-to-speech (with pitchDelta set to -12).

The simulation offers three levels of difficulty, aiming to elicit three different anxiety levels in users. Each level uses a different set of behaviors performed by the VX, which becomes increasingly less friendly as the


Figure 2.2: Examples of different VX behaviors: nodding to the student in set A (a); one of the distracted positions in set B (b); head scratching in set C (c).

level of difficulty increases. Specifically, we defined three sets aimed at making the VX appear respectively friendly (set A), only partially friendly (set B), and unfriendly (set C). Table 2.1 describes all behaviors. Behaviors were chosen based on the indications about their likely effect on users, available in the literature on non-verbal communication [119–123] and virtual agents [124–126].

In set A, the VX maintained a smiling expression to convey agreement [122] and a positive attitude [123–125] (IDLE BEHAVIOR in Table 2.1). The VX performed all other behaviors in set A with a more pronounced smile than IDLE BEHAVIOR to make the perception of the VX facial expression more recognizable, as recommended in [125]. In addition, to further represent agreement and a positive attitude, set A included nodding [122–125] (BEHAVIOR 1) and tilting the head [119] (BEHAVIOR 2 and BEHAVIOR 5). Placing one hand on the opposite hip and the other hand on the opposite shoulder [126] (BEHAVIOR 4) were introduced to represent interest.

In set B, the VX kept a neutral expression on its face, not showing any signs of positive or negative attitude (IDLE BEHAVIOR in Table 2.1). Behaviors performed by the VX in set B to convey a neutral or slightly negative attitude, boredom, or disengagement included shaking its head [124,125] (BEHAVIOR 1 in Table 2.1), turning its head away from participant [119, 124], raising its eyebrows [124] while raising its arms (BEHAVIOR 4).

In set C, the VX maintained a frowning expression to convey a negative

Table 2.1: Description of behaviors in A, B, C. The micro-movements mentioned in the IDLE BEHAVIOR row were obtained through head motion capture of a human actor.

| **Behavior** | **A** | **B** | **C** |
|---|---|---|---|
| IDLE BEHAVIOR | Smiling facial expression with micro-movements of the head and eyes, and eye-blink | Neutral facial expression with micro-movements of the head and eyes, and eye-blink | Annoyed facial expression with micro-movements of the head and eyes, and eye-blink |
| BEHAVIOR 1 | Smiling and nodding (slowly moving the head up and down three times) | Slowly moving the head from right to left and left to right three times while looking at the student with a sad expression | Looking at the student with an annoyed expression, changing the head tilt five times |
| BEHAVIOR 2 | Tilting the head to the left, smiling | Looking distracted, staring at a spot toward the upper left corner of the room | Looking up with a reflective expression, then slowly swinging the head with an annoyed expression |
| BEHAVIOR 3 | Looking at the student, smiling and changing head position and orientation five times | Looking at the student, changing head position and orientation five times | Looking at the student with annoyed expression while rubbing the hands on the thighs |
| BEHAVIOR 4 | Crossing the arms by placing one hand on the opposite hip and the other hand on the opposite shoulder, looking at the student and smiling | Raising the forearms, bringing the hands to shoulders height on the two sides of the body, without changing the height of the elbows, with a doubtful expression | Scratching the head with both hands, with an annoyed expression |
| BEHAVIOR 5 | Tilting the head to the right, smiling | Looking at the student, changing the head tilt five times | Texting on a smartphone kept on the knees under the desk, lowering the head to look at and smile at the phone |
| BEHAVIOR 6 | Looking at the student, changing head tilt five times, smiling | Looking distracted, turning the head to the right, then pausing, then turning the head further to the right | Checking the time on the wristwatch with an annoyed expression |

attitude [125] (IDLE BEHAVIOR in Table 2.1). The VX performed all other behaviors in set C with a more pronounced frowning face than IDLE BEHAVIOR to make the perception of the VX facial expression more recognizable, as recommended in [125]. Behaviors performed by the VX in set C to convey a negative attitude or discomfort included performing a frowning and thoughtful expression and then shaking the head [125] (BEHAVIOR 2 in Table 2.1), scratching the head (BEHAVIOR 4) [127], facing down [125] to use a smartphone (BEHAVIOR 5), checking the time on the wristwatch [123] (BEHAVIOR 6).

Each set also included neutral behaviors (i.e., BEHAVIOR 3 and BEHAVIOR 6 in set A; BEHAVIOR 3 and BEHAVIOR 5 in set B; BEHAVIOR 1 in set C). In these behaviors, the VX changed head position and orientation several times while maintaining a smiling (A), neutral (B), or frowning (C) expression.

Each simulation lasts approximately 12 minutes and consists of three steps:

1. The VX greets the student and tells him/her to get ready to start the exam.

2. The VX asks a set of questions chosen from the list of questions defined in the exam customization section. The questions asked to the student are semi-randomly chosen, ensuring around five questions with different answer times. After asking a question, the VX listens to the participant's answer while performing one of the behaviors of the selected difficulty every 15 seconds, with an 80% chance, until the end of the defined answer time for that question. If the student remains silent for ten consecutive seconds, the VX proceeds to the next question although the question answer time has yet to finish. To detect silence from the participant, the VRE system uses the computer's microphone to capture noise, excluding any noise below the minimum threshold defined in the microphone test.

3. The VX informs the student that the exam is over. Once the exam is finished, the VRE system displays again the exam customization section.

## 2.2 Study 1: feasibility study

For this preliminary study we chose not to use the customizing feature in order to ask exactly the same questions to each participant. In choosing the subject of the questions, we reasoned that different familiarity of participants with the topic could affect the level of anxiety they could experience. To prevent this confounding factor, the questions in the study concerned a topic (Basics of International Law) that was unrelated to the participants' degree curriculum. In addition, we checked that participants were not possibly familiar with the topic for other reasons. Each question was followed by 30 seconds of silence, during which the VX performed three behaviors from the assigned set of behaviors, following the order illustrated in Table 2.2. The timing of events was the same in the three conditions, and is described in detail by Table 2.2.

Table 2.2: Script of the oral exam experience. The VX spoke in participants' language (Italian), all sentences have been translated here into English for reader's convenience.

| **Sequence of events** | **Behaviors performed by the VX** |
|---|---|
| Opening greeting | The VX says *"Good morning. Let's start with the first question"* |
| First question | The VX says *"Describe to me the general principles that are universally recognized by civilized nations"* |
| Wait | The VX performs the following sequence of behaviors from the assigned set (A, B or C):<br>[IDLE BEHAVIOR] for 5 seconds<br>[BEHAVIOR 1] for 10 seconds<br>[IDLE BEHAVIOR] for 5 seconds<br>[BEHAVIOR 2] for 10 seconds<br>[IDLE BEHAVIOR] for 3 seconds |
| Introduction to the next question | The VX says *"Let's proceed with another topic"* |
| Second question | The VX says *"What are the conditions for diplomatic protection?"* |
| Wait | The VX performs the following sequence of behaviors from the assigned set (A, B or C):<br>[IDLE BEHAVIOR] for 5 seconds<br>[BEHAVIOR 3] for 10 seconds<br>[IDLE BEHAVIOR] for 5 seconds<br>[BEHAVIOR 4] for 10 seconds<br>[IDLE BEHAVIOR] for 3 seconds |
| Introduction to the next question | The VX says *"Let's proceed with another question"* |
| Third question | The VX says *"Tell me about the European Convention on Human Rights"* |
| Wait | The VX performs the following sequence of behaviors from the assigned set (A, B or C):<br>[IDLE BEHAVIOR] for 5 seconds<br>[BEHAVIOR 5] for 10 seconds<br>[IDLE BEHAVIOR] for 5 seconds<br>[BEHAVIOR 6] for 10 seconds<br>[IDLE BEHAVIOR] for 3 seconds |
| Ending greeting | The VX says *"That was the last question. It can be enough. Goodbye"* |

### 2.2.1 Hypotheses

We formulated the following hypotheses:

H1. The three conditions would produce three different, increasing values of anxiety because the three sets of behavior of the VX aim at making the VX appear friendly in condition A, only partially friendly in condition B, and unfriendly in condition C. We expect that the increasingly less friendly behavior of the VX elicit a correspondingly increasingly higher level of anxiety in participants.

H2. The three conditions would produce three different, increasingly negative perceptions of VX attitude because the VX has been designed to behave friendly when it performs set A, and to increasingly reduce its friendliness when it performs sets B and C.

H3. The three conditions would produce three different, decreasing (respectively increasing) counts of positive (respectively negative) responses elicited because the VX behavior is more friendly when it performs set A, less friendly when it performs set B, and unfriendly when it performs set C. We thus expect that sets A, B, and C elicit a progressively decreasing (respectively increasing) number of positive (respectively negative) responses in participants.

### 2.2.2 Participants

The study was approved by the Institutional Review Board of the University of Udine. A sample of 32 males was recruited for the study. Participants were volunteers who received no compensation. Their age ranged between 20 and 35 (M=22.28, SD=2.95), and they were recruited through direct contact among undergraduate Computer Science students of the University of Udine. Since the proposed VRE system concerns the simulation of an oral test, we looked for participants who were likely to take an oral test in the immediate future as they could be more representative of the intended users of the system. For this reason, students who had completed all the exams in their curriculum were not considered eligible. One participant was excluded from the analysis because he did not follow the instructions in completing some of the questionnaires. Figure 2.3 shows the participant flow diagram.

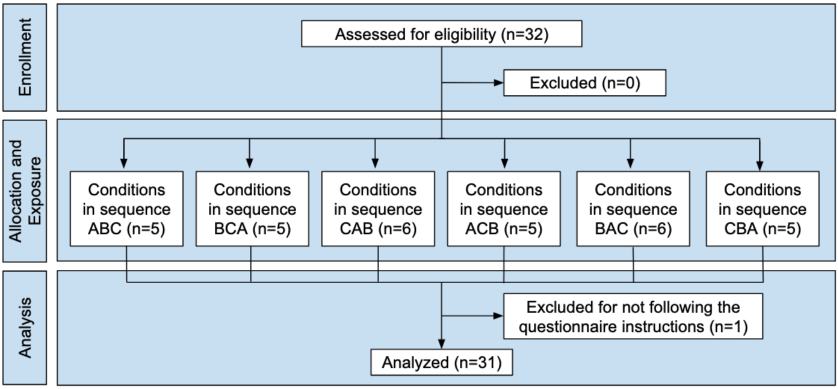



Figure 2.3: Participant flow diagram.

### 2.2.3 Visual Annotation Tool (VAT)

One of the goals of this study was to better understand which aspects of the experience elicit affective responses in participants. To do so, we wanted to identify the exact time instants that had elicited an affective response in participants during exposure, distinguishing between positive responses (i.e., time instants in which participants had felt at ease) and negative responses (i.e., time instants in which participants had felt distressed). Given the central role of the VX in the experience, we were particularly interested in identifying which VX-related factors elicited such responses in participants. To support the achievement of these goals, we developed a Visual Annotation Tool (VAT) as a software-based measuring instrument. The VAT allows one to replay the exposure experience, and to mark any time instant as positive or negative by pressing two different keys on the keyboard, made easier to recognize by a red sticker and a green sticker. Once the replay of the experience is complete, the VAT displays the marked time instants on a timeline, representing positive and negative instants as green and red dots, respectively. Then, each time instant can be selected to replay the corresponding part of the experience and categorize them in a list of pre-defined categories.

In this study, we used the VAT as follows. After experiencing the VRE system, participants used the VAT to replay the three conditions in the same order they had experienced. Participants were instructed to mark the time instants that made them feel at ease or distressed during the first exposure by pressing the green or red key on the keyboard, respectively. Once the replay of all conditions was complete, the VAT displayed green and red dots on the timeline for each condition, allowing the experimenter to identify the specific instants during which the participants respectively felt at ease or distressed. Figure 2.4 shows the timeline of condition C with the positive and negative dots identified by one of the participants.

Subsequently, the experimenter used the three timelines to conduct an interview with participants to investigate the factors that had made them feel distressed or at ease during exposure. To do so, we used as selectable categories in the VAT three VX factors (i.e., facial expression, gaze, and posture), and an additional item "other" that allowed for the inclusion of free descriptive text. The experimenter selected each marked time instant to replay the corresponding part of the experience and asked the participant to indicate which factors had caused the elicited feeling. Based on participant's answer, the experimenter categorized the time instant in the VAT. In Figure 2.4, the experimenter is examining the replay of the second time instant from the left in the timeline (the VAT highlights the selected time instant with a white line below it).

### 2.2.4 Measures

#### Social interaction anxiety

We administered the Italian adaptation of the Social Interaction Anxiety Scale (SIAS) [128] to measure the level of anxiety among participants in general social interactions. The SIAS is a 19-item self-report questionnaire which describes anxious reactions that can occur during social interactions [129]. For each item, respondents rate how true the statement is for them on a 5-point Likert-type scale (0="not at all", 4="extremely").


Figure 2.4: The VAT, as seen by the experimenter. In this screenshot, there is only one time instant in which the participant felt at ease (represented by the first dot from the left in the timeline data visualization, colored green) while all other dots (colored red) identify time instants in which the participant felt distressed. The dot selected by the experimenter is highlighted by a white line. The items listed in the upper right part of the figure allow the experimenter to assign one or more categories to the selected dot after interviewing the participant about that time instant.

Total score ranges from 0 to 76. Higher scores indicate higher social interaction anxiety. Consistently with the literature on VRE systems for social anxiety [66,67,70,130–133], we chose to use the SIAS as an instrument to measure participant's trait anxiety in the specific context of social interaction.

### Anxiety

We used the Visual Analog Scale for Anxiety (VAS-A) as an instrument to assess participants' state anxiety [134,135] during the exposure to each condition. The scale was a 10 cm long line, with "not at all anxious" and "very anxious" printed at its left and right ends, respectively. Participants reported their score by drawing a vertical mark on the scale.

### VX attitude

We administered a 4-item questionnaire (Table 2.3) that asked participants to rate aspects of the VX attitude they perceived on a 7-point Likert-type scale (1="not at all", 7="very").

Table 2.3: VX attitude questionnaire items.

| | |
|---|---|
| 1 | Is it angry, tense, nervous? |
| 2 | Is it serene, relaxed, calm? |
| 3 | Is it rude, distracted? |
| 4 | Is it sociable, kind? |

To calculate the score, the scale of the second and fourth items was inverted, and the answers were averaged. Higher scores indicate a more negative attitude. Cronbach's alpha in the three conditions was respectively 0.70 (A), 0.65 (B), 0.87 (C). We performed an exploratory factor analysis with principal component extraction and Oblimin rotation to evaluate factorial validity. Bartlett's test was significant ($p < 0.001$ in all conditions) and KMO was greater than 0.60 in set A (0.66) and set C (0.75), while it was 0.57 in set B. The analysis confirmed the intended one-factor structure that explained respectively 56.89% of variance in set A, 52.31% in set B, 73.12% in set C.

### Elicited positive and negative responses

We measure the number of times participants reported a feeling of ease (positive responses) or distress (negative responses) with the VAT. For each condition, the two counts were used as an indication of positive and negative responses elicited.

### Factors associated to positive and negative responses

As described in 2.2.3, factors associated to positive and negative responses were collected through the participant's interview and stored in the VAT. The aim was to identify the VX factors (and other possible factors) that had a greater impact in eliciting responses in participants.

### Qualitative interview

After conducting the first part of the interview with the VAT as described in section 2.2.3, we further interviewed participants to gather comments about how to improve the system. The experimenter asked a sequence of structured open-ended questions, following the diagram in Figure 2.5.

The collected interviews were transcribed verbatim from audio recordings. Then, following the method in [136], we performed a thematic analysis to analyze transcripts, identifying and organizing common e prominent themes. The analysis involved:

1. Reading the transcripts multiple times to familiarize with the data.
2. Coding interesting features in the transcripts and collating relevant data for each code.
3. Grouping all codes into potential themes, collecting all pertinent data for each potential theme, organizing themes into levels (e.g., main themes or sub-themes within them), and dividing large or complex themes into one or more sub-themes.
4. Defining the significance of the themes and sub-themes concerning the coded extracts and all transcripts.
5. Refining each theme and sub-theme, generating clear definitions and names.

The experimenter performed the steps described above and coded the data. However, since the process of defining the codes and applying them to the dataset can be biased by subjective interpretation [137], the validity and reliability of the themes must be confirmed through additional coding. Following [138], the data were coded also by an independent external coder, who was not involved in our research, and used a codebook we provided. The codebook listed the themes and sub-themes identified by the thematic analysis. For each code, it provided a label and a complete description with inclusion and exclusion criteria. We also explained to the external coder that he could use multiple codes on the same text fragment. Both coders used Taguette [139], an open-source web-based CAQDAS (Computer Assisted Qualitative Data Analysis Software), to code the data.

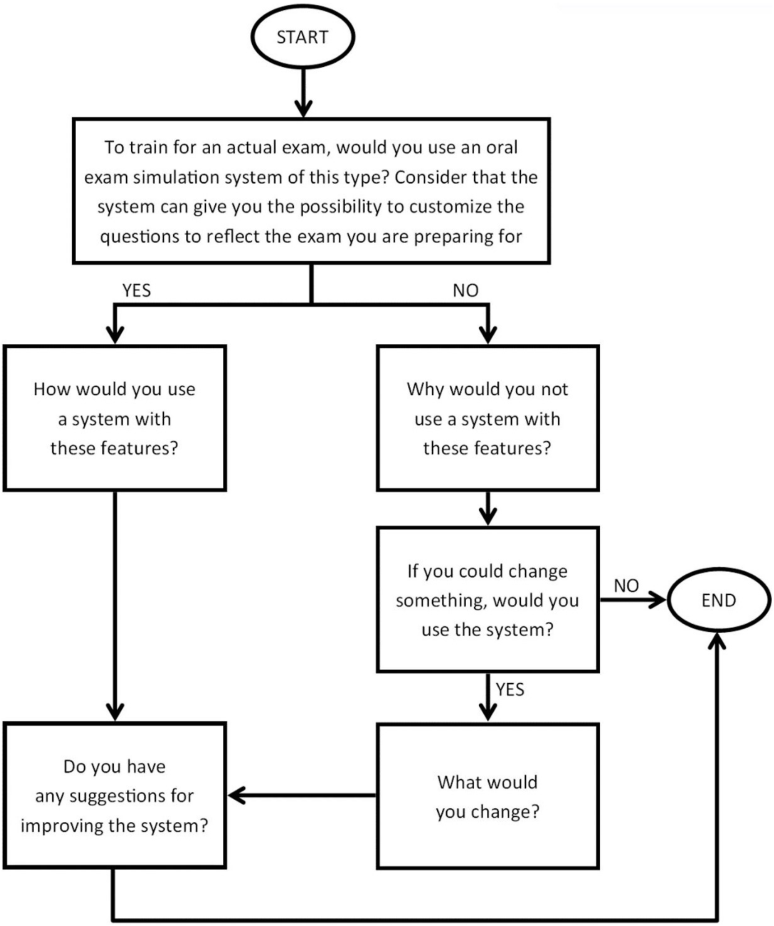



Figure 2.5: Flow diagram of the structured qualitative interview. The interview was conducted in participants' language (Italian), all sentences have been translated here into English for reader's convenience.

### 2.2.5 Procedure

We aimed at maintaining study design consistency with previous studies that assessed the feasibility of VRE systems for social anxiety (e.g., [66,67]), fear of public speaking (e.g., [77,140]) and exam anxiety [55,56]. Participants were tested individually in a 50-minute session that started with filling the SIAS. To prevent the spill over of anxiety elicited by a condition to the next condition, each condition was preceded by a two-minute period during which participants sat in a comfortable position while listening to calm music and watching a series of relaxing images of natural scenarios (such as forests, pools and hills) through a 24-inches desktop monitor positioned in front of them. Participants were asked to imagine being the student in the oral test, and told that they did not need to answer the questions audibly. After exposure to each condition (A, B, C) via the desktop monitor, participants filled the VAS-A and VX perception questionnaires. The same sequence (relax, exposure, questionnaire) was followed for each condition. Participants were exposed to the three conditions in counterbalanced order as illustrated in detail by Figure 2.3. After experiencing all the conditions, participants replayed the conditions in the same order they were presented previously, using the VAT (described in Section 2.2.3) to mark the time instants of the experience in which they remembered to have felt at ease or distressed. Finally, participants were interviewed: for each time instant marked by the participant with the VAT, the experimenter watched the corresponding part of the experience in the VAT together with the participant, and discussed with him/her the reasons why that experience had caused a positive or negative reaction. Then, the experimenter categorized the participant's answer in the VAT. Finally, the experimenter asked a sequence of structured open-ended questions, following the diagram in Figure 2.5.

### 2.2.6 Quantitative results

All analyses were conducted using SPSS version 29.0.0.0. An ANOVA repeated measure was used to compare the effects of the VX behaviors (A, B, C) on VAS-A, VX attitude, and counts of elicited positive and negative responses. If Mauchly's test indicated a violated assumption of sphericity, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. ANOVA results are shown in Table 2.4.

Table 2.4: ANOVA results.

| **Measure** | **A** | **B** | **C** | **Main effect** | | |
|---|---|---|---|---|---|---|
| | **Mean (SD)** | **Mean (SD)** | **Mean (SD)** | $F$ | $p$ | $\eta_p^2$ |
| VAS-A | **1.82** (1.67) | **3.03** (1.62) | **4.15** (2.27) | $F(2,60)=21.93$ | <0.001 | 0.42 |
| NAVX[a] | **2.05** (0.78) | **4.60** (1.16) | **5.54** (1.17) | $F(1.6,59.86)=116.39$ | <0.001 | 0.8 |
| EPR[b] | **3.71** (1.72) | **0.52** (1.00) | **0.06** (0.25) | $F(1.33,39.94)=86.87$ | <0.001 | 0.74 |
| ENR[c] | **0.65** (1.02) | **2.45** (1.43) | **4.77** (2.14) | $F(1.68,50.24)=72.64$ | <0.001 | 0.71 |

[a]NAVX=Negative attitude of VX
[b]EPR=Elicited positive responses
[c]ENR=Elicited negative responses

ANOVA revealed a main effect on state anxiety measured with VAS-A, and Bonferroni post hoc comparison found a significant difference for all pairs of conditions (A vs. B, $p < 0.01$; B vs. C, $p = 0.01$; A vs. C, $p < 0.001$) (Figure 2.6).

ANOVA revealed a main effect of VX attitude, and Bonferroni post hoc comparison found a significant difference for all pairs of conditions (A vs. B, $p < 0.001$; A vs. C, $p < 0.001$; B vs. C $p < 0.01$) (Figure 2.7).

ANOVA revealed a main effect for both counts of elicited positive and negative responses. Bonferroni post hoc comparison found a significant difference for all pairs, both with positive responses (A vs. B, $p < 0.001$; A vs. C, $p < 0.001$; B vs. C, $p = 0.05$) and negative responses (all three pairs, $p < 0.001$). (Figures 2.8 and 2.9).

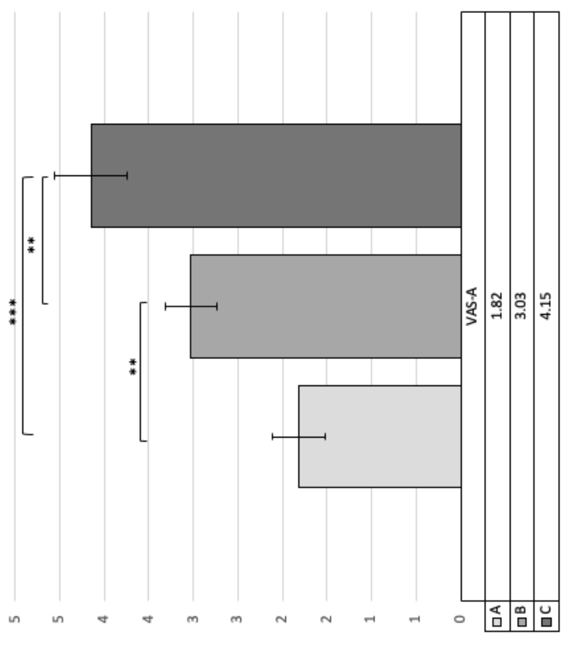



Figure 2.6: Means of VAS-A scores. Capped vertical bars indicate ± SE. The ** and *** signs indicate differences with $p$-values respectively $\leq 0.01$ and $< 0.001$.

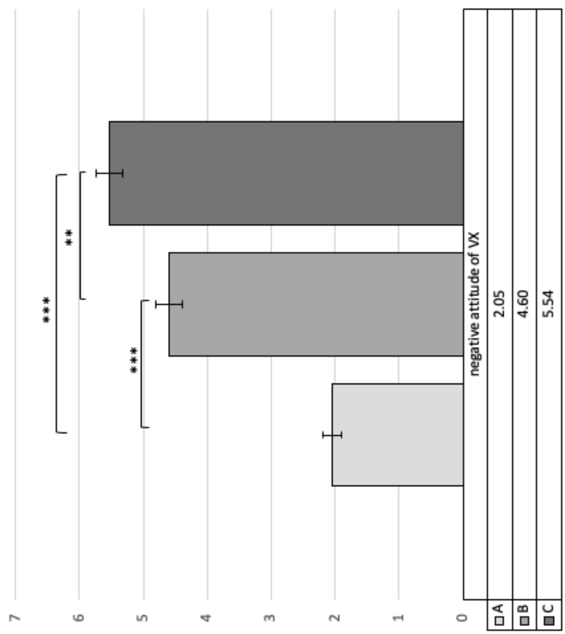



Figure 2.7: Means of VX negative attitude scores. Capped vertical bars indicate ± SE. The ** and *** signs indicate differences with $p$-values respectively $< 0.01$ and $< 0.001$.

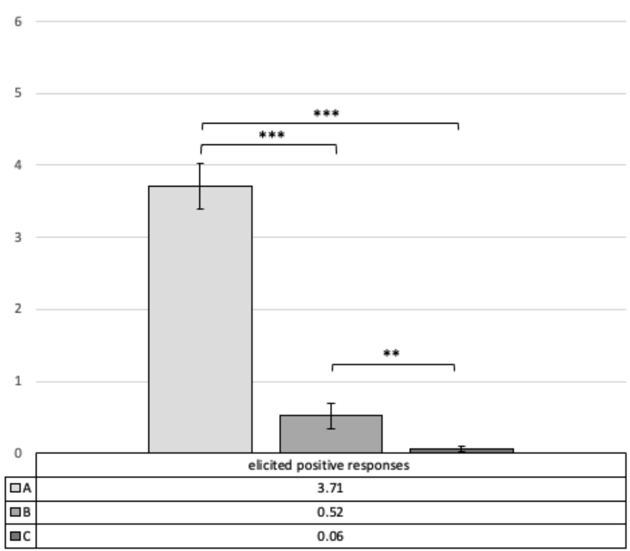



Figure 2.8: Means of counts of positive responses. Capped vertical bars indicate ± SE. The * and *** signs indicate differences with $p$-values respectively $= 0.05$ and $< 0.001$.

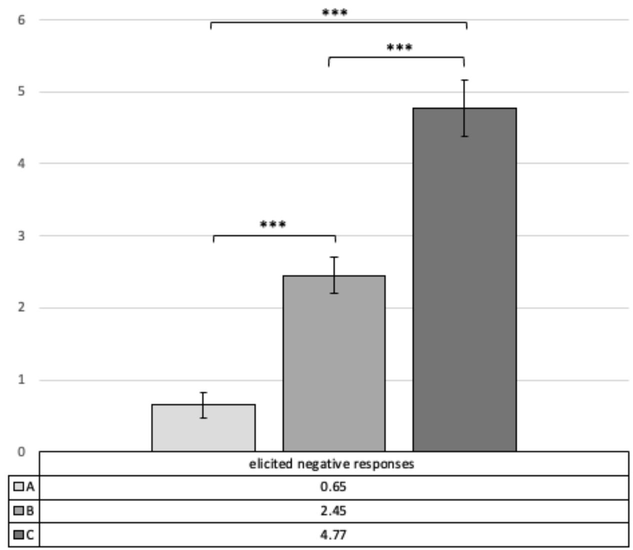



Figure 2.9: Means of counts of negative responses. Capped vertical bars indicate ± SE. The *** signs indicate differences with $p$-values $< 0.001$.

To explore if participants with different levels of social anxiety were affected differently by the proposed system, for each dependent variable we contrasted the group of participants with an above-median score (n=15) with the group of remaining participants (n=16). The median score in the SIAS was 30.00 (M=30.06, SD=11.17, range 13-60). A $3 \times 2$ mixed design ANOVA (between-subjects: above-median vs. below-median SIAS scores, within-subjects: A vs. B vs. C) revealed no significant differences between the two groups for any measure, and no interaction between the two variables.

### 2.2.7 Qualitative results

SPSS version 29.0.0.0 was used to compute Cohen's Kappa to assess inter-rater reliability [141, 142]. The overall kappa coefficient was 0.80, showing a strong level of agreement [143].

Themes fell into three topic areas:

- *Suggestions*: themes capturing suggestions for enhancing the system (Table 2.5).
- *System*: themes related to system features (Table 2.6).
- *Voice*: themes related to the VX voice (Table 2.7).

In the following, we summarize the themes and sub-themes of the three topic areas, also providing sample extracts from the interviews. In the interview extracts, the parts in round brackets are additional questions asked during the interview, while the parts in square brackets are words or phrases added to clarify the sentence.

#### Suggestions

Table 2.5 summarizes the themes and sub-themes that belong to the *Suggestions* topic area.

Table 2.5: Themes and sub-themes of the *Suggestions* topic area.

| **Theme** | **Sub-theme** | **Description** | **Participants** |
|---|---|---|---|
| VX | Animations | Some participants suggested to increase the level of realism of the VX animations | P4, P15, P18, P24, P30 |
| | Artificial Intelligence | Some participants suggested to add artificial intelligence to the VX to make it able to recognize students' correct and wrong answers | P5, P16, P27, P30 |
| | Customization | Some participants suggested to introduce the possibility of customizing the gender, appearance, or behavior of the VX | P12, P16, P20, P24 |
| | Verbal requests | Some participants suggested to make the verbal requests made by the VX more varied, such as "explain in more detail", "make an example", etc. | P2, P6, P10, P11, P21 |
| | Physical appearance | Some participants suggested to increase the variety of behaviors and expressions that the VX can perform | P1, P5, P15, P26, P29 |
| | Set of behaviors | A few participants suggested to increase the variety of behaviors and expressions that the VX can perform | P1, P7, P16 |
| | Voice | Some participants suggested to use a more realistic voice, varying how the VX uses it in the three levels of difficulty | P2, P7, P8, P16, P25 |
| System | Level of difficulty | A few participants suggested to add the ability to manually choose the level of difficulty on which to practice the oral exam | P16, P20 |
| | Immersion | A few participants were interested in trying an immersive version of the system with a headset | P5, P7 |
| VE | Audience | A few participants suggested to add the presence of an audience to the VE | P7, P22, P26 |
| | Realism | A few participants suggested to increase the level of realism of the virtual room | P5, P22 |

Twenty-one participants offered opinions on the VX. Five of them suggested to improve the realism of the VX physical appearance, while five suggested to improve the realism of the animations, e.g.:

> *Make the character's attitude slightly more human. [...] I mean, make it more fluid, less robotic.* – P18

Four participants suggested to expand the VX with artificial intelligence to make it able to understand and possibly correct students' answers, e.g.:

> *[I would like] it to react to my answers [...]. If I give a wrong answer or hesitate, the virtual teacher evaluates me, so I realize that I have to go back to study.* – P5

Four participants suggested to make the VX customizable in terms of appearance or behavior, e.g.:

> *Maybe being able to customize the appearance of the teacher so [...] if, for example, I have in mind a professor that causes me stress [...], being able to give this [virtual] teacher the appearance of the professor I am afraid of...could help me to overcome my anxiety, make me find a way to succeed in passing the exam.* – P16

Three participants suggested to expand the range of actions and emotions that the VX can exhibit, while five others suggested to expand the diversity of its verbal requests to the user, e.g.:

> *Make it a little more engaging, for example by maybe adding questions, I mean, not just questions and answers, but adding questions or remarks between the student's answers. (For example?) Such as, if the student answers in a certain way, the professor might ask him "Give me some examples".* – P21

Five participants suggested to improve the VX synthetic voice using a more realistic one, or vary it among the three sets of behavior, e.g.:

> *There should be a little bit of change in the tone of voice in which the teacher asks the questions. Because [...] in the first test, she was putting [me] a lot at ease and using a certain voice. In the second and third, she did not put people at ease but she used the same voice.* – P7

Four participants offered suggestions about the VRE system. Two of them recommended to provide an immersive mode to use the system with an HMD, while two others suggested to allow users to choose the level of difficulty, e.g.:

> *[I would like] to be able to select whether I want [a more stressful] context, to train myself [to deal with] a professor who tends to be mean and therefore can make me very anxious, or whether [. . . ] I don't want to have the pressure and anxiety of making mistakes, so I want [to train with] an examiner [the VX] who is a little more friendly* – P16

Four participants offered suggestions about the VE. Two of them suggested to increase realism, while three others suggested to include an audience, e.g.:

> *Seeing that there is an audience of other students behind [you], watching and listening, makes it all more anxiety-provoking.* – P7

**System**

Table 2.6 summarizes the themes and sub-themes that belong to the *System* topic area.

Six participants were dissatisfied because the system could not adequately replicate the emotions experienced during a real oral test. Two of them believed that this was due to the fact that they did not have to answer the questions during the experience, e.g.:

> *I did not feel much engaged because I was not actually answering the questions. I mean, I had to watch the VX, but actually the level of anxiety that you feel in a real test when you have to give the answers, and maybe you are not prepared and the professor reacts like that. . . it is a whole different thing.* – P6

Twenty participants said they would use the system and explained why. Six of them would use it as a study technique in preparing for the oral

Table 2.6: Themes and sub-themes of the *System* topic area.

| Theme | Sub-theme | Description | Participants |
|---|---|---|---|
| Critiques | Low emotional involvement | Some participants felt that the system only partially evoked the emotions of a real oral test | P1, P6, P11, P15, P21, P29 |
| Purpose | Emotional preparation | Many participants would use the system to train themselves in managing the emotional aspects of oral exams | P4, P7, P8, P9, P10, P12, P13, P16, P17, P19, P23, P24, P25, P28, P30 |
| | Study method | Some participants would use the system as a study method to prepare for the exam | P2, P4, P6, P20, P26, P27 |
| Intention to use | If improved | A few participants stated that they would use the system if it is improved | P1, P15 |
| | No | A few participants stated they would not use the system in any case | P3, P31 |
| | Yes | Many participants stated that they would use the system | P2, P4, P7, P8, P9, P10, P13, P14, P15, P16, P17, P18, P19, P20, P22, P23, P24, P25, P26, P27, P28, P30 |

exam, e.g.:

> *I would use it more as a training tool to practice and so on, because it is a tool that can be used to review, to prepare for an exam. (So the way you would use the system would be as training for the exam?) Yes exactly, not so much to prepare emotionally, but more to prepare [about] the things that you should know.* – P2

Fifteen participants would use the system to prepare for the different emotional aspects of the exam, e.g.:

> *I would use it mainly to train myself to be more relaxed during the test, because I have noticed that when I am more relaxed, let us say, I can answer with more ease and not get overwhelmed by various emotions during an oral test.* – P9

Three participants would only use the system at the highest difficulty level, e.g.:

> *I would try to use it in the worst-case scenario. . . so with those facial expressions that provoked me the most anxiety, to try to maintain control.* – P8

One participant mentioned that he would utilize the system to train himself to pay greater attention to the examiner during the oral test, e.g.:

> *[. . . ] it can also be useful in case you are not used to looking at the professor during the oral test, in order to develop more attention towards looking at the professor, which is something that is not always obvious.* – P10

Twenty-five participants stated that they would (or would not) generally use the system. Two of them would not use it at all, while other two would use it only after some improvement. Twenty-two participants said they would use the system. Four of them, stated they would only use it occasionally, while other four participants said they would be willing to use the current system while waiting for its more advanced version, e.g.:

> *I would say yes [I would use it]. However, [while] waiting for the developers to introduce new features, I would gladly try it out in the meantime because it still helps to feel heard [. . . ]. It is still a useful thing.* – P30

#### Voice

Table 2.7 summarizes the themes and sub-themes that belong to the *Voice* topic area.
Voice characteristics were discussed by seven participants. Three of them described the voice as flat and without inflection, e.g.:

> *I was unable to feel any particular anxiety or particular tranquility from the tone of the voice. It sounded a little flat to me.* – P25

Table 2.7: Themes and sub-themes of the *Voice* topic area.

| Theme | Sub-theme | Description | Participants |
|---|---|---|---|
| Sound | Flat | A few participants described the VX voice as flat, without inflections | P4, P8, P25 |
| | Neutral | A few participants described the VX voice as neutral | P10, P20 |
| | Robotic | Some participants described the VX voice as robotic, mechanical | P8, P22, P31 |
| Change perception | Strong | Some participants stated that the VX voice was different between (at least two of the three) conditions | P11, P18, P30 |
| | None | Some participants stated that VX voice was the same in the three conditions | P2, P7, P13, P19 |

The VX voice was described as neutral by two participants and as robotic and mechanical by three participants, e.g.:

> *While the movements [of the VX] are quite close to reality, the voice is very robotic so it is difficult to empathize with the situation.* – P8

Seven participants described how the VX voice sounded to them. Three of them felt that the voice changed with the difficulty level. Four participants did not report differences in voice between the three levels of difficulty, e.g.:

> *I found the voice to be somewhat homogeneous in all three tests. And the only thing that gave me a sense of difference between levels was the facial expressions, movements, and posture.* – P13

### 2.2.8 Discussion

Results confirmed the feasibility of a desktop VR system for potential use on VRE for exam anxiety. Indeed, they confirmed our hypotheses on anxiety (H1): conditions A, B, and C elicited different, increasing anxiety levels in participants. The ability of the VRE system to elicit different levels of anxiety extends the findings of other studies of VRE systems for exam anxiety [55,56]. Those systems simulated written exams, used immersive VR, and exposed participants to different conditions in a fixed order. In contrast, our study focused on the simulation of oral exams using desktop VR and exposed participants to different conditions in a counterbalanced order to prevent order effects. Moreover, no significant differences were found between participants with above-median SIAS scores and the remaining participants. This result supports the feasibility of the proposed system in eliciting different levels of anxiety regardless of participants' trait social anxiety.

Results confirmed our hypothesis on perceived VX attitude (H2): the three conditions produced increasing negative perceptions. The ability to elicit increasing levels of anxiety by changing the attitude displayed by the VX extends the results of the literature. Studies of VRE systems for fear of public speaking [140, 144] showed that exposure to a group or crowd of agents with different attitudes can elicit different levels of anxiety in participants. Our study showed that similar results can be obtained with a single virtual agent. A previous study showed that the attitudes of a virtual job interviewer, controlled by the experimenter with a Wizard-of-Oz technique (i.e., the user perceives a direct interaction with the virtual job interviewer, while in fact the experimenter decides and commands the responses of the virtual interviewer in real-time), can elicit different levels of anxiety [66]. Our study showed how similar effects can be obtained in a virtual oral exam administered by an agent that is controlled by sets of predefined behaviors.

Unlike previous virtual agent and VRE studies, we created a virtual annotation tool (VAT) to identify the time instants in which participants felt positive or negative affect. The VAT also supported an interview to collect factors that contributed to participants' positive or negative responses. It is worth noting that information collected through the VAT is also useful for further guiding the development of the system because it helps in identifying the sources of participants' ease or distress. By knowing the different aspects of the VE that cause more at ease or distress in participants, it is possible to further improve the VRE system by inserting the right stressors in the different scenarios of the system. For these reasons, its use could be considered for other VRE studies. The VAT allowed us to obtain information that would not have been collected using the questionnaires alone. Results collected through the VAT confirmed our hypothesis on elicited positive and negative responses (H3): the VX behaviors produced three different, decreasing counts of positive responses elicited and three different, increasing counts of negative responses elicited. Counts of positive and negative responses elicited in participants were consistent with the increasing anxiety elicited by the three conditions (A, B, C), and were also consistent with the increasing negative perception of VX attitude. Results obtained by using the VAT in interviewing participants indicated expression and posture of the VX as particularly influential factors, consistently with previous studies that highlighted the relevance of virtual agent posture and facial expressions in eliciting users' emotions [120]. Several participants remembered they felt a negative response when the VX suspended eye contact with them, consistently with studies that highlighted the importance of eye contact in positive interpersonal interaction [121, 122].

Additional participants' comments revealed that positive responses were often elicited when the VX appeared interested and listening, or calm and relaxed. On the contrary, negative responses were often elicited when the VX appeared distracted, uninterested, judgmental or impatient. Some participants mentioned that their negative responses originated at times from a negative perception of their performance during the simulation rather than from specific VX behavior.

In summary, participants' feedback indicated that the facial expressions, posture, eye contact, level of interest, and general attitude of the VX played a central role in influencing participants' emotional responses.

In the final interview, the system received a high level of consensus: 20 participants stated that they would use the system; a few others would use it after some improvement (n=2). Twenty participants would use it to train for the oral exam (as a study method, n=6; or to learn how to manage emotional aspects, n=15). The fact that most participants would use the system, and the current unavailability of VRE systems for oral exams, motivate the importance of continuing the research, and of further development of the system. The thematic analysis described in Section 2.2.6 allowed to identify several comments and suggestions that can be useful to inform the design of VRE systems for exam anxiety. It is interesting to note that the aspect of the system on which participants reflected more frequently was the VX voice. Five participants suggested to make changes to the VX voice to better reflect the VX attitude. They suggested improving the VX voice to make it more realistic, less flat, and more expressive. In particular, one participant suggested introducing variations in the tone of the VX voice so that when the VX does not put him at ease, its tone of voice also reflects this discomfort. Studies of virtual agents have shown that the emotions they convey through facial expressions, head movements, and voice elicit a greater response in the user while the agent is speaking rather than listening [74]. For this reason, changing VX voice to reflect VX attitude may enhance anxiety elicitation in participants. It is worth noting that three participants believed the VX tone of voice changed between conditions, while it actually did not. Changes they perceived went in the direction of consistency with perceived VX attitude. This can be an interesting aspect for further research because it suggests that VX behavior and facial expressions may cause illusory voice changes that are consistent with participants' expectations derived from behavioral cues.

During the interview, participants identified the three conditions as three different levels of difficulty, where set A was the easiest level, set B was the intermediate level while set C was the most difficult level. The order of difficulty perceived by participants was consistent with the increasingly negative perception of the VX attitude and the level of anxiety increasingly elicited.

During the interview, most participants' suggestions to improve the system were focused on increasing level of realism in different aspects, i.e., quality or variety of animations, appearance of VX or VE, voice or range of expressions of the VX. In particular, four participants suggested that the VX should be customizable not only in terms of behavior but also in appearance, in order to closely resemble the real professor they feared. Allowing students to personalize the appearance of the VX is an aspect that should be considered for future improvements of VRE systems focused on oral exams because a previous study showed that individuals who conducted an oral presentation before virtual agents with an appearance similar to real people known to them experienced higher anxiety than when the presentation was conducted in front of unfamiliar virtual agents [145]. Likewise, a VX resembling the student's real professor in appearance could potentially elicit heightened levels of anxiety, similar to what he/she might experience during an actual oral exam.

This study is the first proposal and feasibility study of a VRE system for exam anxiety that deals with oral exams. While existing VRE systems for exam anxiety simulate predefined written exams whose questions cannot be customized, our system allows users to customize questions and can thus better adapt to students' needs by offering an experience closer to the exam they are preparing for. However, there are some limitations that should be taken into account.

First, the study was conducted on a male sample. Existing literature has extensively explored substantial evidence from several studies that females tend to report higher levels of anxiety [146] and more intense emotional experiences [147, 148] than males. Moreover, females exhibit higher levels of exam anxiety than males [149, 150]. Replicating this study with a female sample might produce more pronounced results.

Second, all participants in the study shared a common academic background in Computer Science, therefore, they all possessed a similar set of skills, academic experiences, and perspectives related to this discipline. This common background might not have fully captured the variability and nuances that could have emerged from participants with different academic backgrounds. As a result, the conclusions of this study might be limited in their generalizability to other fields.

Third, we asked participants to use the VAT to mark the time instants of the experience in which they remembered feeling at ease or distressed during the first exposure. This may introduce a potential recall bias. Participants might have been asked to mark their affective responses during the first exposure, but we ruled out this possibility to maintain their focus on the task and enable them to identify more closely with the role of the student in an oral exam. In this way, they should be more likely to feel emotions similar to those they would experience during a real oral exam. Furthermore, we opted to introduce the VAT only after participants had experienced all three conditions to prevent participants from focusing during exposure on thinking about which emotions they should mark later with the VAT, and thus letting their attention remain primarily on the task.

## 2.3 Study 2: home trial study

The study goal of the current trial was to investigate users' subjective experience with our VRE system over three weeks. In particular, we aimed to gain a deeper understanding of users' perceptions and experience of using our VRE system to mitigate exam anxiety, as well as to identify potential behavioral changes in participants and assess the impact of the VRE system on their well-being.

The choice of a three-week study length was made to allow participants to experience each week a different increasing level of difficulty where the VX behavior was friendly the first week, partially friendly the second week, and unfriendly the third week. The three levels of difficulty used in this trial study correspond to conditions A, B, and C we used in the feasibility study described in Section 2.2.

We conducted the study on undergraduate students who were asked to use the system at home on their computers at least once a week to ensure that they experienced all difficulty levels at least once. Then, they were interviewed at the end of the three weeks. We followed a qualitative rather than quantitative method to obtain insights from participants because open answers can bring to light nuances in how each user experiences the VRE system and help improving the system to make it more useful to users and their well-being.

### 2.3.1 Participants

The study was approved by the Institutional Review Board of the University of Udine, and involved 32 participants (16 males, 16 females). They were recruited through email among undergraduate students of the same university. They were invited to test the VRE system on their own computer for three weeks and were informed the system simulated an oral exam with a virtual examiner, allowing them to customize the questions asked. They were also informed that the system required a Windows or MacOS computer with a microphone and internet connection, and that they could keep the system as compensation for their participation. As with the feasibility study described previously in Section 2.2, participants were sought among those who were going to take an oral exam in the near future, as they could be more representative of the VRE system intended users. To ensure this, students who had already completed all the exams in their curriculum were not recruited. Recruitment resulted in the enrollment of 32 students from different faculties, and each of them consented to participate in the evaluation. After using the system for three weeks, participants were invited via email to the final interview. Eleven participants could not attend the interview for the following reasons: four did not use the VRE system at least once a week as required, due to lack of time; one broke the computer during the evaluation period, and six did not answer the email. The 21 interviewed participants (6 males, 15 females) filled a demographic questionnaire in which they provided information about their gender, age range, and the number of oral exams they took during their university studies (Table 2.8).

Table 2.8: Participants' gender, age range, and number of oral exams experienced during university studies.

| Participant | Gender | Age range | Number of oral exams |
|---|---|---|---|
| P1 | Female | 18-20 | 4 |
| P2 | Male | 21-23 | 4 |
| P3 | Female | 35+ | 16 |
| P4 | Male | 18-20 | 2 |
| P5 | Male | 21-23 | 20 |
| P6 | Male | 21-23 | 12 |
| P7 | Male | 24-26 | 17 |
| P8 | Female | 18-20 | 2 |
| P9 | Male | 21-23 | 6 |
| P10 | Female | 21-23 | 9 |
| P11 | Female | 21-23 | 15 |
| P12 | Female | 21-23 | 5 |
| P13 | Male | 21-23 | 22 |
| P14 | Female | 21-23 | 8 |
| P15 | Female | 35+ | 10 |
| P16 | Male | 18-20 | 6 |
| P17 | Male | 21-23 | 10 |
| P18 | Female | 24-26 | 6 |
| P19 | Male | 21-23 | 5 |
| P20 | Female | 24-26 | 11 |
| P21 | Female | 35+ | 2 |

### 2.3.2 Measures

#### Qualitative interview

We conducted interviews with participants, during which we posed a series of structured open-ended questions outlined in Table 2.9. The interviews resulted in 186 minutes of audio recordings that were transcribed verbatim. Following the methodology detailed in [136], we performed a thematic analysis, adhering to the same procedure as employed in the feasibility study, as explained in Section 2.2.4.

### 2.3.3 Procedure

The study was conducted during the June-July exam period at our university to allow participants using the VRE system in the preparation of

Table 2.9: Interview protocol. The interview was conducted in the participants' language (Italian), all items have been translated here into English for reader's convenience

| | |
|---|---|
| 1 | For three weeks, you freely used the system to simulate oral exams. How did you use the system? |
| 2 | How was your experience with the system? |
| 3 | How did you feel and what did you think when you were using the system? |
| 4 | Do you think that these three weeks of using the system have affected you in any aspect? |
| 5 | In light of the three weeks of use, what is your opinion of the system? |
| 6 | What do you think are the pros and cons of the system? |
| 7 | Is there anything you would change or improve in the system? |
| 8 | The system increased the level of difficulty each week. Have you noticed any changes from week to week? |

the oral exams they intended to take in that session.

Participants sent their signed informed consent via email and received the VRE system through a link provided via email. The study included two meetings with participants, one at the beginning and one at the end of the three weeks of use of the VRE system. To facilitate participation, the meetings were conducted remotely. In the first meeting, participants were individually contacted via video call. They were informed that the study aimed to evaluate a system that virtually exposed students to the oral exam before taking the actual exam with a real professor. They were instructed about how to use the system to experience simulated oral exams conducted in their language, i.e., Italian. They were invited to use the system when they preferred, but asked to ensure a minimum usage of at least once a week as the system automatically advanced its level on a weekly basis, presenting a virtual examiner whose behavior become progressively more challenging to engage with. The participant was also briefed about the anonymity of the collected data. Then, the experimenter assigned the participant a randomly generated ID code and a password to use the VRE system. The ID code was also used to save participants' data and guarantee the trial would increase in difficulty each week, starting from the initial day of system usage over three weeks. The association between a code and the name of the participant was not saved by the experimenter to guarantee participants' privacy. At first access, after entering credentials, participants filled the demographic questionnaire, displayed within the VRE system. Then, they were told that the experimenter remained available to clarify any doubts. They were also informed that, after the three-week period, the VRE system was going to stop working, and they were going to be contacted for the final interview.

At the end of the three-week period, participants were individually video called for the second meeting. Participants were interviewed following a semi-structured approach to gather information about their experience (see Table 2.9 for the interview protocol). If necessary, further questions were asked to examine interesting issues spontaneously raised by participants. After participants' consent, the interviews were recorded and were saved with the assigned ID code to ensure participants' privacy. At the end of the interview, participants were thanked for their participation and their ID code was permanently reactivated as they could keep the VRE system as compensation.

### 2.3.4 Results

The level of agreement among coders was assessed using Cohen's kappa [141,142]. The overall kappa coefficient was 0.72, which indicates substantial agreement [151]. The results of the thematic analysis are organized into three topic areas:

- *System*: themes related to VRE system features (Table 2.10);
- *Influence on the user*: themes related to aspects of the VRE system that have influenced the user in some way (Table 2.11);
- *Suggestions*: themes capturing suggestions for enhancing the VRE system (Table 2.12).

In the following, we summarize the themes and sub-themes of the three topic areas, also providing sample extracts from the interviews. In the interview extracts, the parts in round brackets are additional questions asked during the interview, while the parts in square brackets are words or phrases added to clarify the sentence.

**System**

Table 2.10 summarizes the themes and sub-themes that belong to the *System* topic area.

Twenty participants reported positive aspects of the system. Eleven of them felt that the system was helpful as a study method to prepare for the exam. Six of them felt that the system was also helpful in training them to handle the emotional aspects of exams, e.g.:

> *[The system] makes you practice so it puts you in the same condition as the actual oral exam. There is also the professor, and you have to speak, because if you stay silent, [the VX] changes the question, which is the same thing that also happens in the actual oral exam: you can't just stay silent, and you also have to show and feel yourself confident. [. . . ] The problem is getting questions and having to speak. Very often in oral exams the real problem is having to speak in front of a person rather than knowing the topics, so [in] this, [the system] already helps a lot. (So can it help also on the emotional side?) Yes, in my opinion [it can help] especially on that front.* – P6

Five participants found it beneficial to interact with a virtual human. Among them, one participant mentioned a preference for training with the VX due to feeling embarrassed when repeating in front of someone. Additionally, two others appreciated interact with the VX because it gave the impression of listening to them, e.g.:

> *Certainly, having a person who you know is listening to you [is a benefit] because if I ask [someone] to listen to me when I am preparing for exams [when] maybe I need to repeat, of course I notice that they are not paying attention or they're listening just to be nice to me [. . . ]. Instead, she [the VX] seemed to be listening, [. . . ] more*

Table 2.10: Themes and sub-themes of the *System* topic area.

| Theme | Sub-theme | Description | Participants |
|---|---|---|---|
| Merits | Exam preparation | The VRE system supports exam preparation both as a study method and in managing emotional aspects | P1, P4, P5, P6, P8, P9, P11, P12, P14, P15, P17 |
| | Virtual human | The VX is a virtual human who listens to participants | P3, P5, P13, P15, P19 |
| | Ease of use | The VRE system is simple to use | P7, P10, P18 |
| | Silence detection | The VX moves on to the next question when the participant remains silent | P3, P8, P9, P15, P21 |
| | VX way of speaking | The VX speaks in an appropriate and varied way | P11, P20 |
| | Oral presentation | The VRE system trains participants to give oral presentations | P4, P6, P8, P12, P15, P17, P19, P21 |
| | Customization of questions and answer time | The VRE system supports the customization of questions and answer times | P11, P15, P16, P17, P21 |
| | Random-order questions | The VX asks questions in random order | P1, P3, P10, P13 |
| Critiques | Answer time | Specific aspects of question answer time are criticized (e.g., the length of the maximum answer time) | P1, P4, P6, P14 |
| | Questions balancing | Specific aspects of questions balancing are criticized (e.g., excessive constraints) | P2, P4, P5, P7, P10, P11, P13, P19, P20 |
| | Graphical aspect | Specific graphical aspects are criticized (e.g., the look of the VX) | P9, P20, P21 |
| | Interactivity | The level of interaction with the VX is low | P3, P16 |
| Perception of difficulty level change | VX behavior | Participants perceived that the VX behavior changed over time | P1, P3, P5, P9, P10, P11, P12, P15 |
| | Conversation pace | Participants perceived that the pace of conversation with the VX increased over time | P3, P4, P7, P17 |
| | None | Participants perceived no change in the difficulty level over the three weeks | P2, P6, P8, P13, P16, P18, P19, P20, P21, P14 |

> *participatory, and you felt a little more duty to answer correctly.* – P3

Three participants said that the system was intuitive and easy to use, e.g.:

> *The interface was easy to use, so it was easy to understand how to create new questions.* – P18

Five participants appreciated that the system detected periods of silence. One of them found it useful for not wasting time when she completed her answer before the end of the set answer time. The other four participants found it a helpful way to learn not to remain silent for prolonged time, e.g.:

> *I have seen that [the VX] responds well when I finish [my answer], in the sense that when I finish the topic or if I don't know what to answer, [. . . ] it goes in 10 seconds to the next question, [. . . ] and during the oral exam it seemed even less [seconds], because maybe if I was hesitant for a moment, it seemed like 3 seconds had passed and instead she changed the question. So, this is another hint because it means that I am thinking too long.* – P3

Two participants expressed appreciation for the VX way of speaking, noting its appropriateness and variation in responses, e.g.:

> *I also say a lot of variety in [the VX] transition between questions. It was not repetitive at all so I liked how it was able to stop and move [to another question]. [For example, it said] "Good, let's move on to the next topic" [or] "Very good, let's move on to. . . " [. . . ] that was interesting, I mean, very varied. [. . . ] It was nice and pleasant that it introduced these little sentences and didn't directly jump out to the next question. [. . . ] I liked it.* – P20

Eight participants appreciated the ability to customize the system, with two of them finding the customization of questions particularly useful. The other three participants found benefits in assigning a answer time to each question, e.g.:

*Definitely giving yourself time to answer questions is helpful. It's a task that I have never done, and I realized in trying to answer that [...] it's not so trivial to talk for 4 minutes uninterruptedly about maybe a narrow topic. So, on this [aspect] yes, it definitely helped.* – P17

Moreover, one participant considered the possibility of customizing questions a useful method for realizing what topics the examiner is most likely to ask in the exam, e.g.:

*It helps the repetition [of the subject] but also the study during the last days, for example, because it gives you the conscientiousness and awareness of where your weaknesses are and what questions maybe you could revise. Also, the idea of deciding the questions [is useful], because we insert the questions, not the system itself. It is very useful because it [the system] stimulates you to be on the other side to identify yourself with the examiner and to stop for a moment and think about what parts of the program it might ask more, versus what it might maybe skip.* – P1

Four of them found it useful that the VX asked the questions in random order, e.g.:

*The aspect that I had appreciated the most is the fact that [the VX] re-proposed the questions to me randomly, so I had no idea what question it was going to ask me, and that helped keep me a little bit more, let's say, on my toes or at least with a higher attention threshold because if you already know the order of the questions, obviously you also mentally prepare yourself beforehand to answer them, instead in this case it was not feasible.* – P3

Sixteen participants were dissatisfied with specific aspects of the system. Four participants criticized aspects of answer time of the questions. Two of them found it restrictive to manually set a answer time for each question, while the other two felt that the maximum answer time allowed was too short, e.g.:

> *Sometimes the problem was that some questions [. . . ] took more than 4 minutes [to answer] and so maybe I couldn't always finish [answering the question].* – P6

Nine participants found that the requirement of questions balancing was excessively limiting, e.g.:

> *The fact of having the questions well distributed in terms of time is maybe [. . . ] a little too constraining because maybe one can have a type of questions that are all quite fast [to answer]. So, maybe you have to insert a little bit longer [response] times and then let [the VX] move on to the next question. It's almost, let's say, a trick but it would not be part of the proper use of the system.* – P2

Three participants criticized specific graphical aspects. One of them did not like the system's graphics, and two others claimed the VX or environment were creepy or austere, e.g.:

> *I have to say that I personally found it very austere, very grim as an environment. The teacher was very austere. . . with very dark tints. . . Maybe it was the environment itself that was a bit anxiety-provoking. [. . . ] Teachers are actually like that. But. . . I don't know. . . such an austere figure. . . dressed in that way. . . [. . . ] it's very austere the look of the teacher, however I understand that the teacher's office is not a kindergarten, and you can't put so many colors [in the VE]. This [aspect] stuck out though.* – P20

Two participants complained about the limited interactivity of the VX, e.g.:

> *[The VX] is still a machine and unfortunately does not interact.* – P3

All participants were asked whether they perceived an increase in difficulty over the three weeks of use. Eight participants noted that the VX changed its behavior over time, e.g.:

> *As the weeks went by, obviously the difficulty increased, and you could feel the level of weight [of the oral exam]. I mean, interfacing with someone who makes a weird face at me or puts her hands in her hair is a bit odd. [...] I have noticed the difference in the way the teacher behaved: first, she looked at you, reassured you, then at the third level, she barely looked at you occasionally, [...] I noticed this increase in difficulty also in facial expressions.* – P11

Four participants felt that the pace of conversation with the VX voice rushed over time, e.g.:

> *Especially in the last [week], I saw that [the VX] was a little more concise, in the sense that if I didn't answer quickly, she changed the question immediately and increased the number of questions if I answered faster.* – P4

Ten participants reported not noticing any changes in difficulty, e.g.:

> *I haven't noticed this particular increase in difficulty, maybe it's something related to, I don't know, [...] personal conception of difficulty.* – P2

**Influence of the user**

Table 2.11 summarizes the themes and sub-themes that belong to the *Influence of the user* topic area.

Nineteen participants reported whether they perceived any changes in their behaviors and attitudes during the period of using the system. Six of them reported that the system increased their awareness of their exam preparation level, e.g.:

> *I could tell you that maybe I have more awareness of my degree of preparation because, even though it is a system, a virtual thing, still it is something external that helps me in my study. It's not like asking the question to yourself and then answering yourself. [...] In*

Table 2.11: Themes and sub-themes of the *Influence of the user* topic area.

| Theme | Sub-theme | Description | Participants |
|---|---|---|---|
| Attitudes and behaviors | Awareness | The VRE system helped participants become more aware of their level of preparation for the exam | P1, P4, P11, P12, P14, P15 |
| | Confidence | The VRE system helped participants feel more confident about the exam | P8, P10, P15 |
| | Attention to language production | The VRE system led participants to pay more attention to language production during oral exam simulation | P5, P11 |
| | Encouragement to study | The VRE system motivated participants to study more | P2, P5, P9, P19 |
| | None | Participants perceived no changes in themselves while using the VRE system | P3, P7, P13, P16, P18, P20, P21 |
| Emotions and feelings | Emotions enhancement | While using the VRE system, participants experienced different emotions that improved over time | P8, P11, P15 |
| | Feeling of real exam | When using the VRE system, participants experienced a feeling similar to taking a real exam | P1, P14 |
| | Less anxiety | When using the VRE system, participants experienced reduced anxiety compared to what they feel during a real oral exam | P6, P18 |
| | Time pressure | When using the VRE system, participants felt time pressure in answering the VX questions | P3, P14 |
| | Tranquility | When using the VRE system, participants felt quiet | P2, P10, P11, P17, P20 |
| | Comfort | When using the VRE system, participants felt at ease | P4, P8, P15, P16, P20 |
| | Focus | When using the VRE system, participants were focused on the oral exam simulation | P2, P5, P19 |
| | Oddity of talking with a virtual character | While using the VRE system, participants felt odd talking with the VX | P8, P12 |
| Real experience | | Participants report a real-life experience of how the use of the VRE system has affected their lives | P1, P9, P11 |

*my opinion it's very useful for studying, also because by answering [the questions], I noticed my gaps, the things I understood, the*

> *things [I understood] less, which things I needed to focus more on...* – P14

Three participants experienced an increase in their confidence, e.g.:

> *I needed something to [help me prepare for] exams because I was struggling to prepare on my own, especially for oral exams, [I needed to train myself to] answer in front of the professor. [...] I got along very well [with the system]. [...] It is nice, it is a very nice program, now I feel much more confident.* – P15

Two participants found the system useful in assessing the quality of their responses and, consequently, enhancing their performance, e.g.:

> *[The system] made me think about how I could interact [with the VX] and how I could recheck myself when I was exposing the answers [to questions], so, [for example], not going too fast, trying to be more descriptive, and to focus on what [the VX] is telling me and not what's going on around me. [...] [The system] helped me to be a little more realistic and concrete about a little bit of fear and a little bit of anxiety that I can feel during the exam. Actually, when I study just by reading the book and the notes, I don't have [this feeling] because I ask myself the questions and I don't have any fear or thought about what I'm saying, how I come across.* – P11

Four participants stated that the system pushed them to study more, e.g.:

> *[The system] influenced me in the sense that maybe [before using the system] I tended to study without repeating so much and I studied a lot by transcribing and doing summaries, then I spoke very little. Instead [the system] certainly stimulated me to repeat more because obviously I had to speak.* – P19

Seven participants did not perceive any change, e.g.:

> *No, I don't think so. I haven't noticed any particular change.* – P7

Seventeen participants told how they felt while trying the system. Three of them noted that over time they improved their perceived emotions during exam simulations, e.g.:

> *The first times it felt a little strange because I knew that I was talking by myself, to a certain extent. Ehm... instead after a while, the last week in particular, I felt comfortable, always a little bit under pressure because it was a kind of exam, however, I had no fears or problems of any kind. [...] I felt more confident in the exposure and actually the last time just less pressure than the other times. The first time I was more agitated, from the later ones it [the agitation] went down more and more.* – P8

Two participants stated that they experienced the same feelings as during a real exam, e.g.:

> *I felt a little bit like in the exam, you know? So [I felt] maybe a little bit... in awe let's say. [...] The virtual examiner also simulated the facial expressions that a professor may have during the exam, so I felt just like I was at an exam, maybe not in presence, but maybe at a distance [exam], however, still an exam.* – P1

Two participants felt that they experienced less anxiety than in a real exam, e.g.:

> *I thought I was there though, obviously, with less anxiety [than a real exam]. I think it is also the goal of the system as well, which is to start practicing speaking but with maybe a little bit less anxiety at the beginning, because when you go to do a real oral [exam] you have a lot of it [anxiety].* – P6

Two participants felt that they were in a hurry to answer the question and felt time pressure, e.g.:

> *When I wasn't looking at the screen, I was quiet, whereas if I was staring at her [the VX] who had a little... cold look, and then never*

> *knowing how she was going to answer, I mean, I was a little bit in a hurry to answer faster. [...] When she was asking the random questions, I couldn't remember the [response] time I had put for that question, you know, so [...] I was in a hurry just to tell her everything because I was afraid that maybe it was a one-minute question, and I wouldn't be able to say in short time at least the minimum.* – P3

Five participants felt calm while using the system, e.g.:

> *I was quiet, because obviously at home I had no stress.* – P2

Five of them felt at ease, e.g.:

> *The fact that I knew a little bit about the subject put me at ease with the [exam] simulation. Probably if I hadn't already studied some things, yes, having the teacher there would have had some effect on me.* – P20

Three participants stated they were focused while using the system, e.g.:

> *I was quite focused on what I was saying... it's clearly not like taking a real exam, however, it's also not like repeating yourself.* – P5

Two participants felt strange talking alone, e.g.:

> *Initially, I felt a little uncomfortable because I was talking to something that doesn't actually exist, however, afterwards, I also got unstuck and it was also a way for me to get unstuck in general.* – P12

Three participants shared a real-life experience of how the use of the system has affected their lives. One of them found the system particularly useful for training herself to handle the examiner's different attitudes during the exam, e.g.:

> *[The system] is all super well done. Even the mimicry of the professor's facial expressions. I mean, I just happened [after using*

*the system] on an oral exam to have an examiner who made facial expressions while I was speaking. And so, since I took the high school degree last year, I was used to taking exams with the mask [because of Covid-19], [this exam instead] was different. However, I was already prepared [to handle different examiner moods], you know, under this aspect because indeed with the mask the professor can make as many faces as he wants that you don't see them anyway. Instead [in this case] the system helped me in this thing that, then, actually happened [in the real exam].* – P1

Among the other two participants, one reported achieving a very good grade on the exam after training with the system, while the other received identical questions on the real exam as those he had practiced with using the system, e.g.:

*There was a PowerPoint presentation to be given. Then [. . . ] the professor asked me some questions. I had already prepared about ten of the questions he might ask [. . . ]. I had practiced them in the morning with this system here. [The professor] asked me some of those [questions], not all of them but some.* – P9

**Suggestions**

Table 2.12 summarizes the themes and sub-themes that belong to the *Suggestions* topic area.

Ten participants provided different suggestions to improve the exam simulation section of the system. Two participants proposed the implementation of artificial intelligence in the VX to enable it to ask subsequent question based on the answers received to the previous question, e.g.:

*It would be nice if in the cases where one doesn't know [the answer], so there are those moments of silence, [the VX] asks questions on similar topics or the same [topic][. . . ] because in the end this is what happens in the actual oral exams as well.* – P6

Table 2.12: Themes and sub-themes of the *Suggestions* topic area.

| **Theme** | **Sub-theme** | **Description** | **Participants** |
|---|---|---|---|
| Exam simulation | Artificial intelligence | Add artificial intelligence to the VX to enhance the quality of conversations | P6, P16 |
| | Move on to the next question | Provide alternative ways to move on to the next question | P8, P13 |
| | Timer | Add a visible timer to track the elapsed answer time or the time spent in silence by participants | P3, P9 |
| | Exam length | Increase the exam length | P6, P15 |
| | Answer analysis | Add the recognition of the correctness of participants' answers | P2, P3, P21 |
| | New features | Add specific new features to the exam simulation section | P2, P9, P17 |
| Exam customization | New features | Add specific new features to the exam customization section | P7, P9, P12 |
| | Answer time | Change specific aspects of answer time of questions | P3, P4, P6, P14, P15 |
| | Questions balancing | Make the questions balancing constraints less rigid | P2, P4, P11, P13, P19, P20 |

Two participants suggested alternative methods to move on to the next question after ten seconds of silence or once the assigned answer time for the current question expires, e.g.:

> *I would have appreciated if [the system] recognized that I was still speaking and alerted me instead of moving on to the next question.* – P8

Two participants suggested including a timer in the exam simulation to indicate the seconds of silence already passed or the time already elapsed to answer the current question, e.g.:

> *It would be helpful to have a reference to the allocated time for the question [. . . ] because when I enter the question [in the exam customization section] I can gauge whether I can express the concept within a short time or if I need a few more minutes. But then, when I was saying it [the concept] during the exam, a bit because of tension, I could not remember how many minutes I have set.* – P3

Two participants proposed the extension of the total duration of the exam simulation to better resemble the length of a real exam, e.g.:

> *Make the exam last not 12 minutes but maybe 15 or 30 minutes because usually, in my case at least, a real oral exam takes 30 [minutes] up to an hour.* – P15

Three participants proposed enhancing the system to enable it to differentiate between correct and incorrect responses provided by the user when answering VX questions, e.g.:

> *It might be a good idea to record the [correct answers to] questions [. . . ]. Maybe the system could compare the [correct answer to the] question set with what I am going to say during the exam. Maybe, I don't know, finding keywords or key concepts. This would help so much because I understand if I am going in the right direction, because, obviously, I might have said anything other than the correct answer to the question and it [the system] cannot know that [in the current version].* – P3

Three participants have suggested integrating new features into the oral exam simulation. Specifically, they proposed the inclusion of the capability to record the simulation, the addition of a virtual webcam to emulate an oral exam conducted remotely, and the integration of a chat room where a virtual audience participates remotely to the oral exam, e.g.:

> *If you are, for example, on Teams, one thing that elicits more anxiety is to make it seem as if there are a lot of people connecting [to listen to you].* – P9

Thirteen participants provided suggestions to improve exam personalization. Specifically, two participants would like to resize the system window during customization. Another participant suggested introducing the possibility of inserting correct responses to questions to offer an additional study modality of the system where the VX explains to students the expected answers for each question, e.g.:

> *One useful thing would be the study modality, [...] that is, when you enter the question [in the exam customization section], you also enter the answer [...]. [The VX] asks you the question and [...] reads you the answer.* – P9

Five participants offered suggestions concerning the answer time of the questions. Specifically, one participant expressed a preference for having no limit on the length of answer times, while the other four preferred longer answer times, e.g.:

> *Perhaps I would leave the length [of the answer] more free. In the sense that there could also be longer question answer times than 4 minutes. There are probably subjects for which twice the amount of time may be required.* – P3

Six participants would like more freedom in entering questions because they found the question balancing mechanism too rigid, e.g.:

> *I would eliminate the rigidity [that the system requires] in choosing the [response] times, so if I insert four questions, it [the system] wants one of one minute, one of two [minutes], one of three [minutes], and one of four [minutes]. If I want to put four [questions with answer times] of four minutes, [...] there would be more freedom in choosing the methodology of answering, also based on the professors you have because there are some who want very quick answers and professors who make you talk for even 15 minutes straight.* – P13

In summary, in the final interview, some participants also provided details regarding their usage frequency on the system. They reported using it three times weekly (n=2), twice (n=2), or once a week and a few hours or days before the actual exam (n=2). Some participants also stated that they used the system to prepare for a single exam (n=8) or multiple exams (n=3).

The VRE system received overall positive feedback (n=20). Participants highlighted its utility as a study method (n=11), and its merits for training in oral presentations (n=8) and in coping with emotions during exams (n=6). Positive aspects included customizable questions and answer times (n=5), detection of silence (n=5), interaction with the VX (n=4), user-friendly interface (n=3), and the way the VX spoke (n=2). However, 16 participants also expressed dissatisfaction with certain aspects of the application, citing rigidity of the question balancing requirements (n=9), answer time constraints (n=4), the look of the VX, the VE, or the VRE system (n=3), and limited interactivity of the VX (n=2). Eight participants noted changes in the VX behavior as the difficulty level increased or observed an accelerated conversation pace (n=4), while others observed no change (n=10).

Regarding influence on users, participants noted positive changes in awareness of their preparation level (n=6), study motivation (n=4), self-confidence (n=3), and presentation quality (n=2). Regarding emotions, some users reported feeling calm (n=5), at ease (n=5), or focused (n=3) during VRE system use, with a few experiencing enhanced emotions over time during exam simulations (n=3). A few participants experienced the feeling of a real exam (n=2) or felt less anxiety than during a real exam (n=2). A few felt time pressure (n=2) or found it odd to talk to a virtual character (n=2). Three participants shared experiences in which the VRE system positively affected their real exam performance.

Seventeen participants suggested improvements, including increased flexibility in entering questions (n=6) and answer time (n=5), new features such as recognizing the correctness of user's answers (n=3), alternative ways to move to the next question (n=2), a timer that indicates silences or time elapsed in answering the current question (n=2), integration of artificial intelligence to enhance conversations quality (n=2), and the extension of the total simulation duration (n=2).

### 2.3.5 Discussion

Overall, the analysis of the interviews suggests that the VRE system might provide a valuable assistance to students in developing emotional skills to cope effectively with oral exams. Indeed, several participants reported benefits in this regard (n=9), noting improvements in awareness (n=6) and confidence (n=3). The improvement in participants' confidence and awareness is consistent with findings from a previous study where exposure to various scenarios within a VRE system, including a job interview with a group of virtual interviewers, heightened participants' confidence [131]. In addition, three participants noted an improvement in their emotional state over time during the simulations, moving from an initial agitation to a gradual increase in confidence. This improvement over time was also found in a previous study showing that exposure to verbal interactions with virtual agents leads to increased self-efficacy [133].

The VRE system was also considered effective as a study method (n=11), serving as a valuable tool for practicing oral presentation (n=8), promoting heightened attention to language production (n=2), and fostering increased study motivation (n=4). These aspects support the effectiveness of the VRE system in providing users with exam simulations similar to real-world situations. The customization of questions and their randomized presentation by the VX contributes to this realism, allowing users to train within a safe environment to develop cognitive skills, identify areas for improvement, and address knowledge gaps.

The thematic analysis revealed key themes related to the features of the VRE system: question balancing (n=9), customization of questions and answer times (n=9), silence detection (n=5), and the presence of a virtual human (n=5). Regarding the customization of answer times, three participants found it beneficial to foster time-management skills, trying to answer exhaustively within the set time without exceeding it. In contrast, four participants found it restrictive and suggested either removing (n=1) or extending (n=4) the maximum answer time. Furthermore, nine participants expressed reservations about the question balancing requirement, and six of them would make it more flexible. It should be noted that the both answer time and question balancing were added to the system for experimental purposes to guarantee that each session in the trial contained about five questions in a 12-minute simulation session. The regular version of the VRE system does not need to enforce balancing requirements. However, considering the appreciation of some users regarding the requirement of setting an answer time to each question, it is worth keeping it as an optional feature to enhance user customization.

Participants appreciated the silence detection feature. One participant found it beneficial in preventing wasted time, while the other four highlighted its utility in increasing awareness of silent intervals while thinking of responses. This result is unexpected, as the original goal of silence detection was to facilitate the transition to the next question when the student had nothing more to say and a considerable amount of time remained available for that response. It suggests that silence detection may help improve students' awareness of their silence durations, helping in learning to answer more promptly to show greater preparation and confidence.

Participants also acknowledged the value of a virtual human presence, as they perceived that the VX was listening to them. This perception could foster a sense of trust between the student and the VX, thereby enhancing the student's perception that the VX cares about his/her educational progress, which is important to the student's overall well-being [152].

The emotions and feelings reported by participants, such as focus during the exam simulation, the sense of time pressure, and the feeling of a real exam, suggest that the VRE system can effectively simulate exams by eliciting emotions similar to those experienced during real exams. This is consistent with previous research showing that the VR is a valuable tool for simulating written exams and eliciting emotional responses in students with high test anxiety [55]. Among the various emotions reported by participants, two of them explicitly mentioned experiencing less anxiety than a real exam. This result was expected as it is consistent with the study by Valls-Ratés et al. [88], which showed that a virtual audience in a VRE system for public speaking anxiety elicited lower anxiety levels than a real audience while effectively reducing anxiety levels in participants.

Eight participants noticed that the VX friendliness decreased over time. Since individuals may emotionally react to the behaviors of virtual agents [153], participants' awareness of these changes is noteworthy. The VRE system may influence participant's emotional state by adjusting VX behaviors positively or negatively [74]. Interestingly, four participants perceived that the VX moved quicker to the next question as the difficulty increased, although this did not actually happen. This fact motivates further research because it suggests that the unfriendly behavior of the VX may give participants the illusion that the VX is increasing the conversation pace.

Considering that three participants expressed reservations about the appearance of the VX or the VE, future VRE applications for oral exams might consider allowing users to customize the appearance of the VX and the VE. However, it should be kept in mind that an excessively detailed replication of the real professor's office is unnecessary for the exposure scenario to achieve effective results [154]. The thematic analysis presented in Section 2.2.6 enabled the collection of numerous comments and suggestions that may be valuable in guiding the design of VRE systems for exam anxiety. The gathered suggestions indicate participants' desire for a system providing comprehensive support for exam preparation through an intelligent VX. The VX would access the correctness of participants' answers (n=3) and propose additional questions based on their answers to previous questions (n=2). Other novel features were also suggested, such as recording sessions for later review (n=1), including a visible timer in the VE to monitor elapsed time (n=2), and using a webcam to capture the participant during the exam simulation to intensify the feeling of being observed and elicit a higher level of anxiety, as one can experience during a real remote oral exam (n=2). Another participant suggested adding an always visible chat window in which a virtual audience of students would comment on the participant's oral presentation to amplify him/her anxiety about giving a bad impression (n=1). Overall, the suggestions underscore participants' interest in a system that contributes to their comprehensive oral exam preparation, offering scenarios capable of eliciting anxiety to train them to manage their emotions and thus increase their overall well-being.

Real-life experiences shared by three participants provide further insight. For example, one participant highlighted the VX effectiveness in practicing the handling of negative facial expressions, citing the VX accurate replication of a professor's expressions. She found this training beneficial during a real oral exam where the professor exhibited behaviors similar to those experienced in the VRE system. Another student reported receiving during a real exam some questions identical to the ones he had entered into the application. Another student reported that by using the VRE system, she gradually gained confidence and calmness during exam simulations, and she thought this positively impacted the actual exam result, in which she scored high. This outcome is consistent with an earlier study in which students who practiced for a public presentation by speaking in front of an audience tended to outperform those who practiced alone [155].

To summarize, insights gathered from interviews suggests several features and aspects that participants would like to find in a VRE system for exam anxiety focused on oral exams. Such systems should allow users to customize the pool of questions they wish to use for their training, giving them the freedom to decide whether to set an answer time for each question. These questions should be proposed in random order by the VX, and users should decide the total duration of the simulation, the difficulty level of the behavior the VX should display, and the number of anxiety-provoking elements in the VE (e.g., a readable timer on the VX desk). Finally, the VRE system should include a feature to detect when users remain silent.

A limitation of our study lies in the absence of quantitative data regarding the duration of system usage by each participant. We relied only on spontaneously self-reported data provided by participants during the final interviews. Additionally, the study aimed to gain a deeper understanding of users' perceptions and experience of using the VRE system and did not quantitatively investigate the system's efficacy in reducing participants' exam anxiety and improving their self-efficacy. Therefore, future studies will be based on a quantitative longitudinal design to investigate in depth whether continued use of the VRE system significantly affects these variables. These studies will incorporate automatic logging methods to precisely quantify the VRE system usage and other potential user behaviors. This enhancement should aim to facilitate a more comprehensive assessment of the influence that VRE system usage may have on participants' perceptions and feedback.

## 2.4 Final discussion

Results of the feasibility study described in Section 2.2 and the home trial study described in Section 2.3 suggest that the proposed VRE system has the potential to be a valuable tool for dealing with exam anxiety.

In the feasibility study, findings indicated that the system is able to elicit different levels of anxiety through variations in VX behavior. In addition, the study showed that the three types of VX behavior produced three different, decreasing counts of positive responses elicited and three different, increasing counts of negative responses elicited.

Results of the home trial study extended the analysis of the previous study, revealing that participants perceive the VRE system as a valuable tool for developing emotional skills beneficial in managing oral exams. The customization of questions and their random presentation by the VX contributed to a realistic experience that was greatly appreciated by the participants. In particular, participants reported enhancements in awareness, confidence, and emotional state over time during the simulations.

Both studies identified a demand for customization of the VX, particularly its voice and appearance, to make the experience more tailored to students' individual needs.

In summary, the results from both studies suggest that the proposed

VRE system for oral exam anxiety is promising in supporting students in preparing for oral exams by providing a simulated environment that aims to improve emotions management and cognitive skills to cope effectively with such situations.

# 3

# Immersive VR vs. AR as an exposure method for public speaking anxiety

In Chapter 2 we proposed a VRE system for oral exam anxiety, using a virtual agent as stressor in desktop VR. In this chapter we go one step further by shifting our focus to VRE and ARE systems for another social anxiety-related condition: public speaking anxiety.

This chapter proposes a system that exposes individuals to public speaking scenarios, where they are required to deliver a speech in front of a virtual audience. Although, as we explained in Section 1.1, AR offers advantages over immersive VR, such as allowing individuals to practice public speaking in real-world settings, only VRE systems for public speaking anxiety exist in the literature, while ARE systems for the same purpose are lacking.

Before assessing our exposure system for public speaking anxiety, we conducted a preliminary study to investigate whether the user's perception of the virtual agent changes when shown in immersive VR vs. AR environments. The study results showed that the transition from immersive VR to AR seems to change the perception of virtual agents [156], further motivating the need to compare our system that uses virtual agents in immersive VR and AR.

Similarly to the system described in Chapter 2, we conducted a feasibility study to assess whether our system can effectively elicit anxiety and distress. This assessment involved a comparison between a small and a big virtual audience scenario displayed in both immersive VR and AR environments. To enhance the depth of our analysis, we used eye-tracking to gather data on participants' gaze patterns during public speaking scenarios.

In the following sections, we first introduce our exposure system for public speaking anxiety. Then, we provide a detailed description of the feasibility study we have conducted to assess its efficacy in eliciting anxiety. The results of this study have been documented in [157].

## 3.1 The proposed system

### 3.1.1 The immersive VR and AR environments

The proposed system provides immersive VR and AR display types to see through an HMD the immersive VR and AR environments, respectively.

In immersive VR, the HMD allows users to see a virtual reproduction of the room where the study takes place, with 20 virtual chairs positioned in the center on which a virtual audience is seated. The chairs are placed in staggered rows (i.e., they are arranged in a checkerboard pattern) to minimize the occurrence of virtual agents in the audience being occluded by other agents. The immersive VR environment includes a timer placed on the floor in front of the user, on the right side. The timer is a device that displays a countdown, indicating the remaining time for users to complete their speech. In addition, users can see that they have a neutral human virtual embodiment that moves according to their hands and head tracked by the HMD, as shown in Figure 3.1. The embodiment's feet positioning is determined using inverse kinematics.

In AR mode, the HMD is used in video passthrough mode, allowing users to see the actual room in which they are using the system. Users can see the AR environment that consists of the following virtual elements

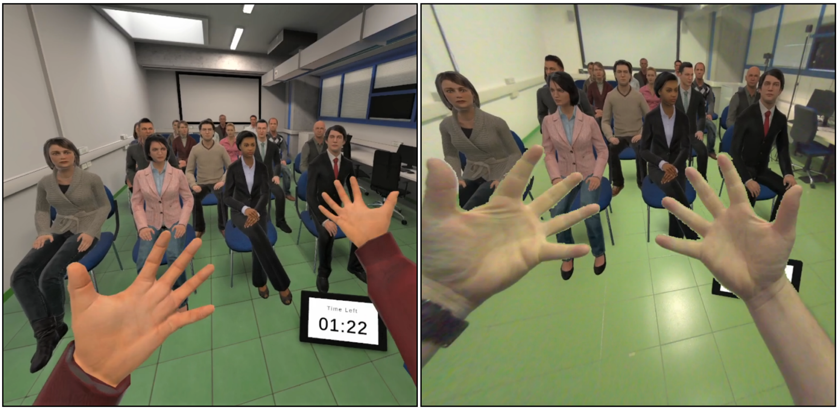



Figure 3.1: Screenshots of the proposed system captured from the left eye of the HMD during an exposure session. The figure on the left displays the big audience in immersive VR mode, where the user sees a neutral human virtual embodiment that moves accordingly to the user's head and hands. The figure on the right displays the big audience in AR mode, where the user can see his/her own body.

superimposed on the real world: the chairs of the audience, the audience, and the timer. They can also see their own body, as shown in Figure 3.1.

The proposed system also offers two different sizes of virtual audience: a small audience of four virtual agents, and a big audience of 16 virtual agents. The immersive VR environment and the AR environment experienced by users are depicted in Figure 3.1.

### 3.1.2 The virtual audience

The virtual audience consists of an equal number of adult female and male virtual agents of different ages, all dressed in business or formal clothes. Table 3.1 shows virtual agents chosen from the characters in the Microsoft Rocketbox library [1] to form the two virtual audience sizes.

Each virtual agent is assigned a fixed position (i.e., in different sessions, every virtual agent is placed on the same chair). In the small audience, each virtual agent is seated in a chair within a distinct row, in a position visible to the user. In the big audience, the arrangement of virtual agents is

Table 3.1: Chosen virtual agents in the Microsoft Rocketbox library [1].

| VA[a] size | Virtual agent name | Virtual agent gender |
|---|---|---|
| Small | “Business_Female_01” | Female |
| | “Business_Female_02” | Female |
| | “Business_Male_01” | Male |
| | “Business_Male_03” | Male |
| Big | “Business_Female_01” | Female |
| | “Business_Female_02” | Female |
| | “Female_Adult_01” | Female |
| | “Female_Adult_02” | Female |
| | “Female_Adult_05” | Female |
| | “Female_Adult_09” | Female |
| | “Female_Adult_14” | Female |
| | “Female_Adult_15” | Female |
| | “Business_Male_01” | Male |
| | “Business_Male_03” | Male |
| | “Business_Male_04” | Male |
| | “Business_Male_06” | Male |
| | “Male_Adult_02” | Male |
| | “Male_Adult_03” | Male |
| | “Male_Adult_07” | Male |
| | “Male_Adult_13” | Male |

[a]VA=Virtual audience

designed to minimize agents being hidden by other agents. In particular, the first three rows are fully occupied by virtual agents, while subsequent rows have a few empty chairs. Figure 3.2 shows the arrangement of virtual agents in both virtual audience sizes.

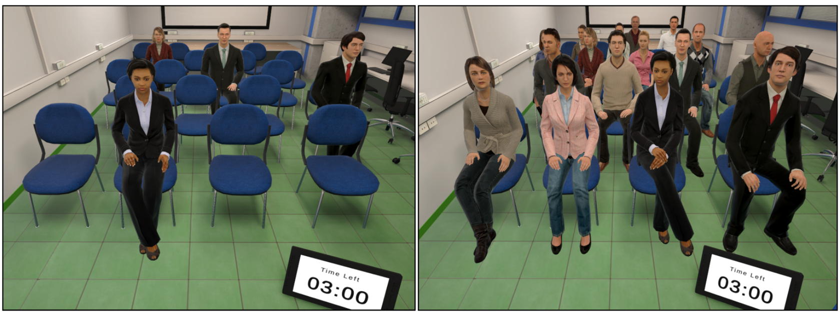



Figure 3.2: Small audience (left) and big audience (right) displayed in immersive VR mode, seen from the user’s viewpoint.

During the exposure session, virtual agents blink their eyes and maintain a sitting idle animation. Specifically, virtual agents can exhibit four distinct sitting idle animations per gender, characterized by variations in hand and leg postures. The use of various idle animations aims to create a diversified and realistic audience, enhancing their resemblance to a real audience.

### 3.1.3 Behavior of the virtual audience

Each virtual agent exhibits one of the following behaviors during the exposure session:

1. Looking at the user throughout the session.
2. Ignoring the user throughout the session.
3. Initially ignoring the user and starting to look at the user for the remainder of the session when the user looks at it for one second.
4. Initially looking at the user, and starting to ignore the user for the remainder of the session when the user looks at it for one second.

In the small audience, each virtual agent exhibits a distinct behavior, as shown in Figure 3.3. In the big audience, virtual agents are divided into four groups, each consisting of four agents, with each group exhibiting a distinct behavior. To ensure homogeneous distribution of the four behaviors within the virtual audience, virtual agents are seated in a way that prevents any row of chairs from having multiple agents performing the same behavior. Figure 3.3 shows the behaviors of virtual agents in small and big audience.

Behaviors (1) and (2) can be considered static behaviors since the virtual agents performing them always look at the user or never look at the user during the session. Conversely, behaviors (3) and (4) can be considered dynamic behaviors, as the virtual agents performing them change their gaze during the session. These four behaviors exhibited by virtual agents can be grouped into two categories: directed behavior and averted behavior, each of which includes a static and dynamic gaze behavior. Thus, directed behavior contains behaviors (1) and (3), while averted behavior contains behaviors (2) and (4).

Figure 3.3: Small audience (left) and big audience (right) in immersive VR mode, seen from the top. The white icons, shaped as a star, circle, triangle, and square, represent respectively the behaviors (1), (2), (3), (4) performed by each virtual agent and described in section 3.1.3.

### 3.1.4 The exposure session and the speech task

Each exposure session of the proposed system is preceded by a calibration phase, during which the system calibrates eye-tracking, ambient noise, and embodiment height.

Eye-tracking calibration ensures the accurate collection of gaze data. Once eye-tracking calibration is done, the user begins to see the environment with the timer device and the empty chairs in the selected display type (immersive VR or AR). This allows the user to familiarize with the environment during subsequent calibrations.

Ambient noise calibration establishes a noise threshold to distinguish user speech from surrounding noise. This threshold is used to measure the user's speech time while excluding ambient noise. The speech time represent the duration, in seconds, that the user spends speaking during the speech task in the exposure session. This calibration involves five seconds of silence, during which the user remains still and silent while the system records ambient noise using a microphone. Then, the system sets the noise threshold at 120% of the highest noise level detected during calibration. This threshold value was chosen empirically to account for additional noise resulting, for example, from footsteps the user might make while giving the speech.

Embodiment height calibration is only required for immersive VR mode and involves adapting the virtual representation of the user's body, i.e., height and proportions, to match the user's actual height.

After the calibration phase, the exposure session begins. The user's view is temporarily filled with a blank black screen for five seconds to prevent the virtual audience from suddenly appearing on the chairs. Once the environment is visible again, the user sees the chairs populated with the virtual audience of the selected size (small or big).

During the exposure phase, the user is required to perform a 3-minute speech task, in which he/she gives a speech to the virtual audience. The speech task begins 30 seconds after the virtual audience is visible, allowing the user to familiarize with its presence. The user's gaze on virtual agents performing dynamic behaviors does not elicit a change in virtual agents' gaze, which will only occur during the speech task. Once the 30 seconds are over, the timer device displays a blank green screen for two seconds, indicating to the user the start of the speech task: the timer device displays a three-minute countdown, during which the user gives a speech to the virtual audience. When the countdown reaches zero, the environment darkens, and the user is surrounded by an empty scenario, indicating the end of the exposure session.

### 3.1.5 Data collection of the proposed system

During each exposure session, the system collects two types of data: session data and eye-tracking data. The session data includes display type and virtual audience size experienced by the user, as well as the speech time. Eye-tracking data is gathered using the embedded eye-tracking system of the HMD. This data is collected at each frame and includes the following information: timestamp, position and orientation of the HMD, orientation of the combined gaze ray from both eyes, and gaze information indicating whether the user's gaze hit a predefined area of interest (AOI) within the environment. The AOI can be the virtual agents' head (AOI heads) or the timer (AOI timer). The presence or absence of a hit on an AOI is determined by checking if the gaze ray intersects with a collider around the AOI. If the user's gaze hits an AOI head, the system logs the name of the corresponding virtual agent and its behavior.

## 3.2 Feasibility study

In the following, we will refer to the group of participants as follows: VR-S group for those who tried the small audience in immersive VR, VR-B for the big audience in immersive VR, AR-S for the small audience in AR, and AR-B for the big-audience in AR.

### 3.2.1 Hypotheses

We formulated the following hypotheses:

H1. The proposed system elicits anxiety and distress levels in both immersive VR and AR modes because VRE systems for public speaking anxiety are able to elicit anxiety in participants [75–79], and ARE systems are able to elicit anxiety in other anxiety disorders such as small animal phobia, claustrophobia, and acrophobia (e.g., [7,8, 158].)

H2. The proposed system elicits comparable anxiety and distress levels in both immersive VR and AR modes because VRE and ARE systems elicit similar levels of anxiety in other types of anxiety such as acrophobia [7], and cockroach and spider phobias [10, 52].

H3. The big audience elicits higher anxiety and distress levels compared to the small audience because a bigger classroom of virtual students induces higher stress levels than a smaller one [84].

H4. Participants' gaze on AOI heads decreases as their public speaking anxiety, anxiety level, and distress level increase because individuals with high levels of public speaking anxiety tend to reduce their gaze toward the audience [76, 91]. Moreover, participants' gaze on the AOI timer increases as their public speaking anxiety, anxiety level, and distress level increase because individuals with higher social anxiety tend to look more at non-social regions than the virtual audience [91]. Additionally, we intend to explore whether there are differences in participants' gaze behavior when exposed to the small vs. big audience.

H5. Participants look more at virtual agents displaying directed behavior compared to those displaying averted behavior because individuals look more at interested or positive pre-recorded audience agents than uninterested or negative pre-recorded audience agents [89, 90].

### 3.2.2 Participants

An a priori power analysis was conducted using the tool G*Power version 3.1.9.4 [159] to determine the minimum sample size required to test the study hypotheses. Results indicated that the required sample size to achieve 90% power for detecting a small effect, at a significance criterion of $\alpha$=0.05, was N=96 for ANOVA repeated measures, within-between interaction.

The study involved a sample of 99 participants (83 males, 14 females, two non-binary). They were undergraduate students in Computer Science at our university, with ages ranging from 20 to 45 (M=22.01, SD=2.78). We asked participants if they were regular VR or AR HMD users: only five participants reported regular usage (hundreds of hours of use). All other participants reported instead between 0 to 50 hours of use (M=1.89, SD=5.86). Finally, we used the 17-item Public Speaking Anxiety Scale questionnaire [160] to assess participants' level of public speaking anxiety (M=48.14,SD=11.78).

Participants were assigned to the four groups in such a way that:

1. Each group had a similar number of participants (VR-S: n=25, VR-B: n=25, AR-S: n=25, AR-B: n=24).

2. The four groups were similar in terms of gender (VR-S: 21M, 4F, 0NB; AR-S: 21M, 3F, 1NB; VR-B: 21M, 3F, 1NB; AR-B: 20M, 4F, 0NB), age, public speaking anxiety scores, and hours of HMD usage of non-regular VR users (Table 3.2).

3. The two non-binary participants were assigned to different groups.

Table 3.2: Means and standard deviations of gender, age, public speaking anxiety scores, and hours of HMD usage of non-regular VR users in VR-S, AR-S, VR-B, and AR-B.

| **Measure** | **VR-S M (SD)** | **AR-S M (SD)** | **VR-B M (SD)** | **AR-B M (SD)** |
|---|---|---|---|---|
| Age | **22.20** (1.53) | **21.80** (1.73) | **21.92** (1.73) | **22.13** (4.91) |
| Public speaking anxiety scores | **48.92** (10.32) | **47.56** (11.68) | **48.72** (14.28) | **47.33** (11.07) |
| Hours of HMD | **1.50** (2.36) | **1.95** (5.10) | **0.96** (1.71) | **3.17** (10.19) |

We confirm the lack of significant differences between the four groups through a Pearson Chi-square test on male and female gender, and a one-way ANOVA on age, hours of HMD usage, and public speaking anxiety score.

### 3.2.3 Materials

The proposed system was developed using the Unity 2020.3.39f game engine and runs on a PC equipped with a 2.5 GHz base clock Intel i9 11900 processor, 32 GB RAM, and an Nvidia 3090 graphic card. The HMD was a Varjo XR-3, a mixed-reality HMD with a horizontal field of view of 115° both in immersive VR and AR. The HMD is equipped with an integrated eye-tracking and hand tracking system.

### 3.2.4 Measures

All questionnaires listed in the following were administered to participants through the PsyToolkit tool [161, 162].

#### Public speaking anxiety

To measure participants' public speaking anxiety, we administered the Public Speaking Anxiety Scale questionnaire (PSAS) [160]. The PSAS is a 17-item self-report scale that assesses cognitive, behavioral, and physiological indicators of public speaking anxiety. Participants rate their degree of agreement with PSAS statements on a 5-point Likert-type scale (1="not at all", 5="extremely"). Total score ranges from 17 to 85, with higher scores indicating higher levels of public speaking anxiety.

#### Anxiety

To measure participants' current transient state of anxiety both before and during their speech task, we administered the State-Trait Anxiety Inventory (STAI-S) [163]. This scale includes 20 items and uses a 4-point Likert-type scale (1="not at all", 4="very much") to rate each response. Total score ranges from 20 to 80, with higher scores indicating higher levels of state anxiety. Before starting the exposure session, participants were asked to provide feedback on their actual state. After the exposure session, they were asked to provide feedback on their highest state anxiety perceived during the speech task.

#### Distress

To measure the level of distress perceived by participants both before and during the speech task, we administered the Subjective Unit of Distress (SUD). The SUD is a single-item self-report distress measure that employs a scale from 0 to 100, with five anchors: 0 (“no distress”), 25 (“mild distress”), 50 (“moderate distress”), 75 (“severe distress”), 100 (“very severe distress”). Before starting the exposure session, participants were asked to provide feedback on their actual distress state. After the exposure session, they were asked to provide feedback on their highest state distress perceived during the speech task.

#### Eye-tracking data

To measure participants’ eye gaze towards the AOIs (i.e., AOI heads and AOI timer) during the exposure session, we analyzed their eye-tracking data and computed two metrics: dwell time and fixations count. Dwell time refers to the amount of time participants spent looking at a specific AOI. Fixations count represents the total number of fixations participants made on the AOIs. We considered a fixation each time participants’ gaze activity lasted 150ms or longer [164]. Dwell time and fixations count were measured for all virtual agents in the four groups.

### 3.2.5 Procedure

Written consent for participation in the study was obtained from participants and the evaluation of the system was approved by the Institutional Review Board of the University of Udine. The experimenter verbally briefed participants about the anonymity of the collected data and informed them that the system involved using an HMD for immersive

VR and AR to do presentations in front of a virtual audience. Participants filled the demographic, PSAS, STAI-S, and SUD questionnaires. Participants were balanced in the VR-S, VR-B, AR-S, and AR-B groups, as described in Section 3.2.2. Then, the experimenter informed participants that they were about to make a public presentation using the system. The experimenter instructed participants to imagine themselves at a big company that is seeking to hire an IT professional. Participants were asked to give a brief self-presentation to the company's IT team, who came to listen to their speech.

Before speaking to the audience, participants had three minutes to prepare their three-minute speech. While preparing their presentation, they were allowed to take written notes, but they could not keep them during the speech task. Once the preparation time elapsed, the experimenter asked participants to stand in a specific place in the room and helped them to wear the Varjo XR-3 HMD.

The experimenter guided participants through the process of calibrating eye-tracking, then asked them to remain silent for five seconds to set the ambient noise threshold. In VR-S and VR-B groups, the embodiment height was also calibrated. After that, the experimenter informed participants that once the session began, they would stand in front of the company's IT team. They were further advised that, after 30 seconds, the timer within the virtual scene would prompt them with a blank green screen to initiate their speech, informing them that the three-minute countdown is started so they can start their speech. The experimenter used the initial 30 seconds of the session to leave the room. Once the timer reached zero, the immersive VR or AR environment was replaced with an empty scenario, indicating the end of the session. The experimenter entered the room and helped participants remove the HMD. Participants filled the STAI-S and SUD questionnaires and where thanked for their participation.

## 3.3 Results

All the analyses were conducted using SPSS version 29.0.0.0. Table 3.3 reports means and standard deviations of public speaking anxiety, anxiety levels, and distress levels measured before the exposure session (pre-session) and during the speech task (during-session) for each group.

Table 3.3: Means and standard deviations of public speaking anxiety, anxiety, and distress.

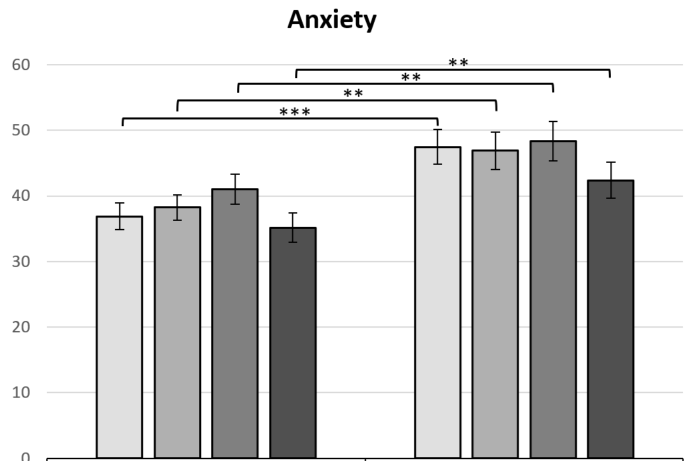
| **Measures** | **VR-S M (SD)** | **AR-S M (SD)** | **VR-B M (SD)** | **AR-B M (SD)** |
|---|---|---|---|---|
| Public speaking anxiety | **48.92** (10.32) | **47.56** (11.68) | **48.72** (14.28) | **47.33** (11.07) |
| Anxiety pre-session | **36.88** (10.20) | **38.24** (9.83) | **41.00** (11.57) | **35.17** (10.91) |
| Anxiety during-session | **47.48** (13.39) | **46.88** (14.25) | **48.36** (15.14) | **42.37** (13.36) |
| Distress pre-session | **15.88** (18.43) | **22.08** (18.72) | **19.48** (20.64) | **16.63** (18.08) |
| Distress during-session | **45.80** (25.40) | **42.00** (26.94) | **40.48** (29.73) | **38.13** (25.38) |

### 3.3.1 Anxiety

STAI-S scores (Figure 3.4) were submitted to a $2 \times 2 \times 2$ mixed design ANOVA, after checking its assumptions were met, in which display type (immersive VR or AR) and audience size (small or big) served as between-subject variables, and time of measurement (pre-session, during-session) served as the within-subject variable.

Statistically significant results revealed a main effect of time of measurement, $F(1, 95) = 40.37$, $p < 0.001$, $\eta_p^2 = 0.30$, while the main effect of display type and audience size, and interactions were not significant.

We thus explored the significant main effect using Bonferroni correction, considering the effects of time of measurement separately for each display type and for each audience size. The difference between pre-session and during-session anxiety was statistically significant in all groups (VR-S: $p < 0.001$; AR-S: $p < 0.01$; VR-B: $p = 0.01$; AR-B: $p = 0.01$), with during-session anxiety significantly higher than pre-session anxiety.



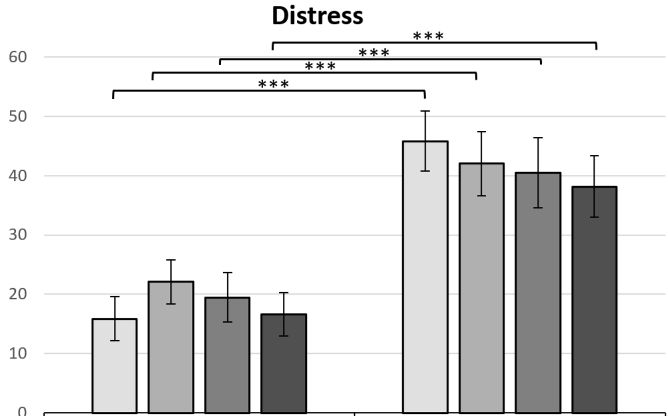

| | pre-session | during-session |
|---|---|---|
| VR-S | 36.88 | 47.48 |
| AR-S | 38.24 | 46.88 |
| VR-B | 41.00 | 48.36 |
| AR-B | 35.17 | 42.37 |

Figure 3.4: Means of the anxiety score at pre-session and during-session. Capped vertical bars indicate ± SE. The **, *** signs indicate differences with $p$-values respectively $< 0.01$, $< 0.001$.

### 3.3.2 Distress

Kolmogorov-Smirnov normality test was performed on SUD scores. Since in some cases they were not normally distributed, the data were subject to square root transformation, as indicated in [165]. The transformed data of SUDS scores (Figure 3.5) were submitted to a $2 \times 2 \times 2$ mixed design ANOVA, in which display type and audience size served as between-subject variables, and time of measurement (pre-session, during-session) served as the within-subject variable.

Statistically significant results revealed a main effect of time of measurement, $F(1, 95) = 77.10$, $p < 0.001$, $\eta_p^2 = 0.45$, while the main effect of display type and audience size, and interactions were not significant. The analysis of simple main effect using Bonferroni correction revealed that the difference between pre-session and during-session distress was statistically significant in all groups ($p < 0.001$), with during-session distress significantly higher than pre-session distress.




| | pre-session | during-session |
|---|---|---|
| VR-S | 15.88 | 45.80 |
| AR-S | 22.08 | 42.00 |
| VR-B | 19.48 | 40.48 |
| AR-B | 16.63 | 38.13 |

Figure 3.5: Means of the distress score at pre-session and during-session. Capped vertical bars indicate ± SE. The *** signs indicate differences with $p$-values respectively $< 0.01$, $< 0.001$.

### 3.3.3 Eye-tracking data

#### AOI timer

Among participants, eight outlier were found based on their dwell time measures on timer, thus they were excluded from the analysis. Kolmogorov-Smirnov normality test was performed on timer dwell time. Since in some cases the values were not normally distributed, they were subjected to a $\log_{10}$ transformation, as indicated in [165]. A $2 \times 2$ ANOVA was conducted with display type and audience size as factors, and timer dwell time as dependent variable. There were no differences in dwell time between immersive VR and AR, and between small and big audience. Moreover, there was no interaction between display type and audience size (Figure 3.6).

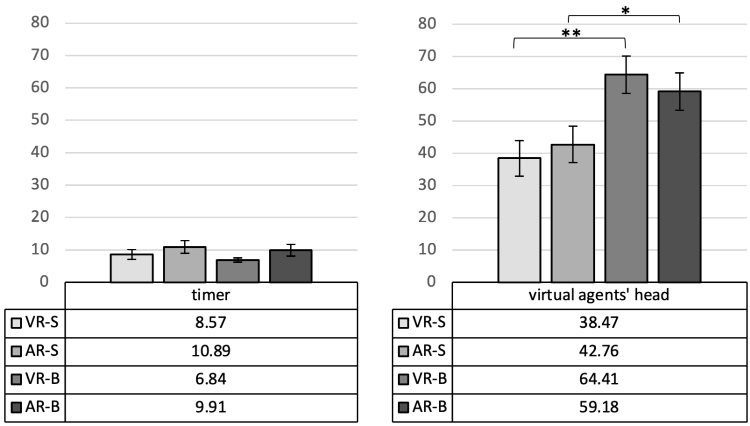



Figure 3.6: Means of AOI timer dwell time (left) and AOI heads dwell time (right) during the speech task. Capped vertical bars indicate ± SE. The * and ** signs indicate differences with $p$-values respectively $< 0.05$, $< 0.01$.

Among the participants, five outliers were found based on their fixations count measures on timer, thus they were excluded from the analysis. Kolmogorov-Smirnov normality test was performed on fixations count measured on timer. Since in some cases the values were not normally distributed, they were subjected to a $\log_{10}$ transformation. A $2 \times 2$ ANOVA was conducted with display type and audience size as factors, and timer fixations count as dependent variable. There were no differences in dwell time between immersive VR and AR, and between small and big audience. Moreover, there was no interaction between display type and audience size (Figure 3.7).

A Pearson correlation was computed to assess possible relationships between timer dwell time and three anxiety measures: public speaking anxiety scores, the difference between during-session anxiety scores and pre-session anxiety scores (STAI-diff), and the difference between during-session distress scores and pre-session distress scores (SUD-diff).

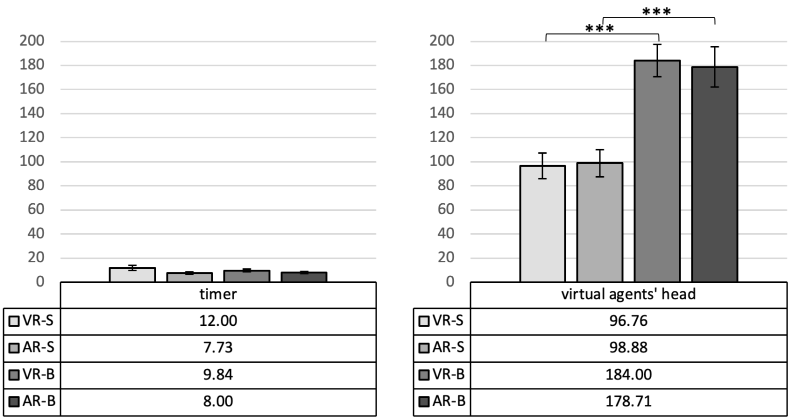



Figure 3.7: Means of AOI timer fixations count (left) and AOI heads fixations count (right) during the speech task. Capped vertical bars indicate ± SE. The *** signs indicate differences with $p$-values $< 0.001$.

The following statistically significant correlations were found:

- A moderate positive correlation between AOI timer dwell time and public speaking anxiety scores in the AR-S group ($r = 0.48$, $n = 23$, $p = 0.02$).
- A large positive correlation between AOI timer dwell time and STAI-diff in the VR-B ($r = 0.58$, $n = 22$, $p < 0.01$) and AR-S groups ($r = 0.5$, $n = 23$, $p = 0.02$).
- A large positive correlation between AOI timer dwell time and SUD-diff in VR-B group ($r = 0.61$, $n = 22$, $p < 0.01$).

**AOI heads**

Dwell time measurements on each virtual agent's head were summed to obtain a single variable indicating the total time participants spent looking at virtual agents' heads during their speech task. Similarly, fixations count measurements on each virtual agent's head were summed to obtain a single variable indicating the total number of fixations participant made on the virtual agents' heads during their speech task.

A $2 \times 2$ ANOVA was conducted with display type and audience size as factors, and AOI heads dwell time as dependent variable. A statistically significant effect of audience size was found, $F(1, 95) = 13.97$, $p < 0.001$, $\eta_p^2 = 0.13$. The analysis of simple main effect using Bonferroni correction revealed that the difference between small (VR-S: M=38.47, SD=27.50; AR-S: M=42.76, SD=28.08) and big audience (VR-B: M=64.41, SD=28.81; AR-B: M=59.18, SD=28.34) was significant in both immersive VR ($p < 0.01$) and AR ($p < 0.05$) (Figure 3.6).

A $2 \times 2$ ANOVA was conducted with display type and audience size as factors, and AOI heads fixations count as dependent variable. A statistically significant effect of audience size was found, $F(1, 95) = 40.30$, $p < 0.001$, $\eta_p^2 = 0.30$. The analysis of simple main effect using Bonferroni correction revealed that the difference between small (VR-S: M=96.76, SD=53.85; AR-S: M=98.99, SD=56.66) and big audience (VR-B: M=184.00, SD=66.52; AR-B: M=178.71, SD=81.87) was significant in both display types ($p < 0.001$) (Figure 3.7).

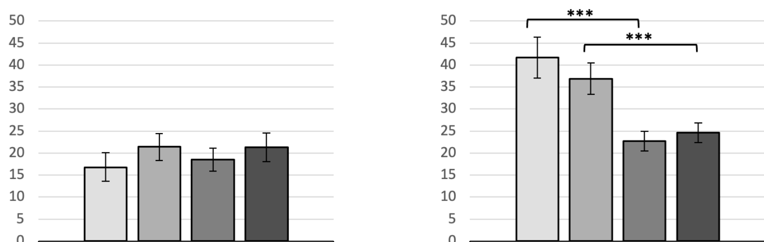
**Participants' gaze behavior on virtual agents**

To investigate whether participants looked at virtual agents that stare at them or ignored them, dwell time measurements on virtual agents with a static or dynamic directed behavior were summed to obtain a single variable indicating the total time participants looked at the heads of the virtual agents with a directed behavior. Similarly, dwell time measurements on virtual agents with a static or dynamic averted behavior were summed to obtain a single variable indicating the total time participants looked at the heads of the virtual agents with an averted behavior.

Among the participants, two outliers were found based on their dwell time measures on the two behavior variables, thus they were excluded from the analysis. Kolmogorov-Smirnov normality test was performed on the two variables. Since in some cases the values were not normally distributed, a $\log_{10}$ transformation was applied.

A $2 \times 2$ ANOVA repeated measures was conducted for each audience size. The analysis used display type as between-subject variable, and virtual agents' behavior as within-subject variable.

In the small audience, no statistically significant main effect of virtual agents' behavior was found. On the contrary, in the big audience, a statistically significant effect of virtual agents' behavior was found, $F(1, 46) = 57.54$, $p < 0.001$, $\eta_p^2 = 0.56$. The analysis of simple main effects using Bonferroni correction revealed that participants significantly looked more at virtual agents with directed behavior (VR-B: M=61.67, SD=23.21; AR-B: M=36.91, SD=17.28) than virtual agents with averted behavior (VR-B: M=22.73, SD=11.03; AR-B: M=24.63, SD=10.63) in both display types ($p < 0.001$) (Figure 3.8).




| | Small audience | Big audience |
|---|---|---|
| VR - Directed | 16.80 | 41.67 |
| AR - Directed | 21.43 | 36.91 |
| VR - Averted | 18.49 | 22.73 |
| AR - Averted | 21.33 | 24.63 |

Figure 3.8: Means of dwell time on virtual agents' heads with directed and averted behaviors in the small audience (left) and in the big audience (right). Capped vertical bars indicate ± SE. The *** signs indicate differences with $p$-values 0.001.

Fixations count on virtual agents with a static or dynamic directed behavior were summed to obtain a single variable indicating the total number of times participants looked at the heads of the virtual agents with a directed behavior. Similarly, fixations count on virtual agents with a static or dynamic averted behavior were summed to obtain a single variable indicating the total number of times participants looked at the heads of the virtual agents with an averted behavior.

A $2 \times 2$ ANOVA repeated measures was conducted for each audience size. The analysis used display type as between-subject variable, and virtual agents' behavior as within-subject variable. In the small audience, a statistically significant effect of virtual agents' behavior was found, $F(1,48) = 7.39$, $p < 0.01$, $\eta_p^2 = 0.13$. The analysis of simple main effect using Bonferroni correction revealed that participants looked at virtual agents with directed behavior (VR-S: M=55.00, SD=43.73; AR-S: M=57.76, SD=38.86) more times than virtual agents with averted behavior (VR-S: M=41.76, SD=23.74; AR-S: M=41.12, SD=24.04), but the difference was statistically significant only in AR display type ($p < 0.05$).

In the big audience, a statistically significant effect of virtual agents' behavior was found, $F(1,47) = 61.91$, $p < 0.001$, $\eta_p^2 = 0.57$. The analysis of simple main effect using Bonferroni correction revealed that participants looked at virtual agents with directed behavior (VR-B: M=112.64,

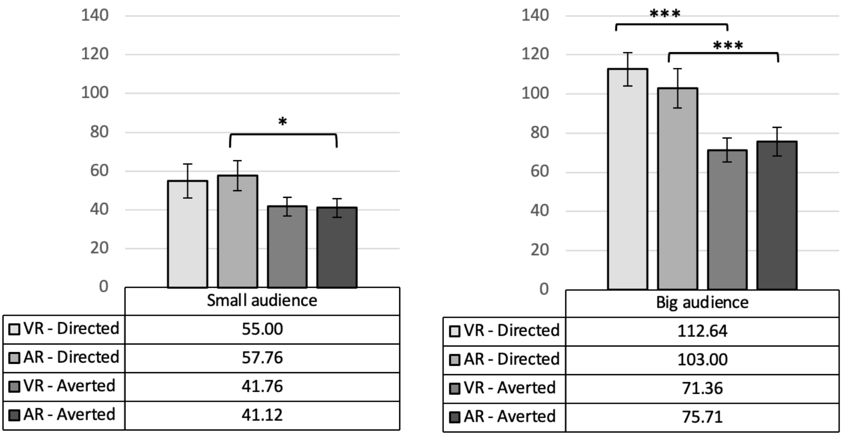



Figure 3.9: Means of fixations count on virtual agents' heads with directed and averted behaviors in the small audience (left) and in the big audience (right). Capped vertical bars indicate ± SE. The * and *** signs indicate differences with $p$-values respectively $< 0.05$ and $< 0.001$.

SD=42.89; AR-B: M=103.00, SD=49.04) significantly more than virtual agents with averted behavior (VR-B: M=71.36, SD=30.96; AR-B: M=75.71, SD=35.86) in both display types ($p < 0.001$) (Figure 3.9).

## 3.4 Discussion

Results have confirmed the first hypothesis H1. The analysis of the STAI-S scores and SUD scores revealed a significant increase in participants' anxiety and distress levels during the speech task compared to before the exposure session. These findings are consistent with previous studies [75–79], where VRE systems for public speaking anxiety were able to elicit anxiety in participants. Moreover, these findings extend the existing research on ARE systems [7,8,158], which were designed for small animals phobia [158], acrophobia [7], and claustrophobia [8]. In contrast, our study focused on public speaking anxiety and results support the feasibility of the proposed system in eliciting anxiety and distress in a public speaking situation in both immersive VR and AR modes.

Results have confirmed the second hypothesis H2. Although the a priori power analysis used in the study was adequate to detect significant differences between the groups, the analysis of STAI-S scores and SUD scores revealed that anxiety and distress experienced by participants in immersive VR and AR modes were similar. These findings are in line with [7], which compared the use of VR and AR modes in a system presenting acrophobic scenarios. The study showed that both immersive VR and AR elicited comparable levels of anxiety in participants. In our study, we obtained similar results within the context of public speaking anxiety, thus contributing to the growing body of research on the effectiveness of immersive VR and AR modes in eliciting anxiety responses. Our results may seem in contrast with the findings in [8], which instead showed that a claustrophobic scenario elicited significantly different anxiety levels in immersive VR than in AR. However, it is important to note that in [8], participants were exposed to the AR claustrophobic scenario inside a real elevator, whereas they experienced the immersive VR scenario displaying a virtual elevator inside a room while seated on a chair. These differences in participants' experiences could have contributed to varying subjective and physiological anxiety levels of participants during the exposure. Our study, instead, ensured that the immersive VR and AR environments were identical and replicated the actual physical environment in which participants experienced them. Indeed, the immersive VR environment was the exact virtual reproduction of the real environment where the study took place, visible through the headset in AR mode (Figure 3.1).

Results have not supported the third hypothesis H3. They showed that the small audience and the big audience elicited similar anxiety and distress levels. These results are consistent with outcomes in [75] that did not find differences in anxiety levels between the different audience sizes. Although in [75] the authors also reported a significant increase in participants' HR in the small audience scenario compared to the other audience sizes, it should be noted that the study used a within-subjects design, and the significant different in HR measures was found only when the small audience was the first scenario experienced by the participants. Our results differ from those in [84], which found that a big audience elicited higher stress levels and HR than a small audience. However, it is important to highlight the several differences between that study and ours. In [84], participants experienced the system twice in two weeks, and at each session, they were randomly assigned to either the big or the small audience. The system simulated a classroom setting, aiming to help preservice teachers become familiar with an audience of adolescent students who were programmed to misbehave during the session (e.g., throwing paper balls, hitting the neighbor). In contrast, our study employed a between-subjects design, where participants were exposed to a single public speaking scenario. Participants delivered a public speech in front of a small or big audience that behaved neutrally, and was displayed in either immersive VR or AR.

Results have provided only partial support for the fourth hypothesis

H4. Indeed, we found no significant correlations between AOI heads dwell time and public speaking anxiety, STAI-diff, or SUDS-diff scores. These findings might suggest that participants with higher anxiety levels might have tended to look at the timer as a coping mechanism to manage their anxiety or as a means of diverting attention away from their anxiety [166]. Unlike the AOI timer, results on AOI heads fixations count and dwell time revealed that participants looked more times at the virtual agents' heads in the big audience than in the small audience in both display modes. These findings might be attributed to the higher number of virtual agents in the big audience compared to those in the small audience: participants have shifted their gaze focus among the different agents, resulting in an overall increase in the total fixations count. Similarly, in the small audience participants have shifted their gaze focus fewer times among the four virtual agents to spend more time on each one during each fixation. However, results obtained on dwell time have shown that participants spent more time looking at virtual agents' heads in the big audience than in the small audience. This suggests that participants have spent more time looking at non-virtual agent objects during their speech task. Since the virtual agents' heads in a public speaking context are likely to be the most anxiety-provoking stimuli of the environment, we can suppose that participants may have employed the coping mechanism of diverting attention away from anxiety-provoking stimuli to manage their anxiety during the speech. Further research is necessary to explore the eye gaze pattern followed by participants during their speech and investigate how this pattern differs between small and big audiences.

Results have confirmed the fifth hypothesis H5. The analysis of dwell time and fixations count on virtual agents' heads with directed behavior and virtual agents with averted behavior revealed that participants looked more at virtual agents with directed behavior than those with averted behavior. Our findings are consistent with a previous study [93] that examined participants' gaze behavior while observing a virtual audience.

The study showed that virtual agents who stared at the participants the whole session were significantly more looked at by participants than those who never looked at participants. We showed that the results are the same when participants were asked to give a speech in front of a virtual audience.

Since positive and negative audiences are generally represented by agents attentive to the speaker and distracted agents, respectively [167], we can consider our results aligned with those in [89] and [90]. In [89], healthy participants looked more at interested audience agents than uninterested ones, while in [90] participants looked more at positive audience agents than neutral agents and more at neutral agents than negative ones, but this difference was found only in participants with low social interaction anxiety levels. In our results, we expected to obtain outcomes more similar to studies conducted on healthy individuals than patients with diagnosed social anxiety disorder. Indeed, our sample did not display pathological public speaking anxiety scores: the PSAS average scores we obtained in the four groups were below 49 (see Section 3.2.2), which is lower than the cutoff score of 73, as defined by the PSAS for diagnosing significant and impairing public speaking anxiety [160]. Moreover, since studies described in [89] and [90] used systems with pre-recorded audiences of real individuals, our study contributed to the existing literature by showing that participants tended to look more at interested audience agents than uninterested ones, also when the audience consisted of 3D computer-generated virtual agents showed in immersive VR as well as AR.

We are the first to propose a system for training public speaking in AR and to assess its impact on participants' anxiety and distress levels. Furthermore, our study is the first to compare participants' anxiety and distress levels experienced in the same environment displayed in immersive VR vs. AR while speaking in front of a virtual audience. However, there are some limitations that should be considered.

First, all participants enrolled in our study were undergraduate students who shared a common academic background in Computer Science, which may limit the generalizability of the findings.

Second, the study was conducted on a predominantly male sample. As mentioned earlier is Section 2.2.8, the possible consequences of a predominant male sample on anxiety measures have been examined in the literature: several studies show that females tend to report greater anxiety [146] and higher intensities of emotional experiences [147, 148] than males. Moreover, females experience higher levels of anxiety in public speaking situations compared to males [168, 169]. Therefore, if male predominance in our sample influenced the results, it might have attenuated the intensity of anxiety. To further validate our findings, a direct comparison between our female sample and a randomly selected male sample of the same size was not feasible due to the limited number of participants in each group. As an alternative approach, we excluded females from our sample and conducted a 2x2x2 mixed design ANOVA on STAI-S scores and SUDS scores. The obtained results aligned with our original findings. Specifically, a main effect of time of measurement was found in both STAI-S ($F(1, 79) = 47.59$, $p < 0.001$, $\eta_p^2 = 0.38$) and SUDS scores ($F(1, 79) = 86.99$, $p < 0.001$, $\eta_p^2 = 0.52$), while the main effect of display type, audience size, and interactions were not significant. Subsequent analysis using Bonferroni correction showed that the difference between pre-session and during-session anxiety and distress were statistically significant in all groups (STAI-S scores: $p < 0.01$ in VR-S, VR-B, and AR-S, $p < 0.001$ in AR-B; SUDS scores: $p < 0.001$ in all groups).

# 4
# Immersive VR vs. desktop VR with biofeedback for relaxation training

In previous chapters, we have focused on VRE systems for anxiety mitigation comparing immersive VR, desktop VR, and AR. In this chapter, we shift our focus to the other category of VR systems for anxiety mitigation, that is VR systems with biofeedback for relaxation training. We present an immersive VR system that uses biofeedback mechanisms. The assessment of this system involved two studies, each serving a specific purpose. The first, aimed to assess the effectiveness of biofeedback mechanisms in reducing anxiety. The second study compared the system in immersive vs. desktop VR.

In the following sections, we first introduce our VR system with biofeedback for relaxation training. Then, we describe the study conducted on healthy individuals, comparing the relaxation effects achieved with real and sham biofeedback. Results of this study have been documented in [170]. Subsequently, we described the second study, which involved the comparison of the system in immersive VR and desktop VR.

## 4.1 The proposed system

This section describes in detail our VR system with biofeedback for relaxation training. First, we illustrate the VR experience, then we analyze the employed biofeedback mechanisms.

### 4.1.1 The VR experience

The VE was developed in Unity 2020.3.27f and is a stylized visual representation of a coastal nature environment with a long, narrow beach facing the sea. Three islands of various sizes are close to the shore, and a windmill stands on the largest island (Figure 4.1 (a)).

Rocks, grass, trees, bushes, and crystals are arranged in the VE following the indication in [171], that compared six natural VEs, each with a different percentage of greenery covering from 10 to 70% of the entire space, showing that the most effective natural environment for reducing negative mood had 10–30% of the scene covered by vegetation. The user sits on a wooden bench with a bush at his/her right (Figure 4.1 (b)). We called the VE "Crystals Archipelago" after the many crystals scattered around it.

The experience begins around sunset with the VE initially covered in fog (Figure 4.1 (c)). To enhance the experience, a variety of environmental sounds are played. A calming background music soundtrack and the sounds of wind, waves, and gears that move the windmill blades are played during the whole experience. Moreover, some environmental sounds are reproduced at specific times, such as when the crystals glow and when the user interacts with the bush.

The VR experience involves the user as the main character of a story narrated by a voice-over. Initially, the voice-over describes the Crystals Archipelago as a magical land that can connect with the user through magical crystals. The voice-over then explains that the user's life spirit is infused into nearby crystals and spread to all the other crystals in the Archipelago, bringing them all in synch with the user and making the

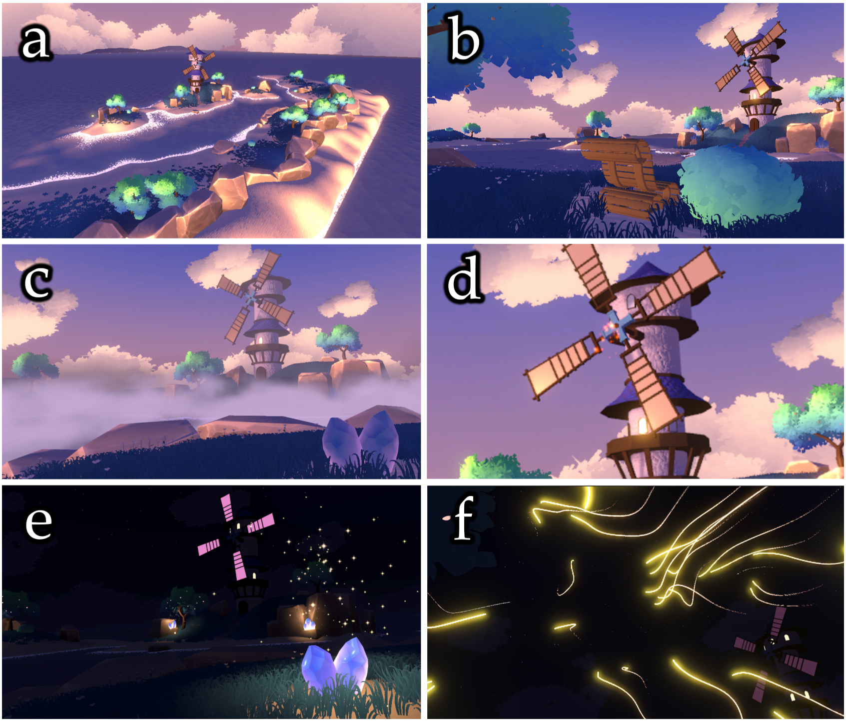



Figure 4.1: (a) Scenery of the Crystals Archipelago; (b) The bench on which the participant sits during the entire experience; (c) Fog in the VE, seen from the user's viewpoint; (d) Jamming of the windmill blades; (e) Fruits on the trees light up and light particles gently rise upward from all the crystals in the finale, from the user's viewpoint; (f) Fireflies flying away and drawing light trails in the finale, from the user's viewpoint.

entire world respond to him/her. The user is told that in this way he/she can explore his/her abilities by first clearing the surrounding fog. After trying that, the user is invited to perform other actions (touching the bush, releasing fireflies into the world) and tasks (making the night fall by relaxing). At the end of the experience, night comes, the fireflies fly away into the sky (Figure 4.1 (f)), and the voice-over concludes the story, informing the user that the experience is over and inviting him/her to return soon to the Archipelago.

### Structure of the VR experience

The VR experience is organized into five phases:

1. System calibration.
2. Clearing the fog.
3. Interaction with the bush.
4. Making the night fall.
5. Finale.

In the first phase, the system needs to be calibrated to the amplitude of the specific user's breathing by detecting the user's maximum and minimum chest expansion values. To this purpose, the experience begins with the voice-over that asks the user to take three deep breaths. Then, the system monitors all user's breaths for 60 seconds, storing the maximum and minimum expansion values detected by the sensor. During this time, two crystals in front of the user keep increasing their brightness until they emit some sparks at the end of the interval. Subsequently, a light trail sparks from the two crystals and moves to reach the different groups of crystals in the VE, brightening them. Then, to bring user's attention to the windmill blades and the foliage, the voice-over asks the user to observe their movement (Table 4.1), which from now on follows user's breathing.

In the second phase, the voice-over instructs the user about how to breathe slowly and deeply with the diaphragm and encourages him/her to do it to clear the VE of fog and maintain it clear. This task lasts for three minutes after which any remaining fog is cleared by the system to allow the user to move to the next phase.

In the third phase, the user hears the sound of rustling leaves coming from the bush at his/her right and the voice-over invites him/her to touch it. As the user touches the bush, fireflies come out of it and slowly fly near the user. If the user does not interact with the bush within 20

Table 4.1: Sentences uttered by the voice-over to direct users' visual attention during the VR experience. The voice-over was in Italian, all sentences have been translated here into English for reader's convenience.

| **VR experience phase** | **Sentence** |
|---|---|
| First phase | *Your life spirit shines in the crystals and has reached every fiber of the environment. Observe the movement of the windmill blades. Also, observe the movement of the leaves of the trees around you.* |
| Second phase | *Breathe slowly and deeply to clear the fog that surrounds you, and try to keep the environment free from fog.* |
| The first time the participant exhibits a respiratory rate above 6 bpm, from the second phase to the end of the fourth phase | *Breathing too quickly damages the windmill.* |
| Third phase | *Also the fireflies are synchronized with you. Your energy flows in everything.* |
| Fourth phase | *Try to relax deeply to let the night fall. Then, try to maintain it.* |

seconds, fireflies automatically come out of the bush to allow the user to move to the next phase.

In the fourth phase, the voice-over informs the user that the sunset time is approaching and asks him/her to deeply relax to allow the night fall and then to keep the world in night conditions. The task lasts for three minutes after which night automatically takes over if the environment is not already in full night conditions.

The fifth phase marks the end of the experience and is meant as a final, emotion-evoking reward. The background music changes from calm to more lively, and the fruits on the trees light up. Additionally, light particles gently rise upward from all the crystals (Figure 4.1 (e)). Finally, the fireflies begin to fly skyward, leaving behind light trails in the night (Figure 4.1 (f)). While they are flying away, the voice-over tells the user they are saying goodbye, then it greets the user and invites him/her to return to the Crystals Archipelago soon.

During the VR experience, the voice-over subtly suggests which elements on the VE the user should pay attention to, but does not mention if and how physiological parameters control the movement or appearance of those elements. Table 4.1 contains the sentences of the voice-over used to direct user's visual attention.

### 4.1.2 Biofeedback mechanisms

The system uses an external application to receive physiological data recorded by a Thought Technology ProComp Infiniti encoder and processes them in real-time with a sampling rate of 10 Hz. An elastic girth sensor is placed over the user's abdomen to measure breathing activity; SC is recorded through a pair of Ag/AgCI electrodes placed in the center of the palm and the carpus of the left hand, respectively; HR is recorded through a photoplethysmograph placed on the distal phalanx of the middle finger of the left hand. Multiple sensors are used to monitor whether the user is in a state of relaxation and capture different aspects of bodily responses. Specifically, breathing activity data indicates whether user's breathing is slow and deep, while electrodermal and cardiac activity data indicate whether user's SC and HR are decreasing, respectively. User's breathing activity, SC and HR affect the way the VE looks and behaves. Table 4.2 provides an overview of the mappings we introduced.

#### Breathing biofeedback

Breathing data are normalized during the first phase of the VR experience so that the maximum and minimum chest expansion correspond to 1 and 0, respectively. If chest expansion is higher than a threshold of 0.7, the inhalation is considered deep; if it is lower than a threshold of 0.45, the exhalation is considered deep. Ten times a second, the rate of change of chest expansion is computed to determine whether user's breathing was slow by taking the two consecutive most recent breathing values, subtracting the older value from the more recent one, and dividing the result by the time elapsed between the two breathing values. If the result

Table 4.2: Mapping of physiological parameters in the VE.

| Parameter | Mapping in the VE | Related references |
|---|---|---|
| Exhalation | From the final part of the first phase until the end of the experience: sound of wind blowing; increase in volume of the sound of wind blowing if the exhalation is deep; increase in leaf swing on trees and bushes; increase in volume and pitch of the sound of windmill gears and rotation speed of windmill blades (the faster the exhalation, the faster the speed and the higher the volume and pitch); windmill blades glow yellow if the exhalation is deep. During the second phase of the experience: increase in amount of fog if the previous inhalation has not been slow or has not been deep; reduction in amount of fog and increase in fog distance from the user if a slow, deep inhalation is followed by a slow, deep exhalation | [26,27,30, 100,108,172, 173] |
| Inhalation | From the final part of the first phase until the end of the experience: decrease in leaf swing of trees and bushes; constant decrease in rotation speed of windmill blades (the longer the inhalation, the lower the speed); windmill blades glow pink if the inhalation is deep | [26,27,108] |
| Respiratory rate | From the second phase until the end of the fourth phase: red sparks come out of the windmill blades and windmill blades wobble if respiratory rate is greater than 6 bpm | [108] |
| Rate of SC change | During the fourth phase: change of sun and moon positions (the lower the rate of SC change, the lower the sun position until it disappears beyond the horizon and the higher the moon position, causing night to fall) | [31] |
| HR | From the third phase until the end of the fourth phase: flashing rate of fireflies | [174] |

is lower than 0.7 during inhalation or 0.8 during exhalation, the user is breathing slowly. The two values were empirically obtained by simulating a respiratory rate of 6 bpm, which is typically used in breathing training systems (e.g., [23,175,176]).

Once the user's minimum and maximum chest expansion values have been detected, the user can control the wind in the VE through his/her breathing activity until the end of the VR experience. Each time the user inhales, the leaves of the trees and bushes slow down their swing. When the user exhales, they swing faster up to a maximum value. If the user holds the breath, the swinging speed decreases until the leaves completely stop. Every exhalation is accompanied by the sound of wind blowing, which increases in volume if the exhalation is deep. The choice to map wind to breathing stems from the fact that when an individual exhales, the produced air displacement feels like a light breeze. The visual feedback of leaves swinging is accompanied by congruent auditory feedback of wind blowing to enhance the sense of relaxation in the VE [172]. The windmill blades behave similarly: the rotation speed of the blades depends on the speed of exhalation (the faster the exhalation, the faster the speed). In addition, as the rotation speed of the windmill blades increases during exhalation, the volume and pitch of the sound of the gears also increase up to a maximum value. On the contrary, during inhalations, the rotation speed of the blades slows down constantly, regardless of the speed of inhalation. This mapping of windmill blades uses the same metaphor of the pinwheel: the air produced by an individual during exhalation makes the pinwheel rotate, but the individual's inhalation does not affect the rotation of the pinwheel, which therefore slows down over time.

To indicate whether the user is performing slow and deep breathing, the windmill blades glow pink during a deep inhalation and yellow during a deep exhalation. In the literature, there are two studies that use color [26] and lighting [27] of elements in the VE to differentiate between inhalations and exhalations, regardless of their depth. In our system, the color of the windmill blades changes only when breathing is deep, and discriminates between inhalation and exhalation, to help the user identify and distinguish deep breaths from more shallow breaths.

If the user is breathing faster than 6 bpm, the windmill blades jam: they slow down abruptly, wobble and emit red sparks (Figure 4.1 (d)). This damage to the windmill blades sends a negative feedback to the user for breathing incorrectly. We have included it in the experience following the operant conditioning paradigm [108]: to motivate users to make more effort in the task, in addition to audio-visual rewards when they maintain slow and deep breathing, this aversive feedback plays the role of an explicit punishment when they do not follow the recommended behavior.

During the second phase of the VR experience, the user performs the task of maintaining slow and deep breathing over time to clear the environment from fog. The system uses two parameters to determine fog behavior: a distance parameter controls the distance of fog from the user, a threshold parameter controls the amount of fog and varies between a minimum and a maximum value corresponding to the maximum amount of fog and the absence of fog, respectively. If the user takes a slow and deep inhalation followed by a slow and deep exhalation, during exhalation the two parameters distance and threshold continuously increase. Thus, the longer the exhalation, the greater the distance of fog from the user and the smaller the amount of fog. If user's breathing is not deep or slow, the threshold parameter decreases thus the amount of fog increases. In the literature, there are three studies that use a similar mapping: in [100], the fog in the VE dissolves if the user maintains the breathing rate between 5.5 and 6 bpm for one minute; in [30], similarly, the sky clears of clouds if user's HRV value increases above a threshold value; in [173], the fog in the VE decreases as user's SC levels decrease and increases as they increase. Our system, as in [173], changes the amount of fog with each user's breath. In this way, the system helps users immediately understand if they are performing breathing properly, allowing them to correct their breathing patterns if necessary.

#### SC biofeedback

During the fourth phase of the VR experience, the user relaxes to make the night fall. The system adopts a biofeedback mechanism similar to [31]. In that system, one scenario consists of an empty VE that only displays a sun high in the sky. As users relax, their SC decreases, causing an increase in the speed at which the sun moves across the sky until it sets beyond the horizon line, bringing the night to fall. Then, further decrease in user's SC causes an increase in the speed at which the moon moves across the sky until it reaches the zenith. If the user does not relax and thus his/her SC increases, the speed at which the sun (and then the moon) moves slows down. Our system uses a similar biofeedback mechanism: a decrease in SC results in a change in the environment light originating from the sun, whose position and intensity decrease until it no longer contributes to VE illumination. At the same time, the moon position and intensity increase.

Unlike the system described in [31], in our system the sun and the moon are not graphically represented by 3D objects but are directional lights whose position and intensity are changed according to user's SC. The position of the sun and the moon are changed according to rate of SC change (which is computed in the same way as the rate of change of chest expansion): if the rate of SC change is negative, the position of the sun lowers towards the horizon until a minimum value, decreasing in intensity and, at the same time, the position of the moon rises towards the zenith until a maximum value, increasing in intensity.

**HR biofeedback**

During the third phase of the VR experience fireflies emerge from the bush and fly around the user until the end of the fourth phase. User's HR is mapped into the flashing rate of fireflies. This type of biofeedback draws on the concept used in many first-person video games where the circular red transparent texture overlay is faded in and out over the entire user's view following the rhythm of a preset heartbeat sound. In [174], this way of displaying user's HR with a flashing texture was the most effective for participants to accurately assess their HR, compared to two alternative visualizations.

## 4.2 Study 1: real vs. sham biofeedback

We conducted a between-subjects study to investigate whether the proposed VR experience has relaxation effects and to explore the possible role played by biofeedback. From now on, we will use the term Biofeedback (BIO) group to refer to the group of participants who tried the system with real biofeedback, i.e., they controlled the VE with their real-time physiological activity. In contrast, the term Sham (SHA) group will refer to participants who tried the system with sham biofeedback, i.e., changes in the VE were controlled by physiological data recordings from a previous user, randomly selected from the BIO group.

### 4.2.1 Hypotheses

We formulated the following hypotheses:

H1. The VR experience relaxes users because the designed VE represents a natural scenario, and exposure to natural environments, both real [177, 178] and virtual [172, 179, 180], can reduce stress and anxiety.

H2. Biofeedback enhances the relaxation effect of the VR experience because participants in the BIO group receive real-time feedback on their physiological activity and this can help them in changing it, potentially achieving greater relaxation compared to users of the SHA group, which do not receive real feedback about their performance.

H3. Biofeedback increases sense of presence in the VE because alterations of the VE through physiological changes in biofeedback may increase attractiveness of feedback [30] and we posit that this could positively influence sense of presence perceived by users. While immersive VR is known for its ability to induce a high sense of presence [181–183], little is known about if biofeedback influences sense of presence. In [32] an immersive VR system with biofeedback achieved a greater sense of presence than immersive and non-immersive VR versions of the same system without biofeedback, but the difference was not statistically significant. Moreover, no study of sense of presence in biofeedback systems has considered a placebo (sham biofeedback) condition.

### 4.2.2 Participants

The study involved a sample of 35 participants (26 males, 9 females), who were volunteers that received no compensation and were recruited through direct contact, email, and the social channels of our university. Their age ranged from 19 to 48 (M=26.60, SD=7.23), and they were undergraduate students from different faculties or university employees. We asked participants if they were regular VR headset users: only two participants reported regular usage (several hundred hours of use). All other participants reported instead between 0 to 80 hours of use. Finally, we asked participants to complete the 20-item STAI-S questionnaire [163] to assess their current transient state of anxiety (M=32.31, SD=5.42) and the 20-item STAI-T questionnaire [163] to assess their relatively stable personality trait related to anxiety (M=40.66, SD=9.38). Participants were assigned to two groups so that:

1. Each group had a similar number of participants (BIO: n=18; SHA: n=17).
2. The two groups were similar in terms of gender (BIO: 14M, 4F; SHA: 12M, 5F), age, regular use of VR HMDs (BIO: 1 "yes", 17 "no"; SHA: 1 "yes", 16 "no"), state anxiety, and trait anxiety scores (Table 4.3).
3. Each of the two participants who were regular VR HMDs users was assigned to a different group.

We confirmed the lack of significant differences between the two groups through a Pearson Chi-square test on gender, and an unrelated *t*-test for age, hours of VR HMDs usage, state and trait anxiety scores. The analysis excluded four participants due to VR HMD shutdown (n=1), artifacts in the recorded data caused by participant's cough (n=1), and unannounced fire alarm tests in the university building during two sessions (n=2). As a result, the analysis was conducted on 31 participants (BIO: n=16; SHA: n=15).

Table 4.3: Means and standard deviations of gender, age, non-regular VR HMD users, state anxiety, and trait anxiety in BIO and SHA groups.

| **Measure** | **BIO M(SD)** | **SHA M(SD)** |
|---|---|---|
| Age | **26.50** (7.14) | **26.71** (7.54) |
| Non-regular use of VR HMDs | **3.06** (7.37) | **5.63** (19.88) |
| State anxiety | **32.72** (6.27) | **31.88** (4.51) |
| Trait anxiety | **40.39** (9.87) | **40.94** (9.13) |

### 4.2.3 Measures

All questionnaires listed in the following were administered to participants through the PsyToolkit tool [161, 162].

#### Anxiety

To measure the possible relaxation effects given by the VR experience, we administered the STAI-S questionnaire [163], described previously in Section 3.2.4, before and after the use of the system.

#### Sense of presence

To measure sense of presence, we administered the Igroup Presence Questionnaire (IPQ) that asks participants to rate 14 items on a 7-point Likert scale, ranging from 0 to 6 [184]. The questionnaire includes a general item related to the sense of "being there", and three subscales for the independent dimensions of "spatial presence" (5 items), "involvement" (4 items), and "experienced realism" (4 items).

### Perceived biofeedback quality and ease of tasks

After the VR experience, we employed a customized biofeedback questionnaire where participants rated eight items on a 7-point Likert-type scale (1="not at all", 7="a lot") to assess perceived biofeedback quality (the first six items in Table 4.4) and the ease of performing the relaxation tasks of clearing the fog and making the night fall (the two last items in Table 4.4). Participants' ratings of the first six items were averaged to form a reliable scale of perceived biofeedback quality (Cronbach's $\alpha$=0.76). Participants' rating of the last two items were averaged to form a reliable scale of ease of performing the relaxation tasks (Cronbach's $\alpha$=0.60).

Table 4.4: Biofeedback questionnaire items. The questionnaire was filled out in Italian, all items have been translated here into English for reader's convenience.

| | |
|---|---|
| 1 | During the experience, I had the impression that I was controlling the wind with my breathing. |
| 2 | During the experience, I had the impression that the leaves of trees and bushes moved according to my breathing. |
| 3 | During the experience, I had the impression that the movement of the windmill blades followed my breathing. |
| 4 | During the experience, I had the impression that the flashing of fireflies was slower when I was more relaxed. |
| 5 | During the phase of the experience in which I had the task of clearing the fog, I had the impression that the fog cleared if I breathed slowly and deeply. |
| 6 | During the phase of the experience in which I had the task of making night fall, I had the impression that night fell when I relaxed. |
| 7 | During the phase of the experience in which I had the task of clearing the fog, I found it easy to maintain slow, deep breathing. |
| 8 | During the phase of the experience in which I had the task of making night fall, I found it easy to relax. |

### Experienced emotions

The Positive and Negative Affect Schedule (PANAS) contains two scales, each comprising 10 items that measure the presence of positive (PANAS-PA) and negative emotions (PANAS-NA) [185]. For each of seven specific moments of the VR experience they had tried, we asked participants to rate, on a 5-point Likert-type scale (1="very little or not at all", 5="a lot"), to what extent each PANAS item described how they felt. A different screenshot was shown for a specific moment to help users recall it: beginning of the experience in the Crystals Archipelago, brightening of the crystals, task of clearing the fog, jamming of the windmill, appearance of fireflies, task of making night fall, final departure of the fireflies. We used the difference between the PANAS-PA score and the PANAS-NA score during the two relaxation tasks of clearing the fog (Task1, hereinafter) and making the night fall (Task2, hereinafter) to assess whether the experience primarily elicited positive emotions.

#### Interview

As a last step, each participant was interviewed. In the first part of the interview, the experimenter showed participants a grid of images containing all the screenshots used with the PANAS questionnaire and invited them to provide any comment they wished to make about the events depicted ("These are the images you saw in the questionnaire. Is there anything you would like to say about the events depicted?"). The second part of the interview focused on exploring the strengths and weaknesses of the experience, following the sequence shown in Figure 4.2.

#### Objective measures

The analysis of physiological data focused on Task1 and Task2. We used the Ledalab SC analysis software [186] to divide SC data into the tonic and phasic components, corresponding to skin conductance level (SCL) and skin conductance response (SCR), respectively. We then applied a Butterworth low-pass filter from the NeuroKit [187] Python library to remove any artifacts. As suggested by [188], we calculated the mean value of SCL and the number of spikes per minute of SCR, considering an amplitude greater than 0.05 $\mu$S. Additionally, we used NeuroKit to compute respiratory rate (RR) from the recorded girth sensor data. As mentioned in Section 4.1.2, the system record physiological data through a Thought Technology ProComp Infiniti encoder using an external appli-

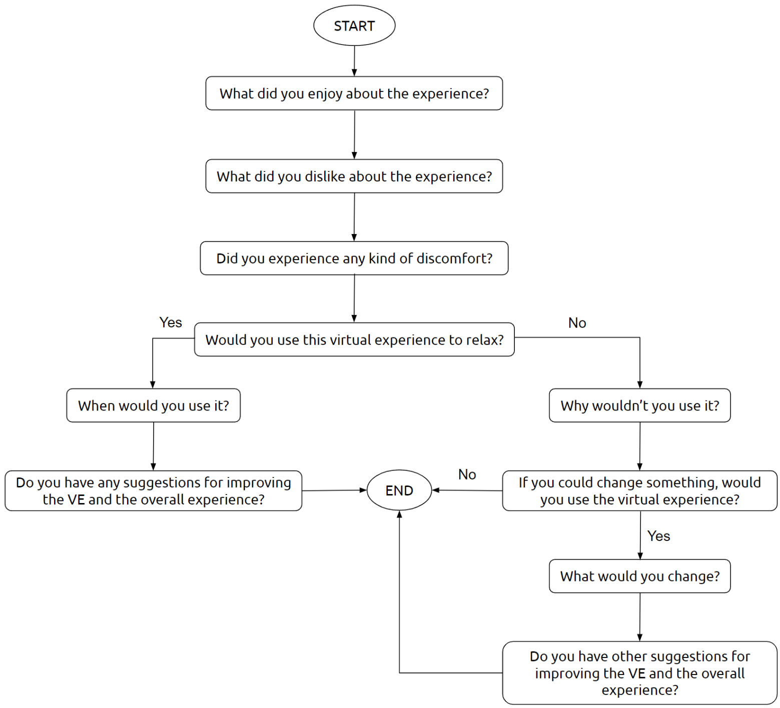



Figure 4.2: Flow diagram of the second part of the interview. The interview was conducted in Italian, all sentences have been translated here into English for reader's convenience.

cation that acquires the data in real-time with a sampling rage of 10 Hz. Since a sampling rate of 500 Hz is recommended, and lower sampling rates can cause inaccuracy of HRV analysis [189], we also recorded cardiac activity at 500 Hz using a blood volume pulse finger clip sensor placed on the distal phalanx of the index finger of the left hand and connected to a BioSignalsPlux encoder. To compute HRV for analysis, we used the OpenSignals software. In particular, we derived the following three highly correlated measures [190]:

1. The square root of the mean squared difference of successive NN intervals (RMSSD).

2. The number of interval differences of successive NN intervals greater than 50 ms (NN50).

3. The proportion derived by dividing NN50 by the total number of NN intervals (pNN50).

### 4.2.4 Procedure

Written consent for participation in the study was obtained from participants and the evaluation of the system was approved by the Institutional Review Board of the University of Udine. The experimenter verbally briefed participants about the anonymity of the collected data and informed them that the VR experience involved using a HMD and physiological sensors. Then, participants sat on a swivel chair and filled the demographic, STAI-S, and STAI-T questionnaires. Participants were balanced in the BIO and SHA groups as described in Section 4.2.2.

Figure 4.3: Virtual living room where baseline of participants' physiological activity was recorded for three minutes.

The experimenter applied physiological sensors to participants, helped them wear an Oculus Quest 2 HMD, and gave them one Quest Touch controller to hold with their right hand.

Participants were asked to choose a comfortable position. Then, they were immersed in a neutral VE representing a living room (Figure 4.3), while the baseline of physiological activity was recorded for three minutes. Afterwards, they tried the Crystals Archipelago experience that lasted 11 minutes.

After the experience, the experimenter helped participants remove the VR HMD and physiological sensors and asked them to fill the following questionnaires: STAI-S, IPQ, biofeedback questionnaire, and PANAS. Finally, they were interviewed and thanked for their participation.

### 4.2.5 Results

All the analyses were conducted using SPSS version 28.0.1.0. Table 4.5 reports means and standard deviation of all self-reported measures for each group.

Table 4.5: Mean and standard deviation of all self-reported measures for each group.

| **Measure** | **BIO M(SD)** | **SHA M(SD)** |
|---|---|---|
| STAI-S before VR experience | **32.31** (6.54) | **31.93** (4.15) |
| STAI-S after VR experience | **24.75** (3.57) | **28.53** (5.71) |
| **Sense of presence** | | |
| IPQ total | **4.16** (0.66) | **3.58** (0.80) |
| Being there | **5.31** (0.87) | **4.60** (1.24) |
| Spatial presence | **4.86** (0.77) | **4.15** (1.33) |
| Involvement | **4.75** (1.07) | **4.38** (1.17) |
| Experienced realism | **2.39** (1.18) | **1.82** (0.89) |
| **Biofeedback questionnaire** | | |
| Perceived biofeedback quality | **5.57** (0.80) | **4.22** (1.15) |
| Ease of performing the relaxation tasks | **4.84** (1.15) | **4.93** (1.05) |
| **PANAS** | | |
| Difference between PANAS-PA and PANAS-NA in Task1 | **17.31** (9.44) | **9.73** (7.94) |
| Difference between PANAS-PA and PANAS-NA in Task2 | **16.50** (12.19) | **13.80** (9.32) |

Table 4.6 reports mean and standard deviation of all objective measures collected during baseline, and during the execution of the two relaxation tasks.

Table 4.6: Mean and standard deviation of all objective measures for baseline, Task1, Task2.

| **Measure** | **Baseline M(SD)** | **Task1 M(SD)** | **Task2 M(SD)** |
|---|---|---|---|
| RR | **16.92** (2.75) | **7.41** (2.04) | **11.64** (4.36) |
| SCL | **8.35** (5.00) | **8.06** (5.70) | **7.50** (5.66) |
| SCR | **3.30** (1.14) | **2.00** (0.59) | **1.81** (0.67) |
| RMSSD | **39.03** (15.73) | **48.94** (25.84) | **43.71** (20.12) |
| NN50 | **42.10** (27.20) | **50.06** (33.19) | **40.61** (29.01) |
| pNN50 | **0.19** (0.13) | **0.22** (0.15) | **0.18** (0.13) |

### Anxiety

STAI-S scores were submitted to a $2 \times 2$ mixed design ANOVA, after checking its assumptions were met, in which group served as the between-

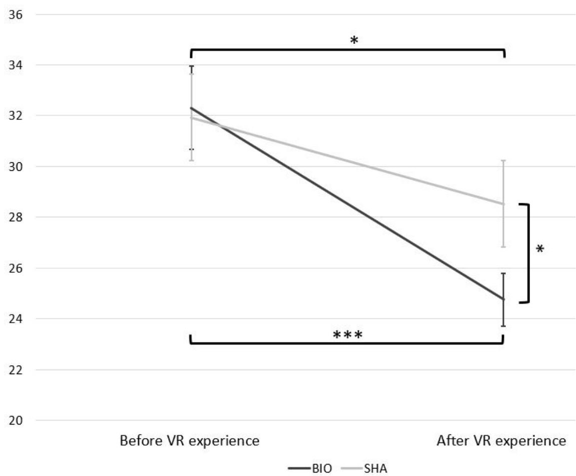



Figure 4.4: Group by time of measurement interaction in STAI-S scores. Capped vertical bars indicate $\pm$ SE. The *, *** signs indicate statistically significant differences with $p$-values respectively $< 0.05$, $< 0.001$.

subject variable, and time of measurement (before and after the VR experience) served as the within-subject variable. Statistically significant results revealed a main effect of time of measurement, $F(1, 29) = 31.46$, $p < 0.001$, $\eta_p^2 = 0.52$, and a group by time of measurement interaction, $F(1.29) = 4.54$, $p < 0.05$, $\eta_p^2 = 0.14$, as illustrated in Figure 4.4.

As suggested in [165], we analyzed each simple effect using Bonferroni correction, considering the effects of time of measurement separately for each group and the effects of group separately at each time of measurement. STAI-S results showed a significant decrease in scores after the experience in both groups (BIO: $p < 0.001$; SHA: $p < 0.05$), and a significant difference in scores between the two groups after the experience ($p < 0.05$).

**Objective measures**

Kolmogorov-Smirnov normality test was performed on RR, SCL, SCR, RMSSD, NN50 and pNN50 data collected during baseline, Task1, and Task2. Since in some cases RMSSD was not normally distributed, that data were subjected to a $\log_{10}$ transformation, as indicated in [165, 191]. Physiological data collected during Task1 and Task2 were submitted to two distinct $2 \times 2$ mixed design ANOVA in which group served as the between-subject variable, and time of measurement (baseline and relaxation task) served as the within-subject variable.
In Task1, statistically significant results revealed a main effect of time of measurement on:

- RR: $F(1, 29) = 240.07$, $p < 0.001$, $\eta_p^2 = 0.89$, Figure 4.5 (a).
- SCR: $F(1, 29) = 59.77$, $p < 0.001$, $\eta_p^2 = 0.67$, Figure 4.5 (b).
- NN50: $F(1, 29) = 4.67$, $p < 0.05$, $\eta_p^2 = 0.14$, Figure 4.5 (c).
- RMSSD: $F(1, 29) = 9.14$, $p < 0.01$, $\eta_p^2 = 0.24$, Figure 4.5 (d).

In Task2, statistically significant results revealed a main effect of time of measurement on:

- RR: $F(1,29) = 53.44, p < 0.001$, $\eta_p^2 = 0.65$, Figure 4.6 (a).
- SCR: $F(1,29) = 46.60$, $p < 0.001$, $\eta_p^2 = 0.62$, Figure 4.6 (b).
- SCL: $F(1,29) = 4.49$, $p < 0.05$, $\eta_p^2 = 0.13$, Figure 4.6 (c).
- RMSSD: $F(1,29) = 4.29$, $p < 0.05$, $\eta_p^2 = 0.13$, Figure 4.6 (d).

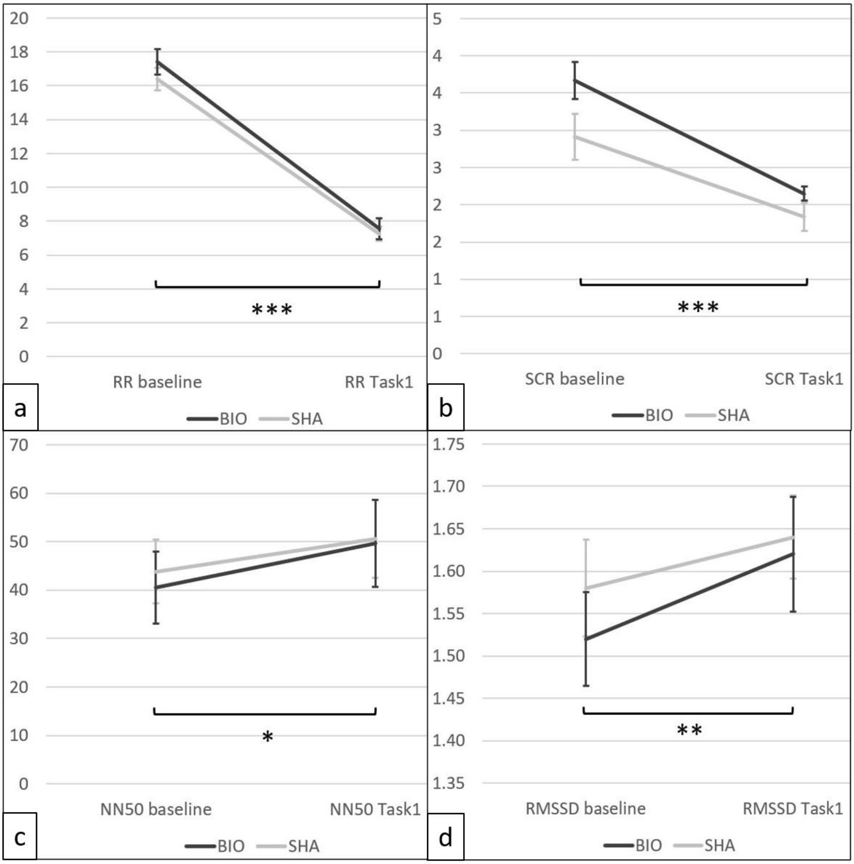



Figure 4.5: Group by time of measurement in objective measures with statistically significant results in Task1. Capped vertical bars indicate $\pm$ SE. The *, **, and *** signs indicate statistically significant differences with $p$-values respectively $< 0.05$, $< 0.01$, and $< 0.001$. (a) RR (a); (b) SCR; (c) NN50; (d) RMSSD.

Moreover, a group by time of measurement interaction was found for RMSSD, $F(1,29) = 4,43$, $p < 0.05$, $\eta_p^2 = 0.13$. The analysis of simple effects with Bonferroni correction revealed a significant difference between baseline and Task2 but only in the BIO group ($p < 0.01$) while no statistically significant differences were found between groups (Figure 4.6 (f)).

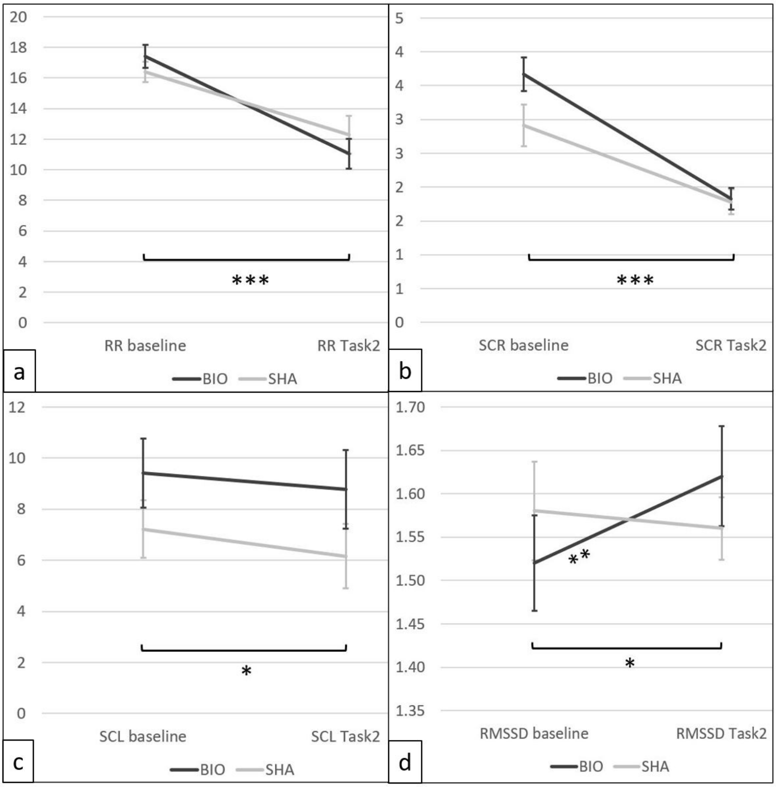
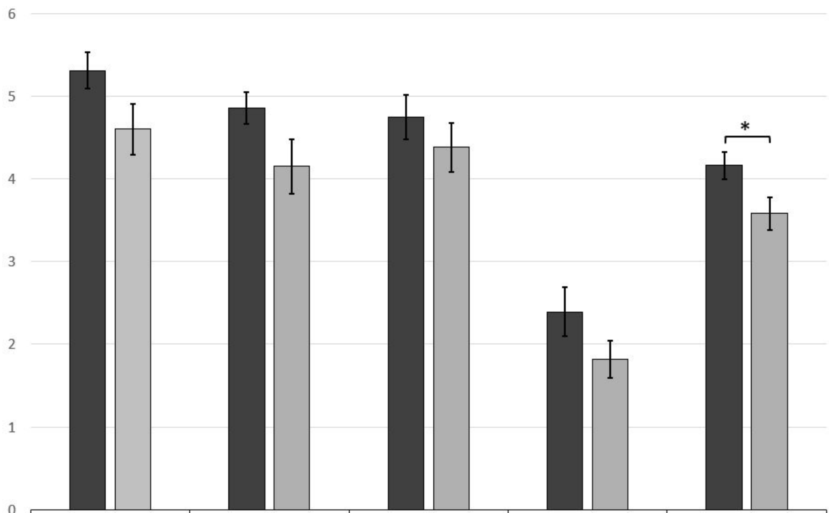



Figure 4.6: Group by time of measurement in objective measures with statistically significant results in Task2. Capped vertical bars indicate ± SE. The *, **, and *** signs indicate statistically significant differences with $p$-values respectively $< 0.05$, $< 0.01$, and $< 0.001$. (a) RR; (b) SCR; (c) SCL; (d) RMSSD; the ** sign on the BIO group line indicates a significant difference between baseline and Task2.

**Sense of presence**

An unrelated $t$-test was used to compare sense of presence in the two groups (Figure 4.7). The BIO group had higher IPQ total values than the SHA group. The difference between means was significant in the total IPQ score, $t(29) = 2.2$, $p < 0.05$, two-tailed, Cohen's $d = 0.79$. No statistically significant differences were found for the subscales.




| | Being there | Spatial presence | Involvement | Experienced realism | IPQ Total |
|---|---|---|---|---|---|
| BIO | 5.31 | 4.86 | 4.75 | 2.39 | 4.16 |
| SHA | 4.60 | 4.15 | 4.38 | 1.82 | 3.58 |

Figure 4.7: Means of sense of presence. Capped vertical bars indicate ± SE. The * sign indicates statistically significant difference with $p$-value $< 0.05$.

**Perceived biofeedback quality and ease of tasks**

A Shapiro-Wilk normality test on the two scales of the biofeedback questionnaire data indicated that data followed a Gaussian distribution. An unrelated $t$-test performed on the scale "perceived biofeedback quality" and the scale "ease of performing the relaxation tasks" revealed that the BIO group perceived a higher biofeedback quality than the SHA group, $t(29) = 0.21$, $p < 0.001$, two-tailed, Cohen's $d = 1.37$ (Figure 4.8). No

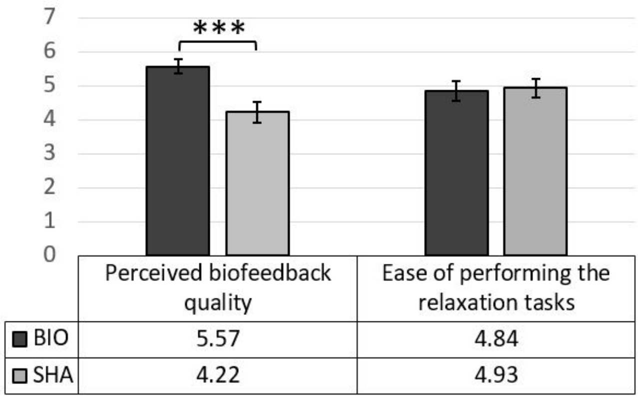



Figure 4.8: Means of scores of the biofeedback questionnaire. Capped vertical bars indicate ± SE.

statistically significant differences were found between the groups on the ease of performing the relaxation tasks.

**Experienced emotions**

The mean scores of PANAS-PA felt by participants during the seven specific moments of the VR experience were higher than the mean scores of PANAS-NA. Negative emotions were generally higher in the SHA group than in the BIO group. All specific moments but one were shown to all participants during the experience: the exception is the windmill jamming that could occur from the second phase until the end of the fourth phase only if the recording of physiological activity indicated fast and shallow breathing (BIO: n=7; SHA: n=4).

An unrelated *t*-test on the difference between PANAS-PA and PANAS-NA during the two tasks of relaxation revealed that in Task1 the BIO group and the SHA group were significantly different, $t(29) = 2.41$, $p < 0.05$, two-tailed, Cohen's $d = 0.87$ (Figure 4.9). No statistically significant differences were found for Task2.

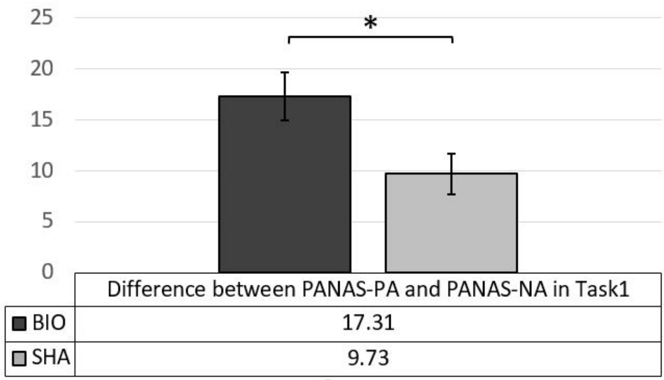



Figure 4.9: Difference between PANAS-PA and PANAS-NA in Task1. Capped vertical bars indicate ± SE. The * sign indicates statistically significant difference with $p$-value $< 0.05$.

### Interview

During the first part of the interview, seven participants stated that they particularly enjoyed the fireflies. Three of them reported they felt sad to see the fireflies fly away at the end of the experience. One participant in the BIO group reported experiencing frustration and insecurity while performing the two relaxation tasks: although he thought he was following instructions well, he felt that the VE was not changing as he expected and had the impression that the VE was reflecting his own insecurity.

In the second part of the interview, participants expressed appreciation for the following aspects of the experience: scenery (BIO: n=12; SHA: n=12), the two relaxation tasks (BIO: n=8; SHA: n=5), colors (BIO: n=1; SHA: n=5), fireflies (SHA: n=4), storytelling (BIO: n=1; SHA: n=2) and virtual hand (BIO: n=2).

The aspects of the experience that participants did not appreciate were instead: a feeling of not controlling the environment (SHA: n=4), a lack of interactivity (SHA: n=3), low realism of the VE (BIO: n=2; SHA: n=1), the slow pace of Task1 (BIO: n=3; SHA: n=1) or of the whole experience (BIO: n=1; SHA: n=1), limited use of the hand controller (BIO: n=1), and having only one hand in the VE instead of a full embodiment (BIO: n=1;

SHA: n=1). Some participants experienced discomfort when trying slow, deep diaphragmatic breathing because they found it difficult (BIO: n=1) or because they did not understand whether they were performing it well (BIO: n=1; SHA: n=3).

Twenty-six participants would use the VR experience as a tool to relax (BIO: n=15; SHA: n=11). In particular, some of them would use it in the evening (BIO: n=8; SHA: n=7), to calm down and not think about problems (BIO: n=1; SHA: n=3), or before an exam or a deadline (BIO: n=1; SHA: n=1). Three participants said they would not use it (BIO: n=1; SHA: n=2), while two participants of the SHA group would use it only after some changes that is a faster experience (n=1) or a multiplayer mode (n=1).

The most frequently mentioned suggestions for improvement were the following: allow moving around the VE (BIO: n=1; SHA: n=2), add more interactions (SHA: n=3), decrease the duration of the experience (SHA: n=2), and change the appearance of the bush (BIO: n=1; SHA: n=1).

### 4.2.6 Discussion

Results have confirmed our hypothesis that the VR experience relaxes users (H1). The analysis of the STAI-S scores revealed that participants significantly enhanced their level of relaxation after the VR experience. This is aligned with the results on [178, 192], where spending time in nature had positive effects on stress levels. Similarly, [193, 194] found that natural VEs induced relaxation and restore attentional resources. The presence of “clouds, sunsets, and leaves moving in a breeze” in the VR experience may have contributed to participants’ relaxation. Indeed, according to [192], such natural elements can stimulate fascination in individuals, thus improving the effectiveness of restorative experiences. The significantly improved level of relaxation in participants after the VR experience was also found in results obtained on objective measures: RR and SCR significantly decreased, and RMSSD and NN50 significantly increased during Task1, while RR, SCR, and SCL significantly decreased, and RMSSD significantly increased during Task2. These results obtained on breathing, cardiac, and electrodermal parameters are in line with those found in previous studies, where the exposure to VR systems with biofeedback for relaxation training significantly increased HRV parameters [24, 30] or significantly decreased in HR [30, 99]. Since diaphragmatic breathing stimulates relaxation [195, 196], diaphragmatic breathing taught to participants during Task1 may have enhanced the overall level of relaxation achieved by them. Previous studies also showed that diaphragmatic breathing improves mood [15], reducing stress and anxiety levels [15, 16, 196]. In our study, STAI-S scores highlighted a significant reduction in anxiety levels after the VR experience. This is in line with results found in previous studies on VR systems with biofeedback for relaxation training [24, 27, 99, 116], where the comparison of participants' self-reported state anxiety before and after exposure to the system resulted in a significant decrease in self-reported state anxiety. Finally, since in the final interview participants reported that they would use the system to relax (n=26), we can conclude that they perceived the relaxation effect induced by the VR experience, and this motivated their willingness to use the system again.

Results have confirmed our hypothesis that biofeedback enhances the relaxation effect of the VR experience (H2). More precisely, the analysis of the STAI-S scores revealed that the BIO group achieved a significantly higher level of relaxation than the SHA group after the VR experience. Moreover, the analysis of the objective measures found that RMSSD significantly increased in the BIO group but not in the SHA group during Task2. When evaluating biofeedback systems for relaxation training, comparing real biofeedback with a placebo condition is necessary to determine the role actually played by biofeedback. However, in the literature, only two previous studies have evaluated a VR system with biofeedback for relaxation training by comparing it with a sham biofeedback [103, 106]. One of the studies used non-immersive VR and focused on comparing two algorithms for stress detection, only one of which performed significantly better than sham biofeedback [103]. The study did not analyze relaxation effects and performed statistical analysis only on a questionnaire that assessed the biofeedback quality perceived by participants. In our study, the biofeedback quality perceived by participants was positively influenced by real biofeedback. This finding can be attributed to the selection of the visual and auditory feedback used in the VR experience. Such feedback facilitated participants in the BIO group to perceive the effects on their physiological activity in the VE, which is important for reinforcement learning [197]. Ease of performing the relaxation tasks was instead perceived similarly by both groups, which achieved relatively high values. The second study that considered a placebo condition focused on relaxation [106], and concluded that sham biofeedback was better than real biofeedback, resulting in a lower HR. The disparity between this finding and ours could be explained by the extreme differences between their system and our system. In [106], the system consisted of an empty immersive VE that only displayed a cloud whose movements towards and away from the user corresponded to the user's breathing, whereas in the placebo condition, the cloud moved back and forth every three seconds. In contrast, as illustrated in detail in Section 4.1, our system is a much more complex, natural VE where rich biofeedback influences different elements of the VE through various mechanisms and using multiple physiological parameters. While simple forms of biofeedback stimuli may carry the risk of becoming monotonous [198, 199], this risk can be reduced by feedback displayed in various changes to a natural VE, which could improve motivation and reduce distractions [30]. Moreover, proceeding through the various steps of the narrative that our system provides could keep user's curiosity and attention alive, and also be rewarding. In particular, our system associated Task1 and Task2 with two different rewards: achieving a clear sky and a beautiful nighttime setting, respectively. Overall, the VR experience elicited positive emotions in participants, as shown by the PANAS results. Although the windmill jamming was an element of the experience designed as negative feedback, the PANAS results found that seeing the windmill jamming during the VR experience (n=11) made participants feel mostly attentive and alert. This finding is in line with the operant conditioning paradigm [108]: to prevent the occurrence of the aversive feedback (the windmill jamming), participants might have improved their concentration to better perform the recommended behavior (slow breathing in this case). The difference between PANAS-PA and PANAS-NA experienced during Task1 was significantly greater for the BIO group than the SHA group. Since breathing can be consciously controlled and was used to perform the task of clearing the fog, participants in the SHA group were likely less helped because they could not see stimuli in the VE that fully corresponded to their real breathing activity. This may have provoked greater negative emotions and fewer positive emotions compared to the BIO group. Indeed, the final interview highlighted that sham biofeedback decreased the degree of satisfaction with the VR experience because all participants who expressed dissatisfaction with poor interaction with the VR environment (n=3) or felt unable to control the VR environment (n=4) belonged to the SHA group. Additionally, the two participants who would have preferred a shorter duration of the experience also belonged to the SHA group. This supports the possibility that the SHA group experienced more feelings of boredom than the BIO group, and this difference is due to the absence of real biofeedback that allows for greater motivation and focus during the VR experience [26]. In addition, three out of four participants who said they had difficulty understanding if they were performing the task correctly belonged to the SHA group. This is consistent with the role of biofeedback in helping participants improve their sense of control over the VR environment and their physiological parameters.

Results have confirmed our hypothesis that biofeedback increases sense of presence in the VR experience (H3). We found a statistically significant result regarding sense of presence, where the IPQ total score was higher with real biofeedback than with sham biofeedback. In both groups, sense of presence scores were high (between 4 and 6 in a 0 to 6 scale) for the sense of being there as well as for spatial presence and involvement subscales. Only realism scores were low in the two conditions, but this was expected, given the design choice to use stylized fantasy graphics. The present study extends previous results on biofeedback and presence by considering an immersive VR placebo condition. A previous study found that an immersive VR system with biofeedback obtained higher sense of presence than the same system without biofeedback in both immersive and desktop VR, but the difference in sense of presence felt by participants with the three versions of the system was not statistically significant [32]. In our study, the higher sense of presence felt by the BIO group than the SHA group may have also influenced the result obtained with the PANAS, where the difference between PANAS-PA and PANAS-NA experienced was significantly higher in the BIO group than the SHA group. Since the BIO group experienced a greater sense of presence, and affective state can be influenced by presence [200], this may have led to a greater emotional impact on the BIO group than the SHA group.

In summary, our study showed that the VR experience improved the level of relaxation in both groups. Specifically, real biofeedback produced better results than sham biofeedback in terms of both relaxation and sense of presence. These results highlight the importance of the role that biofeedback can play in VR systems for relaxation training to increase sense of presence felt by users and to achieve higher levels of relaxation than those that would be obtained using the same system without biofeedback.

However, it should be noted that our study was not focused on identify which specific aspect of the VR experience had the greatest impact on our outcomes. Further research should investigate the individual contributions of the visual and auditory elements of the VR experience, the two relaxation tasks, and the role of storytelling, to better understand how each of these aspects influenced the relaxation effects experienced by participants.

Furthermore, our study did not assess the duration of the relaxation effect after the VR experience and did not explore its potential longitudinal effects. A previous study has shown that users can improve their diaphragmatic breathing over time, thereby achieving deeper relaxation [26]. Additionally, Bossenbroek et al. [115] found that the use of the VR system with biofeedback described in [27] induced relaxation in a clinical sample that persisted for approximately two hours. Future studies should therefore extend our single-session VR experience to a study with multiple training sessions to investigate the duration of relaxation effects and other possible long-term effects.

## 4.3 Study 2: desktop vs. immersive VR

After having assessed with the study described in Section 4.2 that the system relaxes and biofeedback enhances the level of relaxation, we want to compare our system in immersive versus desktop VR. In this study, we evaluated a sample of healthy individuals who experienced our system with biofeedback in desktop VR. We then compared the results of this study with those collected in the previous study by the BIO group who tested the system with biofeedback in immersive VR. From now on, we will use the term DSK group to refer to the group of participants who tried the system in desktop VR, and the term IMM group to refer to the group of participants who tried the system in immersive VR.

### 4.3.1 Hypotheses

We formulated the following hypotheses:

H1. The desktop version of the VR experience relaxes users because the previous study, as described in Section 4.2, showed that the natural scenery of the designed VE relaxes users when experienced in immersive VR. Furthermore, previous studies have shown that videos featuring natural settings experienced through desktop monitors can have a significant effect on users' relaxation by improving their mood [201], increasing positive emotions [202], and reducing stress [203]. Thus, we expect the VR experience relax users also when tried in desktop VR. As shown in Table 1.3, four desktop VR systems in the literature use biofeedback mechanisms, but studies conducted on these systems have not analyzed the effects of relaxation [103–105]. Only in [101] the HRV of participants was analyzed, with the RMSSD indicating an improvement in relaxation levels during system usage (see Table 1.4).

H2. The immersive version of the VR experience increases sense of presence in the VE because immersive VR systems induce a greater sense of presence than desktop VR [181, 183].

H3. As the sense of presence increases, the level of relaxation increases because in VRE systems designed to induce anxiety in users, the sense of presence is positively correlated with perceived anxiety [204]. Therefore, in our system for relaxation training, a higher sense of presence should correspond to a higher level of relaxation.

H4. Immersive VR relaxes more than desktop VR because the greater sense of presence perceived in immersive VR leads to achieve higher levels of relaxation than in desktop VR. The heightened sense of presence perceived in immersive VR should make the participants feel engaged, enabling them to be more focused on Task1 and Task2, and this can improve their relaxation. In [32], a VR system with biofeedback in immersive VR achieved greater relaxation levels than the same system without biofeedback in desktop VR. Nevertheless, no study has yet explored the variations in relaxation achieved through a VR system with biofeedback when comparing the same system in both desktop VR and immersive VR.

### 4.3.2 Participants

The study involved a sample of 16 participants (10 males, 6 females), who were volunteers that received no compensation and were recruited through direct contact. They were undergraduate students from different faculties or workers who were former university students. Their age ranged from 22 to 29 (M=26.19, SD=2.37).

We asked participants if they regularly played first-person video games: five participants answered positively, declaring an average of 2.75 hours of gaming per week (SD=2.31). All other participants reported playing first-person video games for an average of 1.96 hours of gaming per week (SD=0.93).

All 16 participants were assigned to the DSK group. We compared the results obtained by the DSK group with those obtained by the IMM group, previously identified as the BIO group in Section 4.2.

We confirmed the lack of significant differences between the DSK and IMM groups through a Pearson Chi-square test on gender and an unrelated *t*-test for age, state anxiety scores, and trait anxiety scores (Table 4.7).

Table 4.7: Means and standard deviations of age, state anxiety, and trait anxiety in the DSK and IMM groups.

| **Measure** | **DSK M(SD)** | **IMM M(SD)** |
|---|---|---|
| Age | **26.19** (2.37) | **26.44** (7.59) |
| State anxiety | **33.13** (9.37) | **32.31** (6.54) |
| Trait anxiety | **39.81** (8.44) | **39.56** (10.09) |

### 4.3.3 Measures

In this study we employed the same subjective and objective measures used in the previous study, as described in Section 4.2.3. Questionnaires STAI-S, IPQ, biofeedback questionnaire, and PANAS were administered to participants through the PsyToolkit tool [161, 162].

**Desktop VR system with biofeedback**

The VR experience in desktop VR is identical to the original experience in immersive VR except for two aspects. First, the user sees the VR experience through a 17-inches desktop monitor and moves his/her view using a mouse. The user can perform a 360° horizontal rotation and a 45° vertical rotation upward and downward to reproduce the same movements that he/she could perform with an HMD. Second, since in desktop VR the user does not have a controller, during the third phase of the VR experience the user can interact with the bush by pressing a button that shows up in that specific phase (Figure 4.10 (a)). Once pressed the button, it disappears, and a virtual hand, identical to the one employed in immersive VR, appears and touches the bush, thus fireflies come out of it and slowly fly near the user (Figure 4.10 (b)). If the user does not press the button within 20 seconds, the virtual hand automatically appears, touches the bush, releasing fireflies to allow the user to move on to the next phase.

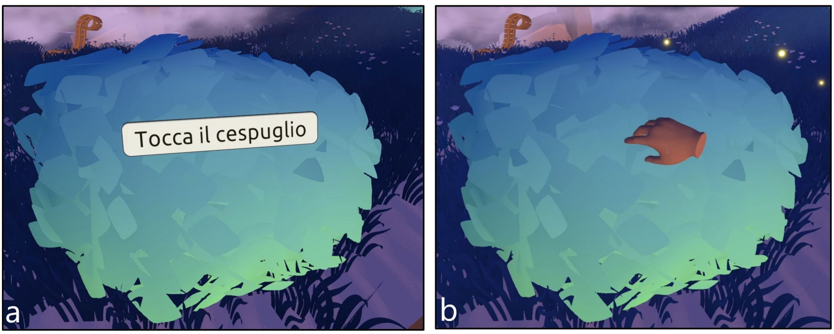



Figure 4.10: Third phase of the VR experience in desktop VR. (a) The button appears, inviting the user to press it; (b) the virtual hand touches the bush and fireflies come out from it.

### 4.3.4 Procedure

Written consent for participation in the study was obtained from participants and the evaluation of the system was approved by the Institutional Review Board of the University of Udine. The experimenter verbally briefed participants about the anonymity of the collected data and informed them that the VR experience involved using a desktop monitor and physiological sensors. Then, participants sat on a chair and filled the demographic, STAI-S, and STAI-T questionnaires. The experimenter applied physiological sensors to participants, and provided them with a mouse to use with their right hand (all participants were right-handed) to control the viewpoint within the VE. Participants were asked to choose a comfortable position. Then, they watched the same virtual living room used in the study described in Section 4.2, using a 17-inches desktop monitor positioned in front of them. The baseline of their physiological activity was recorded for three minutes. Afterwards, they tried the Crystals Archipelago experience that lasted 11 minutes.

After the experience, the experimenter helped participants remove the physiological sensors and asked them to fill the following questionnaires: STAI-S, IPQ, biofeedback questionnaire, and PANAS. Finally, they were interviewed and thanked for their participation.

### 4.3.5 Results

All the analyses were conducted using SPSS version 29.0.0.0. Table 4.8 reports means and standard deviations of all self-reported measures for each group. Table 4.9 reports means and standard deviations of all objective measures collected during baseline, and during the execution of the two relaxation tasks.

#### Anxiety

STAI-S scores were submitted to a $2 \times 2$ mixed design ANOVA, after checking its assumptions were met, in which group (IMM and DSK)

Table 4.8: Means and standard deviations of all self-reported measures for each group.

| Measure | DSK M(SD) | IMM M(SD) |
|---|---|---|
| STAI-S before VR experience | **33.13** (9.37) | **32.31** (6.54) |
| STAI-S after VR experience | **29.13** (5.21) | **24.75** (3.57) |
| **Sense of presence** | | |
| IPQ total | **3.07** (1.12) | **4.16** (0.66) |
| Being there | **3.69** (1.58) | **5.31** (0.87) |
| Spatial presence | **3.64** (1.39) | **4.86** (0.77) |
| Involvement | **3.52** (1.20) | **4.75** (1.07) |
| Experienced realism | **1.77** (1.22) | **2.39** (1.18) |
| **Biofeedback questionnaire** | | |
| Perceived biofeedback quality | **5.41** (1.22) | **5.57** (0.80) |
| Ease of performing the relaxation tasks | **4.84** (1.14) | **4.84** (1.15) |
| **PANAS** | | |
| Difference between PANAS-PA and PANAS-NA | **10.85** (6.27) | **15.69** (6.54) |

Table 4.9: Mean and standard deviation of all objective measures for baseline, Task1, Task2.

| Measure | Baseline M(SD) | Task1 M(SD) | Task2 M(SD) |
|---|---|---|---|
| RR | **17.04** (3.11) | **7.53** (2.12) | **10.53** (3.53) |
| SCL | **8.53** (4.61) | **9.65** (5.62) | **8.23** (5.25) |
| SCR | **3.47** (1.11) | **2.99** (1.24) | **2.18** (1.02) |
| RMSSD | **37.28** (17.52) | **52.69** (27.83) | **43.38** (21.96) |
| NN50 | **38.53** (29.04) | **50.69** (31.78) | **37.16** (30.26) |
| pNN50 | **0.18** (0.16) | **0.23** (0.16) | **0.17** (0.14) |

served as the between-subject variable, and time of measurement (before and after the VR experience) served as the within-subject variable. Statistically significant results revealed a main effect of time of measurement, $F(1,30) = 29.82$, $p < 0.001$, $\eta_p^2 = 0.50$, as illustrated in Figure 4.11.

As suggested in [165], we analyzed each simple effect using Bonferroni correction, considering the effects of time of measurement separately for each group and the effects of group separately at each time of measurement. STAI-S results showed a significant decrease in scores after the experience in both groups (DSK: $p < 0.05$; IMM: $p < 0.001$), and a significant difference in scores between the two groups after the experience ($p < 0.05$).

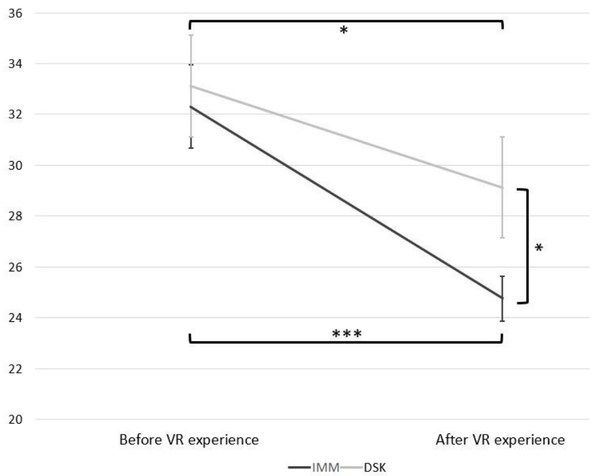



Figure 4.11: Means of STAI-S scores before and after the VR experience. Capped vertical bars indicate ± SE. The *, *** signs indicate statistically significant differences with *p*-values respectively $< 0.05$, $< 0.001$.

### Objective measures

Kolmogorov-Smirnov normality test was performed on RR, SCL, SCR, RMSSD, NN50, and pNN50 data collected during baseline, Task1, and Task2. Since in some cases RMSSD and pNN50 were not normally distributed, that data were subjected to a $\log_{10}$ transformation, as indicated in [165, 191]. Physiological data collected during Task1 and Task2 were submitted to two distinct $2 \times 2$ mixed design ANOVA in which group served as the between-subject variable, and time of measurement (baseline and relaxation task) served as the within-subject variable.

In Task1, statistically significant results revealed a main effect of time of measurement on:

- RR: $F(1,30) = 166.44$, $p < 0.001$, $\eta_p^2 = 0.85$, Figure 4.12 (a).
- SCL: $F(1,30) = 5.28$, $p < 0.05$, $\eta_p^2 = 0.15$, Figure 4.12 (b).
- RMSSD: $F(1,30) = 20.71$, $p < 0.001$, $\eta_p^2 = 0.41$, Figure 4.12 (c).

- NN50: $F(1,30) = 6.12$. $p < 0.05$, $\eta_p^2 = 0.17$, Figure 4.12 (d).

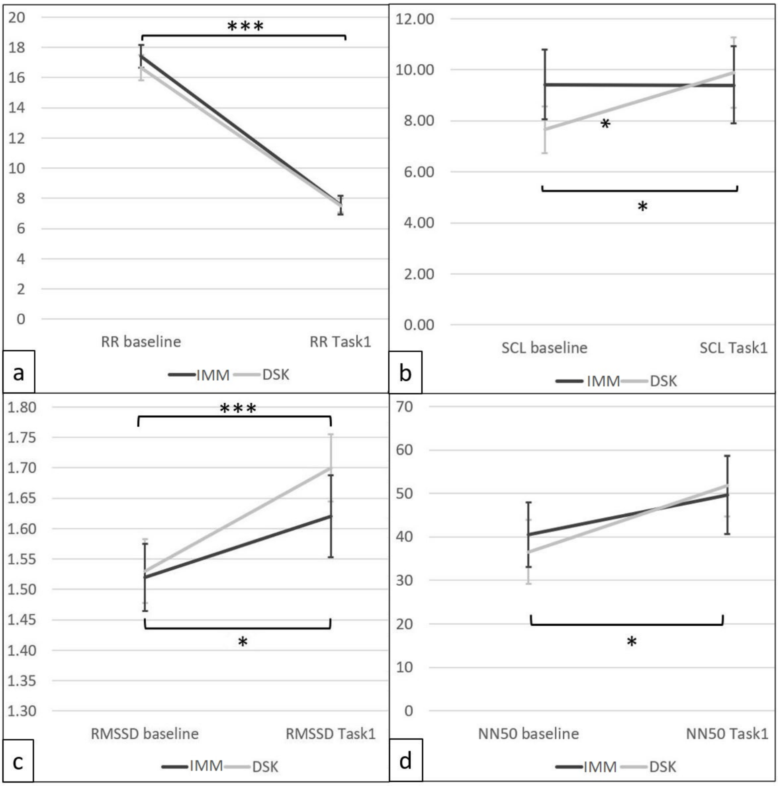



Figure 4.12: Group by time of measurement in objective measures with statistically significant results in Task1. Capped vertical bars indicate ± SE. The * and *** signs indicate statistically significant differences with $p$-values respectively $< 0.05$, and $< 0.001$. (a) RR; (b) SCL; the * sign on the DSK group line indicates a significant difference between baseline and Task1; (c) RMSSD; (d) NN50.

In Task2, statistically significant results revealed a main effect of time of measurement on:

- RR: $F(1,30) = 677.69$, $p < 0.001$, $\eta_p^2 = 0.72$, Figure 4.13 (a).
- SCR: $F(1,30) = 39.33$, $p < 0.001$, $\eta_p^2 = 0.57$, Figure 4.13 (b).

- RMSSD: $F(1,30) = 4.17$, $p = 0.05$, $\eta_p^2 = 0.12$, Figure 4.13 (c).

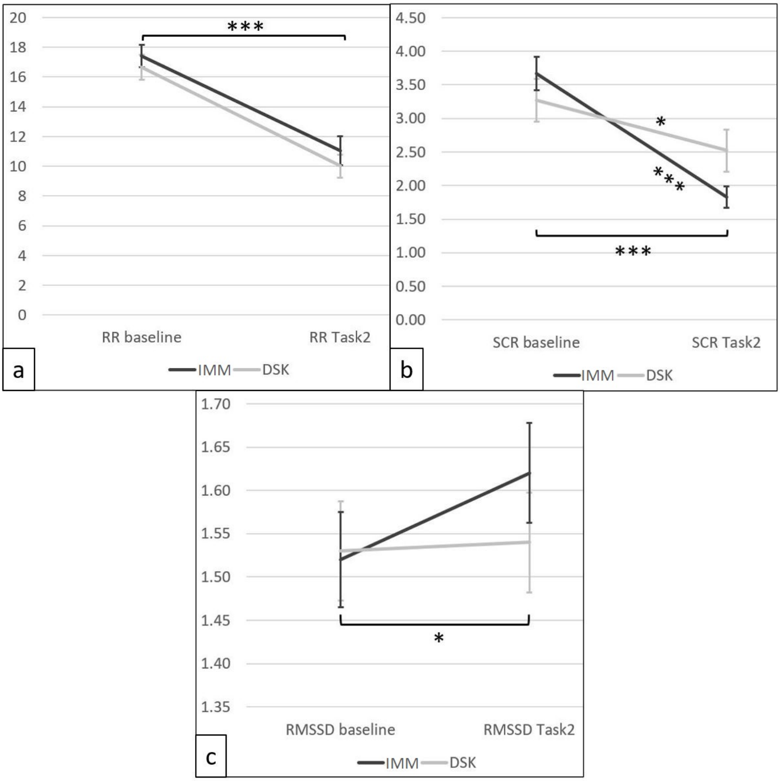



Figure 4.13: Group by time of measurement in objective measures with statistically significant results in Task2. Capped vertical bars indicate ± SE. The * and *** signs indicate statistically significant differences with $p$-values respectively $< 0.05$, and $< 0.001$. (a) RR; (b) SCR; the * on the DSK group line and the sign *** on the IMM group line indicate a significant difference between baseline and Task2; (c) RMSSD.

Moreover, the following group by time of measurement interactions were found:

- SCL: $F(1,30) = 5.35$, $p < 0.05$, $\eta_p^2 = 0.15$. The analysis of simple effects with Bonferroni correction revealed a significant difference between baseline and Task 1 but only in the DSK group ($p < 0.01$)

while no statistically significant differences were found between groups (Figure 4.12).

- SCR: $F(1, 30) = 6.93$, $p < 0.05$, $\eta_p^2 = 0.19$. The analysis of simple effects with Bonferroni correction revealed a significant difference between baseline and Task 2 in both groups (IMM:$p < 0.001$; DSK:$p < 0.05$). Moreover, in Task2 the difference between groups was close to significance ($p = 0.056$) (Figure 4.13).

#### Sense of presence

An unrelated $t$-test was used to compare sense of presence in the two groups (Figure 4.14).

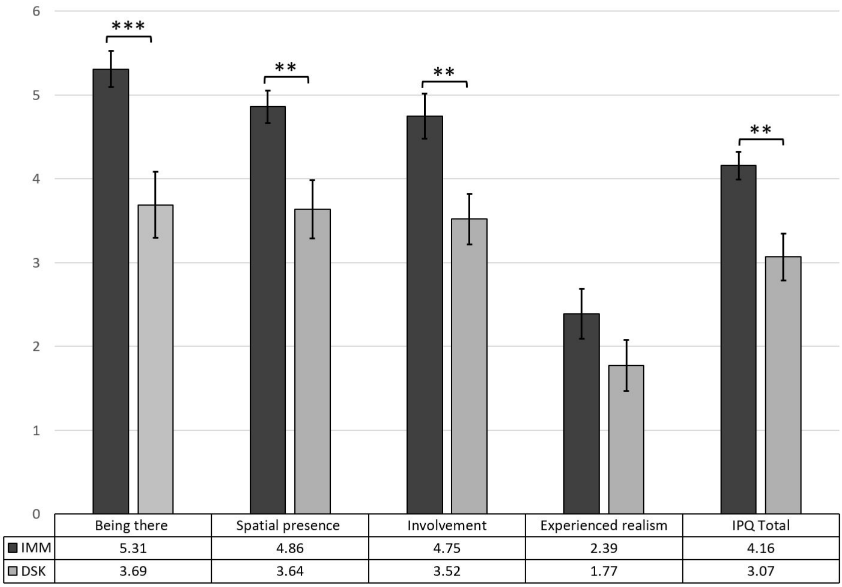



Figure 4.14: Means of sense of presence. Capped vertical bars indicate ± SE. The **, *** signs indicate statistically significant differences with $p$-values respectively $< 0.01$, $= 0.001$.

The analysis revealed the following statistically significant differences between the two groups:

- Being there: $t(30) = 3.60$, $p = 0.001$, two-tailed, Cohen's $d = 1.27$.
- Spatial presence: $t(30) = 3.09$, $p < 0.01$, two-tailed, Cohen's $d = 1.09$.
- Involvement: $t(30) = 3.07$, $p < 0.01$, two-tailed, Cohen's $d = 1.09$.
- IPQ total score: $t(30) = 3.35$, $p < 0.01$, two-tailed, Cohen's $d = 1.19$.

A Pearson correlation was computed to assess possible relationships between sense of presence and the difference between STAI-S score measured before the VR experience and the STAI-S score measured after the VR experience. The statistically significant correlations found in the DSK group are shown in Table 4.10, while no statistically significant correlations were found in the IMM group.

Table 4.10: Statistically significant correlations between sense of presence and the difference between STAI-S score measured before the VR experience and the STAI-S score measured after the VR experience.

| **IPQ scale** | $r$ | $n$ | $p$ |
|---|---|---|---|
| Being there | 0.57 | 16 | $< 0.05$ |
| Spatial presence | 0.65 | 16 | $< 0.01$ |
| Involvement | 0.72 | 16 | $< 0.01$ |
| Experienced realism | 0.72 | 16 | $< 0.01$ |
| IPQ total | 0.79 | 16 | $< 0.001$ |

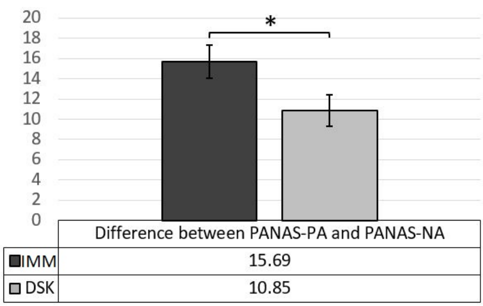
**Perceived biofeedback quality and ease of tasks**

A Shapiro-Wilk normality test on the two scales of the biofeedback questionnaire data indicated that data followed a Gaussian distribution. An unrelated $t$-test performed on the scale "perceived biofeedback quality" and the scale "ease of performing the relaxation tasks" revealed that the IMM group and the DSK group perceived similar biofeedback quality. Moreover no statistically significant differences were found between the groups on the ease of performing the relaxation tasks.

### Experienced emotions

The mean scores of PANAS-PA felt by participants during the seven specific moments of the VR experience were higher than the mean of scores of PANAS-NA. Positive emotions were higher in the IMM group than in the DSK group, while negative emotions were similar in the two groups. All specific moments but one were shown to all participants during the experience: the exception is the windmill jamming that could occur from the second phase until the end of the fourth phase only if the recording of physiological activity indicated fast and shallow breathing (IMM: n=7; DSK: n=7).

An unrelated $t$-test on the difference between PANAS-PA and PANAS-NA of the sum of all seven specific moments of the VR experience revealed that the IMM group and the DSK group were significantly different, $t(30) = 2.14$, $p < 0.05$, two-tailed, Cohen's $d = 0.76$ (Figure 4.15). No statistically significant differences were found for the difference between PANAS-PA and PANAS-NA during the two tasks of relaxations Task1 and Task2.




.

Figure 4.15: Difference between PANAS-PA and PANAS-NA. Capped vertical bars indicate $\pm$ SE. The * sign indicates statistically significant difference with $p$-value $< 0.05$.

### Interview

During the first part of the interview, five participants mentioned that they relaxed. Ten participants expressed their enjoyment of various aspects of the scenery, including light trail, crystals, movement of the leaves, sea view, and night. Additionally, five participants particularly appreciated the fireflies.

During the second part of the interview, participants expressed appreciation for the following aspects of the experience: scenery (IMM: n=6; DSK: n=5), relaxation tasks (IMM: n=3; DSK: n=6), synchronization with the breath (IMM: n=5; DSK: n=3), the system's ability to induce relaxation (IMM: n=4; DSK: n=3) and isolate from the real world (IMM: n=2; DSK: n=3), fireflies (DSK: n=5), windmill (IMM: n=1; DKS: n=4), voice-over (IMM: n=1; DSK: n=3), background music (DSK: n=2), and sound of the wind (DSK: n=2).

In contrast, the aspects of the experience that participants did not like were: the slow pace of Task 1 (IMM: n=4, DSK: n=3) or both relaxation tasks (IMM: n=1; DSK: n=2), the graphical style of the VE (IMM: n=1; DSK: n=2), navigation method using the mouse (DSK: n=2), the windmill blades (DSK: n=1), Task2 (DSK: n=1), and little interaction (IMM: n=1). In addition, one participant stated that she felt distressed when she saw the fireflies flying away at the end of the experience because it signified she had to return to the real world. Some participants experienced discomfort due to the length of the experience (IMM: n=1; DSK: n=2) because they found it difficult to remain relaxed for the required duration, or when trying to clear the fog (DSK: n=2).

Out of 32 participants (IMM: n=16, DSK: n=16), 26 would use the system as a tool to relax. Specifically, some of them would use it before bedtime (IMM: n=7; DSK: n=7), when they want to calm down (IMM: n=2; DSK: n=5) and not think about problems (IMM: n=1; DSK: n=1), when they are stressed (IMM: n=1; DSK: n=1) to lower their heart rate (DSK: n=1), or after work to relieve tension (DSK: n=1). Four participants said they would not use the system (IMM: n=1; DKS: n=3), while two participants of the DSK group would consider using it only after some modification that would be to make the scenario more realistic.

Participants suggested the following improvements: changing the appearance of the bush (IMM: n=1; DSK: n=1), allowing users to move around the VE (IMM: n=1; DSK: n=1), adding more tasks (DSK: n=1), and introducing a virtual animal companion that sleeps or moves around based on the user's level of relaxation (n=1).

### 4.3.6 Discussion

The results confirmed our hypothesis H1 that the VR experience relaxes users when tried in desktop VR. Analysis of STAI-S scores revealed that participants in the DSK group significantly improved their level of relaxation after the VR experience. This is in line with the findings of our previous study, discussed in Section 4.2.6, which showed the system's capability to relax participants through the representation of virtual natural scenarios experienced through immersive VR. Previous studies on desktop VR systems with biofeedback (see Table 1.4 for details) have not analyzed the level of relaxation achieved [103–105]. Only Sonne and Jensen [101] reported relaxation effects based on HRV measurements. Our current study aligns with their results, indeed we found improvement of relaxation in participants after the VR experience also on objective measures: RR significantly decreased, and RMSSD and NN50 significantly increased during Task1, while RR and SCR significantly decreased, and RMSSD significantly increased during Task2. Finally, since in the final interview participants in the DSK group reported they would use the system to relax, immediately (n=11) or after some modification (n=2), we can conclude that they felt the effect of relaxation induced by the VR experience, which motivated their intention to use the system again.

Results have confirmed our hypothesis H2 that the VR experience in immersive VR enhances the sense of presence in the VE. We found a statistically significant difference in the general item being there, in subscales spatial presence and involvement, and in the IPQ total score, where scores were higher with immersive VR than with desktop VR. In contrast, no significant difference was found regarding realism, as both groups scored low in this subscale, which was expected due to our design choice of using stylized fantasy graphics. To the best of our knowledge, this study is the first to compare the same system using biofeedback mechanisms in both immersive VR and desktop VR. A previous study [32] compared a VR system with biofeedback in immersive VR and desktop VR without finding significant differences in the sense of presence. However, in that study, the system in desktop VR did not include biofeedback. The higher sense of presence experienced by the IMM group compared with the DSK group might have also influenced the result obtained with PANAS, where the difference between PANAS-PA and PANAS-NA scores was significantly higher in the IMM group than in the DSK group. As explained in the previous study (Section 4.2.5), since the IMM group experienced a greater sense of presence, and the affective state can be influenced by presence [200], this may have led to a greater emotional impact on the IMM group than the DSK group.

Results have confirmed our hypotheses H3 that as the sense of presence increases, the level of relaxation increases. The results showed that the DSK group has a strong positive correlation between the sense of presence and the level of relaxation achieved after the VR experience. This result indicates that the sense of presence represents a significant factor in relaxation systems, as it contributes to enhancing the overall relaxation effect. Conversely, the IMM group did not display correlations between the sense of presence and the level of relaxation achieved after the VR experience. This outcome may be attributed to the fact that in the IMM group, the sense of presence scores were high (between 4 and 6 on a scale of 0 to 6) for the sense of being there, as well as for the subscales of spatial presence and involvement. As shown in the data in Table 4.8, the sense of presence in the IMM group showed minimal variability, which limited the ability to identify significant correlations. This is in contrast to the DSK group, where scores were moderate (between 3 and 4) except for experienced realism.

Results have confirmed our hypothesis H4 that immersive VR relaxes more than desktop VR. More specifically, the analysis of the STAI-S scores revealed that the IMM group achieved a significantly higher level of relaxation than the DSK group after the VR experience. Kosunen [32] showed that the system without biofeedback in desktop VR relaxed significantly less than the same system with biofeedback in immersive VR. Our study extends these findings by showing that the difference in relaxation levels between the two groups persists also when the biofeedback mechanisms is present in both display types. The results obtained in the biofeedback questionnaire show no differences between the two groups regarding the perceived quality of biofeedback and the ease of the tasks. This finding suggests that the difference in the level of relaxation achieved by the two groups comes from the different levels of immersion of the system. The IMM group, having experienced the system in immersive VR, achieved higher levels of sense of presence, which in turn, significantly contributed to the enhancement of the effect of relaxation.

In summary, this study showed that the VR experience improved the level of relaxation in both groups. Specifically, immersive VR produced better results than desktop VR in terms of relaxation and sense of presence. These findings suggest that a higher level of immersion contributes to a better user experience. Future research on relaxation systems should prefer the use of HMD, as they have the potential to enhance the relaxation effects of the system.

It is important to note that all participants in the DSK group were exclusively recruited through direct contact, in contrast to participants in the IMM group who were recruited through direct contact, email and social channels of our university. This recruitment strategy was adopted to ensure the inclusion of individuals with characteristics balanced with those of the IMM group participants. The difference in the recruitment process might have affected the representativeness of the sample, as the DSK group could be considered a convenience sample. Consequently, this limitation could impact the generalizability of the results.

## 4.4 Final discussion

Results of Study 1 and Study 2, described in Sections 4.2 and 4.3, respectively, confirm that the VR experience induces relaxation among users, as supported by both subjective and objective measures. The studies show the positive impact of biofeedback on enhancing relaxation levels, highlighting the benefits of combining biofeedback with immersive VR.

Regarding the sense of presence, results reveal that the VR system with biofeedback significantly enhances the overall sense of presence compared to the VR system with sham biofeedback and that immersive VR significantly increases the sense of presence compared to desktop VR.

In summary, the two studies show that the VR system induces relaxation, biofeedback plays a crucial role in improving relaxation levels, and integrating immersive VR with biofeedback yields better results than integrating desktop VR with biofeedback.

# 5

# Clinical trial of the therapeutic effects of immersive VR with biofeedback on patients with fibromyalgia

In Chapter 4 we showed that our VR system with biofeedback for relaxation training relaxes (Section 4.2) and that the use of immersive VR enhances relaxation effects (Section 4.3). This thesis concludes by bringing such developed knowledge in a real clinical setting to assess the effectiveness of our system as an Immersive Therapeutic system (i.e., the application of advanced immersive technologies, such as immersive VR and AR, for therapeutic purposes in the medical field). Thus, we conducted a five-session longitudinal study in which we used the system with biofeedback in immersive VR on a sample of patients with fibromyalgia.

In the following sections, we first present a review of the current literature about studies on immersive and desktop VR systems for fibromyalgia. Then, we provide detailed description of the clinical trial we have carried out on patients. Results will be reported also in a paper to submit to an international journal.

## 5.1 Related work: fibromyalgia

Fibromyalgia is a chronic debilitating rheumatologic condition with unknown etiology, characterized by chronic widespread pain, sore spots in muscles and soft tissues, and extensive musculoskeletal pain. It is often accompanied by a multitude of concurrent clinical symptoms and comorbidities that make its diagnosis difficult. This disease presents several symptomatic manifestations that encompass physiological, psychological, and social aspects including sleep disturbances, fatigue, muscle stiffness, mood changes, and cognitive impairments [205,206]. In particular, anxiety and depressive disorders are the most frequent psychiatric comorbidities among adult patients with fibromyalgia [207]. These symptoms have a negative impact on patients' quality of life and lead to a reduced ability to perform everyday activities, potentially resulting in physical and psychosocial disabilities [206].

Positive emotions can help counteract the negative impact that chronic pain has on mood because they influence cognitive processes, well-being, and health [208, 209]. Due to biased perceptions, this syndrome has a female predominance but unbiased studies do not support this claim (Wolfe et al., 2018). Nevertheless, the majority of studies have been conducted on female samples (Wolfe et al., 2018).

The most accepted intervention for fibromyalgia is a multidimensional approach that includes pharmacological and nonpharmacological treatments [210]. The latter include psychological interventions, such as education and cognitive behavioral therapy (CBT) [211], and physical exercises [212]. A meta-analysis study classified psychological interventions for fibromyalgia into six categories [213]: CBT, relaxation, educational interventions, behavioral treatments, mindfulness-based programs, and other treatments.

In recent years, there has been an increase in the use of VR for therapeutic purposes to treat psychological and physical disorders (e.g., [211, 214, 215]) or to induce relaxation in patients with fibromyalgia(e.g., [102, 211]). VR is used for both physical rehabilitation and pain management according to the principle of distraction: these systems are able to direct users' attention to the VE, shifting their focus away from the real world and reducing cognitive resources dedicated to pain processing [216].

In the literature, immersive and desktop VR systems used for the purpose of enhancing the quality of life of patients with fibromyalgia fall into two categories: exergame systems and relaxation systems.

### 5.1.1 Exergame systems

Exergame systems, in the context of fibromyalgia, are used to engage patients in aerobic and muscle-strengthening activities, as well as to improve their flexibility, mobility, and balance. Such systems improve their quality of life, reducing anxiety, depression, pain, and stress responses [217].

VirtualEx-FM is a non-immersive exergame system specifically designed to enhance the physical condition and functional capabilities of patients with fibromyalgia. The system comprises three VEs that enable participants to improve postural control, coordination, balance, aerobic condition, mobility, and limb strengths [218]. Studies conducted on patients with fibromyalgia who used VirtualEx-FM have shown it is an effective treatment as it has reduced pain levels and fibromyalgia symptoms, and has enhanced the overall participants' quality of life [218–221].

Other studies have explored the use of exergame systems for the management of fibromyalgia, employing commercial motion-controlled video games (MCVG) consoles, Nintendo Wii Fit Plus, or Microsoft Xbox Kinect video games. Two separate studies used commercial MCVG games involving a comparison of three different consumer consoles [222, 223]. However, results showed that while participants found the game sessions enjoyable and distracting from pain, there was no improvement in symptoms or reduction in pain. In contrast, studies using video games from the Nintendo Wii Fit Plus [224, 225] and the Microsoft Xbox Kinect [212] have shown their potential to mitigate the impact of fibromyalgia. These studies reported enhancements in pain threshold perceptions [224], improvements in movement ability [225], and when combined with exercise training programs, they are found to be valuable in enhancing the quality of life of participants [212].

In addition to the non-immersive VR systems previously described, an immersive VR system has also been evaluated in the existing literature [215]. This system presents two VEs where users had to strike balls approaching them using their hands and feet, or had to dodge guillotines-like obstacles. The assessment of the system showed its efficacy as an adjunctive therapy in conjunction with exercise training programs, due to its positive outcomes, such as the reduction of pain, alleviation of fatigue, and enhancement in quality of life.

### 5.1.2 Relaxation systems

Three studies have added to the traditional CBT the use of the non-immersive VR system Engaging Media for Mental Health Applications (EMMA's) World [211, 214, 226]. This system allows for customization of images, sounds, and narratives to meet specific needs and offers five naturalistic VEs designed to elicit emotional responses. Botella et al. [211] used EMMA's World to provide instructions and guidance for slow breathing, resulting in a reduction in pain and depressive symptoms, as well as an increase in positive emotions [211]. Other two studies used the system to induce positive emotions and promote motivation and self-efficacy in patients with fibromyalgia [214, 226]. Results showed enhancements in emotional state, motivation, and self-efficacy [214], along with improved fibromyalgia symptoms and perceived quality of life [226].

Furthermore, it is worth noting a study in the literature involving a system that consists of two identical immersive VEs, one dedicated to audio-guided meditation without biofeedback and the other instructing users to engage in slow, deep breathing at a rate of six bpm, with visual biofeedback mechanisms displaying inhalations and exhalations (for additional details regarding this system and study, see Table 1.3 and Table 1.4) [102]. However, it is important to highlight that this study did not exclusively target individuals with fibromyalgia but involved a broader spectrum of patients with chronic pain disorders. Results of the study showed a reduction in chronic pain levels in both VEs and a significant reduction in anxiety levels within the VE without biofeedback.

## 5.2 Study: clinical trial on patients with fibromyalgia

In this study, we aim to put our immersive VR system with biofeedback to a more complex evaluation. Unlike previous studies, our primary objective is to achieve a more ambitious outcome: we seek to achieve heightened relaxation that leads to a reduction in anxiety levels which, in turn, mitigate the perception of pain.

We conducted a five-session longitudinal study in which we used our system on a sample of patients with fibromyalgia. We compared a treatment group (TR group, hereinafter) and a waitlist group (WL group, hereinafter). The TR group tried the VR system with biofeedback for five days. Conversely, the WL group remained in a waiting phase until the conclusion of the treatment period of the TR group before trying the same treatment. Both groups were also evaluated in a follow-up conducted 15 days after the conclusion of the treatment period.

### 5.2.1 Hypothesis

We hypothesized that, compared with a waitlist control group, participants in the treatment group reduce pain levels and alleviate fibromyalgia symptoms because stress, anxiety, and depression strongly influence participants' perceived pain [227–229]. The study described in Section 4.2 showed that the system effectively enhances relaxation levels. Therefore, the increased relaxation achieved using the system could potentially lead to a reduction in perceived pain levels.

### 5.2.2 Participants

The study involved a sample of 24 female participants with fibromyalgia. They were volunteers that received no compensation and were recruited through direct contact and phone calls. Their age ranged from 21 to 54 (M=44.85, SD=7.58), and they were patients at the Rheumatology Clinic of Santa Maria della Misericordia Hospital in Udine. They were divided into two groups, which had an equal number of participants (TR: n=12; WL: n=12), and were similar in gender and age (TR: M=43.87, SD=8.81; WL: M=46.08, SD=5.84).

We confirmed the lack of significant differences in terms of age between the two groups using an unrelated $t$-test. Two participants (TR: n=1; WL: n=1) experienced four instead of five sessions due to illness or difficulties in scheduling an appointment at the clinic. Four participants, two from each group, dropped out of the study for the following reasons: two encountered difficulties in scheduling their presence at the clinic for five days to try the system, one experienced fatigue in focusing on the images observed through the HMD, one experienced excessive discomfort due to the elastic girth placed on her abdomen. As a result, the analysis was conducted on 20 participants (TR: n=10; WL: n=10).

### 5.2.3 Measures

#### Pain

We administered the Short-Form McGill Pain Questionnaire (SF-MPQ) to assess patients' pain [230]. The SF-MPQ assesses the individual's perceived pain using a set of 15 items, each rated on a scale from 0 ("none") to 3 ("severe"), resulting in a total score derived from the sum of all item scores (ranging from 0 to 45). Higher total scores on this scale are indicative of more severe pain. The questionnaire includes two subscales for the independent dimensions of "sensory" and "affective" descriptors.

The SF-MPQ also includes the Present Pain Intensity (PPI) index and a Visual Analogue Scale for Pain (VAS Pain) to assess participants' pain intensity. The PPI is recorded as a number from 1 to 5 (1="mild", 2="discomforting", 3="distressing", 4="horrible", 5="excruciating"). The VAS Pain is a 10-centimeter long line, with "no pain at all" and "the most intense pain" printed at its left and right ends, respectively. Participants reported their pain severity by drawing a vertical mark on the scale.

The SF-MPQ was administered to both groups, before and after the five-session treatment period. Moreover, participants also filled a VAS Pain questionnaire after experiencing the third treatment session to assess the progression of perceived pain intensity throughout the course of the treatment. Additionally, VAS Pain was measured at 15 days post-treatment to assess whether any improvements in pain perception were maintained over time.

**Fibromyalgia symptoms**

We administered the Fibromyalgia Impact Questionnaire (FIQ) to assess patients' fibromyalgia symptoms [231]. The FIQ comprises ten items (i.e., physical function, feel good, work missed, job ability, pain fatigue, morning tiredness, stiffness, anxiety, and depression), each with a maximum score of 10. Total scores can vary from 0 to 100, and higher values indicate a stronger influence on the patient's quality of life. The FIQ was administered to both groups, before and after the five-session treatment period, and in a 15-day follow-up.

### 5.2.4 Procedure

The study was approved by the Institutional Review Board of the University of Udine. Participants provided written consent for their participation in the study. Subsequently, the experimenter verbally apprised them about the anonymity of the collected data and informed them that the VR experience involved using a VR HMD and physiological sensors. Then, participants sat in a chair and filled the SF-MPQ, and FIQ questionnaires.

Subsequently, the TR group began the treatment, which comprised five different sessions of the VR experience in five days scheduled to be completed over a period of maximum ten days. Meanwhile, the WL group remained in a waiting period until the completion of the treatment of the TR group.

At each session of the treatment the experimenter applied physiological sensors to participants, helped them wear an Oculus Quest 2 HMD, and gave them a Quest touch controller to hold with their right hand. Participants were asked to choose a comfortable seated position. Then, they were immersed for three minutes in the same virtual living room used in the study described in Section 4.2. After that, participants performed diaphragmatic breathing training for one minute, guided by a voice-over that provided instructions and invited them to observe the movements of a lotus flower placed in the center of the virtual living room. This lotus flower opened and closed its petals with each inhalation and exhalation of participants, respectively (Figure 5.1).

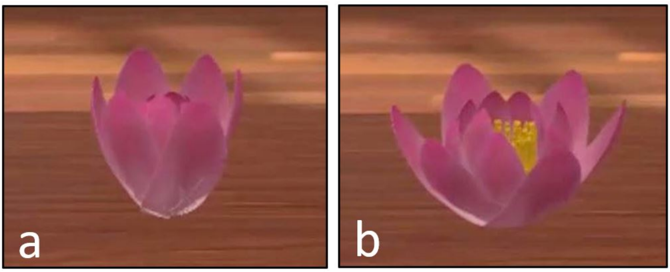



Figure 5.1: Lotus flower in the diaphragmatic breathing training before the VR experience. (a) The more the user exhales, the more the flower closes its petals; (b) The more the user inhales, the more the flower opens its petals.

Then, participants tried the VR experience of the Crystals Archipelago, which lasted for 11 minutes. After the VR experience, the experimenter helped participants remove the HMD and physiological sensors. At the end of the third session, after the VR experience, the TR group filled out the VAS Pain questionnaire. At the end of the fifth treatment session, both groups filled the SF-MPQ, and FIQ questionnaires. At this point, the WL group started the same five-session treatment, following the same procedure followed for the TR group. Thus, participants of the WL group also filled the VAS Pain questionnaire at the end of the third session, and the SF-MPQ, and FIQ questionnaires after the five-session treatment. Fifteen days after the end of treatment, participants in both groups completed the VAS Pain and FIQ questionnaires. Our longitudinal study is illustrated in Figure 5.2.

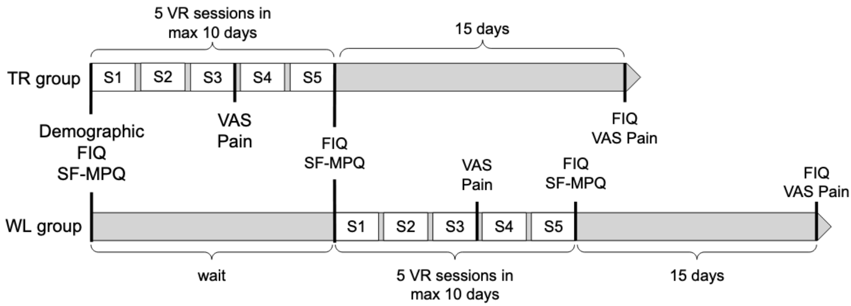



Figure 5.2: Design of the longitudinal study.

### 5.2.5 Results

**Between-subjects study**

All the analyses were conducted using SPSS version 29.0.0.0. The design of the between-subject study is shown in Figure 5.3.

SF-MPQ scores were submitted to a $2 \times 2$ mixed design ANOVA, after checking its assumptions were met, in which group (TR and WL) served as the between-subject variable, and time of measurement (before and after the treatment period of the TR group) served as the within-subject variable. Table 5.1 reports means and standard deviations of SF-MPQ and FIQ scores for each group.

The analyses revealed the following statistically significant differences between the two groups:

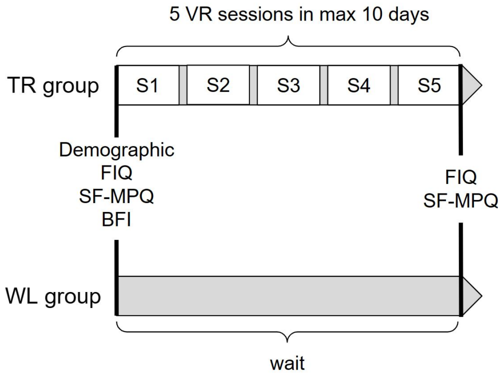



Figure 5.3: Design of the between-subjects study.

Table 5.1: Means and standard deviations of SF-MPQ and FIQ scores for each group.

| Measure | First measurement | | Second measurement | |
|---|---|---|---|---|
| | **TR group M(SD)** | **WL group M(SD)** | **TR group M(SD)** | **WL group M(SD)** |
| **SF-MPQ** | | | | |
| Sensory | **14.30** (5.85) | **12.50** (7.55) | **9.70** (4.30) | **13.50** (6.98) |
| Affective | **6.80** (3.16) | **5.60** (2.80) | **4.10** (0.99) | **5.50** (3.03) |
| McGill total | **21.00** (8.50) | **18.10** (9.70) | **13.80** (4.34) | **19.00** (9.65) |
| PPI | **2.40** (0.84) | **2.80** (0.79) | **1.90** (0.74) | **2.70** (1.16) |
| VAS Pain | **58.00** (12.06) | **57.50** (25.74) | **34.50** (13.63) | **60.50** (25.44) |
| FIQ | **68.10** (12.12) | **59.20** (16.45) | **42.30** (13.17) | **60.20** (17.35) |

- Affective: $F(1,18) = 5.65$, $p < 0.05$, $\eta_p^2 = 0.24$ (Figure 5.4).
- VAS Pain: $F(1,18) = 12.15$, $p < 0.01$, $\eta_p^2 = 0.40$ (Figure 5.5).

Moreover, results revealed the following group by time of measurement interactions:

- Affective: $F(1,18) = 4.87$, $p < 0.05$, $\eta_p^2 = 0.21$ (Figure 5.4). We analyzed each simple effect using Bonferroni correction, considering the effects of time of measurement separately for each group and the effect of group separately at each time of measurement. SF-MPQ scores on affective descriptor showed a significant decrease in scores after the treatment only in the TR group ($p < 0.01$).

- VAS Pain: $F(1, 18) = 20.31$, $p < 0.001$, $\eta_p^2 = 0.53$ (Figure 5.5). Simple effects using Bonferroni correction showed a significant decrease in scores after the treatment only in the TR group ($p < 0.001$), and a significant difference in scores between the two groups after the treatment ($p < 0.05$).

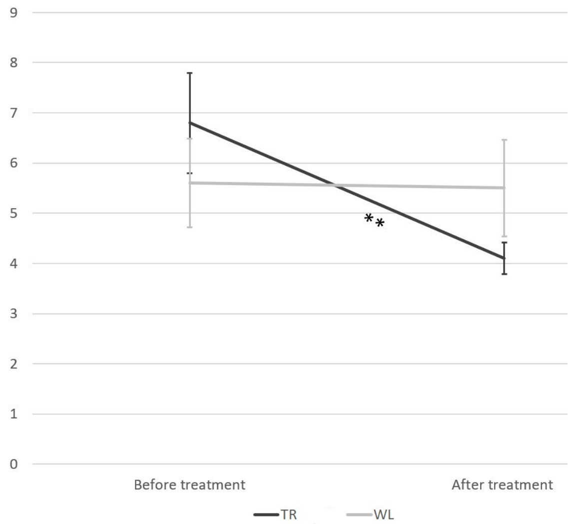



Figure 5.4: Group by time of measurement interaction in affective descriptor of the SM-MPQ scores. Capped vertical bars indicate ± SE. The ** sign indicates statistically significant difference with $p$-value $< 0.01$.

FIQ scores were submitted to a 2 × 2 mixed design ANOVA, in which group served as the between-subject variable, and time of measurement served as the within-subject variable (Table 5.1).

Statistically significant results revealed a main effect of time of measurement, $F(1, 18) = 24.91$, $p < 0.001$, $\eta_p^2 = 0.66$, and a group by time of measurement interaction, $F(1, 18) = 40.77$, $p < 0.001$, $\eta_p^2 = 0.69$, as illustrated in Figure 5.6. We analyzed each simple effect using Bonferroni

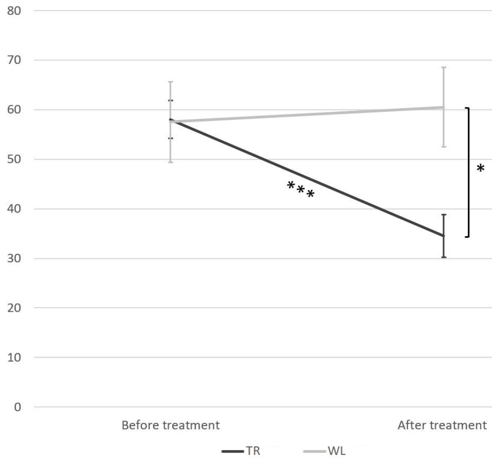



Figure 5.5: Group by time of measurement interaction in VAS Pain scores. Capped vertical bars indicate ± SE. The * and *** signs indicate statistically significant differences with $p$-values respectively $< 0.05$, $< 0.001$.

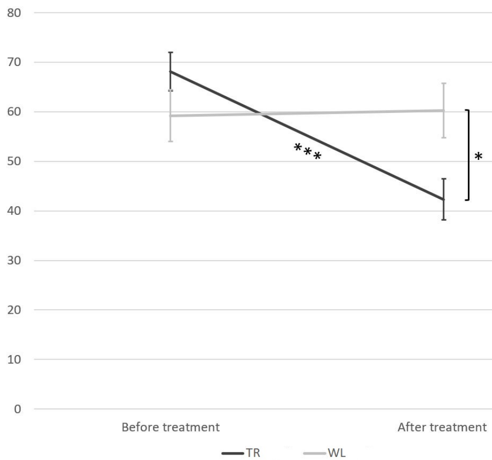



Figure 5.6: Group by time of measurement interaction in FIQ scores. Capped vertical bars indicate ± SE. The * and *** signs indicate statistically significant differences with $p$-values respectively $< 0.05$, $< 0.001$.

correction, considering the effects of time of measurement separately for each group and the effects of group separately at each time of measurement. FIQ results showed a significant decrease in scores after the treatment in the TR group ($p < 0.001$), and a significant difference in scores between the two groups after the treatment ($p < 0.05$).

### Within-subjects study

Once we found that treatment was effective for patients, we aggregated data from the two groups and treated them as a single group for analysis. The design of the within-subject study is shown in Figure 5.7.

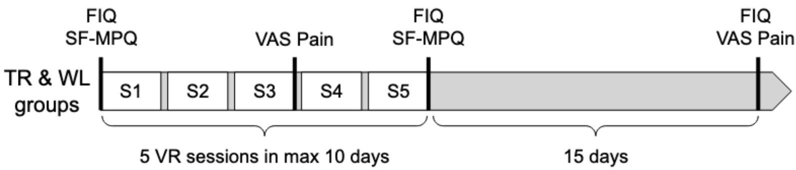



Figure 5.7: Design of the within-subjects study.

A repeated-measure ANOVA was used to compare the scores of the questionnaires. Specifically, two time points were assessed for the SF-MPQ, three time points were assessed for the FIQ, and four time points were assessed for the VAS Pain score. Tables 5.2 , 5.3, and 5.4 reports means and standard deviations of SF-MPQ, FIQ, and VAS Pain scores.

Table 5.2: Means and standard deviations of SF-MPQ scores.

| **Measure** | **Before treatment M(SD)** | **After treatment M(SD)** |
|---|---|---|
| Sensory | **13.85** (6.28) | **8.90** (4.80) |
| Affective | **6.15** (3.08) | **3.65** (1.87) |
| McGill total | **20.00** (8.91) | **12.55** (6.10) |
| PPI | **2.55** (1.00) | **2.10** (0.79) |

Analyses revealed the following statistically significant results:

- Sensory: $F(1,19) = 13.52$, $p < 0.01$, $\eta_p^2 = 0.42$ (Figure 5.8).

Table 5.3: Means and standard deviations of FIQ scores.

| Measure | Before treatment M(SD) | After treatment M(SD) | Follow-up M(SD) |
|---|---|---|---|
| FIQ | **64.16** (15.12) | **43.90** (18.54) | **50.40** (17.85) |

Table 5.4: Means and standard deviations of VAS Pain.

| Measure | Before treatment M(SD) | After third session M(SD) | After treatment M(SD) | Follow-up M(SD) |
|---|---|---|---|---|
| VAS Pain | **59.25** (19.42) | **49.75** (25.52) | **37.75** (21.67) | **52.50** (22.15) |

- Affective: $F(1,19) = 17.34$, $p < 0.001$, $\eta_p^2 = 0.48$ (Figure 5.8).

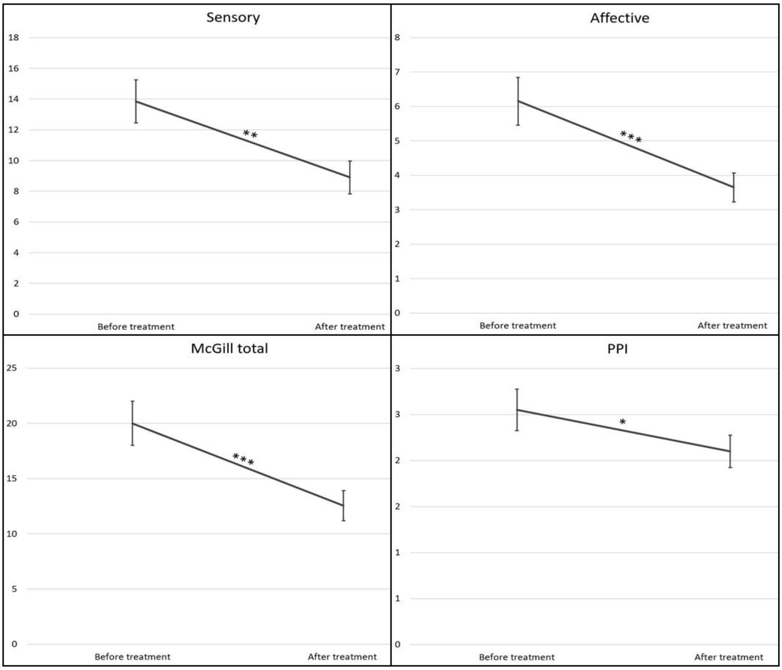



Figure 5.8: Means of SF-MPQ scores. Capped vertical bars indicate ± SE. The *, **, and *** signs indicate statistically significant differences with $p$-values respectively $< 0.05$, $< 0.01$, $< 0.001$.

- McGill total: $F(1, 19) = 16.26$, $p < 0.001$, $\eta_p^2 = 0.46$ (Figure 5.8).
- PPI: $F(1, 19) = 5.94$, $p < 0.05$, $\eta_p^2 = 0.24$ (Figure 5.8).
- FIQ: $F(2, 38) = 19.67$, $p < 0.001$, $\eta_p^2 = 0.51$ (Figure 5.9).
- VAS Pain: $F(1.98, 37.55) = 10.34$, $p < 0.001$, $\eta_p^2 = 0.35$ (Figure 5.10).

In FIQ scores, Bonferroni post hoc comparison found the following significant differences:

- Before treatment vs. after treatment: $p < 0.001$.
- Before treatment vs. follow-up:$p < 0.01$

In VAS Pain scores, Bonferroni post hoc comparison found the following significant differences:

- Before treatment vs. after treatment: $p < 0.001$.
- After third session vs. after treatment: $p < 0.01$.
- After treatment vs. follow-up: $p < 0.05$.

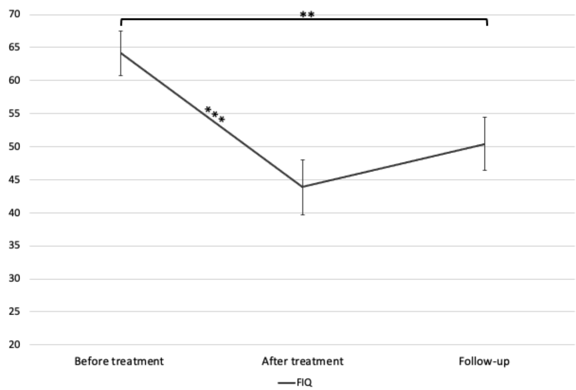



Figure 5.9: Means of FIQ. Capped vertical bars indicate ± SE. The **, and *** signs indicate statistically significant differences with $p$-values respectively $< 0.01$, $< 0.001$.

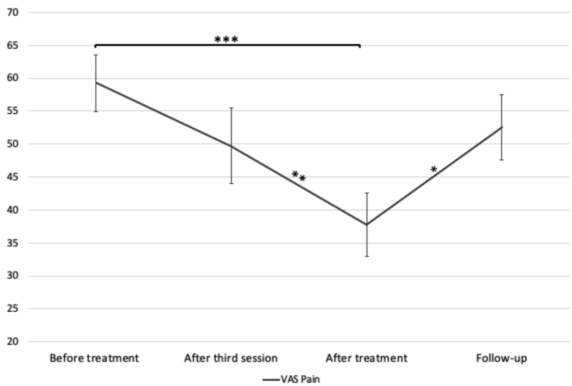



Figure 5.10: Means of VAS. Capped vertical bars indicate ± SE. The *, **, and *** signs indicate statistically significant differences with $p$-values respectively $< 0.05$, $< 0.01$, $< 0.001$.

## 5.2.6 Discussion

Results have confirmed our hypothesis that VR system with biofeedback for relaxation training reduces pain levels and alleviates fibromyalgia symptoms. Analysis of the SF-MPQ and FIQ questionnaires revealed that participants in the treatment group significantly lowered pain levels and improved fibromyalgia symptoms compared with the waitlist control group. After the five-session treatment with our system, the TR group significantly lowered the scores of the VAS Pain, the affective descriptor of the SF-MPQ, and the FIQ. Findings also revealed an interaction between the two groups that showed that the affective descriptor significantly decreased after treatment only in the TR group, and that the scores of VAS Pain and FIQ were significantly different between the two groups after treatment.

Considering the treatment data collected first from the TR group, and then from the WL group as one single group, data analysis showed better results. Scores of sensory and affective descriptors, McGill total, PPI, VAS Pain, and FIQ significantly decreased after the treatment compared to before the treatment.

These results are in line with previous studies that have shown the effectiveness of exergame and relaxation systems in reducing pain levels [102, 211, 215, 218] and alleviating fibromyalgia symptoms [219, 226]. In [218], the authors conducted a longitudinal study on a group of female patients with fibromyalgia that used the VirtualEx-FM system, resulting in reduced pain levels and fewer fibromyalgia symptoms when compared to a waitlist control group. This non-immersive VR system was designed to enhance the physical condition and daily functioning of individuals with fibromyalgia, and provided immediate visual feedback to users on the quality of their movements. In contrast, our VR system aimed to train participants in slow, deep breathing and relaxation using real-time biofeedback mechanisms to show them the quality of their breathing and their level of relaxation.

Gulsen et al. [215] carried out a longitudinal study comparing a group that followed a physical exercise program accompanied by the use of their immersive exergame VR system to a control group that solely engaged in the physical exercise program. Results of the study showed that both groups achieved significant improvements in pain, kinesiophobia (i.e., fear of movement resulting from past physical trauma), fatigue, and the mental component of quality of life, but the treatment group obtained significantly better results. In our study, we exclusively employed our system without combining it with other nonpharmacological treatments, and we observed significant differences in pain reduction and fibromyalgia symptoms only in the treatment group. Our immersive VR system would also be usable by individuals with fibromyalgia in situations where acute pain might discourage them from performing a physical training program, while still obtaining the same pain reduction benefits.

In the studies conducted by Botella et al. [211] and Garcia-Palacios et al. [226], longitudinal research evaluated the use of EMMA's World system in conjunction with a CBT program. In [211], the EMMA's World system was used for relaxation and mindfulness purposes. Participants were encouraged to practice relaxation and mindfulness exercises also at home using a CD that recalled sounds and images from EMMA's

World. Follow-up results after six months showed a significant reduction in pain and depression. In [35], a treatment group that received CBT alongside the system customized to elicit positive emotions was compared with a waitlist control group. Results did not show a reduction in perceived pain, but improvements were observed in quality of life, depression, and fibromyalgia symptoms. In [211], EMMA's World was customized to present instructions for slow breathing and mindfulness practice, guiding participants to observe the different elements offered by the VE. In our system, similarly, the voice-over trained slow, deep breathing, but also incorporated a biofeedback mechanism, allowing participants to receive immediate feedback on their breathing. Although our system was not designed to teach mindfulness meditation, the voice-over prompted participants to observe different elements of the VE without mentioning whether and how physiological parameters controlled the movement or appearance of those elements (Table 4.1). In addition, the voice-over accompanied participants through all five phases, describing the surroundings and inviting them to perform the two tasks of clearing the fog and making the night fall. Participants may have focused on these two tasks, which may have helped divert their attention from the pain.

In the existing literature, only one study has explored the use of an immersive VR system with biofeedback on patients with chronic pain [102]. In this study, participants experienced the two VEs of the immersive VR system. In one VE, they performed a guided meditation, while in the other VE, which was visually identical to the first, they followed the instructions of a guided voice that invited them to maintain a breathing rate of six bpm. This second VE used biofeedback mechanisms to show users their inhalations and exhalations. As described in Table 1.3, the system represented breathing through 2D circles that grew and shrank with each inhalation and exhalation. In our system, breath was mapped to multiple elements of the VE, as outlined in Table 4.2. Additionally, the system used the color of the windmill blades to indicate whether the user was taking deep inhalations and exhalations. Notably, while Venuturupalli et al. exclusively used biofeedback for breath, which was captured through a microphone, our system employed the Thought Technology ProComp Infinity encoder, equipped with sensors to record respiratory, cardiac, and electrodermal activity of participants. Furthermore, while in [102] the study obtained a reduction in the level of pain and anxiety within one session, evaluating a sample of patients with chronic pain without a control group for comparison, our study involved a treatment group and compared it to a waitlist control group over a period of five sessions experienced in five different days, strengthening the reliability of the observed outcomes.

The 15-day follow-up revealed an increase in pain and fibromyalgia symptoms when compared to the end of treatment. In particular, while the VAS Pain scores at follow-up were not significantly different from pre-treatment scores, the FIQ scores remained significantly better than those measured before the treatment. These findings suggest that the initial positive impact on pain may not be sustained over time after the cessation of treatment, emphasizing the necessity for further research to investigate solutions that can maintain long-term benefits. A similar trend was observed in the improvement in fibromyalgia symptoms, although this effect persists for a longer duration, resulting in scores that remain significantly better than pre-treatment scores.

Our results are in line with studies in the literature regarding specific treatments for fibromyalgia [210, 232]. In Altan et al. [232], a comparison was made between a pilates treatment and a treatment involving relaxation and stretching exercises performed at home. Although post-treatment results showed significantly better outcomes in pain and fibromyalgia symptoms in pilates group, the 12-week follow-up did not show significant differences between the two groups. In another study [210], Häuser et al. conducted a meta-analysis of randomized controlled clinical trials to assess the effectiveness of multicomponent therapy treatments (i.e., therapies that included at least one educational or psychological therapy and at least one exercise therapy). Meta-analysis results showed a reduction in pain at post-treatment. However, follow-ups of 3-4 months and 6-12 months did not show significant effects on pain.

It should be noted that the sample of our study consisted exclusively of women. This decision was influenced by the documented gender differences in pain perception [233, 234], but it limits the generalizability of our findings. It is noteworthy that all previous studies in the literature examining the impact of VR systems on patients with fibromyalgia have also focused solely on female samples [211,212,214,218–226]. The potential applicability of our VR system with biofeedback on male individuals with fibromyalgia remains unexplored.

# Conclusions

This PhD thesis explored the use of desktop, immersive, and augmented reality to assess their effectiveness in VR systems for anxiety mitigation. The thesis investigated the capacity of the different display types to elicit and mitigate anxiety in individuals through VRE systems for anxiety mitigation as well as VR systems with biofeedback for relaxation training.

In Chapter 2 the thesis proposed a VRE system for exam anxiety that deals with oral exams. In the three scenarios offered by the system, a VX conducts the oral exam, performing behaviors from one of three predefined sets, which differ in how friendly the VX behaves and define three different levels of difficulty. This is the first feasibility study of a VRE system for exam anxiety that deals with oral exams. Results of the quantitative study show that the three difficulty levels of the system are able to elicit three different levels of increasing anxiety. The thematic analysis of the interview conducted with participants provided further insights into the aspects of the system that contributed to eliciting positive or negative responses in them, which can help in improving the design of VRE systems for exam anxiety. Following the assessment of the feasibility of the VRE system, a trial was conducted where a sample of students freely used the system at home for three weeks, and were interviewed at the end of the trial. Thematic analysis of the interviews suggests that the VRE system can be a valuable tool for helping students develop emotional and cognitive skills to cope successfully with oral exams. The VRE system provides a safe environment that can positively impact users' confidence, awareness of their preparation, and encouragement to study. Experiences shared by participants reinforce the usefulness of the assessed VRE system. Based on the encouraging results of the feasibility and trial studies, future research will take into account participants' suggestions to further improve the VRE system. Then, a quantitative study, which includes a waitlist control group, will be conducted to assess in detail the effects of VRE system on students' exam anxiety, self-efficacy, and overall well-being.

In Chapter 3 the thesis assessed the feasibility of a system for public speaking anxiety in eliciting anxiety and distress in participants when facing a small vs. big virtual audience displayed in immersive VR vs. AR. Additionally, it analyzed participants' gaze behavior during a public speech. This study is the first to assess the feasibility of an ARE system for public speaking anxiety and compare the effectiveness of immersive VR and AR in public speaking context. Findings suggest that both immersive VR and AR are equally effective in eliciting anxiety and distress when using the system. This means that AR could be a viable alternative to VR, enabling individuals to practice public speaking while achieving the same benefits of VR over IVE, such as easy access to feared stimuli and complete control over the exposure process. Furthermore, AR provides additional benefits over VR, including the ability of individuals to train themselves to speak in front of a virtual audience displayed directly within the real environment in which they might give a public speech, such as the meeting room of their workplace. Analysis of participants' gaze behavior found they looked more at virtual agents' heads in the big audience than the small audience. These findings suggest a potential relationship between audience size and participants' gaze behavior, highlighting the need for further research to explore possible gaze patterns followed by participants during their speech task. A deeper understanding of the impact of individuals looking at small and big audiences in immersive VR and AR would contribute to enhancing the effectiveness, engagement, and overall user experience of VRE systems for public speaking anxiety. While this feasibility study showed the capability of the proposed system to elicit anxiety and distress, future research will focus on assessing its impact on individuals' performance and public speaking skills.

In Chapter 4 the thesis proposed a VR system with biofeedback for relaxation training. The system aims to teach users how to perform slow and deep diaphragmatic breathing by immersing them in a natural VE that changes based on multiple physiological measurements. A narrative that unfolds in multiple stages involves the user as the main character of a story that evolves through the performance of two main activities required to advance the narrative. In the first study, we assessed the actual contribution of biofeedback by comparing the results of a group of participants who used the system with real biofeedback to a placebo group who used the system with sham biofeedback. Results showed that the system helped participants to relax in both groups. Moreover, outcomes suggested that biofeedback led to greater relaxation as well as sense of presence than sham biofeedback. These findings underline the value of adding biofeedback into VR systems for relaxation training, while also reiterating the importance of including placebo control conditions in studies evaluating systems with biofeedback. Following the assessment of the feasibility of the system in mitigating anxiety and the confirmation of the actual contribution of biofeedback in improving relaxation levels, the thesis investigated the influence of immersive VR on relaxation effects of the system when compared to desktop VR. The study confirmed that immersive VR improves both relaxation effects of the VR system with biofeedback and the sense of presence. This means that while the system in desktop VR can relax users, the use of immersive VR is preferable when aiming to maximize the level of relaxation achieved. The VR experience of the system provided two relaxation activities controlled by biofeedback based on user's breathing and SC, while it did not include relaxation activities controlled by biofeedback based on cardiac parameters. Research has indicated that biofeedback using HRV is effective in preventing and treating anxiety and stress [235]. Previous studies on systems using HR biofeedback [116] or HRV biofeedback [24, 30] have shown that such VR systems with biofeedback for relaxation training effectively relaxed participants. Future research will consider enriching the VR experience by designing a new relaxation task controlled by biofeedback based on

HR or HRV, taking into account existing research based on cardiac parameters [24,30,99,105,114,116] to further enhance the relaxing effect of the proposed system.

In Chapter 5, the thesis concluded by bringing the developed knowledge in a real clinical setting to assess the effectiveness of the proposed VR system with biofeedback for relaxation training as an Immersive Therapeutic system. The clinical trial involved a sample of patients with fibromyalgia in a treatment consisting of multiple sessions to assess the potential long-term effects of relaxation. Results showed a significant reduction in pain and fibromyalgia symptoms among patients, supporting the fact that the proposed rich and varied VR experience could promote the prolonged use of the system. This, in turn, might assist users in voluntarily changing their physiological parameters until, over time, these changes persist without the use of the system. This study is the first to evaluate the effects of an immersive VR system with biofeedback using multiple physiological parameters through a longitudinal approach on patients with fibromyalgia. Given the substantial impact of fibromyalgia on quality of life, the use of VR systems with biofeedback for relaxation to reduce anxiety and pain may make a difference in enhancing the lives of individuals coping with fibromyalgia.

# Bibliography

[1] M. Gonzalez-Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano, S. Orts-Escolano, *et al.*, "The rocketbox library and the utility of freely available rigged avatars," *Frontiers in virtual reality*, vol. 1, p. 20, 2020.

[2] M. Davis and M. Powers, "Exposure therapy for anxiety: Principles and practice," 2011.

[3] E. Carl, A. T. Stein, A. Levihn-Coon, J. R. Pogue, B. Rothbaum, P. Emmelkamp, G. J. Asmundson, P. Carlbring, and M. B. Powers, "Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials," *Journal of anxiety disorders*, vol. 61, pp. 27–36, 2019.

[4] J. Fernández-Álvarez, D. Di Lernia, and G. Riva, "Virtual reality for anxiety disorders: rethinking a field in expansion," *Anxiety Disorders: Rethinking and Understanding Recent Discoveries*, pp. 389–414, 2020.

[5] T. D. Parsons and A. A. Rizzo, "Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis," *Journal of behavior therapy and experimental psychiatry*, vol. 39, no. 3, pp. 250–261, 2008.

[6] M. B. Powers and P. M. Emmelkamp, "Virtual reality exposure therapy for anxiety disorders: A meta-analysis," *Journal of anxiety disorders*, vol. 22, no. 3, pp. 561–569, 2008.

[7] M. C. Juan and D. Pérez, "Using augmented and virtual reality for the development of acrophobic scenarios. comparison of the

levels of presence and anxiety," *Computers & Graphics*, vol. 34, no. 6, pp. 756–766, 2010.

[8] C.-F. Tsai, S.-C. Yeh, Y. Huang, Z. Wu, J. Cui, L. Zheng, *et al.*, "The effect of augmented reality and virtual reality on inducing anxiety for exposure therapy: a comparison using heart rate variability," *Journal of Healthcare Engineering*, vol. 2018, 2018.

[9] M. Wrzesien, J.-M. Burkhardt, M. Alcañiz, and C. Botella, "How technology influences the therapeutic process: a comparative field evaluation of augmented reality and in vivo exposure therapy for phobia of small animals," in *Human-Computer Interaction–INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part I 13*, pp. 523–540, Springer, 2011.

[10] C. Suso-Ribera, J. Fernández-Álvarez, A. García-Palacios, H. G. Hoffman, J. Bretón-López, R. M. Banos, S. Quero, and C. Botella, "Virtual reality, augmented reality, and in vivo exposure therapy: a preliminary comparison of treatment efficacy in small animal phobia," *Cyberpsychology, Behavior, and Social Networking*, vol. 22, no. 1, pp. 31–38, 2019.

[11] C. Botella, M. Á. Pérez-Ara, J. Bretón-López, S. Quero, A. García-Palacios, and R. M. Baños, "In vivo versus augmented reality exposure in the treatment of small animal phobia: a randomized controlled trial," *PloS one*, vol. 11, no. 2, p. e0148237, 2016.

[12] C. Botella, J. Bretón-López, S. Quero, R. Baños, and A. García-Palacios, "Treating cockroach phobia with augmented reality," *Behavior therapy*, vol. 41, no. 3, pp. 401–413, 2010.

[13] C. Chandler, E. Bodenhamer-Davis, J. M. Holden, T. Evenson, and S. Bratton, "Enhancing personal wellness in counselor trainees using biofeedback: An exploratory study," *Applied Psychophysiology and Biofeedback*, vol. 26, pp. 1–7, 2001.

[14] S. H. Kim, S. M. Schneider, L. Kravitz, C. Mermier, and M. R. Burge, "Mind-body practices for posttraumatic stress disorder," *Journal of Investigative Medicine*, vol. 61, no. 5, pp. 827–834, 2013.

[15] V. Perciavalle, M. Blandini, P. Fecarotta, . A. Buscemi, D. Di Corrado, L. Bertolo, F. Fichera, and M. Coco, "The role of deep breathing on stress," *Neurological Sciences*, vol. 38, no. 3, pp. 451—458, 2017.

[16] Y.-F. Chen, X.-Y. Huang, C.-H. Chien, and J.-F. Cheng, "The effectiveness of diaphragmatic breathing relaxation training for reducing anxiety," *Perspectives in psychiatric care*, vol. 53, no. 4, pp. 329–336, 2017.

[17] Y. Hayama and T. Inoue, "The effects of deep breathing on "tension–anxiety" and fatigue in cancer patients undergoing adjuvant chemotherapy," *Complementary Therapies in Clinical Practice*, vol. 18, no. 2, pp. 94–98, 2012.

[18] S.-D. Kim and H.-S. Kim, "Effects of a relaxation breathing exercise on fatigue in haemopoietic stem cell transplantation patients," *Journal of clinical nursing*, vol. 14, no. 1, pp. 51–55, 2005.

[19] M. Good, G. C. Anderson, M. Stanton-Hicks, J. A. Grass, and M. Makii, "Relaxation and music reduce pain after gynecologic surgery," *Pain Management Nursing*, vol. 3, no. 2, pp. 61–70, 2002.

[20] M. Good, M. Stanton-Hicks, J. A. Grass, G. C. Anderson, C. Choi, L. J. Schoolmeesters, and A. Salman, "Relief of postoperative pain with jaw relaxation, music and their combination," *Pain*, vol. 81, no. 1-2, pp. 163–172, 1999.

[21] M. Mikolasek, J. Berg, C. M. Witt, and J. Barth, "Effectiveness of mindfulness-and relaxation-based ehealth interventions for patients with medical conditions: a systematic review and synthesis," *International journal of behavioral medicine*, vol. 25, pp. 1–16, 2018.

[22] R. P. Brown and P. L. Gerbarg, "Sudarshan kriya yogic breathing in the treatment of stress, anxiety, and depression: part i—neurophysiologic model," *Journal of Alternative & Complementary Medicine*, vol. 11, no. 1, pp. 189–201, 2005.

[23] L. Chittaro and R. Sioni, "Evaluating mobile apps for breathing training: The effectiveness of visualization," *Computers in Human Behavior*, vol. 40, pp. 56–63, 2014.

[24] J. Blum, C. Rockstroh, and A. S. Göritz, "Heart rate variability biofeedback based on slow-paced breathing with immersive virtual reality nature scenery," *Frontiers in psychology*, vol. 10, p. 2172, 2019.

[25] A. Michela, J. M. van Peer, J. C. Brammer, A. Nies, M. M. van Rooij, R. Oostenveld, W. Dorrestijn, A. S. Smit, K. Roelofs, F. Klumpers, *et al.*, "Deep-breathing biofeedback trainability in a virtual-reality action game: A single-case design study with police trainers," *Frontiers in Psychology*, vol. 13, p. 806163, 2022.

[26] C. Rockstroh, J. Blum, and A. S. Göritz, "A mobile vr-based respiratory biofeedback game to foster diaphragmatic breathing," *Virtual Reality*, vol. 25, pp. 539–552, 2021.

[27] M. Van Rooij, A. Lobel, O. Harris, N. Smit, and I. Granic, "Deep: A biofeedback virtual reality game for children at-risk for anxiety," in *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, pp. 1989–1997, 2016.

[28] "Association for applied psychophysiology and biofeedback." `https://aapb.org/About_BioFeedback`. (Accessed: 5th of September, 2023).

[29] S. Bouchard, F. Bernier, É. Boivin, B. Morin, and G. Robillard, "Using biofeedback while immersed in a stressful videogame increases the effectiveness of stress management skills in soldiers," *PloS one*, vol. 7, no. 4, p. e36169, 2012.

[30] C. Rockstroh, J. Blum, and A. S. Göritz, "Virtual reality in the application of heart rate variability biofeedback," *International Journal of Human-Computer Studies*, vol. 130, pp. 209–220, 2019.

[31] C. Shaw, D. Gromala, and M. Song, "The meditation chamber: towards self-modulation," in *Metaplasticity in virtual worlds: Aesthetics and semantic concepts*, pp. 121–133, IGI Global, 2011.

[32] I. Kosunen, M. Salminen, S. Järvelä, A. Ruonala, N. Ravaja, and G. Jacucci, "Relaworld: neuroadaptive and immersive virtual reality meditation system," in *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pp. 208 – 217, 2016.

[33] J. L. Maples-Keller, B. E. Bunnell, S.-J. Kim, and B. O. Rothbaum, "The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders," *Harvard review of psychiatry*, vol. 25, no. 3, p. 103, 2017.

[34] S. Bandarian-Balooch, D. L. Neumann, and M. J. Boschen, "Exposure treatment in multiple contexts attenuates return of fear via renewal in high spider fearful individuals," *Journal of behavior therapy and experimental psychiatry*, vol. 47, pp. 138–144, 2015.

[35] A. Garcia-Palacios, C. Botella, H. Hoffman, and S. Fabregat, "Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias," *Cyberpsychology & behavior*, vol. 10, no. 5, pp. 722–724, 2007.

[36] P. Caserman, A. Garcia-Agundez, A. Gámez Zerban, and S. Göbel, "Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook," *Virtual Reality*, vol. 25, no. 4, pp. 1153–1170, 2021.

[37] D. Saredakis, A. Szpak, B. Birckhead, H. A. Keage, A. Rizzo, and T. Loetscher, "Factors associated with virtual reality sickness in

head-mounted displays: a systematic review and meta-analysis," *Frontiers in human neuroscience*, vol. 14, p. 96, 2020.

[38] A. D. Souchet, D. Lourdeaux, A. Pagani, and L. Rebenitsch, "A narrative review of immersive virtual reality's ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload," *Virtual Reality*, vol. 27, no. 1, pp. 19–50, 2023.

[39] D. Opriş, S. Pintea, A. García-Palacios, C. Botella, Ş. Szamosközi, and D. David, "Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis," *Depression and anxiety*, vol. 29, no. 2, pp. 85–93, 2012.

[40] T. F. Wechsler, F. Kümpers, and A. Mühlberger, "Inferiority or even superiority of virtual reality exposure therapy in phobias?—a systematic review and quantitative meta-analysis on randomized controlled trials specifically comparing the efficacy of virtual reality exposure to gold standard in vivo exposure in agoraphobia, specific phobia, and social phobia," *Frontiers in psychology*, p. 1758, 2019.

[41] N. Morina, H. Ijntema, K. Meyerbröker, and P. M. Emmelkamp, "Can virtual reality exposure therapy gains be generalized to real-life? a meta-analysis of studies applying behavioral assessments," *Behaviour research and therapy*, vol. 74, pp. 18–24, 2015.

[42] C. Botella, J. Fernández-Álvarez, V. Guillén, A. García-Palacios, and R. Baños, "Recent progress in virtual reality exposure therapy for phobias: a systematic review," *Current psychiatry reports*, vol. 19, pp. 1–13, 2017.

[43] T. Horigome, S. Kurokawa, K. Sawada, S. Kudo, K. Shiga, M. Mimura, and T. Kishimoto, "Virtual reality exposure therapy for social anxiety disorder: a systematic review and meta-analysis," *Psychological medicine*, vol. 50, no. 15, pp. 2487–2497, 2020.

[44] K. P. Wong, C. Y. Y. Lai, and J. Qin, "Systematic review and meta-analysis of randomised controlled trials for evaluating the effectiveness of virtual reality therapy for social anxiety disorder," *Journal of Affective Disorders*, 2023.

[45] I. L. Kampmann, P. M. Emmelkamp, and N. Morina, "Meta-analysis of technology-assisted interventions for social anxiety disorder," *Journal of anxiety disorders*, vol. 42, pp. 71–84, 2016.

[46] R. K. Chesham, J. M. Malouff, and N. S. Schutte, "Meta-analysis of the efficacy of virtual reality exposure therapy for social anxiety," *Behaviour Change*, vol. 35, no. 3, pp. 152–166, 2018.

[47] P. M. Emmelkamp, K. Meyerbröker, and N. Morina, "Virtual reality therapy in social anxiety disorder," *Current psychiatry reports*, vol. 22, pp. 1–9, 2020.

[48] M. H. Lim, V. Aryadoust, and G. Esposito, "A meta-analysis of the effect of virtual reality on reducing public speaking anxiety," *Current Psychology*, vol. 42, no. 15, pp. 12912–12928, 2023.

[49] R. Reeves, D. Curran, A. Gleeson, and D. Hanna, "A meta-analysis of the efficacy of virtual reality and in vivo exposure therapy as psychological interventions for public speaking anxiety," *Behavior Modification*, vol. 46, no. 4, pp. 937–965, 2022.

[50] M. B. Toffolo, J. R. Fehribach, C. P. Van Klaveren, I. Cornelisz, A. Van Straten, J.-L. Van Gelder, and T. Donker, "Automated app-based augmented reality cognitive behavioral therapy for spider phobia: Study protocol for a randomized controlled trial," *Plos one*, vol. 17, no. 7, p. e0271175, 2022.

[51] A. Zimmer, N. Wang, M. K. Ibach, B. Fehlmann, N. S. Schicktanz, D. Bentz, T. Michael, A. Papassotiropoulos, and D. J. de Quervain, "Effectiveness of a smartphone-based, augmented reality exposure

app to reduce fear of spiders in real-life: A randomized controlled trial," *Journal of anxiety disorders*, vol. 82, p. 102442, 2021.

[52] M. Wrzesien, J. Bretón-López, C. Botella, J.-M. Burkhardt, M. Alcañiz, M. Á. Pérez-Ara, and A. R. Del Amo, "How technology influences the therapeutic process: evaluation of the patient-therapist relationship in augmented reality exposure therapy and in vivo exposure therapy," *Behavioural and cognitive psychotherapy*, vol. 41, no. 4, pp. 505–509, 2013.

[53] M. Palau-Batet, J. Bretón-López, J. Grimaldos, L. Díaz-Sanahuja, and S. Quero, "Improving the efficacy of exposure therapy using projection-based augmented reality for the treatment of cockroach phobia: a randomised clinical trial protocol," *BMJ open*, vol. 13, no. 5, p. e069026, 2023.

[54] C. Botella, J. Breton-Lopez, S. Quero, R. M. Baños, A. Garcia-Palacios, I. Zaragoza, and M. Alcañiz, "Treating cockroach phobia using a serious game on a mobile phone and augmented reality exposure: A single case study," *Computers in Human Behavior*, vol. 27, no. 1, pp. 217–227, 2011.

[55] I. Alsina-Jurnet, C. Carvallo-Beciu, and . J. Gutiérrez-Maldonado, "Validity of virtual reality as a method of exposure in the treatment of test anxiety," *Behavior Research Methods*, vol. 39, no. 4, pp. 844–851, 2007.

[56] J. H. Kwon, N. Hong, K. Kim, J. Heo, J.-J. Kim, and E. Kim, "Feasibility of a virtual reality program in managing test anxiety: a pilot study," *Cyberpsychology, Behavior, and Social Networking*, vol. 23, no. 10, pp. 715–720, 2020.

[57] D. Luo, X.-l. Deng, Y.-w. Luo, and G.-x. Wang, "Design and implementation of virtual examination system based on unity 3d," in *Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing*, pp. 1–8, 2019.

[58] L. A. Burke-Smalley, "Using oral exams to assess communication skills in business courses," *Business and Professional Communication Quarterly*, vol. 77, no. 3, pp. 266–280, 2014.

[59] H. Gharibyan, "Assessing students' knowledge: oral exams vs. written tests," *ACM SIGCSE Bulletin*, vol. 37, no. 3, pp. 143–147, 2005.

[60] M. Huxham, F. Campbell, and J. Westwood, "Oral versus written assessments: A test of student performance and attitudes," *Assessment & Evaluation in Higher Education*, vol. 37, no. 1, pp. 125–136, 2012.

[61] S. Pastore and M. Pentassuglia, "Teachers' and students' conceptions of assessment within the italian higher education system," *Practitioner Research in Higher Education*, vol. 10, no. 1, pp. 109–120, 2016.

[62] R. Innes, "Italian education: Between reform and restoration," *Education and Europe: The politics of austerity*, pp. 55–85, 2013.

[63] R. Ehrlich, "Giving bonus points based on oral exams," *American Journal of Physics*, vol. 75, no. 4, pp. 374–376, 2007.

[64] L. R. Murillo-Zamorano and M. Montanero, "Oral presentations in higher education: a comparison of the impact of peer and teacher feedback," *Assessment & Evaluation in Higher Education*, vol. 43, no. 1, pp. 138–150, 2018.

[65] J. R. Sparfeldt, D. H. Rost, U. M. Baumeister, and O. Christ, "Test anxiety in written and oral examinations," *Learning and Individual Differences*, vol. 24, pp. 198–203, 2013.

[66] D. Hartanto, I. L. Kampmann, N. Morina, P. G. Emmelkamp, M. A. Neerincx, and W.-P. Brinkman, "Controlling social stress in virtual reality environments," *PloS one*, vol. 9, no. 3, p. e92804, 2014.

[67] N. Morina, W.-P. Brinkman, D. Hartanto, and P. M. Emmelkamp, "Sense of presence and anxiety during virtual social interactions between a human and virtual humans," *PeerJ*, vol. 2, p. e337, 2014.

[68] J. H. Kwon, C. Alan, S. Czanner, G. Czanner, and J. Powell, "A study of visual perception: Social anxiety and virtual realism," in *Proceedings of the 25th Spring Conference on Computer Graphics*, pp. 167–172, 2009.

[69] I. L. Kampmann, P. M. Emmelkamp, D. Hartanto, W.-P. Brinkman, B. J. Zijlstra, and N. Morina, "Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial," *Behaviour research and therapy*, vol. 77, pp. 147–156, 2016.

[70] S. Bouchard, S. Dumoulin, G. Robillard, T. Guitard, E. Klinger, H. Forget, C. Loranger, and F. X. Roucaut, "Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial," *The British Journal of Psychiatry*, vol. 210, no. 4, pp. 276–283, 2017.

[71] N. H. Zainal, W. W. Chan, A. P. Saxena, C. B. Taylor, and M. G. Newman, "Pilot randomized trial of self-guided virtual reality exposure therapy for social anxiety disorder," *Behaviour research and therapy*, vol. 147, p. 103984, 2021.

[72] O. for Economic Co-operation and Development, *Are Students Happy?: PISA 2015 Results: Students' Well-being*. OECD Publishing, 2017.

[73] M. B. Powers, N. F. Briceno, R. Gresham, E. N. Jouriles, P. M. Emmelkamp, and J. A. Smits, "Do conversations with virtual avatars increase feelings of social anxiety?," *Journal of anxiety disorders*, vol. 27, no. 4, pp. 398–403, 2013.

[74] C. Qu, W.-P. Brinkman, Y. Ling, P. Wiggers, and I. Heynderickx, "Conversations with a virtual human: Synthetic emotions and hu-

man responses," *Computers in Human Behavior*, vol. 34, pp. 58–68, 2014.

[75] F. Mostajeran, M. B. Balci, F. Steinicke, S. Kühn, and J. Gallinat, "The effects of virtual audience size on social anxiety during public speaking," in *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 303–312, IEEE, 2020.

[76] H. Kim, J. E. Shin, Y.-J. Hong, Y.-B. Shin, Y. S. Shin, K. Han, J.-J. Kim, and S.-H. Choi, "Aversive eye gaze during a speech in virtual environment in patients with social anxiety disorder," *Australian & New Zealand Journal of Psychiatry*, vol. 52, no. 3, pp. 279–285, 2018.

[77] D.-P. Pertaub, M. Slater, and C. Barker, "An experiment on public speaking anxiety in response to three different types of virtual audience," *Presence*, vol. 11, no. 1, pp. 68–78, 2002.

[78] M. Slater, D.-P. Pertaub, C. Barker, and D. M. Clark, "An experimental study on fear of public speaking using a virtual environment," *CyberPsychology & Behavior*, vol. 9, no. 5, pp. 627–633, 2006.

[79] O. D. Kothgassner, A. Felnhofer, H. Hlavacs, L. Beutl, R. Palme, I. Kryspin-Exner, and L. M. Glenk, "Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment," *Computers in human behavior*, vol. 62, pp. 124–135, 2016.

[80] P. L. Anderson, E. Zimand, L. F. Hodges, and B. O. Rothbaum, "Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure," *Depression and anxiety*, vol. 22, no. 3, pp. 156–158, 2005.

[81] M. Takac, J. Collett, K. J. Blom, R. Conduit, I. Rehm, and A. De Foe, "Public speaking anxiety decreases within repeated virtual reality training sessions," *PloS one*, vol. 14, no. 5, p. e0216288, 2019.

[82] P. Lindner, J. Dagöö, W. Hamilton, A. Miloff, G. Andersson, A. Schill, and P. Carlbring, "Virtual reality exposure therapy for public speaking anxiety in routine care: a single-subject effectiveness trial," *Cognitive Behaviour Therapy*, vol. 50, no. 1, pp. 67–87, 2021.

[83] P. Premkumar, N. Heym, D. J. Brown, S. Battersby, A. Sumich, B. Huntington, R. Daly, and E. Zysk, "The effectiveness of self-guided virtual-reality exposure therapy for public-speaking anxiety," *Frontiers in psychiatry*, vol. 12, p. 694610, 2021.

[84] Y. Huang, E. Richter, T. Kleickmann, and D. Richter, "Class size affects preservice teachers' physiological and psychological stress reactions: An experiment in a virtual reality classroom," *Computers & Education*, vol. 184, p. 104503, 2022.

[85] S. R. Harris, R. L. Kemmerling, and M. M. North, "Brief virtual reality therapy for public speaking anxiety," *Cyberpsychology & behavior*, vol. 5, no. 6, pp. 543–550, 2002.

[86] M. D. Nazligul, M. Yilmaz, U. Gulec, M. A. Gozcu, R. V. O'Connor, and P. M. Clarke, "Overcoming public speaking anxiety of software engineers using virtual reality exposure therapy," in *Systems, Software and Services Process Improvement: 24th European Conference, EuroSPI 2017, Ostrava, Czech Republic, September 6–8, 2017, Proceedings 24*, pp. 191–202, Springer, 2017.

[87] S. Kahlon, P. Lindner, and T. Nordgreen, "Virtual reality exposure therapy for adolescents with fear of public speaking: a non-randomized feasibility and pilot study," *Child and adolescent psychiatry and mental health*, vol. 13, no. 1, pp. 1–10, 2019.

[88] Ï. Valls-Ratés, O. Niebuhr, and P. Prieto, "Unguided virtual-reality training can enhance the oral presentation skills of high-school students," *Frontiers in Communication*, vol. 7, p. 196, 2022.

[89] M. Rubin, S. Minns, K. Muller, M. H. Tong, M. M. Hayhoe, and M. J. Telch, "Avoidance of social threat: Evidence from eye movements during a public speaking challenge using 360-video," *Behaviour research and therapy*, vol. 134, p. 103706, 2020.

[90] M. Lin, S. G. Hofmann, M. Qian, S. Kind, and H. Yu, "Attention allocation in social anxiety during a speech," *Cognition and Emotion*, vol. 30, no. 6, pp. 1122–1136, 2016.

[91] N. T. M. Chen, L. M. Thomas, P. J. F. Clarke, I. B. Hickie, and A. J. Guastella, "Hyperscanning and avoidance in social anxiety disorder: The visual scanpath during public speaking," *Psychiatry research*, vol. 225, no. 3, pp. 667–672, 2015.

[92] F. Palmas, J. Cichor, D. A. Plecher, and G. Klinker, "Acceptance and effectiveness of a virtual reality public speaking training," in *2019 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 363–371, IEEE, 2019.

[93] P. Raimbaud, A. Jovane, K. Zibrek, . C. Pacchierotti, M. Christie, L. Hoyet, J. Pettre, and A.-H. Olivier, "The stare-in-the-crowd effect in virtual reality," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 281–290, IEEE, 2022.

[94] B. Yu, M. Funk, J. Hu, Q. Wang, and L. Feijs, "Biofeedback for everyday stress management: A systematic review," *Frontiers in ICT*, vol. 5, p. 23, 2018.

[95] R. Jerath, M. W. Crawford, V. A. Barnes, and K. Harden, "Self-regulation of breathing as a primary treatment for anxiety," *Applied psychophysiology and biofeedback*, vol. 40, no. 2, pp. 107–115, 2015.

[96] V. Busch, W. Magerl, U. Kern, J. Haas, G. Hajak, and P. Eichhammer, "The effect of deep and slow breathing on pain perception, autonomic activity, and mood processing—an experimental study," *Pain Medicine*, vol. 13, no. 2, pp. 215–228, 2012.

[97] S. I. Hopper, S. L. Murray, L. R. Ferrara, and J. K. Singleton, "Effectiveness of diaphragmatic breathing for reducing physiological and psychological stress in adults: a quantitative systematic review," *JBI Evidence Synthesis*, vol. 17, no. 9, pp. 1855–1876, 2019.

[98] J. Blum, C. Rockstroh, and A. S. Göritz, "Development and pilot test of a virtual reality respiratory biofeedback approach," *Applied Psychophysiology and Biofeedback*, vol. 45, pp. 153–163, 2020.

[99] A. Gorini, F. Pallavicini, D. Algeri, C. Repetto, A. Gaggioli, and G. Riva, "Virtual reality in the treatment of generalized anxiety disorders," *Annual Review of Cybertherapy and Telemedicine 2010*, pp. 39–43, 2010.

[100] V. G. Prabhu, L. M. Stanley, C. Linder, and R. Morgan, "Analyzing the efficacy of a restorative virtual reality environment using hrv biofeedback for pain and anxiety management," in *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, pp. 1–4, IEEE, 2020.

[101] T. Sonne and M. M. Jensen, "Chillfish: A respiration game for children with adhd," in *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, pp. 271–278, 2016.

[102] R. S. Venuturupalli, T. Chu, M. Vicari, A. Kumar, N. Fortune, and B. Spielberg, "Virtual reality–based biofeedback and guided meditation in rheumatology: a pilot study," *ACR open rheumatology*, vol. 1, no. 10, pp. 667–675, 2019.

[103] L. Chittaro and R. Sioni, "Affective computing vs. affective placebo: Study of a biofeedback-controlled game for relaxation training," *International Journal of Human-Computer Studies*, vol. 72, no. 8-9, pp. 663–673, 2014.

[104] E. A. Schoneveld, M. Malmberg, A. Lichtwarck-Aschoff, G. P. Verheijen, R. C. Engels, and I. Granic, "A neurofeedback video game (mindlight) to prevent anxiety in children: A randomized controlled trial," *Computers in Human Behavior*, vol. 63, pp. 321—333, 2016.

[105] A. A. Schuurmans, K. S. Nijhof, I. P. Vermaes, R. C. Engels, and I. Granic, "A pilot study evaluating "dojo," a videogame intervention for youths with externalizing and anxiety problems," *Games for health journal*, vol. 4, no. 5, pp. 401–408, 2015.

[106] A. M. Tinga, I. Nyklíček, M. P. Jansen, T. T. de Back, and M. M. Louwerse, "Respiratory biofeedback does not facilitate lowering arousal in meditation through virtual reality," *Applied psychophysiology and biofeedback*, vol. 44, pp. 51–59, 2019.

[107] J. T. Reason and J. J. Brand, *Motion sickness.* Academic press, 1975.

[108] J. E. Staddon and D. T. Cerutti, "Operant conditioning," *Annual review of psychology*, vol. 54, no. 1, pp. 115–144, 2003.

[109] A. Lobel, M. Gotsis, E. Reynolds, M. Annetta, R. C. Engels, and I. Granic, "Designing and utilizing biofeedback games for emotion regulation: The case of nevermind," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1945–1951, 2016.

[110] J. Z. Goldenberg, M. Brignall, M. Hamilton, J. Beardsley, R. D. Batson, J. Hawrelak, B. Lichtenstein, and B. C. Johnston, "Biofeedback for treatment of irritable bowel syndrome," *Cochrane Database of Systematic Reviews*, no. 11, 2019.

[111] J. Greenhalgh, R. Dickson, and Y. Dundar, "Biofeedback for hypertension: a systematic review," *Journal of hypertension*, vol. 28, no. 4, pp. 644–652, 2010.

[112] J. Greenhalgh, R. Dickson, and Y. Dundar, "The effects of biofeedback for the treatment of essential hypertension: a systematic review.," *Health technology assessment (Winchester, England)*, vol. 13, no. 46, pp. 1–104, 2009.

[113] S. Woodward, C. Norton, and P. Chiarelli, "Biofeedback for treatment of chronic idiopathic constipation in adults," *Cochrane Database of Systematic Reviews*, no. 3, 2014.

[114] H. Scholten, M. Malmberg, A. Lobel, R. C. Engels, and I. Granic, "A randomized controlled trial to test the effectiveness of an immersive 3d video game for anxiety prevention among adolescents," *PloS one*, vol. 11, no. 1, p. e0147763, 2016.

[115] R. Bossenbroek, A. Wols, J. Weerdmeester, A. Lichtwarck-Aschoff, I. Granic, M. M. van Rooij, *et al.*, "Efficacy of a virtual reality biofeedback game (deep) to reduce anxiety and disruptive classroom behavior: single-case study," *JMIR mental health*, vol. 7, no. 3, p. e16066, 2020.

[116] C. Repetto, A. Gaggioli, F. Pallavicini, P. Cipresso, S. Raspelli, and G. Riva, "Virtual reality and mobile phones in the treatment of generalized anxiety disorders: a phase-2 clinical trial," *Personal and Ubiquitous Computing*, vol. 17, pp. 253–260, 2013.

[117] L. Chittaro and M. Serafini, "Desktop virtual reality as an exposure method for test anxiety: quantitative and qualitative feasibility study," *Multimedia Tools and Applications*, 2023.

[118] M. Serafini and L. Chittaro, "Trial of a desktop virtual reality application as a method of exposure for test anxiety: a qualitative study," *Behaviour & Information Technology*, pp. 1–13, 2024.

[119] P. E. Bull, *Posture & gesture.* Pergamon Press, 1987.

[120] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15–33, 2012.

[121] A. Mehrabian, "Relationship of attitude to seated posture, orientation, and distance.," *Journal of personality and social psychology*, vol. 10, no. 1, p. 26, 1968.

[122] L. Tickle-Degnen and R. Rosenthal, "The nature of rapport and its nonverbal correlates," *Psychological inquiry*, vol. 1, no. 4, pp. 285–293, 1990.

[123] K.-A. Veljaca and R. M. Rapee, "Detection of negative and positive audience behaviours by socially anxious subjects," *Behaviour Research and therapy*, vol. 36, no. 3, pp. 311–321, 1998.

[124] M. Chollet and S. Scherer, "Perception of virtual audiences," *IEEE computer graphics and applications*, vol. 37, no. 4, pp. 50–59, 2017.

[125] Y. Glémarec, J.-L. Lugrin, A.-G. Bosser, b. A. Collins Jackson, C. Buche, and M. E. Latoschik, "Indifferent or enthusiastic? virtual audiences animation and perception in virtual reality," *Frontiers in Virtual Reality*, vol. 2, p. 666232, 2021.

[126] N. Kang, W.-P. Brinkman, M. B. Van Riemsdijk, and M. Neerincx, "The design of virtual audiences: noticeable and recognizable behavioral styles," *Computers in Human Behavior*, vol. 55, pp. 680–694, 2016.

[127] H. M. Rosenfeld, "Instrumental affiliative functions of facial and gestural expressions.," *Journal of Personality and Social Psychology*, vol. 4, no. 1, p. 65, 1966.

[128] C. Sica, I. Musoni, L. Chiri, B. Bisi, V. Lolli, and C. Sighinolfi, "Social phobia scale (sps) and social interaction anxiety scale (sias): Italian translation and adaptation," *Bollettino di Psicologia Applicata*, vol. 252, pp. 59–71, 2007.

[129] R. P. Mattick and J. C. Clarke, "Development and validation of measures of social phobia scrutiny fear and social interaction anxiety," *Behaviour research and therapy*, vol. 36, no. 4, pp. 455–470, 1998.

[130] A. Felnhofer, H. Hlavacs, L. Beutl, I. Kryspin-Exner, and O. D. Kothgassner, "Physical presence, social presence, and anxiety in participants with social anxiety disorder during virtual cue exposure," *Cyberpsychology, Behavior, and Social Networking*, vol. 22, no. 1, pp. 46–50, 2019.

[131] H. E. Kim, Y.-J. Hong, M.-K. Kim, Y. H. Jung, S. Kyeong, and J.-J. Kim, "Effectiveness of self-training using the mobile-based virtual reality program in patients with social anxiety disorder," *Computers in Human Behavior*, vol. 73, pp. 614–619, 2017.

[132] H.-J. Kim, S. Lee, D. Jung, J.-W. Hur, H.-J. Lee, S. Lee, G. J. Kim, C.-Y. Cho, S. Choi, S.-M. Lee, *et al.*, "Effectiveness of a participatory and interactive virtual reality intervention in patients with social anxiety disorder: longitudinal questionnaire study," *Journal of medical Internet research*, vol. 22, no. 10, p. e23024, 2020.

[133] N. Morina, W.-P. Brinkman, D. Hartanto, I. L. Kampmann, and P. M. Emmelkamp, "Social interactions in virtual reality exposure therapy: A proof-of-concept pilot study," *Technology and Health Care*, vol. 23, no. 5, pp. 581–589, 2015.

[134] R. Abend, O. Dan, K. Maoz, S. Raz, and Y. Bar-Haim, "Reliability, validity and sensitivity of a computerized visual analog scale measuring state anxiety," *Journal of behavior therapy and experimental psychiatry*, vol. 45, no. 4, pp. 447–453, 2014.

[135] H. M. Davey, A. L. Barratt, P. N. Butow, and J. J. Deeks, "A one-item question with a likert or visual analog scale adequately measured current anxiety," *Journal of clinical epidemiology*, vol. 60, no. 4, pp. 356–360, 2007.

[136] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[137] G. Guest, K. M. MacQueen, and E. E. Namey, *Applied thematic analysis*. sage publications, 2011.

[138] J. T. DeCuir-Gunby, P. L. Marshall, and A. W. McCulloch, "Developing and using a codebook for the analysis of interview data: An example from a professional development research project," *Field methods*, vol. 23, no. 2, pp. 136–155, 2011.

[139] R. Rampin and V. Rampin, "Taguette: open-source qualitative data analysis," *Journal of Open Source Software*, vol. 6, no. 68, p. 3522, 2021.

[140] M. Slater, D.-P. Pertaub, and A. Steed, "Public speaking in virtual reality: Facing an audience of avatars," *IEEE Computer Graphics and Applications*, vol. 19, no. 2, pp. 6–9, 1999.

[141] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[142] J. Cohen, "Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit.," *Psychological bulletin*, vol. 70, no. 4, p. 213, 1968.

[143] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[144] D. P. Pertaub, M. Slater, and C. Barker, "An experiment on fear of public speaking in virtual reality," in *Medicine Meets Virtual Reality 2001*, pp. 372–378, IOS Press, 2001.

[145] A. Barrett, A. Pack, D. Monteiro, and H.-N. Liang, "Exploring the influence of audience familiarity on speaker anxiety and performance in virtual reality and real-life presentation contexts," *Behaviour & Information Technology*, pp. 1–13, 2023.

[146] C. P. McLean and E. R. Anderson, “Brave men and timid women? a review of the gender differences in fear and anxiety,” *Clinical psychology review*, vol. 29, no. 6, pp. 496–505, 2009.

[147] M. Grossman and W. Wood, “Sex differences in intensity of emotional experience: a social role interpretation.,” *Journal of personality and social psychology*, vol. 65, no. 5, p. 1010, 1993.

[148] U. Hess, S. Senécal, G. Kirouac, P. Herrera, P. Philippot, and R. E. Kleck, “Emotional expressivity in men and women: Stereotypes and self-perceptions,” *Cognition & Emotion*, vol. 14, no. 5, pp. 609–642, 2000.

[149] M. I. Núñez-Peña, M. Suárez-Pellicioni, and R. Bono, “Gender differences in test anxiety and their impact on higher education students’ academic achievement,” *Procedia-Social and Behavioral Sciences*, vol. 228, pp. 154–160, 2016.

[150] D. L. Bandalos, K. Yates, and T. Thorndike-Christ, “Effects of math self-concept, perceived self-efficacy, and attributions for failure and success on test anxiety.,” *Journal of educational psychology*, vol. 87, no. 4, p. 611, 1995.

[151] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *biometrics*, pp. 159–174, 1977.

[152] A. J. Cavanagh, X. Chen, M. Bathgate, J. Frederick, D. I. Hanauer, and M. J. Graham, “Trust, growth mindset, and student commitment to active learning in a college science course,” *CBE—Life Sciences Education*, vol. 17, no. 1, p. ar10, 2018.

[153] E. Klinger, S. Bouchard, P. Légeron, S. Roy, F. Lauer, I. Chemin, and P. Nugues, “Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study,” *Cyberpsychology & behavior*, vol. 8, no. 1, pp. 76–88, 2005.

[154] G. M. Lucas, J. Mell, J. Boberg, F. Zenone, E. J. de Visser, C. Tossell, and T. Seech, "Customizing virtual interpersonal skills training applications may not improve trainee performance," *Scientific Reports*, vol. 13, no. 1, p. 78, 2023.

[155] T. E. Smith and A. B. Frymier, "Get 'real': Does practicing speeches before an audience improve performance?," *Communication Quarterly*, vol. 54, no. 1, pp. 111–125, 2006.

[156] M. Serafini and L. Chittaro, "Perception of virtual agents as communicators in virtual vs. augmented reality by a male sample," in *International Conference on Persuasive Technology*, pp. 36–49, Springer, 2023.

[157] M. Serafini and L. Chittaro, "Augmented reality vs. virtual reality in exposing users to public speaking scenarios," *Submitted to journal*, 2023.

[158] N. A. De Witte, S. Scheveneels, R. Sels, G. Debard, D. Hermans, and T. Van Daele, "Augmenting exposure therapy: Mobile augmented reality for specific phobia," *Frontiers in Virtual Reality*, p. 8, 2020.

[159] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior research methods*, vol. 39, no. 2, pp. 175–191, 2007.

[160] E. M. Bartholomay and D. D. Houlihan, "Public speaking anxiety scale: Preliminary psychometric data and scale validation," *Personality and individual differences*, vol. 94, pp. 211–215, 2016.

[161] G. Stoet, "Psytoolkit: A software package for programming psychological experiments using linux," *Behavior research methods*, vol. 42, pp. 1096–1104, 2010.

[162] G. Stoet, "Psytoolkit: A novel web-based method for running online questionnaires and reaction-time experiments," *Teaching of Psychology*, vol. 44, no. 1, pp. 24–31, 2017.

[163] C. D. Spielberger, *Manual for the State-Trait Anxiety Inventory (self-evaluation questionnaire)*. Consulting Psychologists Press, 1970.

[164] G. Westheimer, "Eye movement responses to a horizontally moving visual stimulus," *AMA archives of ophthalmology*, vol. 52, no. 6, pp. 932–941, 1954.

[165] B. H. Cohen, *Explaining psychological statistics*. John Wiley & Sons, 2008.

[166] S. Q. Hussain, N. Akhtar, N. Shabbir, N. Aslam, and S. Arshad, "Causes and strategies to cope english language speaking anxiety in pakistani university students," *Humanities & Social Sciences Reviews*, vol. 9, no. 3, pp. 579–59, 2021.

[167] N. Kang, W.-P. Brinkman, M. B. van Riemsdijk, and M. A. Neerincx, "An expressive virtual audiencewith flexible behavioral styles," *IEEE Transactions on Affective Computing*, vol. 4, no. 4, pp. 326–340, 2013.

[168] R. R. Behnke and C. R. Sawyer, "Anticipatory anxiety patterns for male and female public speakers," *Communication education*, vol. 49, no. 2, pp. 187–195, 2000.

[169] A. Gaibani and F. Elmenfi, "The role of gender in influencing public speaking anxiety," *British Journal of English Linguistics*, vol. 2, no. 3, pp. 7–13, 2014.

[170] L. Chittaro, M. Serafini, and Y. Vulcano, "Virtual reality experiences for breathing and relaxation training: the effects of real vs. placebo biofeedback," *Submitted to journal*, 2023.

[171] T. Gao, T. Zhang, L. Zhu, Y. Gao, and L. Qiu, "Exploring psychophysiological restoration and individual preference in the different environments based on virtual reality," *International journal of environmental research and public health*, vol. 16, no. 17, p. 3102, 2019.

[172] M. Annerstedt, P. Jönsson, M. Wallergård, G. Johansson, B. Karlson, P. Grahn, Å. M. Hansen, and P. Währborg, "Inducing physiological stress recovery with sounds of nature in a virtual reality forest—results from a pilot study," *Physiology & behavior*, vol. 118, pp. 240–250, 2013.

[173] D. Gromala, X. Tong, A. Choo, M. Karamnejad, and C. D. Shaw, "The virtual meditative walk: virtual reality therapy for chronic pain management," in *Proceedings of the 33rd Annual ACM conference on human factors in computing systems*, pp. 521–524, 2015.

[174] S. Gradl, M. Wirth, T. Zillig, and B. M. Eskofier, "Visualization of heart activity in virtual reality: A biofeedback application using wearable sensors," in *2018 IEEE 15th international conference on wearable and implantable body sensor networks (BSN)*, pp. 152–155, IEEE, 2018.

[175] A. Parnandi and R. Gutierrez-Osuna, "Partial reinforcement in game biofeedback for relaxation training," *IEEE Transactions on Affective Computing*, vol. 12, no. 1, pp. 141–153, 2018.

[176] A. Parnandi and R. Gutierrez-Osuna, "Visual biofeedback and game adaptation in relaxation skill transfer," *IEEE Transactions on Affective Computing*, vol. 10, no. 2, pp. 276–289, 2017.

[177] R. Berto, "The role of nature in coping with psycho-physiological stress: A literature review on restorativeness," *Behavioral sciences*, vol. 4, no. 4, pp. 394–409, 2014.

[178] S. Kaplan, "The restorative benefits of nature: Toward an integrative framework," *Journal of environmental psychology*, vol. 15, no. 3, pp. 169–182, 1995.

[179] A. P. Anderson, M. D. Mayer, A. M. Fellows, D. R. Cowan, M. T. Hegel, and J. C. Buckey, "Relaxation with immersive natural scenes presented using virtual reality," *Aerospace medicine and human performance*, vol. 88, no. 6, pp. 520–526, 2017.

[180] D. Villani and G. Riva, "Does interactive media enhance the management of stress? suggestions from a controlled study," *Cyberpsychology, Behavior, and Social Networking*, vol. 15, no. 1, pp. 24–30, 2012.

[181] F. Buttussi and L. Chittaro, "Effects of different types of virtual reality display on presence and learning in a safety training scenario," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 2, pp. 1063–1076, 2017.

[182] J. J. Cummings and J. N. Bailenson, "How immersive is enough? a meta-analysis of the effect of immersive technology on user presence," *Media psychology*, vol. 19, no. 2, pp. 272–309, 2016.

[183] G. Makransky, T. S. Terkildsen, and R. E. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning," *Learning and instruction*, vol. 60, pp. 225–236, 2019.

[184] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 266–281, 2001.

[185] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the panas scales.," *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.

[186] M. Benedek and C. Kaernbach, "Decomposition of skin conductance data by means of nonnegative deconvolution," *psychophysiology*, vol. 47, no. 4, pp. 647–658, 2010.

[187] D. Makowski, T. Pham, Z. J. Lau, . J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. A. Chen, "Neurokit2: A python toolbox for neurophysiological signal processing," *Behavior research methods*, pp. 1–8, 2021.

[188] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.

[189] M. Merri, D. C. Farden, J. G. Mottley, and E. L. Titlebaum, "Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability," *IEEE transactions on biomedical engineering*, vol. 37, no. 1, pp. 99–106, 1990.

[190] H. R. Variability, "Standards of measurement, physiological interpretation, and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Eur Heart J*, vol. 17, no. 3, pp. 354–381, 1996.

[191] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, *Using multivariate statistics*, vol. 6. pearson Boston, MA, 2013.

[192] R. Kaplan and S. Kaplan, *The experience of nature: A psychological perspective*. Cambridge university press, 1989.

[193] H. Ohly, M. P. White, B. W. Wheeler, A. Bethel, O. C. Ukoumunne, V. Nikolaou, and R. Garside, "Attention restoration theory: A systematic review of the attention restoration potential of exposure to natural environments," *Journal of Toxicology and Environmental Health, Part B*, vol. 19, no. 7, pp. 305–343, 2016.

[194] D. Valtchanov, K. R. Barton, and C. Ellard, "Restorative effects of virtual nature settings," *Cyberpsychology, Behavior, and Social Networking*, vol. 13, no. 5, pp. 503–512, 2010.

[195] R. Fried, *The psychology and physiology of breathing: In behavioral medicine, clinical psychology, and psychiatry*. Springer Science & Business Media, 1993.

[196] X. Ma, Z.-Q. Yue, Z.-Q. Gong, H. Zhang, N.-Y. Duan, Y.-T. Shi, G.-X. Wei, and Y.-F. Li, "The effect of diaphragmatic breathing on attention, negative affect and stress in healthy adults," *Frontiers in psychology*, vol. 8, p. 234806, 2017.

[197] A. Gaume, A. Vialatte, A. Mora-Sánchez, C. Ramdani, and F.-B. Vialatte, "A psychoengineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback," *Neuroscience & Biobehavioral Reviews*, vol. 68, pp. 891–910, 2016.

[198] H. Huang, S. L. Wolf, and J. He, "Recent developments in biofeedback for neuromotor rehabilitation," *Journal of neuroengineering and rehabilitation*, vol. 3, pp. 1–12, 2006.

[199] F. Soyka, M. Leyrer, J. Smallwood, C. Ferguson, B. E. Riecke, and B. J. Mohler, "Enhancing stress management techniques using virtual reality," in *Proceedings of the ACM symposium on applied perception*, pp. 85–88, 2016.

[200] G. Riva, F. Mantovani, C. S. Capideville, A. Preziosa, F. Morganti, D. Villani, A. Gaggioli, C. Botella, and M. Alcañiz, "Affective interactions using virtual reality: the link between presence and emotions," *Cyberpsychology & behavior*, vol. 10, no. 1, pp. 45–56, 2007.

[201] A. E. Van den Berg, S. L. Koole, and N. Y. van der Wulp, "Environmental preference and restoration:(how) are they related?," *Journal of environmental psychology*, vol. 23, no. 2, pp. 135–146, 2003.

[202] F. S. Mayer, C. M. Frantz, E. Bruehlman-Senecal, and K. Dolliver, "Why is nature beneficial? the role of connectedness to nature," *Environment and behavior*, vol. 41, no. 5, pp. 607–643, 2009.

[203] Y. A. de Kort, A. L. Meijnders, A. A. Sponselee, and W. A. IJsselsteijn, "What's wrong with virtual trees? restoring from stress in a mediated environment," *Journal of environmental psychology*, vol. 26, no. 4, pp. 309–320, 2006.

[204] Y. Ling, H. T. Nefs, N. Morina, I. Heynderickx, and W.-P. Brinkman, "A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders," *PloS one*, vol. 9, no. 5, p. e96144, 2014.

[205] F. Wolfe, D. J. Clauw, M.-A. Fitzcharles, D. L. Goldenberg, R. S. Katz, P. Mease, A. S. Russell, I. J. Russell, J. B. Winfield, and M. B. Yunus, "The american college of rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity," *Arthritis care & research*, vol. 62, no. 5, pp. 600–610, 2010.

[206] M. Kösehasanoğullari, N. E. Gündüz, and E. Akalin, "Is fibromyalgia syndrome a neuropathic pain syndrome?," *Archives of rheumatology*, vol. 34, no. 2, p. 196, 2019.

[207] P. Fietta, P. Fietta, and P. Manganelli, "Fibromyalgia and psychiatric disorders," *Acta Biomedica-Ateneo Parmense*, vol. 78, no. 2, p. 88, 2007.

[208] A. M. Isen, "A role for neuropsychology in understanding the facilitating influence of positive affect on social behavior and cognitive processes," *Handbook of Positive Psychology*, 2009.

[209] B. L. Fredrickson, "The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions.," *American psychologist*, vol. 56, no. 3, p. 218, 2001.

[210] W. Häuser, K. Bernardy, B. Arnold, M. Offenbächer, and M. Schiltenwolf, "Efficacy of multicomponent treatment in fibromyalgia syndrome: a meta-analysis of randomized controlled

clinical trials," *Arthritis Care & Research*, vol. 61, no. 2, pp. 216–224, 2009.

[211] C. Botella, A. Garcia-Palacios, Y. Vizcaíno, R. Herrero, R. M. Baños, and M. A. Belmonte, "Virtual reality in the treatment of fibromyalgia: a pilot study," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 3, pp. 215–223, 2013.

[212] M. Polat, A. Kahveci, B. Muci, Z. Günendi, and G. Kaymak Karataş, "The effect of virtual reality exercises on pain, functionality, cardiopulmonary capacity, and quality of life in fibromyalgia syndrome: a randomized controlled study," *Games for Health Journal*, vol. 10, no. 3, pp. 165–173, 2021.

[213] J. A. Glombiewski, A. T. Sawyer, J. Gutermann, K. Koenig, W. Rief, and S. G. Hofmann, "Psychological treatments for fibromyalgia: a meta-analysis," *PAIN®*, vol. 151, no. 2, pp. 280–295, 2010.

[214] R. Herrero, A. García-Palacios, D. Castilla, G. Molinari, and C. Botella, "Virtual reality for the induction of positive emotions in the treatment of fibromyalgia: a pilot study over acceptability, satisfaction, and the effect of virtual reality on mood," *Cyberpsychology, Behavior, and Social Networking*, vol. 17, no. 6, pp. 379–384, 2014.

[215] C. Gulsen, F. Soke, K. Eldemir, Y. Apaydin, C. Ozkul, A. Guclu-Gunduz, and D. Akcali, "Effect of fully immersive virtual reality treatment combined with exercise in fibromyalgia patients: A randomized controlled trial," *Assistive Technology*, vol. 34, no. 3, pp. 256–263, 2022.

[216] N. Ahmadpour, H. Randall, H. Choksi, A. Gao, C. Vaughan, and P. Poronnik, "Virtual reality interventions for acute and chronic pain management," *The international journal of biochemistry & cell biology*, vol. 114, p. 105568, 2019.

[217] A. J. Busch, S. C. Webber, M. Brachaniec, J. Bidonde, V. D. Bello-Haas, A. D. Danyliw, T. J. Overend, R. S. Richards, A. Sawant, and C. L. Schachter, "Exercise therapy for fibromyalgia," *Current pain and headache reports*, vol. 15, pp. 358–367, 2011.

[218] D. Collado-Mateo, F. J. Dominguez-Muñoz, J. C. Adsuar, M. A. Garcia-Gordillo, and N. Gusi, "Effects of exergames on quality of life, pain, and disease effect in women with fibromyalgia: a randomized controlled trial," *Archives of physical medicine and rehabilitation*, vol. 98, no. 9, pp. 1725–1731, 2017.

[219] D. Collado-Mateo, F. J. Dominguez-Muñoz, J. C. Adsuar, E. Merellano-Navarro, and N. Gusi, "Exergames for women with fibromyalgia: A randomised controlled trial to evaluate the effects on mobility skills, balance and fear of falling," *PeerJ*, vol. 5, p. e3211, 2017.

[220] J. P. Martín-Martínez, S. Villafaina, D. Collado-Mateo, J. Pérez-Gómez, and N. Gusi, "Effects of 24-week exergame intervention on physical function under single-and dual-task conditions in fibromyalgia: A randomized controlled trial," *Scandinavian journal of medicine & science in sports*, vol. 29, no. 10, pp. 1610–1617, 2019.

[221] S. Villafaina, D. Collado-Mateo, F. J. Dominguez-Munoz, N. Gusi, and J. P. Fuentes-Garcia, "Effects of exergames on heart rate variability of women with fibromyalgia: A randomized controlled trial," *Scientific reports*, vol. 10, no. 1, p. 5168, 2020.

[222] J. Mortensen, L. Q. Kristensen, E. P. Brooks, and A. L. Brooks, "Women with fibromyalgia's experience with three motion-controlled video game consoles and indicators of symptom severity and performance of activities of daily living," *Disability and Rehabilitation: Assistive Technology*, vol. 10, no. 1, pp. 61–66, 2015.

[223] E. Petersson Brooks and A. L. Brooks, "Engagement in game-based rehabilitation for women with fibromyalgia syndrome," in *Universal*

*Access in Human-Computer Interaction. Aging and Assistive Environments: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part III 8*, pp. 359–367, Springer, 2014.

[224] M. S. d. Carvalho, L. C. Carvalho, F. d. S. Menezes, A. Frazin, E. d. C. Gomes, and D. H. Iunes, “Effects of exergames in women with fibromyalgia: a randomized controlled study,” *Games for Health Journal*, vol. 9, no. 5, pp. 358–367, 2020.

[225] M. S. de Carvalho, L. C. Carvalho, R. d. S. Alves, F. d. S. Menezes, E. d. C. Gomes, A. Frazin, and D. H. Iunes, “Analysis of the muscular activity, peak torque in the lower limbs, and static balance after virtual rehabilitation in women with fibromyalgia: a randomized controlled study,” *Games for Health Journal*, vol. 10, no. 3, pp. 190–197, 2021.

[226] A. Garcia-Palacios, R. Herrero, Y. Vizcaíno, M. A. Belmonte, D. Castilla, G. Molinari, R. M. Baños, and C. Botella, “Integrating virtual reality with activity management for the treatment of fibromyalgia,” *The Clinical journal of pain*, vol. 31, no. 6, pp. 564–572, 2015.

[227] A. R. Moore, S. Straube, J. Paine, C. J. Phillips, S. Derry, and H. J. McQuay, “Fibromyalgia: moderate and substantial pain intensity reduction predicts improvement in other outcomes and substantial quality of life gain,” *Pain*, vol. 149, no. 2, pp. 360–364, 2010.

[228] A. Michaelides and P. Zis, “Depression, anxiety and acute pain: links and management challenges,” *Postgraduate medicine*, vol. 131, no. 7, pp. 438–444, 2019.

[229] E. Pilchowska-Ujma and J. Krakowiak, “Genetic aspects of pain and its variability in the human population,” *population*, vol. 28, no. 4, pp. 569–574, 2021.

[230] R. Melzack, "The short-form mcgill pain questionnaire," *Pain*, vol. 30, no. 2, pp. 191–197, 1987.

[231] C. S. Burckhardt, S. Clark, , R. M. Bennett, *et al.*, "The fibromyalgia impact questionnaire: development and validation," *J rheumatol*, vol. 18, no. 5, pp. 728–33, 1991.

[232] L. Altan, N. Korkmaz, Ü. Bingol, and B. Gunay, "Effect of pilates training on people with fibromyalgia syndrome: a pilot study," *Archives of physical medicine and rehabilitation*, vol. 90, no. 12, pp. 1983–1988, 2009.

[233] A. M. Castro-Sánchez, G. A. Matarán-Peñarrocha, M. M. López-Rodríguez, I. C. Lara-Palomo, L. Arendt-Nielsen, and C. Fernández-de-las Peñas, "Gender differences in pain severity, disability, depression, and widespread pressure pain sensitivity in patients with fibromyalgia syndrome without comorbid conditions," *Pain Medicine*, vol. 13, no. 12, pp. 1639–1647, 2012.

[234] M. J. Lami, M. P. Martínez, A. I. Sánchez, E. Miró, F. N. Diener, G. Prados, and M. A. Guzmán, "Gender differences in patients with fibromyalgia undergoing cognitive-behavioral therapy for insomnia: preliminary data," *Pain Practice*, vol. 16, no. 2, pp. E23–E34, 2016.

[235] V. C. Goessl, J. E. Curtiss, and S. G. Hofmann, "The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analy- sis," *Psychological medicine*, vol. 47, no. 15, pp. 2578–2586, 2017.