SimulEase: A VR-Assisted Approach to Overcoming Social Anxiety

Delia-Elena Ungureanu
"Alexandru Ioan Cuza" University of
Iasi, Faculty of Computer Science
St. General Henri Mathias Berthelot 16, Iași
deliaungureanu2001@yahoo.com

Adrian Iftene
"Alexandru Ioan Cuza" University of
Iasi, Faculty of Computer Science
St. General Henri Mathias Berthelot 16, Iași
adiftene@gmail.com

ABSTRACT

The aim of this project and thesis is to develop and analyze the results of SimulEase. This virtual reality (VR) program replicates real-life situations that cause anxiety in a controlled, repeatable, and safe setting. Through immersive exposure and AI stress detection (from audio and text user input), the software seeks to assist users in better managing the symptoms of social anxiety and stress associated with public speaking. The participants can become tolerant and learn coping behaviors in a step-by-step process without having to endure the consequences of their behavior in real life by mimicking anxiety-evoking events in a well-controlled, risk-free VR environment. By putting forth an interactive solution that combines immersive technology with real-time emotion recognition for individualized exposure therapy and self-guided support, our study seeks to advance the expanding field of digital mental health solutions. Research shows that VRET (VR Exposure Therapy) is a promising supplementary treatment for SAD (Social Anxiety Disorder), especially when combined with CBT (Cognitive Behavioral Therapy).

Author Keywords

Virtual reality, Exposure therapy, Social anxiety disorder.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces. J.3. Life and Medical Sciences. I.3.7. Three-Dimensional Graphics and Realism

General Terms

Human Factors; Design; Measurement.

DOI: 10.37789/icusi.2025.7

INTRODUCTION

Among the world's most prevalent mental conditions, anxiety disorders have the potential to substantially interfere with a person's capacity for daily functioning, engagement with others, or soothing themselves in mundane situations such as face-to-face communication or speaking publicly. Although effective, traditional treatment methods such as CBT may be restricted by stigma, availability, or the ability of the user to cope with complicated situations in daily life. Our objective in working with SimulEase is to span the gap between therapy intent and actual practice using interactive technology. VRET has also been shown to be an effective treatment for SAD, as research indicates in recent times [1]. VRET consistently reports significant short- and long-term reductions in symptoms of anxiety, a systematic review of VR therapies for SAD showed. Eighteen studies were included in this review, most of which simulated increasingly challenging social encounters using unique VR settings. These therapies, which were normally given between a single and fourteen treatment sessions, demonstrated strong evidence that VR is capable of helping participants face their problems in a progressive, controlled, yet realistic setting. The use of ML algorithms in improving diagnostic accuracy and guiding treatment plans has made artificial intelligence (AI) a game-changer in anxiety detection and treatment. AI platforms are able to accurately detect symptoms of anxiety from diverse modalities such as wearable sensors, smartphone behavior, and physiological signals, research indicates. This extends the use of conventional psychological tests [2-4].

RELATED WORK

Existing Applications

In the past few years, many VR applications have been developed to assist users in keeping their anxiety level down, calming their response, and overall improving their mental state. The applications utilize different virtual environments, which generally replicate real-life situations that can be anxiety-inducing. *oVRcome*[1] is a VR smartphone application that provides exposure therapy through the simulation of environments like public speaking, flying, parties, and heights. Developed with user-friendliness in mind, oVRcome allows users to utilize a smartphone and a basic VR headset, allowing it to be easily used at home. It contains specially developed programs, specifically tailored for children and teenagers, as an attempt to treat phobias in age-appropriate environments. *XRHealth*[2] combines remote clinical monitoring with VR therapy. Contrary to independent experiences, the method of XRHealth is through sessions in which the users practice cognitive behavior and stress relief techniques under the supervision of a therapist or clinician. The platform also uses AI to adjust sessions in real-time to provide optimal user engagement and optimize treatment

[1] https://www.ovrcome.io

[2] https://www.xr.health/us/services/patients/

benefits. *TRIPP*[3] is very different in that it uses meditation and mindfulness in engaged environments. It is a program that offers over 40 richly visual, customizable meditative sessions for increasing concentration, mood, and emotional regulation. It also has features like mood tracking and setting daily mindfulness goals, thus being more of a lifestyle tool than one that is clinically designed as a treatment.

These are some of the applications that demonstrate the growing potential of VR for mental health treatment, even though most current ones either lean toward general mindfulness (e.g., TRIPP) or exposure therapy without physiological feedback (e.g., oVRcome). XRHealth does incorporate clinician involvement, but still lacks real-time adaptation based on biometric data. This thesis aims to address these gaps by integrating real-time physiological stress monitoring within immersive VR environments.

Emotion and Stress Detection Technologies

More precise identification of human emotions and stress levels is now possible in a faster way because of recent developments in wearable technology and affective computing. Recent research, such as [5], demonstrates the integration of biosensors and ML to objectively classify anxiety levels within VR environments. Their system captures real-time physiological signals—including heart rate, electrodermal activity (EDA), and frontal brain activity—through wearable sensors while participants perform a VR-based version of the Emotional Stroop Task[4], a known psychological assessment method for emotional interference. The study emphasizes how VRET and sophisticated emotion detection methods can be combined to produce evidence-based and adaptive therapeutic experiences. Another growing area is voice-based emotion recognition. Changes in vocal parameters—such as pitch, intonation, tone, and pauses—can reflect emotional states like nervousness, stress, or calmness [7]. Voice analysis is useful in social anxiety applications, as it provides a non-invasive, context-aware way of evaluating how the user responds in simulated social scenarios. Several studies and APIs (e.g., Affectiva[5] or open-source models) have demonstrated high accuracy in detecting stress through vocal analysis. Motivated by this strategy and research found in this area, this project uses signals from voice and text input to identify anxiety indicators in VR social scenarios.

TECHNICAL ARCHITECTURE

System Overview

SimulEase applies a user-centered design approach to build immersive anxiety management scenarios through the combination of real-time anxiety detection and VR development. The project aimed to recreate real-life anxiety triggers within a safe environment to enable users to build their coping skills and resilience over time. The system combines several state-of-the-art technologies, such as realistic 3D animation, AI-based conversational agents, and speech recognition/synthesis to mimic real-world social settings and deliver interactive practice for users. The system architecture is modular and distributed, with three core subsystems: the VR world, the AI conversational interface agent, the emotion recognition AI service, and the speech interaction platform. These communicate predominantly through RESTful APIs to enable real-time responsiveness and scalability. The system design of our Unity-based VR social anxiety therapy is shown in Fig. 1.

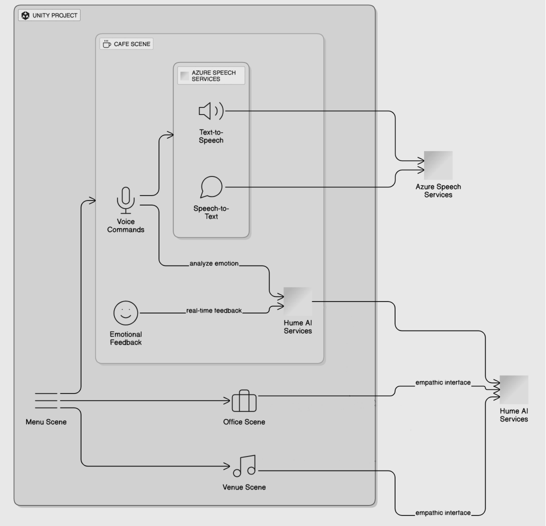

Figure 1. System Architecture

Unity[6] game engine served as the development platform for the application core, while scenes depicting common anxiety triggers like public speaking and casual conversations were created from both custom-made and free assets. Blender[7] together with Unity Animator, Mixamo[8], AccuRig[9], and blendshapes[10] handle the animation of the characters present in the scenes. We used real-life situations commonly reported to be SAD triggers when designing scenes in SimulEase. Each scene individually depicts a different level of difficulty, but perhaps more significantly, the user can choose his/her level of difficulty for each of the situations, allowing for a graded exposure approach. Each Unity scene is standalone but

[3] https://www.tripp.com/product/

[4] A psychological test used to assess attentional bias toward emotionally charged stimuli [6].

[5] https://www.affectiva.com

[6] Unity: https://unity.com

[7] Blender: https://www.blender.org

[8] Mixamo's technologies simplify the character animation process, using ML techniques.

[9] AccuRIG allows you to concentrate on model creation while still getting excellent rigging.

[10] Blendshapes are a 3D animation technique used for facial expressions.

shares access to centralized AI services. Voice, emotion, and conversation AI are real-time integrated. Azure and Hume AI services are invoked with API calls, allowing for flexibility and cross-platform functionality. The Unity Project's foreground scenes are the system foundation:

- **Menu Scene**: This is where the user is entering and being directed.
- **Office Scene**: An environment intended to place public speakers under pressure from non-player characters.
- **Venue Scene**: Distracting as well as anxiety-inducing sounds for a large crowd simulation.
- **Cafe Scene**: This is the most interactive with real-time AI dialogue.

Populated with animated non-player characters (NPCs) that provide a range of behaviors to elicit realistic social pressure, the environments themselves are replicated in all their textures and nuances. These behavioral cues are deliberately nuanced—from subtle eye movements and inattentiveness to impatience—to evoke the complex social dynamics users might face in real-world settings. Through this replication, the system aims at creating desensitization to social triggers that is both engaging and therapeutically effective. The core element of the system, Alina, is an AI-powered conversational assistant that perfectly exemplifies the application of state-of-the-art natural language processing in the virtual reality space. With support from a cloud architecture by Microsoft Azure and OpenAI's GPT-4 language model, this agent can have dynamic, contextually relevant conversations with the user. The most essential feature of the most exciting experience has been the conversational AI; besides having actual social interaction, including maintaining eye contact, it has the additional virtue of giving instant feedback that encourages proper social behavior. This is a new approach to resolving social anxiety by using AI-driven interaction within an immersive environment based on strengths that combine VR exposure with intelligent dialogue systems. To optimize exposure therapy effectiveness in VR, the NPCs must both appear realistic and act believably. Describing Figure 2, we can see the realism of character was given priority in the office scenario, not only in terms of body detail but with slight behavioral cues: gaze, scan motion, fidgeting, and interest/disinterest facial expressions (e.g., a manager looking at the watch or an employee typing away on a laptop). Micro-animations were instrumental in replicating those minor social signals that trigger anxiety in real-life gatherings and presentations. The venue scene, while having a greater number of people, uses crowd simulation and random animation to maintain the experience of social pressure even when the number is large. The cafe scene required much more realism, particularly in the case of AI character Alina, who takes on the role of the focal interlocutor. Alina not only coexists with facial and idle movement but also utilizes text-to-speech (TTS) synchronization to simulate mouth movement throughout dialogue.

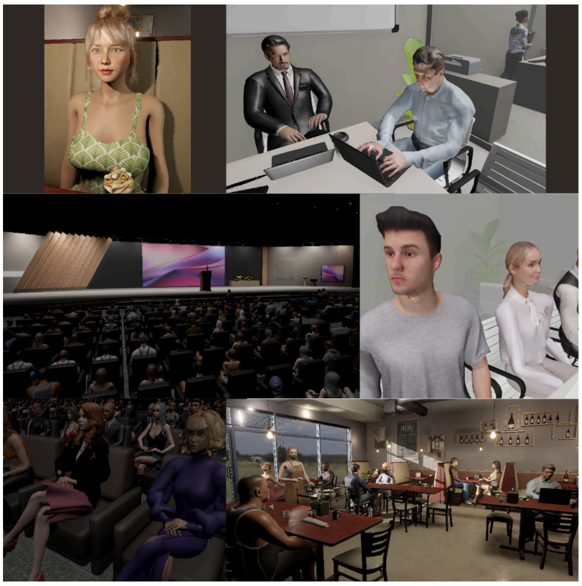

Figure 2. Characters in SimulEase

Eye gaze and perception are also provided, with requests to the user to continue social interaction. Such subtle details are conducive to social presence, a principal metric in VR interaction, which is necessary to enable real emotional reaction. Photorealistic textures, uniform lighting, and animation response all work together to immerse the user in the simulation and minimize cognitive dissonance.

Voice-Based Stress Analysis

Hume AI[11] is a sophisticated affective computing platform whose goal is to deconstruct human emotions from multimodal inputs such as voice, text, and facial expression. In the current project, Hume AI plays a central role in detecting the emotional state of the user through voice signals. In contrast to other sentiment analysis methods, which examine only semantic content, the voice model in Hume AI measures more subtle paralinguistic features (e.g., pitch modulation, vocal tension, prosody, and speech rate) to forecast nuanced affective states like stress, anxiety, and nervousness. It is specifically designed for social anxiety contexts, where emotional shifts may not be verbally coded but are generally open to being detected in the way something is said, rather than in what is said. When heightened stress levels are found (e.g., strain in the voice, rising pitch, or jitter), the system can offer subtle feedback, like in-screen hints or adaptive modifications of Alina's answers. The effective responsive feedback cycle allows the system to dynamically support the social interaction of the user without stopping to challenge him or her in a regulated and individualized therapeutic setting. Hume AI's emotional voice model, trained on a large corpus of emotionally labeled speech samples, builds on this principle with deep learning techniques to distinguish between

[11] https://www.hume.ai/

affective states such as “nervous,” “tense,” or “calm” with high granularity. With the integration of Hume AI within the VR environment of Unity, the system can now pick up on and respond to emotional states in real-time, to an empathetic and adaptive model of digital therapy.

Azure Speech & Open AI Services

Microsoft Azure Speech Services[12] is a feature-rich, cloud-based solution offering real-time speech-to-text (STT) recognition, TTS synthesis, and language translation via robust deep neural network-based models. In this VR-based therapy solution, Azure Speech Services plays a core role in enabling natural, fluid communication between the user and AI-driven virtual agents, particularly within the Cafe Scene as the user is interacting with Alina, a conversational AI avatar. Through Unity’s microphone input pipeline, the system receives the user's voice as input for STT and streams it to Azure’s STT API endpoint. For high-emotion and acoustically dynamic VR contexts, Azure’s automatic speech recognition (ASR) engine can interpret spoken words into text in real-time, even accounting for accent or ambient noise. In addition to triggering Alina’s natural language reply, the text being analyzed is independently verified by Hume AI for emotional characteristics and emotional tone. Voiced Alina’s speech is synthesized from GPT-produced text replies via Azure Speech Services’ TTS functionality. As reported by [8], Azure features over 400 neural TTS voices across more than 140 languages and dialects, such as affectively rich voices of ”Jenny Multilingual” and ”Aria.” The voice used for the project to speak with Alina is warm, sympathetic, and natural; this adds to the sense of social presence and reality. Real-time latency is kept to a minimum to facilitate realistic and quick back-and-forth communication because Azure’s cloud processing is incredibly quick. One of the significant aspects of this project’s interactive experience is the use of GPT-4, being hosted via Azure OpenAI Service (also known as Azure Foundry[13]). Hosting the GPT-4 enables real-time natural language processing and generation for the AI character Alina within the cafe scenario. In end-user dialogue, what they say gets transcribed by the Azure STT service and forwarded to GPT-4, which analyzes the message and returns a conversational, natural, and emotionally empathetic reply. This is then spoken out loud through Azure’s TTS, creating a natural and fluid conversational cycle. The system is also provided with enterprise-level reliability, data privacy, and scalability with GPT-4 in Azure’s secure cloud. The GPT-4 model facilitates understated, context-heavy conversation that has therapeutic intent combined with emotional veracity. It maintains social presence and encourages constructive interaction, both of which are crucial in social anxiety exposure therapy.

USER EXPERIENCE DESIGN

User Journey and Flow Diagram

Following installation, the user is taken to the Scenario Library, a repository of fear-evoking virtual scenarios crafted with graded exposure therapy principles[9]. Scenarios range from being in front of a small or large group, initiating conversation in a social setting, like a coffee shop, or conducting a work interview. The user may choose a scenario and customize the exposure time as well as the level of difficulty, which will begin in static or less intense exposure modes, appropriate for initial acclimation. Following the activation of the chosen scenario, the user will be submerged in a full VR experience. The session is constantly monitored with AI emotion recognition to make sure the user is safely experiencing the SimulEase scene. The user is provided with a performance report after each session. This includes statistics such as session duration and a bar chart of the recognised emotions in their speech. All these give feedback to the user to reflect on their improvement and know their anxiety patterns better. As the user navigates the application, increasingly difficult levels of scenarios are revealed, which instill a feeling of accomplishment and motivation to keep going. This technique of gradual exposure makes sure that users are not beyond a zone of comfort, a principle referred to as the ”therapeutic window” [10]. SimulEase is also compatible with collaborative care. Individuals with assistance from a coach, counselor, or therapist can share the safety reports of progress with professionals for proper direction and adjustments in plans with appropriate treatment. As illustrated in Figure 3, the user chooses a scenario and level of difficulty on the first start of the app, after which they can enter one of the customized experiences. Each scenario incorporates affective monitoring systems, user input-dependent branching logic, and dynamic ambient features (e.g., AI interactions or animated crowds). The provision of assistance prompts and optional breaks is initiated if anxiety signs are observed (e.g., via affect detection or eye contact). The flow promotes either ongoing involvement or temporary disengagement, allows for self-regulation by the user, and concludes with an application exit or scenario completion.

Scenario Selection

SimulEase’s experiential therapy is all based on a set of very carefully developed VR situations. It focuses on each specific social situation usually encountered in social anxiety. The Office Presentation Scene is one of the key virtual environments in SimulEase, designed to elicit workplace-specific social anxiety, such as performance evaluation and public speaking in front of colleagues. In this case, the user is standing at the head of a conference table while preparing to deliver a presentation to a group of ten seated virtual characters. The spatial arrangement and body language of these characters have been deliberately

[12] https://azure.microsoft.com/en-us/products/ai-services/ai-speech/

[13] https://azure.microsoft.com/en-us/products/ai-services/openai-service

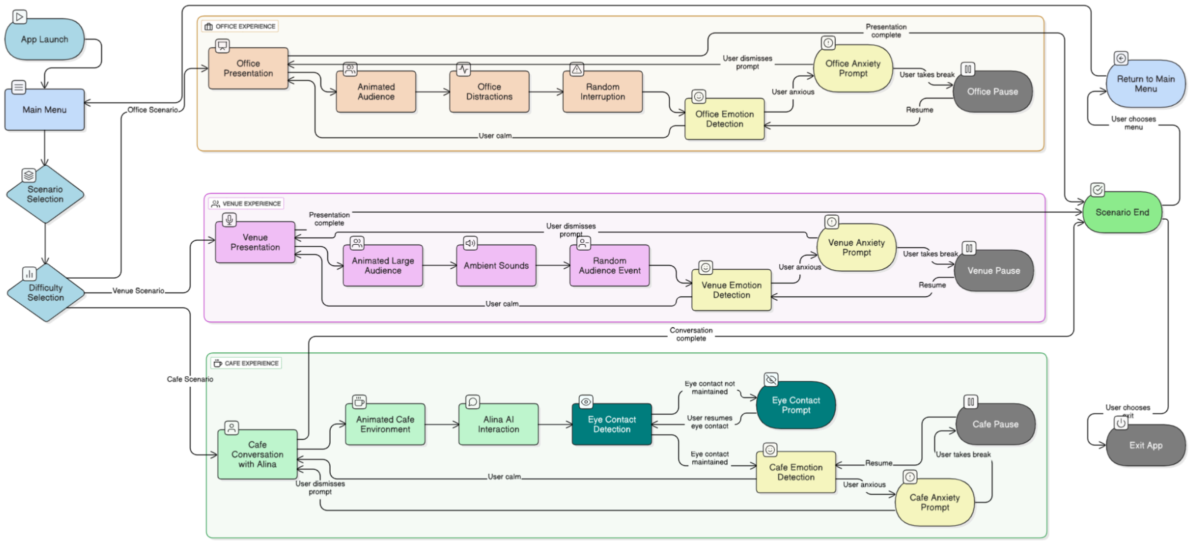

Figure 3. User Journey Diagram

crafted to evoke realistic and subtle sources of social pressure, consistent with the cognitive-behavioral model of social anxiety, which emphasizes perceived judgment and negative evaluation [11]. To enhance the immersive anxiety-inducing quality of the environment, the characters exhibit a range of socially provocative behaviors. Several avatars maintain direct eye contact or visibly scan the user from head to toe, again contributing to the sense of being observed. One character seems detached, typing away at a laptop throughout the presentation, illustrating inattentiveness, a common social intimidation within socially anxious individuals. Another avatar, positioned at the end of the table and modeled as a ”boss” figure, intermittently checks their watch with a visible expression of impatience, subtly conveying time pressure and possible judgment. Additionally, an office colleague character visible through a glass wall randomly walks past the office door, with loud footstep sound effects, creating the possibility of unexpected external interruptions. This not only simulates real-world distractions but also introduces uncertainty, which has been shown to exacerbate anxiety responses during exposure tasks [12]. The second scene SimulEase provides is supposed to recreate an experience of great tension associated with public speaking in front of an audience, this being another source of social anxiety for most people. The Large Venue Presentation Scene puts the user on stage in front of a large, spacious auditorium with about 200 virtual attendees. The audience is designed to be semi-interactive, seated but subtly animated to create a realistic and socially dynamic atmosphere. Avatars shift in their seats, lean to whisper to one another, and occasionally appear distracted, simulating a wide range of audience behaviors that may be interpreted as critical or disengaged by socially anxious individuals [13]. To add to the immersive fidelity of the experience, ambient sound serves a very important function. There is a low murmur of conversation among the people in the venue, the ringing of cell phones, finger tapping on touchscreens, and the odd microphone static, all of which introduce a feeling of unpredictability and exposure. These elements are used to represent realistic distractions commonly experienced in live public speaking settings that may heighten anticipation, anxiety, and cognitive load [14]. Adding to the unpredictability of the environment, a random scripted animation is included in which a woman stands up mid-session and walks out of the auditorium. This moment is intended to simulate a common catastrophic interpretation in individuals with social phobia, that disengagement from others is a direct reflection of their failure [11]. The most socially interactive environment within our project, SimulEase, is the Cafe Conversation Scene, designed to simulate an informal social setting that targets fears related to small talk, maintaining attention, and one-on-one interpersonal engagement. In this specific scenario, the user sits at a cafe table across from a virtual character named Alina, an AI-driven conversational partner who responds contextually to the user’s speech and behavior, mimicking a friend of the user. This configuration is a reflection of a common scenario for individuals with SAD, which can be a major cause of distress through perceived social assessment or avoidance of negative appraisal [15]. Alina is modeled to exhibit realistic idle behaviors such as subtle head movements and blinking, and her speech is accompanied by animated lip-syncing and expressive gestures, enhancing verisimilitude and engagement. The cafe itself is populated with other virtual people around, creating ambient environmental complexity. Characters are seen chatting at nearby tables, one individual is working quietly on a laptop, and animated waitstaff walk through the space intermittently. These background details help reinforce the immersion and ecological validity of the simulation while introducing minor distractions to mirror a real-life setting. What sets this scene apart is its emphasis on real-time behavioral feedback and subtle nonverbal social dynamics. If the user fails to keep eye contact with Alina for a long time, this is a usual reaction from socially anxious people who try to avoid perceived attention [16], then a gentle on-screen hint calls their attention to refocus the user's gaze again.

Post-Session Reports

After each session, the system also generates a post-session report to provide the user with constructive criticism of his/her interaction in the virtual environment. When the session comes to a close, emotion predictions are collected and filtered against a relevance threshold to locate the most applicable emotions experienced throughout the conversation. The final report includes a bar chart with the leading emotions and corresponding relative percentage scores (e.g., Joy 60%, Calm 32%, Interest 18%). This gives the user an understanding of his or her affective state as interpreted by the system, facilitating self-awareness and emotion recognition. Finally, this screen enables session report exportability to be downloaded as a shareable card (a .png file). The particular purpose is to make the feedback portable and reusable. The users will be able to share it or retain their conversational data to use it therapeutically, or monitor for personal reasons, by transforming the post-session report into an exportable static format.

LIMITATIONS

The validity, realism, and generalization of the VR application are all compromised by a number of limitations, despite being innovatively designed and intended. One of the most significant limitations is in recognizing emotions with accuracy through text and audio analysis. Although voice-based emotion recognition of Hume AI provides a non-invasive alternative to physiological sensors, it may not be entirely accurate in reflecting the nuances of emotional states, especially when speakers contribute infrequently, flatly, or with minimal vocabulary. This can lead to anxiety levels being misjudged or mislabeled under important social stressors, diminishing the utility of real-time adaptive input. The variety and realism of virtual environments are yet another drawback. Iterative exposure to fixed situations will eventually cause the flattening of reactions, whereby users will come to anticipate events and reduce anxiety responses, not using actual adaptation, but by habituation to the simulation, despite the immersiveness created through the animated characters and ambient sound (e.g., audience murmurs, ringing phones) in the settings. Besides, although the three contexts: office, public speaking setting, and cafe conversation, are heterogeneous in the difficulties they represent, their therapeutic applicability may be limited as they do not account for the entire range of socially anxious situations (e.g., dating, interviewing for a job, or group conversation).

TARGET GROUP

This project's primary target audience is represented best by people who struggle in ordinary social situations, such as public speaking, job interviews, workplace conferences, or even casual social encounters. This includes people medically diagnosed with SAD, but also people wanting to improve their socialization skills, or maybe unaware that they might be suffering from a deeper illness. SimulEase is especially geared towards young adults and working professionals (ages 18–40), a demographic often affected by high-performance pressure and more likely to develop mental disorders because of this. To ground this definition in practical terms and ensure that design choices addressed real user needs, we developed a set of personas representing typical users within the target group.

Persona 1: "Maria, 22, university student" - Experiences intense anxiety during class presentations and avoids group work. Wants a safe way to practice public speaking before real audiences. Finds gradual VR exposure less intimidating than live rehearsal.

Persona 2: "Alex, 29, software engineer" - Struggles with networking events and workplace meetings. Wants to improve conversational confidence for career growth. Needs real-time feedback to recognize when anxiety affects his speech clarity. These personas informed scenario design, ensuring that SimulEase addressed challenges ranging from public speaking to informal conversations. They also guided feature selection; for example, real-time emotion recognition was prioritized for users like Alex, who struggle to monitor their own stress signals.

EVALUATION AND TESTING

Usability testing and performance measurement of the SimulEase scenes were designed to establish how effectively the system can deliver a socially and emotionally convincing virtual experience, particularly for users struggling with social anxiety. A small group of subjects underwent the cafe scenario through a standalone VR build and provided formative feedback in an extensive survey. The overall purpose of the testing procedure was to gather qualitative and quantitative user opinions from unfamiliar users in an attempt to assess its intuitiveness, emotional valence, and perceived use for exposure scenarios relevant to anxiety. Our primary target population consists of young adults (18–30) experiencing mild-to-moderate Social Anxiety Disorder, particularly in academic and professional settings where public speaking and networking are frequent. The subjects were provided with a standalone executable build of the cafe simulation, which they ran on their own compatible Windows machines. The build included the complete immersive environment: interactive dialogue with Alina, ambient animation, and voice-controlled emotion recognition feedback. Users were invited to interact in the scene normally, establishing conversation, speaking aloud, and engaging with the character as one would in a normal, casual encounter. Participants were led to a structured Google Forms survey to record their comments and experiences after going through the simulation. Likert-scale[14] questions (such as ”How natural did Alina's responses feel?”), open-ended commentary statements, and perceived realism, usability, emotional engagement, and immersion questions were all included in the survey. To enable better interpretation of answers, some general demographic and technical context data (e.g., age group, prior experience with VR) were also collected. This approach yielded insightful initial results about how users view the virtual world and its ability to mimic socially significant interaction.

User Feedback

The demographics information showed that most participants were in the 18–34 age group, with a close gender parity between male and female participants. Most participants had some previous experience with either virtual reality or conversational AI interfaces like ChatGPT or Siri, creating a relatively well-informed test population. Generally, the response was extremely positive. Over 80% of the respondents scored their overall experience at the highest rating, with many using terms to characterize the virtual interaction with Alina as natural, smooth, and emotionally sensitive. All subjects reported that Alina reacted normally to their input, and most also reported that the setting was natural and comfortable. Interestingly, users reported that the system was emotionally supportive, a principal goal of the simulation, pointing out that Alina's presence and ambient elements of the scene instilled a sense of calmness and a feeling of welcome. The criticism also revealed opportunities for improvement. Some users noted that Alina's delayed responses sometimes broke the illusion of being an ongoing conversation, and some users noted that her lip-sync and face animations were not always perfectly synchronized with the words that were being said. Some users also observed that looping background animation or music could be repetitive or even slightly unrealistic in the long term. The following are suggested to enhance facial expressiveness, response timing, adjust UI prompt placement to minimize intrusiveness, and provide more scene diversity or emotion richness in voice from the AI. Many participants indicated the realistic nature of the contact, the feeling of emotional understanding, and the spontaneous course of the conversation, even in response to these criticisms. As an aspect of the system's perceived emotional intelligence, a user reported, ”The AI knew I was talking about something sad even though I tried to sound happy”. Others praised the cafe environment's tendency to encourage relaxed interaction and a natural response style. This feedback adds to the main design objective of the project: providing a socially interactive, emotionally secure, and realistic virtual environment where users can practice and become confident in scenarios that trigger their anxiety. Similarly, feedback on conversational AI latency prompted adjustments to reduce response delays, as users reported that breaks in dialogue reduced immersion and heightened anxiety.

Strengths and Weaknesses of the System

Among the greatest assets of the system is that it can integrate a range of AI-driven services such as speech recognition (Azure), emotion detection (Hume AI), and conversational language models (OpenAI) to create a natural, emotionally empathetic virtual conversation. Features like gaze simulation and UI feedback behavioral cues support user attention and immersion, particularly useful where social training, anxiety reduction, or naturalistic conversation practice is performed. Modularity of the system is also a technical advantage. Randomized animation chaining and ambient behaviors also contribute to the world's believability, making the world feel less robotic or repetitive, a characteristic of much VR software. The system is not, however, unlimited. Emotion recognition based on 5-second audio samples introduces latency in the system response relative to user input, disrupting conversational flow and responsiveness perceptions. In addition, voice input is subject to disruption by stuttering speech or in noisy conditions, challenges that adversely impact Azure's transcription, especially. Emotion recognition introduced a minimum delay of 5.4s (5s window + 0.4s classification). The reuse of animation assets from various sources (Mixamo, AccuRig) sometimes caused jerky or unnatural avatar motion. Participants rated perceived responsiveness with qualitative comments indicating that delays above 3s were disruptive to conversational flow. Emotional cues can be detected and utilized, but true human emotional intelligence and its subtlety are normally missing. Overall, the system is a breathtaking prototype of affect-sensitive virtual interaction but remains at the juncture of promise and limitation, both indicating the potential and current boundaries of affective computing in immersive worlds.

CONCLUSIONS AND FUTURE WORK

SimulEase has delivered an effective, emotionally intelligent, interactive VR simulation designed to help and test user engagement in socially difficult situations, such as those associated with social anxiety. By developing a highly interactive virtual cafe environment and presenting Alina, a virtual human character capable of engaging in natural speech, facial animation, and basic emotional responsiveness, the project demonstrates how novel AI services can be leveraged for the simulation of complicated human interactions. From scripted small talk to behavioral subtlety like gaze monitoring and adaptive tone of voice, the system shows the potential for immersive digital agents to deliver naturalistic, emotionally engaging interaction. The driving force behind this project was a much-documented and emergent mental health need: the treatment of social anxiety and associated disorders.

[14] A Likert scale allows respondents to express how much they agree or disagree with a particular idea

Traditional treatments such as CBT, as powerful as it has proven to be, are limited by such barriers as stigma, barriers to access, and real-world uncertainty that can undermine patient compliance. To bridge this gap, the project sought to examine how immersive technologies, and more particularly VRET, can simulate anxiety-provoking social situations in a safe and controlled virtual environment where individuals can learn to manage incrementally without the social costs of real exposure. The unification of many next-generation AI services into one cohesive and dynamic interaction pipeline is a major project milestone. OpenAI's language model gave Alina her conversational powers, enabling her to create natural-sounding, contextually apt responses that mirror the user's speech both in tone and content. The input of the user can be quickly transcribed because of the TTS service of Azure, and the technology provides a natural-sounding voice that goes with the facial expressions of Alina. The ability of the project to blend real-time AI output with a meticulously crafted VR world is an exemplary case of the way thoughtful system design can increase the emotional realism and usability of virtual agents in complex social simulations. Usability testing showed participants valued SimulEase's immersive environment and natural conversations, but issues with motion capture, delayed emotion prediction, and noisy speech recognition highlighted areas for improvement. The project also addressed ethical concerns around emotional inference, data privacy, and user well-being in sensitive scenarios like social rejection.

SimulEase shows strong potential but requires further development to enhance its therapeutic impact. Key improvements include integrating real-time physiological indicators for objective emotion tracking, expanding VR scenarios beyond coffee shop and public speaking to cover diverse social anxiety contexts, and making scenes dynamically adjustable to user progress. Enhancements in character realism, response timing, and usability are also planned. Future controlled trials will be crucial to evaluate its effectiveness compared to CBT or therapist-led VRET.

REFERENCES

1. S. Shahid, J. Kelson, and A. Saliba. Effectiveness and user experience of virtual reality for social anxiety disorder: Systematic review. JMIR Mental Health, 11:e48916, 2024.
2. R. Pal, D. Adhikari, M. B. B. Heyat, B. Guragai, V. Lipari, J. Brito Ballester, I. De la Torre Diez, Z. Abbas, and D. Lai. A novel smart belt for anxiety detection, classification, and reduction using iiomt on students' cardiac signal and msy. Bioengineering, 9(12), 2022.
3. S. Choudhary, N. Thomas, S. Alshamrani, G. Srinivasan, J. Ellenberger, U. Nawaz, and R. Cohen. A machine learning approach for continuous mining of non-identifiable smartphone data to create a novel digital biomarker detecting generalized anxiety disorder: Prospective cohort study. JMIR Med Inform, 10(8):e38943, Aug 2022.
4. R. Lemos, S. Areias-Marques, P. Ferreira, P. O'Brien, M. E. Beltrán-Jaunsarás, G. Ribeiro, M. Martín, M. del Monte-Millán, S. López-Tarruella, T. Massarrah, F. Luís-Ferreira, G. Frau, S. Venios, G. McManus, and A. J. Oliveira-Maia. A prospective observational study for a federated artificial intelligence solution for monitoring mental health status after cancer treatment (faith): study protocol. BMC Psychiatry, 22(1):817, 2022.
5. D. Mevlevioğlu, S. Tabirca, and D. Murphy. Real-time classification of anxiety in virtual reality therapy using biosensors and a convolutional neural network. Biosensors, 14(3), 2024.
6. J. M. G. Williams, A. Mathews, and C. MacLeod. The emotional stroop task and psychopathology. Psychological Bulletin, 120(1):3–24, 1996.
7. A. Madison, M. Vasey, C. F. Emery, and J. K. K.-G. and. Social anxiety symptoms, heart rate variability, and vocal emotion recognition in women: evidence for parasympathetically-mediated positivity bias. Anxiety, Stress, & Coping, 34(3):243–257, 2021.
8. M. Azure. Azure speech service documentation, 2023. Available at: https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/.
9. 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, 25(3):103–113, 2017.
10. E. B. Foa and M. J. Kozak. Emotional processing of fear: Exposure to corrective information. Psychological Bulletin, 99(1):20–35, 1986.
11. D. M. Clark and A. Wells. A cognitive model of social phobia. In Social Phobia: Diagnosis, Assessment, and Treatment, pages 69–93. Guilford Press, 1995.
12. D. W. Grupe and J. B. Nitschke. Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. Nature Reviews Neuroscience, 14(7):488–501, 2013.
13. C. R. Hirsch and D. M. Clark. Information-processing bias in social phobia. Clinical Psychology Review, 24(7):799–825, 2004.
14. J. Ayres, T. Hopf, and A. Will. Are reductions in ca an experimental artifact? A reply to finn. Communication Education, 49(3):289–296, 2000.
15. M. R. Leary. Social anxiousness: The construct and its measurement. Journal of Personality Assessment, 47(1):66–75, 1983.
16. F. R. Schneier, T. L. Rodebaugh, C. Blanco, H. Lewin, and M. R. Liebowitz. Fear and avoidance of eye contact in social anxiety disorder. Comprehensive Psychiatry, 52(1):81–87, 2011.