Audience simulation and perception in virtual reality.

Yann Glémarec

COLLEGE MATHSTIC
DOCTORAL BRETAGNE
BRETAGNE OCEANE

ÉCOLE NATIONALE D'INGÉNIEURS DE BREST

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE
D'INGÉNIEURS DE BREST

ÉCOLE DOCTORALE N° 644
Mathématiques et Sciences et Technologies de l'Information et de la Communication en Bretagne Océane
Spécialité : *Informatique*

Par

Yann GLÉMAREC

Audience simulation and perception in virtual reality.

Application to public speaking training in an academic environment.

Thèse présentée et soutenue à Plouzané, le 27 Juin 2023
Unités de recherche : Lab-STICC UMR 6285, Interaction, Universität Würzburg, Informatik IX, HCI group
Thèse N° : 2023ENIB0001

Rapporteures avant soutenance :

Magalie Ochs — Maîtresse de Conférences HDR, Aix Marseille Université, France
Maud Marchal — Professeure des Universités, Université de Rennes, France

Composition du Jury :

Président :	Maud Marchal	Professeure des Universités, Université de Rennes, France
Examinateurs :	Domitile Lourdeaux	Professeure des Universités, Université de technologie de Compiègne, France
	David Panzoli	Maître de conférences HDR, Institut National Universitaire J.F. Champollion, France
Dir. de thèse :	Cédric Buche	Professeur des Universités, ENIB, IRL 2010 Crossing, CNRS, Australia
	Marc Erich Latoschik	Professor Doktor, Universität Würzburg, Informatik IX, HCI group, Germany
Co-encadrants :	Anne-Gwenn Bosser	Maîtresse de Conférences HDR, ENIB, France
	Jean-Luc Lugrin	Akademischer Oberrat, Universität Würzburg, Informatik IX, HCI group, Germany

Remerciements Acknowledgements

Premièrement, je souhaiterai remercier mon équipe d'encadrements, Anne-Gwenn Bosser, Cédric Buche, Jean-Luc Lugrin et Marc Latocshik de m'avoir donné la chance de travailler sur mon sujet de recherche durant ces 4 années, ainsi que pour leurs conseils et l'expérience partagés. Tout particulièrement merci à Anne-Gwenn et Jean-Luc pour le temps investi et le soutien journalier qu'ils m'ont consacrés notamment dans les moments difficiles de ma thèse de doctorat. Également merci aux membres de mon CSI, Thierry Duval et Fred Charles pour leurs conseils et encouragements.

Je remercie également l'ensemble des étudiants qui ont participés à mes travaux de recherche avec qui j'ai adoré travailler, Paul, Lucas, Fergal, Erwan et Jessica.

Je souhaiterai également remercier mes collègues et amis qui ont pu directement ou indirectement participer à la réalisation de ma thèse et qui ont su me soutenir pendant plus de 4 ans.

Mille mercis à Aryana, Pierre, Anais, Hugo, Jean-Victor, Amélie et Jean-Michel, Anne du CERV qui ont passé des heures à m'écouter me plaindre et ont toujours été à l'écoute. Merci, aux collègues et amis du HCI Group qui m'ont accueilli royalement et directement considéré comme l'un des leurs pendant les mois difficiles du confinement en Bavière. En espérant rapidement vous revoir pour vous remercier en personne, Andreas, Jinghuai, Murat, Larissa, David, Florian, Maximilian, Andrea, Fabian, Mounsif, Peter, Franziska, Lucas, Chris et Kristof. Un grand merci à mes colocataires Paul et Becky avec qui j'ai beaucoup ri et partagé de beaux moments à Würzburg. Vielen Dank!

Danke, an die Kollegen und Freunde der HCI-Group, die mich während der schwierigen Monate in Bayern königlich aufgenommen und direkt als einen der ihren angesehen haben. In der Hoffnung, euch bald wiederzusehen, um euch persönlich zu danken, Andreas, Jinghuai, Murat, Larissa, David, Florian, Maximilian, Andrea, Fabian, Mounsif, Peter, Franziska, Lucas, Chris und Kristof. Ein großes Dankeschön an meine Mitbewohner Paul und Becky, mit denen ich viel gelacht und schöne Momente in Würzburg geteilt habe. Vielen Dank!

Je n'aurai jamais terminé mes travaux sans le soutien de ma famille et amis proches, qui sont toujours restés derrière moi pour m'encourager. Je ne saurai jamais comment remercier mes parents et mon frère pour tout ce qu'ils ont fait. Un grand merci également, à Baptiste, Guillaume, Benoit, Marine, Paul, Sarah, Louis, Ewen, Orlane, Benjamin, Aymeric, Chloé C.,

Maxime D., Coralie, Maxime B., Noémie et David C. qui pendant toutes ces années ont été comme une famille pour moi. Je me dois de remercier la personne qui m'a le plus aidé et qui a commencé à me soutenir bien avant le début de mon doctorat, sans qui je n'aurai sans doute pas réussi, merci pour tout Chloé et merci pour ton amour durant toutes ces années.

Je souhaiterai à nouveau remercier Guillaume, Paul, Marine et Benoit pour leur soutien durant cette 4ème année qui a été particulièrement difficile pour moi sur un plan personnel, mais qui ont su trouver les mots pour me remonter le moral. Enfin, je dois remercier Coraline sans qui je n'aurai jamais réussi à terminer cette ultime année de recherche et qui a su être l'amie dont j'avais besoin dans le moment le plus difficile de mes études et sans qui, j'en suis sûr, j'aurai abandonné. Merci à toi tu es une personne incroyable et unique, sans doute la plus belle rencontre que j'ai faite durant ma thèse.

Merci à tous pour votre amour, amitié et soutien.

PREAMBLE

This thesis was under joint supervision between the ENIB (École Nationale d'Ingénieurs de Brest) and the Julius-Maximilians-Universität Würzburg. Besides these two academic actors, the supervision agreement involves two research teams: the COMMEDIA team from Interaction research departement, which supervises the research work carried out by the PhD student in France (Lab-STICC, UMR 6285), and the HCI group (chair for Human-Computer Interaction, Computer Science IX) from the University of Würzburg supervising the PhD student's research work in Germany.

This work was supported by French government funding managed by the National Research Agency under the investments for the Future program (PIA) grant ANR-21-ESRE-0030 (CONTINUUM).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

We frequently have to speak in public in our everyday life, for instance, during professional meetings or business presentations. Similarly, students have to talk in front of audiences, whether for an oral exam or presentation. We practise a lot during our studies and even beyond college. For example, the French baccalauréat now has a compulsory Grand Oral, where the students must demonstrate their ability to speak in public, clearly and convincingly. The jury values the solidity of the student's knowledge, their ability to argue and link knowledge, their critical mind, the precision of their expression, the clarity of their speech, their commitment to their words, and their strength of conviction [Ministere de l'Education Nationale et de la Jeunesse, 2020].

However, public speaking can be stressful. Students might fear judgement on their personality. It is much more intrusive and engaging than written exams where the criticism is deferred. The lawyer Bertrand Perrier explains in an interview how the exercise can be an ordeal for students, paralysing depending on the context, and sometimes disabling [Iribarnegaray, 2021]. In some extreme cases, these symptoms are due to a social disorder directly related to public speaking. Social anxiety disorder is one of the most common mental disorders. It represents about 4% of mental disorders cases in all countries and can affect about 14% of people in Canada and Scandinavia [Kringlen et al., 2001; D. J. Stein et al., 2017; M. B. Stein et al., 2000]. General social anxiety disorders have an onset in adolescence with about 80% of untreated individuals, but can lead to further impairment in adulthood and cause the development of generalised social anxiety [Wittchen and Fehm, 2003]. Public speaking anxiety (PSA) is a specific form of social disorder that can cause a fear of negative evaluations of others and feelings of embarrassment or humiliation in social situations [American Psychiatric Association, Association, et al., 2013]. Yet, when treated for PSA, patients experience a reduction in their generalised social anxiety, which might lead to fewer societal and personal costs of the disease [Hofmann, 2004].

As a result, a number of applications propose training or therapeutic services, whether for academic, research or commercial uses. These commercial applications provide public

speaking training sessions in virtual reality (VR) to improve users' speaking skills or reduce their anxiety [Straightlabs GmbH & Co.KG, 2023; VirtualSpeech, 2022; VRSpeaking, LLC, 2022]. They let users face a simulated audience mimicking human behaviour and reaction to the speech [VirtualSpeech, 2022]. Some applications dedicated to training propose feedback features to help clients improve their speaking skills [VRSpeaking, LLC, 2022]. All these applications simulate audiences for the users to practise, and they all share the same critical characteristics. Thus, this chapter will introduce the underlying concepts of virtual audience simulation, the limitations VR technology brings, and the challenges it remains to take on for audience simulation in VR.

1.1 Virtual Audience Simulation

To introduce the concept of a virtual audience, we use the definition from Pertaub et al., 2002, outlining a virtual audience as a group of virtual characters situated in the same environment that mimics a public speaking situation. The two main characteristics of this definition are that multiple virtual characters populate the audience and replicate human behaviour.

Creating such virtual humans requires some core capabilities, each with a different level of significance [Swartout et al., 2006]. In our case, the ability to create communicative characters with emotional behaviours or humanlike attitudes is the first feature to build a plausible audience. Various research works use non-verbal behaviour and language communication to emulate these behaviours and convey emotions [Wang and Ruiz, 2021], e.g. with gestures, facial expressions or gaze. However, such complexity implies that groups of virtual humans can process coherent collective behaviour to stay believable, especially when users interact with the group [Prada and Paiva, 2005].

To sum up, in order to enact believable behaviours, virtual audiences rely on a complex behaviour models and animation pipelines and thus depend on virtual reality technology to render the virtual environment.

1.1.1 Audience Simulation Requirements

Many virtual reality applications use virtual audiences to treat PSA or to provide a suitable environment for speech training. Indeed, multiple studies indicate that virtual reality technology significantly contributes to reducing patient anxiety [P. L. Anderson

et al., 2005; Rothbaum et al., 2000] or practising speaking skills [Chollet et al., 2015]. However, the ability of VR to provide a safe and effective environment depends on the VR application's efficiency and the simulation's credibility.

Virtual reality Scalability

Simulating a virtual audience full of communicative virtual humans with rich social signals has a cost on the VR performance. When poor, it can significantly decrease the user's experience quality. In some cases, when low efficacy reduces the latency and lowers the number of frames per second, users can feel sick and experience cybersickness, which can induce nausea, stomach consciousness, headache and associated symptoms. There are multiple sources of performance drop in VR applications, like the rendering of a complex 3D environment and realistic virtual characters or the time to process behaviours requiring heavy computation, e.g. compound non-verbal behaviours to express an emotion. As a consequence, large crowds are not always feasible and need adapting content to maintain high and constant performances.

Virtual Humans Design

The characters populating the virtual audience are of primary importance, and their 3d models do not necessarily require being photo-realistic. On the contrary, it does not necessarily benefit VR users compared to cartoon-style characters. Realistic characters are more demanding regarding VR performances than others which are simpler 3D models and thus have a faster rendering time. Detailed virtual human models contain hundreds of thousands of triangles, whilst VR-ready characters only comprise hundreds of triangles. However, their communication capabilities are far more crucial than the appearance and realism in VR applications to root the user in the simulation. Many studies indicate that virtual environments populated by virtual characters which can interact with each other or with the user improve the user experience. Several criteria greatly enhance the audience's believability, such as language, group behaviours, reactions to an event, and any communicative signals using non-verbal behaviours. These communication abilities amplify certain underlying VR concepts that we will further introduce in Chapter 2, such as the feeling of presence, the audience's credibility or the simulation's acceptability.

1.1.2 Virtual Human Behaviours

Therefore, it is necessary to consider performance and scalability constraints when designing audiences in VR. But, the audience's behaviour remains the most crucial element in their simulation, notably for the audience to enact emotions on the user.

Characters' Non-verbal Behaviours

The virtual humans composing the audience mimic spectators listening to a speech or a presentation and show behaviours and reactions a real audience would display. Each virtual spectator from the audience requires body signals to display their current attitude toward the speech and the user. The resulting attitude arises from the association of various non-verbal behaviours such as facial expressions, head movements, and postures or gestures. Numerous studies investigate how these social signals influence our perception and how they are associated with specific characteristics, e.g. emotions, moods, attention or intentions. Consequently, virtual audience behaviours are modelled according to these characteristics and their associated non-verbal behaviours. For instance, the two models from Kang et al., 2016 and Chollet and Scherer, 2017 we used as foundations for this work define a set of audience behaviour types ranging from very attentive and positive to very negative and disengaged thanks to a list of non-verbal behaviours influencing our perception of the overall audience behaviour.

Behaviour animation

Usually, animators use 3D software to design virtual characters' animations. They are sometimes extracted from motion tracking data and then post-treated. This complexity makes the entire process very time-consuming and expensive, considering that multiple characters populate virtual audiences, and they all need several animations. Because the behaviour models are parametric, the virtual humans necessitate compound 3D animations, which display multiple non-verbal behaviours to make a more complex one carrying the characters' attitudes. Thus, the characters require a pipeline animation able to blend the different non-verbal behaviour according to the given parameters from the virtual audience behaviour model. Such a pipeline can also use procedural animations to adapt the displayed animation, i.e. to build behaviours from various rules adapting the resulting one to the simulation. For example, to create a bored spectator from procedures ruling the agent's facial expression, posture, gaze or leg position at runtime.

The displayed behaviours must reflect the intended attitude and match the one users perceived. It is even more critical for interactions and context-related behaviours. For instance, Lugrin et al., 2016 use disruptive behaviours for classroom management pedagogical scenarios to let the trainees face specific situations. These events and reactions to the user's activities broaden the range of possible behaviours and complicate their triggering.

1.2 Users' Perception in Virtual Reality

1.2.1 User Perception

Studies indicate differences between VR head-mounted displays and traditional computer screens. Technical differences may alter users' perceptions. An egocentric point of view, like in VR, defines the position, orientation, and movements of objects according to the body representation, while an exocentric point of view, such as desktop screens, is not affected by the body location [Bowman et al., 2005]. Thus, a VR environment where the user's position is dynamic might be different from a static exocentric one. Furthermore, including the user's body representation in the environment may alter the virtual audience perception compared to a third-person viewpoint.

The added value of VR is to benefit from a virtual environment that reproduces real ones, e.g. by its visual field of view, the audio and the behaviours of the characters populating the simulation. Therefore, the following section introduces the underlying concepts of users' audience's behaviour perception in VR in comparison to real life or videos.

1.2.2 Virtual Audience Behaviour Perception

Pertaub et al., 2002 studied whether the audience's behaviour influences the users in VR. Their work indicates that users negatively evaluated facial expressions like frowns and positively evaluated smiles similarly as we would do in real life. These results suggest that it is possible to reproduce non-verbal behaviours interpretation in virtual reality that we usually have. Also, one may wonder how all the audience's behaviours are perceived together. Experiments aiming at building behaviour models to parameterise audiences and control the user's perceived attitude studied non-verbal behaviours perception. For instance, Kang et al., 2016 studied audience non-verbal behaviours according to different psycho-cognitive dimensions such as valence and arousal. Chollet and Scherer, 2017 in-

vestigated the impact of each behaviour with respect to other spectators' behaviour, but also according to their location in the environment.

These studies are complex and challenging to conduct in virtual reality because they require plenty of participants to compare all conditions. Therefore, these model evaluations often turn toward online studies to test different modalities and correlate user perception with specific behaviour. Alternatively, many VR systems rely on experts' knowledge of their application domain. For instance, Breaking Bad Behaviours manages the audience's behaviour with the help of classroom management experts [Lugrin et al., 2016], and Kahlon et al., 2019 ask secondary school students to validate the audience's behaviour simulating a high school class. Accordingly, we distinguish in the literature the virtual reality applications that rely on domain experts and those that depend on behaviour models. For example, Fukuda et al., 2017 use Paul Ekman's emotions to build five unique classroom atmospheres [Ekman, 1999], whilst Kang et al., 2016 tend to automate behaviour changes based on models.

These two approaches help simulate audiences capable of expressing several attitudes by modifying their virtual spectators' behaviours. Nevertheless, both of these approaches produce behaviour models limited by the number of behaviours and the complexity to create interaction between the virtual agents the users. Virtual reality applications requiring different strategies to control the virtual audience are even more affected by these limitations, e.g. to follow a scenario or display context-specific behaviours.

1.3 Virtual Audience Controls

Various VR applications emerge from these two approaches. We can distinguish Wizard of Oz applications, which are manually controlled and require a human in the loop, from the autonomous systems, which depend on behaviour models.

1.3.1 Wizard of Oz Applications

The Wizard of Oz VR applications benefit from context-specific behaviours and rely on experts in the domain to trigger and supervise the simulation. It suits well VR training systems because it lets the instructors personalise and adapt the ongoing session to the pedagogical plan and the trainee's actions. All the same for therapeutic application, which in some situations needs to fine-tune the audience's behaviour, especially if patients suffer

from public speaking anxiety or a phobia. However, these applications require experts for each session and cannot be used outside their supervision, e.g. it can't be used to practise with follow-up exercises. Wizard of Oz systems can also put a heavy workload on the experts, who in a majority are not computer science experts and primarily focus on their trainees and patients to personalise their session.

1.3.2 Autonomous Applications

Although, in autonomous applications, it is a program which changes the audience's behaviour. This program sometimes relies on behaviour models like in the training applications from Chollet et al., 2022 or on experts' knowledge to build a dedicated model like in Kahlon et al., 2019 with their therapeutic application for secondary school pupils.

When the application requires interactions or reactions to the user's actions, it is once again the model which decides when to intervene and what behaviour to display. For example, the Cicero application uses a user performance analysis system with physiological capture to adapt the virtual audience to the user's presentation [Chollet et al., 2022]. Alternatively, Delamarre, 2020 rely on decision trees established by pedagogy experts via UML diagrams to establish all possible interactions and actions during a training session. However, unlike Wizard of Oz systems, these systems do not allow direct interaction with the scenario in progress or with the virtual audience.

1.3.3 Usability and Acceptance

Consequently, Wizard of Oz systems' usability might be lower than the autonomous ones because of the elicited workload and the complexity of uses. Mouw et al., 2020 evaluated BBB's usability and concluded that its control interface is a suitable tool for establishing classroom strategies, yet it remains hard to get to know the interface and to be able to use it. On the contrary, autonomous applications do not allow intervention at run-time, which can lower to system's acceptance for the users. Therapists may want to intervene in the session to fix the audience's behaviour and adapt the scenario to the ongoing simulation. For instance, the IVT-T application lets the experts design their scenario beforehand but does not allow them to alter it or adapt it during the training session [Delamarre, 2020]. This lack of interactivity may be unattractive to instructors and therapists who wish to fine-tune the simulation.

1.4 Contributions

In this research work, we develop a behaviour model evaluated for VR and validate its performance and deployment in professional applications. Thus, we tackled multiple challenges regarding characteristics of audience simulation and VR technology, e.g. guaranteeing high performances in VR is restraining and complex to ensure with large agent groups. Then, providing a parametric model is tough, especially when it can be extended and when it requires high-level controls for non-experts users, for instance, instructors authoring the system without programming knowledge. Therefore, proposing an adequate user interface for the instructors is difficult if we plan to provide advanced features along with the high usability of the tool.

The first challenge we took on to propose a new behaviour model for VR so that the generated audience displays rich social signals to show various attitudes:

- Since none of the parametric audience behaviour models we found in the literature was directly evaluated in VR, we wondered if user perception might change depending on the media. Therefore, we built a behaviour model based on two existing models providing non-verbal behaviours to display audience attitudes according to a dimensional approach relying on valence and arousal [Chollet and Scherer, 2017; Kang et al., 2016].
- The models used as references are detailed enough to be reproduced. Moreover, they obtained conclusive results with users differentiating various levels of arousal and valence with videos. Thus, we implemented our model into a VR application, to reproduce a similar evaluation in VR and compare the user perception to it. Similarly, as in the studies with videos, our model triggers various non-verbal behaviours and blends them to create more complex ones, e.g. facial expressions, head movements, postures, and gaze direction. In order to display different audience attitudes and compare our results we parameterise the model according to the two same dimensions, i.e. valence and arousal.
- Finally, we benchmarked the model's characteristics in VR, such as the animation pipeline, the audience's size or the 3D characters' style impact on the performances to validate its ability to simulate an entire audience without breaking the VR performances and lowering the user's experience.

After developing our behaviour model and validating its performance in VR, we investigated its ability to create a believable audience that VR users could recognise the

attitude. Following this perception study, we continued our evaluations in various application contexts to study our system acceptance and usability in VR, which led us to work on model control issues for users without programming skills.

- First, we established a protocol to evaluate the user perception of the audience behaviour in VR.
- Then, we ran two user perception studies that confirmed the model's capability to generate different types of audience attitudes, e.g. bored, enthusiastic, interested or indifferent.
- We also included this model in a VR public speaking training application (STAGE) at the University of Würzburg to confirm our results in an academic context with lecturers and bachelor students.
- The STAGE was the opportunity to evaluate the usability and acceptance of the system to follow the training plan designed by lecturers for the students to face specific presentation situations, e.g. a phone ringing in the audience or a spectator coming in late. This last evaluation helped us to design a graphical user interface to control the virtual application with scenarios and disruptive behaviours.

Finally, we deployed the STAGE application in a professional context to study the feasibility of using such a model in a VR exposure therapy context.

1.5 Thesis Structure

This manuscript is structured as followed:

Chapter 1: This first chapter introduces the thesis motivations and the different challenges to take on to successfully simulate virtual audiences in VR. This section highlights the limitations related to VR technologies and the issues for audience behaviour simulation and scenario supervision for training and therapeutic applications.

Chapter 2: Chapter 2 provides a review of the concepts involved in virtual audience simulation. It introduces the different audience behaviour models and approaches to evaluate the user perception and validate models. Finally, this section reviews the different application domains for virtual audience simulations in VR, such as public speaking training or exposure therapy.

Chapter 3: Accordingly, chapter 3 describes the techniques we adopted to create a virtual audience able to display complex non-verbal behaviours based on a given attitude. We first present the system architecture implemented to generate the virtual audience's behaviour. Then we outline the animation pipeline developed for the agents to display compound animations that result in expressive behaviours. Finally, we provide a VR scalability benchmark for our implementation to grasp the performance limitations. The overall chapter documents the audience behaviour operating system and provides enough information to reproduce the whole system.

Chapter 4: Then chapter 4 describes the methodology used to build our behaviour model and reports the VR perception studies we ran to validate it. Then this section provides a detailed explanation of our model evaluations. We first studied the perception of non-verbal behaviour in VR and then explored how the resulting attitude is perceived. Finally, we discuss the results and limitations and provide the guidelines extracted from our statistical analysis.

Chapter 5: In chapter 5, we provide more insight into the issues related to virtual audience controls. We picture these issues with the use case of VR public speaking training. Then, we provide an overview of the virtual reality training system developed and evaluated at the University of Würzburg called STAGE. Furthermore, we designed this educational application with user-centred development with an iterative method. Finally, we describe the whole process and demonstrate the feasibility of integrating such a model into a VR application.

Chapter 6: This ultimate chapter canvasses the model's improvements as well as the future STAGE's architecture (see chapter 6). Thus, it depicts the future developments and user studies already planned for improving the current system and issues all the limitations pinpointed during our user studies. Finally, this chapter summarises the thesis findings, publications, and contributions to virtual reality training, therapeutic applications and audience behaviour simulation in VR.

Chapter 2

State of the art

2.1 Introduction

In this chapter, we will first describe critical concepts for virtual audiences and VR, namely crucial notions related to immersive technologies involved in the users' perception and experience, such as Presence. These definitions illustrate the complexity of the challenges to take on. In the second section, we will focus on how VR users perceive virtual human behaviour through different perception theories to highlight their significant role in behaviour model implementation. We will use various studies to depict and analyse the pros and cons of each method. Finally, after a review of the related VR applications using virtual audiences, we will conclude the chapter by emphasising the limitation of existing techniques to generate and control the virtual audience to support our hypothesis and stress that our suggestions can benefit VR applications for training and therapeutic purposes.

2.2 Presence and Perception in Virtual Reality

2.2.1 Presence

The feeling of presence is known to be the moment when the virtual environment replaces the real one or when we have the feeling of being in the virtual environment through our senses [Minsky, 1980], which testify to the virtual environment's efficacy [Sheridan et al., 1992; Slater, Lotto, et al., 2009; Witmer and Singer, 1998]. The concept comes from Minsky, 1980 with the idea of telepresence. It describes the feeling an operator can experience when interacting through a remote system, i.e. the sense of being in another location with a machine being the body of the user. The concept of presence requires the virtual environment to include the users and their bodies provided with parallax from the user's viewpoint when rotating the head to maintain the illusion of being immersed in a

3D space. On the psycho-cognitive level, the feeling of presence may lead to a paradoxical construct where you consciously and unconsciously respond to a virtual environment as if it was real, e.g., if one walks on a wooden plank on top of a skyscraper, his heart bit rate may accelerate, and he could feel some fear of heights even if he is in a virtual environment [Meehan et al., 2002]. Yet, the 3D rendering does not perfectly match reality. Thus the feeling of presence does not only rely on the virtual environment's visual fidelity. Some body-centred approaches to the feeling of presence state it is grounded in the ability to "do" in the virtual environment [Schubert et al., 2001; Slater et al., 1998], i.e. there is a correspondence between kinaesthetic proprioception and sensory inputs.In other terms, being able to move and interact with the virtual environment increases the feeling of presence.

The body representation seem to play a role in the feeling of presence as well, manifestly, a fully modelled body strengthens more the sensation of Presence than a simple geometric representation [Biocca, 1997]. Including body representation when considering the feeling of presence introduces the concept of self-presence as the perception of one's person in the virtual environment [K. M. Lee, 2004], but more importantly, it contributes to the sense of being in the virtual location [Slater, Spanlang, and Corominas, 2010].

Visual realism is not the main factor responsible for the feeling of presence, and various studies measuring it with questionnaires [Slater et al., 1999] cannot differentiate high-fidelity environments from low ones [Lugrin et al., 2015; Zimmons and Panter, 2003]. On the contrary, some sensory information like sound or touch cue a tremendous role in this feeling [Hendrix and Barfield, 1996]. Experimental studies reported a stronger sense of Presence on spatial sound [Poeschl et al., 2013], which can eventually compensate for the lack of visual fidelity and help in the process of self-motion perception [Väljamäe et al., 2004]. In some VR applications, when virtual objects' location matches with real ones, VR users experience haptic feedback, which increases the feeling of presence too, e.g. with fear of height studies, where the user walks on wooden planks represented in the virtual environment [Meehan et al., 2002]. The configuration of the matching objects frequently induces static haptic feedback, unlike general haptic feedback, which requires advanced third-party devices to reproduce the sensation. Moreover, the feeling of presence can arise from a well-designed virtual environment with a high agency degree supplying an immersive user experience [Jicol et al., 2021].

2.2.2 Immersion

The notion of immersion should not be confused with Presence since it has multiple definitions. According to Nilsson et al., 2016, it can refer to at least four characteristics. A far-reaching one for VR technology is the "system immersion" as a measurable technology characteristic intervening in the experience, i.e. the immersion level is more significant when the body tracking and the display are of higher fidelity [Slater, 2003]. Unlike the feeling of presence, immersion is related to the system's ability to maintain an equivalent fidelity to real-world sensations and not to the user's reaction.

Some definitions classify immersion as a perceptual response. Witmer and Singer, 1998 associate it with the feeling of being enveloped by, included in, and interacting with the virtual environment, considering that immersion can be pertinent to responses to narratives or challenges. For instance, Arsenault, 2005 describe fictional immersion as the process of being mentally gripped by characters, stories or entire worlds. This definition echoes their definition of systemic immersion as the mental absorption experienced when challenged by one's capabilities, including those when spectating and not participating directly. In the same fashion McMahan, 2013 link with engagement through the example of games where immersion is the state of being engrossed by the game, for instance, by winning and earning rewards or planning strategies. Yet, in our work, we refer to the immersion as the "system immersion" and not as the fictional one to avoid confusion.

However, to successfully elicit a sense of Presence and provide a satisfactory immersive VR experience, VR applications must fulfil multiple prerequisites.

2.3 Presence Prerequisites in virtual reality

2.3.1 Technical Constrains

The definitions used to introduce the concept of Presence in section 2.2.1 indicate that it is a fundamental trait of VR applications. Though the feeling of presence is not always straightforward, VR applications require to put some attention to technical prerequisites and design characteristics to elicit this sensation. Different studies state that a low latency with head tracking and a wide stereoscopic field of view enhances the feeling of presence and performance for varied tasks in VR [Arthur et al., 1993; Hale and Stanney, 2006; C. Lee et al., 2010; Lugrin et al., 2013]. With a VR head-mounted display (HMD), the user has an egocentric point of view and is immersed in the virtual environment, thus

being able to control a virtual body through good body tracking can increase the feeling of presence too [Cummings and Bailenson, 2016].

However, VR technologies come with exposure side effects and various symptoms gathered into what is commonly called cybersickness [Stanney et al., 1997] that can affect the feeling of presence [Weech et al., 2019]. Unlike simulation sickness, which is also related to motion sickness, disorientation seems to be prevalent compared to oculomotor symptoms and nausea, ranging from headache, sweating, and increased salivation to stomach awareness and vomiting [LaViola Jr, 2000; Stone III, 2017]. Additionally, researchers have extensively studied the causes of these effects, which some VR users seem to endure better. For instance, the sensory mismatch theory [Oman, 1990; Reason and Brand, 1975; Stanney et al., 1997] exposes the role of mismatches between observed and expected sensory signals. But there are other theories to the cause of VR sickness, such as the poison theory, which states that the illness would come from a biological defence to prevent us from being in such a situation similar to when we digest toxic substances [Treisman, 1977]. There are other popular theories, such as the postural instability theory, where the sickness comes from our lack of natural strategy to comply and maintain stable postural stability [Riccio and Stoffregen, 1991] or the rest frame theory, where the symptoms are due to response to linear and angular acceleration detected by our ear and eyes [Virre, 1996]. Overall, cybersickness depends on numerous requirements such as latency [McCauley and Sharkey, 1992], self-motion, visual display characteristics [Moss and Muth, 2011] and user experience [Gamito et al., 2008; Knight and Arns, 2006]. For instance, a larger field of view can increase such symptoms when combined with motion or when heavy latency induces a sensory mismatch. On the whole, Reben et al. refer to these factors of cybersickness as design factors, displays and rendering modes.

In our work, we placed some attention to the system's latency and the overall VR performance to avoid any cybersickness and breaks in Presence. The reason for this focus is that populating an audience with intelligent virtual characters can drastically reduce VR performance. A drop in the frame rate due to a longer rendering time can lead to cyber sickness. For this reason, the VR community provides guidelines to avoid such drawbacks,e.g. HMD manufacturers recommend having an average frame rate above 60 frames per second (FPS) and a maximal 20 milliseconds motion-to-photon latency [Oculus VR LLC, 2017].

Therefore, it is essential to design a VR application by considering the cost virtual humans have on the system. The rendering time is longer because VR technology needs

multiple view-ports. VR applications traditionally use a two-pass stereo rendering where all virtual entities are rendered for one eye and then for the second one, or with a single-pass stereo rendering where each entity is rendered on each eye one after the other. Thus, drawing detailed virtual humans can be harmful to performance due to the approaches to rendering a stereoscopic view. Unfortunately modern rendering techniques based on fovea to reduce the pixel density and reduce the rendering time are not yet available [Matthews et al., 2020]. The following section tackles the users' representation, which is of primary importance to elicit the feeling of presence as well.

2.3.2 Embodiment and Avatars

In section 2.2.1, we highlighted the importance of body representation for the feeling of presence to root the user in the virtual environment. Therefore, this section introduces the concept of embodiment and how VR users' avatars play a role in the users' perception of themselves and increase the feeling of presence and immersion.

The term avatar refers to the virtual body representation of the user [Biocca, 1997], and the role it has in VR applications is crucial for users. The avatar is the user's alter-ego in the virtual world one owns and controls. Avatars play a tremendous role in their immersion into the virtual environment and are the medium through which they interact with their surroundings. The concept of embodiment is of primary importance and refers to notions from cognitive science questioning how the brain represents the body and how it can be transformed under certain circumstances [Graziano and Botvinick, 2002; Metzinger, 2009]. As for VR technology, Biocca, 1997 evidenced the relation embodiment has with the users' sensation of being located in the virtual environment to increase it. If we look at the related literature, the embodiment is often associated with three notions: the sense of self-location, agency and body ownership [Kilteni et al., 2012], which we detail below.

Self-Location designates one's spatial experience of being inside a body and does not refer to the spatial experience of being inside a world from [Slater, Pérez Marcos, et al., 2009]. It is a core aspect of bodily self-consciousness [Blanke, 2012], together with self-identification [Serino et al., 2013] and first-person perspective [Blanke and Metzinger, 2009], which define the moment when one's self matches with one's body through a mental construct in the same manner as in the rubber hand experimentation [Botvinick and Cohen, 1998], where participants associate the rubber hand with their body schema. The

self-location is more significant for an egocentric visual perspective than a third-person point of view [K. M. Lee, 2004], especially when the virtual body matches the real one and haptic or tactile interaction occur.

Agency refers to the sense of having precise global motor control, which includes the subjective experience of action, control, intention, motor selection, and the conscious experience of will [Blanke and Metzinger, 2009]. If the consequences of a movement do not align with the user's expectations (synchronous visuomotor correlation), it can disturb the feeling of embodiment or the sense of being the cause of the action [Kilteni et al., 2012]. Although when the virtual body differs from the real one, the embodied user can tolerate a virtual space recalibration or motor learning [Clower and Boussaoud, 2000]. Moreover, it appears that high-quality tracking movements increase this feeling, therefore to avoid the users feeling like passive observers, applications should provide avatars with high interaction capabilities and good movement tracking [Argelaguet et al., 2016].

Body Ownership is associated with one's self-attribution of a body [Tsakiris et al., 2006]. Synchronous visuoproprioceptive correlations and morphological similarity condition this feeling, yet artificial bodies can elicit these feelings. The body representation can alter the illusion of body ownership. Latoschik et al., 2017 studied avatar realism and found that a realistic one evokes a greater virtual body ownership acceptance. Furthermore, in some applications, VR users conform to the behaviour they believe others would expect them to have, depending on their avatar [Yee and Bailenson, 2007]. This construct is called the Proteus effect and can lead to changes in the users' behaviour when users take over their avatar characteristics, e.g. altering the avatar's gender [Slater, Spanlang, Sanchez-Vives, et al., 2010] or skin colour [Peck et al., 2013] can modify users' behaviour and even temporarily change their body schema and self-image such as the body size [Buche and Bigot, 2018].

Through these definitions, we can see the role of avatars in VR to root the users in a virtual environment. These concepts define prerequisites for VR applications to let the users thoroughly enjoy the VR experience. In the VR research field, the notions like the feeling of presence characterise the user's VR experience. They determine how great this experience is for the user. Some of these concepts mainly rely on the user experience with the virtual environment and the range of interactions with it, whilst some depend on virtual characters' behaviour and interactions with the user's avatar to elicit other feelings,

such as the feeling of co-presence. The following section will introduce this feeling and develop its relationship with virtual humans.

2.3.3 Co-Presence

Co-presence depends on the others and the feeling of being with them. The term "co-presence" refers to the feeling of presence when related to the perception of being located with others in the same virtual environment. It was initially called social presence and described as the sensation of being in the company of other people in the virtual environment [Biocca et al., 2003]. The term "people" can refer to humans or artificial intelligence embodying a virtual character. As with the feeling of presence, Social Presence is a continuum in which others can be more or less present in the virtual environment [Short et al., 1976].

Goffman, 1963 present co-presence as the consciousness of others conveyed by senses, considering the body of others as a communication channel. Because the body can express rich communicative signals, co-presence occurs when people sense that they are close enough to be perceived and aware of being perceived by others. Biocca, 1997 stresses the importance of the sense of intelligence to increase this feeling when noticing behaviours that suggest the presence of another intelligence, meaning that a behaviour showing a human-like reaction increases this feeling. Similar definitions related to psychological involvement refer to the concepts of intimacy, immediacy or mutual understanding. Where immediacy is how intense and direct an interaction is between two persons [Mehrabian, 1972], and intimacy is a function of various parameters such as proximity, personal topics of conversation, and nonverbal behaviour like eye contact and smiling [Argyle and Dean, 1965]. Mutual understanding refers to the ability to make oneself understood in an interaction lacking communicative signals [Savicki and Kelley, 2000], i.e. the perceived similarity in emotions and attitudes, to measure social presence. A relatively recent definition from Palmer, 1995 of social presence involves behavioural engagement such as eye contact, non-verbal behaviour or turn-taking. They emphasise interaction and interactivity through multichannel exchanges of behaviours that virtual humans can achieve from within a VR environment.

Moreover, the technical prerequisites described in section 2.3.1, such as low latency and a good head-tracking or a wide stereoscopic field of view, can significantly enhance performances related to VR tasks [Steed et al., 2016] and communications between users if added to a better sensation of immersion and natural interactions with the virtual

environment or other users [Narayan et al., 2005]. Recent studies have explored how to leverage rich social signals and behaviour in social VR applications to achieve feelings of co-presence, immersion, and interaction by adding co-located characters and embodied avatars for the user to interact with [Latoschik et al., 2019]. Their results indicate a positive effect on the quality of user experience when the number of virtual characters increases. These outcomes are decisive for virtual audience simulation applications populated with virtual humans if provided with communicative capabilities.

2.4 Virtual Humans

When a program operates 3D-rendered characters, they are called virtual agents. In some cases, virtual agents are able to simulate personalities or traits that users can recognise [Cafaro et al., 2012]. Usually, they apply non-verbal behaviours like gaze direction or facial expressions to display these emotions, but virtual agents can also gain from context-related behaviours, the users' interpersonal distance or natural language. Since cultural background influences non-verbal behaviour, virtual agents can mimic certain non-verbal traits [Ting-Toomey and Dorjee, 2018]. But there is more than just agents' non-verbal behaviour, e.g. some applications studying crowds use groups of virtual agents to simulate navigation behaviours and social interactions in public spaces.

2.4.1 Crowd Simulation

Crowd simulation is distinguishable from virtual audiences by its focus on navigation and the evaluation of a scenario [Ulicny and Thalmann, 2001], whilst virtual audiences primarily focus on the communication and the emotions to display the reactions to the user's actions. Applications are often dedicated to a single crowd behaviour, and can sometimes represent the crowd as a flow or a network [Chenney, 2004]. On the contrary, systems simulating crowds' individuals rely on rules-based and behavioural systems, like in crowd behaviour simulation for training applications [S. Lee and Son, 2008; Pelechano Gómez et al., 2005; Varner et al., 1998]. Thus, applications range from motion simulation for architecture [Bouvier and Guilloteau, 1996], emergency evacuation conditions [Braun et al., 2003; Thompson and Marchant, 1995] or even sociological simulations [Braun et al., 2003].

In all of these applications, the main feature is the crowd's motion, which is not of

primary importance for virtual audiences. Yet there are critical characteristics that can benefit audience simulation as well. Some systems examine how virtual humans should respond to external events or other characters and what is the appropriate way to model this perception for many characters [Niederberger and Gross, 2003; Ulicny and Thalmann, 2002]. Furthermore, a virtual audience simulation would focus on the characters' expressiveness and would not refer to the same type of event or decision-making but rather to emotional contagion and reactions to the user's actions. There is another type of virtual human, called embodied conversational agent, which has a high communication ability and is very interesting for virtual audience simulation because they can express believable emotions.

2.4.2 Embodied Conversational Agents

When equipped with language capabilities, these virtual agents form the subgroup of embodied conversational agents (ECA). They blend natural language processing with non-verbal behaviour simulation to mimic human-like communication. Unlike chatbots, ECAs have a humanoid representation.

With natural language processing ability, ECAs can generate meaningful sentences. A common way to proceed is to convert the user voice into text and then analyse its semantics with a natural language processing service which feeds a second one able to generate a sentence [Poggi et al., 2005]. Some ECA use context-based synthesis to generate dialogues from past events, the agent's future tasks, and the current state of the conversation and to select the appropriate association of gesture and verbal answer [Morency et al., 2005]. Niewiadomski et al., 2009 favour a planning system to anticipate the speaker and listener's intent based on audio and visual input from services detecting the speaker's facial expressions. But there are many approaches for language processing, e.g. Flipper 2.0 use a hierarchical tree-based structure to build conversation [van Waterschoot et al., 2018], OpenDial uses probabilistic rules, and Pydial uses statistical networks to create ontologies. Some others use machine learning techniques, such as Huang et al., 2019, who use recurrent neural networks to generate conversations according to an annotated corpus. These techniques can adapt the ECA's behaviour, such as Dermouche and Pelachaud, 2019, who use Long Short-term Memory to adapt the agent's behaviour regarding its past actions and the user's current behaviour. Various works implemented agents with neural networks, e.g. to produce gestures from an input speech or from its semantics [Kucherenko et al., 2019; Kucherenko et al., 2020].

From these generated conversations, text-to-speech programs can read out loud the answer with an appropriate turn-taking mechanism which a utility-based algorithm inspired by cognitive psychology models drives to avoid interrupting the speaker's utterance [Janowski and André, 2019]. For instance, Morency et al., 2006 preferred users' eye gestures during interaction to detect the moment when the user is waiting for the ECA to take its turn.

Their 3D representation allows ECAs to convey more information by associating non-verbal behaviours with text-to-speech. In a dialogue, the listener's role is decisive in bringing meaning to dynamic communications [Garrod and Pickering, 2004]. Moreover, a conversation is not only a speech since multiple non-verbal behaviours come with the interaction and are part of the communication process. Thus, nonverbal behaviour can convey feelings, opinions or judgement representing the agent's stance. Such behaviour is studied to create a mutual posture for agents to share, for example, to display politeness or agreement [Prepin et al., 2012]. At a lower scale, non-verbal behaviour or backchannels can prevent virtual agents from freezing while listening to look more natural. These backchannels can influence the user depending on the displayed behaviour [Bevacqua, Hyniewska, et al., 2010] and convey a virtual agent's interest in a conversation or its opinion towards the speech [Bevacqua, Pammi, et al., 2010]. Yngve, 1970 defines backchannels as "non-intrusive acoustic and visual signals provided during the speaker's turn", e.g. a head nod with a stressed "M-hm" to express the agent's agreement. When voice and non-verbal behaviours are associated with human-agent interactions, ECAs can simulate advanced socioemotional behaviours, e.g. Potdevin et al., 2021 successfully built a tourism counsellor with intimate behaviours, where others worked on empathy or the feeling of rapport [Krämer et al., 2018; Lisetti et al., 2013].

Therefore, researchers use different approaches to model and implement the ECA's emotions. Ochs et al., 2008 provide a rule-based emotion model called SIP (Semantic-Interpretation-Principal), inferring the user's intention from dialogues and then inferring the empathetic emotional attitude toward the user. With another method, K. Anderson et al., 2013 used the Greta agent too but modelled its emotions with a multi-dimensional emotional one to create the agent's mood. It uses physiological signals from the users' body language, such as their gestures or posture and measured stress, with Bayesian networks to adapt the ECA's signals to display.

2.4.3 Virtual Agent Believability

A central concept for virtual human design and even more for ECA is believability. A virtual human is not considered believable only due to a realistic physical appearance but thanks to the emotions and personality they convey [Aylett, 2004; Lester et al., 1997]. A shared definition is that "believable" means "to accept as real", i.e. that the virtual human's "actions and communication ought to appear similar enough to those of real people" [Allbeck and Badler, 2001]. Thus, a virtual agent's behaviour must be consistent regarding its appearance, non-verbal behaviour and speech in the case of ECA [De Rosis et al., 2003]. Moreover, De Rosis et al., 2003 state that in the context of ECA, a believable agent must manage its behaviour according to the ongoing interaction. To echo the appraisal theories, Becker et al., 2005 stress the importance of virtual agents to adapt to the situation for the users to perceive them as more "human being-like". For instance, Niewiadomski et al., 2010 evaluated a virtual assistant's believability and evidenced that its behaviour should be "socially adapted" regarding the socio-cognitive factors of warmth and competence. These results testify to the complexity of evaluating virtual agents' behaviour regarding the number of parameters involved in the perception mechanism. Hence, most existing applications are tightened to a given domain and specialised for a specific context.

For instance, ECAs' complex abilities to display believable behaviours are applied to healthcare applications to back medical decision support [DeVault et al., 2014] or to coaching applications for job interview simulation [K. Anderson et al., 2013]. In a totally different context, agents or bots populating video games must be believable to provide engaging and human-like opponents to players, but the game characteristics define the bots' believability and not their socioemotional behaviours. Even et al. highlighted the importance of the judge's experience to evaluate the agent's believability which in their case was first-person shooter games. These results testify to the implication contextual behaviours have on agent believability perception. However, in these applications, the users do not share the virtual environment with the agents, unlike to virtual reality applications where users embody their own virtual human and can interact with their environment and the virtual agents. Yet, they convey rich social signals and express complex emotions through verbal and non-verbal behaviour and can significantly improve the user experience and performance, as explained in section 2.2.1. Therefore, the following section 2.5 introduces the theories of emotion perception and their implementation in behaviour models are discussed.

2.5 Emotion Perception

In the section 2.2.1, we emphasised the role the feeling of presence and co-presence can have on VR users. It is worth studying emotions perception and how rich social signals and interactions between virtual humans reinforce the co-presence.

2.5.1 Perception Theories

There are multiple emotion models with various approaches for virtual agents to communicate emotions. Unlike research about embodied conversational agents, we do not focus on verbal communication but only on non-verbal modalities since the audiences are primarily silent when listening and not engaged in a conversation. A way to work with these emotions is to use models that can be categorised into three groups: categorical, dimensional and appraisal-based [Gunes et al., 2011].

A mainstream categorical one is from P. Ekman and his colleagues, in which they gather a collection of facial expressions that can elicit basic emotions. It regroups fear, anger, disgust, sadness, contempt for the negative emotions and amusement, pride in achievement, satisfaction, relief and contentment for the positive ones [Ekman, 1999; Ekman and Friesen, 1975]. This model stresses the difference in the appraisal that might exist between positive and negative emotions and distinguishes emotions from affective phenomena, i.e. using the term "basic" with a list of characteristics to discriminate them, such as the duration, the distinctive physiological signals or if it is present in other primates [Ekman, 1999].

In contrast to categorical models, continuous models map an individual's emotional states along dimensions [Mehrabian, 1996]. It favours complex and multi-modal non-verbal signals to reflect the complexity and subtlety of affective states. A central model is the "Circomplex of Affect" from Mehrabian, 1972 presumes that each basic emotion represents a bipolar entity being a part of the same emotional continuum. This model proposes a dyad to describe emotions through the Valence and the Arousal, respectively opposing pleasant against unpleasant and relaxed against aroused emotions. Hence, its representation places emotions along these axes on a graph ranging from -1 to 1, as in Figure 2.1. Another continuous model is the PAD for pleasure-arousal-dominance, which places emotion within a 3D space adding the dominance – submissiveness axis, see Figure 2.2.

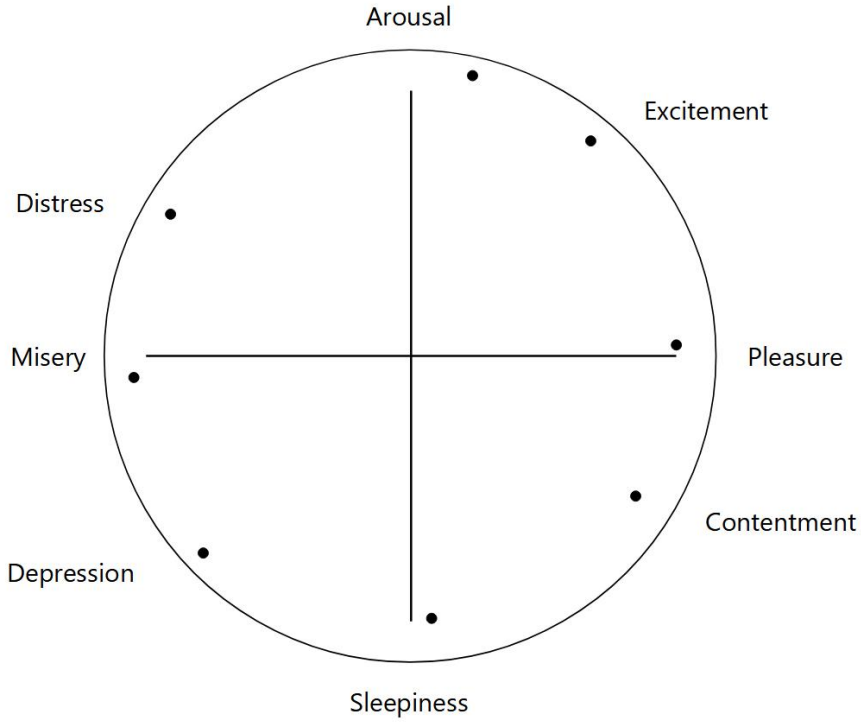


Figure 2.1 – The Circumplex of Affect from James A. Russel [Russell, 1980]

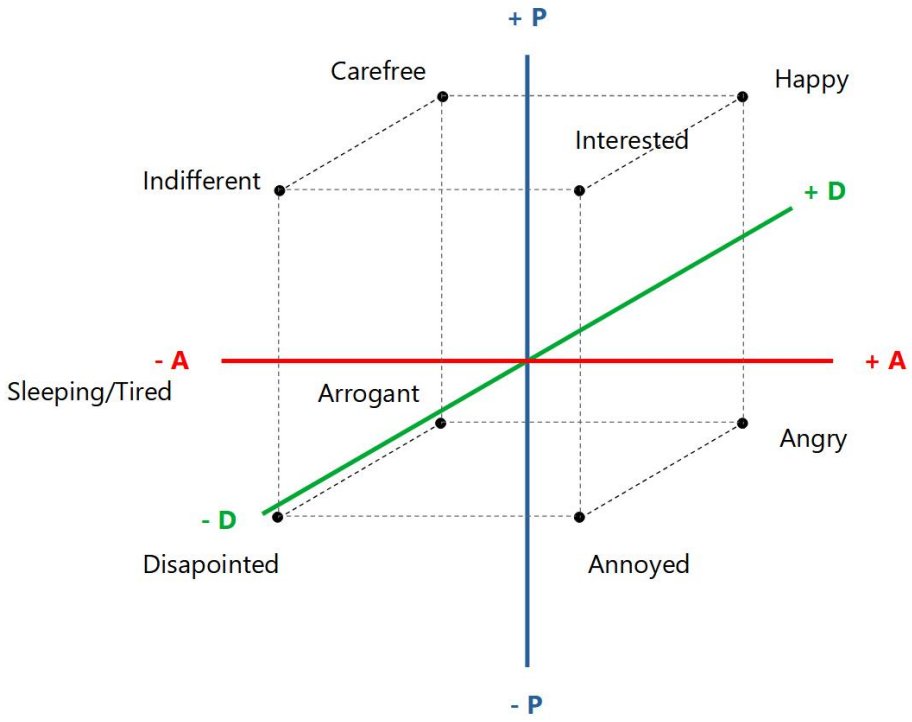


Figure 2.2 – An example of a PAD representation where facial expressions are mapped accordingly from Heudin, 2004

In the research for a complete description of affect and emotion perception, models can add a fourth or even fifth dimension to include expectation [Fontaine et al., 2007] and intensity [Grandjean et al., 2008]. Such a model uses appraisal theories that state the importance of the evaluation and the interpretation of an event to explain an individual's emotions [Roseman, 1991], i.e. it takes into account internal and external states [Scherer et al., 2001].One can say that the event's subjective evaluation in relation to the agent's goals and needs is responsible for emotion. This approach remains complex to implement, especially because emotional processes are elicited and dynamically patterned as people continually and recursively appraise behaviours or events [Sander et al., 2005]. However, it can be reduced to a dimensional model like virtual audience continuous models [Ortony et al., 1988] and then implemented for ECA to display emotions from a low-dimensional model [Ochs et al., 2005].

2.5.2 Multi-modal Signals

Previous works provide various approaches to modelling emotional perception and offer convergent proposals for the visual signals involved in these emotions. Facial expressions are the most shared non-verbal behaviours among these models. For example, categorical models like Ekman, 1999 express feelings through discrete facial expressions, whilst continuous models depict the relationship between emotional dimensions, e.g. valence and arousal. Moreover, these models include head movements and posture to refine their set of behaviours to display emotions [Lhommet and Marsella, 2015]. For instance, studies consider posture as a static behaviour where the orientation is crucial to describe openness or closeness [Mehrabian, 1972] and can be completed with the trunk lean [Harrigan, n.d.]. When a movement conveys communicative information, intentionally or not, the term gesture describes such a movement. In the same fashion, studies evidenced the role of each non-verbal behaviour, such as head movements which can be associated with different emotions, whether with dominant or inferior emotions [Mignault and Chaudhuri, 2003] or with positive and negative ones [Gunes and Pantic, 2010].

It is worth mentioning the backchannels that are powerful multi-modal behaviours conveying rich emotional signals. These behaviours are tremendous for ECA to "exhibit appropriate behaviour while speaking and listening" [Bevacqua, Pammi, et al., 2010]. The term backchannel has been introduced to describe "non-intrusive acoustic and visual signals provided during the speaker's turn" [Yngve, 1970]. These signals provide primary communicative functions during a conversation, such as attention, interest or attitude toward the speaker [Allwood et al., 1992; Poggi, 2007]. Yet, there are massively used with ECAs to increase their range of emotions during conversations. Still, since a virtual audience is in a listening situation it is interesting to keep in mind backchannel showing agreement, interest or understanding signals like a head nod associated with a para-verbal "mmhmm" [Bevacqua, Pammi, et al., 2010]. Interestingly virtual audience behaviour models are extensively inspired by these models and perceptual studies and thus strongly influence their implementation.

2.6 Virtual Audience Behaviour Models

Since virtual humans populate virtual audiences, the same models and approaches condition their build, implementation and evaluation by nature. In the same way, as in the previous section, the model depends on the perception theory employed. Hence, we can

find the three same approaches inspired by categorical, dimensional and appraisal-based models.

2.6.1 Knowledge Based Models

The most common approach regroups categorical models and those based on heuristics built from the professionals' expertise. Models tighten to a specific application domain rely on experts' knowledge of the given discipline, such as lecturers for a classroom simulation [Lugrin et al., 2016] or therapists for public speaking anxiety [Kahlon et al., 2019; Lindner et al., 2021]. This method is compatible with categorical theories and can be used to design behaviour models [Fukuda et al., 2017]. Thus, the resulting models give some space to domain-specific behaviours since they are dedicated to given situations, e.g. students falling asleep to depict a bored audience. It is the case in the Breaking Bad Behaviour application [Lugrin et al., 2016], where such disruptive behaviours are triggered to display different level of discipline. Delamarre, 2020 proposes another classroom management training system to drive the virtual audience behaviour according to scenarios designed by experts in classroom management. Education experts provided a list of scenarios from which they could extract non-verbal behaviours classified into three groups (neutral, off-task and aggressive behaviours) subdivided into low and high intensity behaviours [Delamarre, 2020, page 65]. A slightly different method is asking the final users to validate the behaviours. Kahlon et al., 2019 developed a scenario that inspired multiple culturally and age-appropriate Norwegian classrooms and adjusted their behaviours after feedback from testing by four Norwegian adolescents.

Nevertheless, all these applications require experts to validate the virtual audiences, whether in educational applications [Delamarre et al., 2021; Lugrin et al., 2016], therapeutic applications [Lindner et al., 2021], or when aiming for commercial use [Ovation [1] VR-Speaking, LLC, 2022]. The resulting models are specialised, and only experts' inputs can extend them. Moreover, since this approach is categorical, it does not provide fine-tuning of the audience. This method results in neglecting a wide range of non-verbal behaviours and backchannels, e.g. behaviour frequency, phatic expressions, gesture and their blending.

Unlike the categorical approach, dimensional models offer to parameter the audience to associate various non-verbal behaviours and build more complex ones to vary the perceived audience's attitude. Since no experts validate the model, they need to run perception

1. Ovation Application, https://www.ovationvr.com/, [Accessed March 29, 2022]

studies to evaluate the resulting audience. The following sections provide insight into two methods to build and assess dimensional virtual audience behaviour models.

2.6.2 Behavioural Styles

To start with, the works of Ni Kanga, Willem-Paul Brinkman, M.Birna van Riemsdijk, and Mark Neerincx [Kang et al., 2016; Kang et al., 2013] provide a solid ground for the configuration of virtual audience behaviours according to various styles and different dimensions to model the users' perception. These studies bring two interesting results about the users: the first is their ability to differentiate audience attitudes and the second is their ability to notice a different level of arousal and valence.

Background

The motivation for this research was to create a model which can generate a virtual audience that is realistic, adaptable and expressive to display different attitudes [Kang et al., 2013]. To do so, Ni Kang and her colleagues created a parameterised behaviour model they implemented following the architecture introduced by Norvig and Russel, 2002, which divides it into three modules, the mind, the perception and the behaviour module. Two approaches were available to design such a system: the first was to use the existing theories on behaviour generation consisting of extracting the information from the literature, and the second was the statistical approach where the behaviour model relies on real-life observation information. Different studies used the theoretical approach in which the literature, cognitive psychology theories, and expert knowledge about the application contexts are used to establish the procedures to generate the appropriate behaviours. This is the case in Bevacqua et al., 2012, where behaviour rules implement an embodied conversational agent, or in Lugrin et al., 2016, where the virtual audience is designed based on expert knowledge to simulate classroom behaviour. As for the statistical analysis, models try to predict virtual agents' behaviours [Chollet, Ochs, et al., 2014]. At that time, Kang et al., 2013 built a statistical audience behaviour model based on real-life observations.

Perception Studies

To build their nonverbal behaviours library, Kang et al., 2013 observed a real-life audience and coded their non-verbal behaviour depending on four types of presentation and

different questionnaires. The four types of attitudes were positive, neutral, bored and critical. Each presentation had a different topic, ranging from software design to philosophy. Only PhD students attempted the presentations, so the observed audience was only based on students' behaviours and no other socio-professional category. In the positive condition, spectators had rewards to motivate them, whilst in the critical, the audience received complaints about PhD work and behaviour. After annotating the videos, they linked behaviours to moods or personalities with two standardised questionnaires. First, they used the SAM questionnaire to correlate behaviours with three emotional dimensions, valence, arousal and dominance and then the IPIP-NEO to measure the Big Five personality traits (extroversion, agreeableness, conscientiousness and neuroticism). Finally, they designed a statistical model that generates parameterised audiences to display various attitudes based on the subjective data extracted from the questionnaires.

This model requires a perception evaluation to confirm the correlation between behaviours, mood and personality hence investigating if the participant can perceive different attitudes and how to discern the model's dimension, e.g. valence, arousal and neuroticism. Kang and her colleagues generated different audiences to include all parameter variations to know if changes in the nonverbal behaviours were significantly perceived and correctly associated with these dimensions. The statistical analysis reveals that users can significantly recognise different attitudes with free description, but results cannot confirm they can still identify them when manipulating the model's parameters. At this point, the lack of facial expressions and the observed interaction between valence and arousal can explain the study's limits.

Therefore, they conducted a second study to investigate this specific point and further detail how participants perceive five audience categories with a neutral attitude as a baseline [Kang et al., 2016]. These audience settings were related to explicit contexts such as lectures, business proposals, budget cuts, not-funny shows and more, clustered through statistical analysis. As a result, different levels of attentiveness were significantly perceived. Finally, thanks to these analyses the new statistical model was evaluated and provided valuable results regarding the postures, the frequency of body movements and reactions to disruptive events. The benefits are twofold: the studies provide insight into the user perception of audience attitudes and a first parameterised audience behaviour model. Even if this model is detailed enough to generate different attitudes it is still unclear how they are perceived in VR.

2.6.3 Crowd-sourced Model

To follow up with these perception studies, Chollet and Scherer, 2017 proposed another model built on the literature and then evaluated it with a statistical approach. They also submitted a valence arousal model since Kang et al., 2016 showed that it is possible to recognise different attitudes according to these two dimensions. Their strategy was to use a statistical approach to evaluate the model upon crowd-sourced data.

Background: This research aims at providing insight into the impact the audience's non-verbal behaviours have on the user perception and how spectators individually influence the overall attitude. Such understanding of the audience perception allows for developing a model with nonverbal behaviours fine-tuned to generate the desired attitude. They decided to use a web interface to collect crowd-sourced data and let the participants design the behaviours by manipulating different parameters, comparably as in Ochs et al., 2013.

Methodology With this web interface, participants could design an agent's behaviour according to a paired value of valence and arousal through seven nonverbal parameters extracted from the literature:

— Posture
— Facial Expression
— Face Frequency
— Head movements
— Head Frequency
— Gaze
— Gaze away Frequency

The valence and arousal dimensions were respectively described as the opinion and the engagement toward the speech or the speaker. Each of these parameters had multiple choices: rarely, sometimes and often for the behaviour frequencies and rarely, about half the time, most of the time and always for the gaze away frequency. The facial expressions and head movements included a "none" item to consider the absence of behaviour and comprised the smile, frown, eyebrows raised for the facial expressions and nod and shake for the head. As for the posture, they were composed of two elements to make six unique ones. The first component is the agent's relaxation, and the second is how lean the agent was, which gives three possibilities, backward, upright and forward and six arms position,

hands behind the head, arms crossed, hands-on lap, self-hold, chin on fist and hands together. Hence, participants were able to design the behaviour of a single agent and thus investigate the influence of these two dimensions and how these nonverbal behaviour are associated with them.

They designed another web interface where users had to identify the audience's state by rating it through five-point scales, i.e. one for the valence and one for the arousal. The motivation was to confirm their model's correctness through a second crowd-sourcing campaign. Besides, they investigated how the audiences were perceived, depending on the number of virtual spectators that display a particular behaviour, and if they can predict the engagement and arousal scores of a virtual audience from the displayed nonverbal behaviours.

Crowd-sourcing Perception Study: Results provide rich details on how virtual audiences are perceived. It first reveals how a higher arousal value leads to more frequent behaviour and closer postures, whilst valence is associated with nonverbal behaviours like frowning for the negative valence and smiling for the positive one. For instance, the gaze away frequency is the most influential behaviour regarding the perception of the audience's engagement, while head movements seem to be the most influential behaviour for the perception of the audience's opinion. Mathieu Chollet and his colleagues evaluated the model for an audience of 10 agents placed in two rows.

For this configuration, the location of individual agents had no significant effect on the users' identification rate. Finally, the perceived audience's attitude varied depending on the number of agents displaying a specific behaviour. For a given ten agents virtual audience, three agents showing a negative behaviour will likely trigger the perception of a negative audience, whereas the threshold goes up to 6 agents displaying a positive behaviour for the perception of a positive audience. To perceive the audience's engagement from at least four virtual spectators displaying a given level of engagement among the ten agents. Nonetheless, three different facial expressions, two types of head movements and six different postures are sufficient to allow a user to perceive different audience attitudes. These behaviours remain generic since they did not evaluate this model for a specific context. On this account, the statistical analysis provides enough information to design virtual audiences based on different levels of valence and arousal but also indicates how changing one agent's behaviour at a time can alter the attitude. With this study, they also detailed how contradictory behaviours are perceived and how prominent behaviours

affect the global perception of the audience. For instance, behaviour signals related to negative valence seem to be twice as influential as positive ones. Finally, it appears that agents' location in the audience does not significantly influence the user perception, whilst the proportion of agents with congruent behaviour with the intended generated attitude seems to be related to the capacity to recognise the audience's attitude. Table 2.1 and Table 2.2 list the non-verbal behaviours of Chollet and Scherer, 2017 and Kang et al., 2016 according to different levels of Valence and Arousal.

Table 2.1 – Summary of the non-verbal behaviours used in Kang et al., 2016's model.

		Posture	Facial Expressions	Gaze Away Direction
Valence	Negative	Hands open on desk; one hand touching or holding the other hand; standard position, both feet flat on the floor; leg joggling or tapping on the floor; with more frequent posture shifts	Lip Corners down; inner brows down	
	Neutral	Clenched hands resting on desk; hands open on desk one hand touching the neck, with the other resting on the front torso; folded arms	No facial expression	
	Positive	Supporting the head; one or two hands tap on the desk continuously; crossed or twisted legs; with less frequent posture shifts	Smile (mouse close) with lip corners up	
Arousal	Low	Torso forward	Lowered head	Upright, lowered head, tilted head
	Medium	Torso backward;		Upright, tilted position, facing the front
	High	Torso upright; torso forward; torso backward		Upright

Table 2.2 – Summary of the non-verbal behaviours used in Chollet and Scherer, 2017's model.

		Posture	Head Movement	Facial Expressions	Gaze Away Direction
Valence	Negative	Closed postures	Shakes	Frowns	
	Neutral	Upright		Eyebrows raising	
	Positive	Relaxed postures	Nods	Smiles	
Arousal	Low	Low postural proximity	Low frequency	Low frequency	High gaze aversion frequency
	Medium	Medium postural proximity	Medium frequency	Medium frequency	Medium gaze aversion frequency
	High	High postural proximity	High frequency	High frequency	Low gaze aversion frequency

With an overview of different theories (Section 2.5) and approaches to model virtual audience it is also interesting to review how they are implemented for different virtual reality applications. Therefore, the next section 2.7 will provide a non-exhaustive introduction to the different application domains for virtual audience simulation in VR and the limitations they bring.

2.7 Applications Domains

The simulation of virtual audiences remains complicated, generally speaking, for all virtual humans. It is particularly true since their behaviour modelling implies considering many socio-cognitive, cultural and contextual criteria. Hence, depending on the application domain, studies focused on specific features for virtual audiences to fulfil particular purposes.

2.7.1 Virtual Reality Exposure Therapy

The ability VR applications have to help and assist therapists in psychological disorders treatment is now studied for nearly thirty years since the first controlled studies in the early 90s [Williford et al., 1993], and the first study investigating VR effectiveness with social phobia from North et al., 1998 VR seems to reproduce similar results as in exposure therapy. This method consists of repeatedly exposing a patient to varying degrees of a feared stimulus to modify a behavioural or cognitive response [P. L. Anderson et al., 2005; Rothbaum et al., 2000]. For this promise to hold for VR applications where the social aspect is key such as public speaking anxiety treatment, fine control of stimulus is paramount for rooting the user in the virtual scene and providing therapeutic virtual environments. As such, the feeling of presence or telepresence was already recognised to be a key factor of VR exposure therapy [North et al., 1998; Pertaub et al., 2002; Slater et al., 2006]. Hence, it led to the study of VR therapeutic intervention for fear of public speaking [Slater et al., 2006] where a particular focus was put on the patient's response to the virtual audience's behaviour which evidenced that VR users are affected by the different types of virtual audiences. This 'presence response' refers to experienced anxiety by a patient within the virtual environment when sufficiently similar to a real-life experience in a comparable situation [Slater et al., 2006]. Therefore, if virtual audiences provide interpersonal interaction as well as a sensory awareness of the agents or other users the

feeling co-presence can be elicited [Slater et al., 2000].

Moreover, one of the main advantages of using a virtual environment is its ability to reproduce situations that are almost unfeasible or unsafe during in vivo simulation. A VR application provides an ecological environment that empowers therapists with fine control over the fear stimulus, like in acrophobia therapy where it is unfeasible and unsafe to practise on top of a building. Such a virtual environment supplies the therapists with stronger stimuli which can enhance the desensitisation due to their greater intensity [Williford et al., 1993]. For instance, with the fear of public speaking, gathering a crowd in an auditorium is barely feasible for successive therapy sessions, even if we do not consider the fact that some patients are not willing to participate in group therapy as well as the difficulty to control the crowd's behaviour and feedback [P. Anderson et al., 2003], i.e. the stimulus to which the patient is exposed.

Since then, various studies explored how different application domains could benefit from the same effect. Slater, 2009 summarised it as if you feel that what is happening in the simulation is happening to you, hence you are more likely to respond as if it were real. Therefore, multiple studies investigated how this can be transferred to other applications like training ones. In the context of this thesis, because the range of studies regarding VR exposure therapy is widespread, we narrow down the focus on social simulation applications, such as those for public speaking training or public speaking anxiety (PSA) therapy.

For instance, in cognitive-behavioural therapy, PSA is defined as a social anxiety disorder expressed by the fear of negative evaluation of others in social situations and feeling embarrassed or humiliated [American Psychiatric Association, Association, et al., 2013]. The use of virtual agents as a media to mitigate PSA [P. L. Anderson et al., 2013; Wallach et al., 2009] or to improve public speaking skills [Batrinca et al., 2013] became commonly used with VR. The reason is that virtual audiences can elicit stress or anxiety similar to a real audience and can be used in a training system [Kelly et al., 2007].

Hence, therapeutic or training systems require the VR environment to be populated with groups of virtual spectators like in public speaking skills [Batrinca et al., 2013; Chollet, Sratou, et al., 2014], audience management training [Delamarre et al., 2021; Fukuda et al., 2017; Hayes et al., 2013; Lugrin et al., 2016; Shernoff et al., 2020], or forms of social anxiety disorders treatment applications [P. L. Anderson et al., 2013; Kahlon et al., 2019; Wallach et al., 2009].

2.7.2 Virtual Reality Training Application

The key feature remains the virtual audience's behaviour that is supposed to elicit the user's response, i.e. the stimulus. Hence, in these studies and applications, the behaviour of the audience is manipulated so that virtual agents display various attitudes. For instance, to elicit anxiety in public speaking training and therapeutic applications or to match a specific situation for the user to face during a training session.

2.7.3 Virtual Audience Controls

Different techniques are used to represent and control the VA. One approach for representing the virtual audience is to use videos that can be embedded into the virtual environment or directly displayed as immersive content. In this configuration, the video clips change to display a different agent behaviour [Gallego et al., 2011; Klinger et al., 2005,Ovation [2]]. It implies having video records for each behaviour a spectator can display, this is why this kind of representation will not be further described in this chapter since it mainly relies on huge libraries of behaviours actors have played that are then displayed on a screen. In fact, this might be very expensive and time-consuming to film actors compared to software-generated animations. Thus, the second approach is to use 3D models animated by a script, e.g. within a game engine. Still, whether or not it relies on 3D models or videos, both are meant to display virtual agents' non-verbal behaviours.

There are several challenges in the design of social skills training or PSA treatment systems including interactive virtual agents. A critical one is to control a virtual audience to follow a training plan, whilst allowing it to react to the user's behaviours and interactions. As an example, behaviours models like those introduced in the previous section 2.6 provide exhaustive tools but require to be compatible with instructors requirements.

Wizard Of Oz Applications

A significant limitation of existing systems is their tendency to rely on a Wizard of Oz (WoZ) approach to drive the audience in reaction to the user's performance [Fukuda et al., 2017; Harris et al., 2002; Pertaub et al., 2002]. For instance, Chollet et al., 2015 used a WoZ method to provide the speaker's performance input to the system and adapt the virtual audience behaviour according to such data. In *Breaking Bad Behaviours*, an instructor drives each virtual agent's behaviour to tailor it to the trainee's actions at

2. Ovation Application, https://www.ovationvr.com/, [Accessed March 29, 2022]

run-time [Lugrin et al., 2016]. Such systems seem to elicit a heavy cognitive load for the instructors when it comes to following a classroom strategy and manually authoring each virtual agent [Mouw et al., 2020]. These applications always face the same limitation: they require a human in the loop or an expert to feed the application with relevant data to adapt the ongoing simulation, i.e. the audience's behaviour.

Autonomous Applications

On the opposite, Delamarre et al., 2021 provide another classroom management training system to drive virtual audiences' behaviour according to scenarios designed by experts in classroom management. As a result, because the virtual audience is scripted, it has no scenario flexibility and might suffer from simulation repetitiveness, but it provides a high-level authoring tool for users without knowledge of scripting languages. Models analysing the VR users' behaviour and actions can drive such a fully autonomous system to adapt it to the virtual audience [Chollet et al., 2022]. Both designs commit to the system's ability to successfully adapt the VA's behaviour to the users' actions. Applications like CICERO continuously adapt the VA's attitude and perfectly suit training applications that aim at completing traditional teaching or personal training. But it comes at a price [Batrinca et al., 2013]: first, it relies on physiological data and audio or video recording, which can be invasive, expensive and complex to manipulate. Then, it removes the experts from the training supervision. Therefore, such a system does not suit the therapeutic applications we are aiming for, where replacing tutor expertise with an autonomous component is not desirable, e.g. VR therapy and training could need real-time adjustments and temporary fine control of the environment. As far as we know, there is not yet an application that successfully combines AI-driven applications with therapeutic scenarios and experts' direct intervention.

2.7.4 Synthesis

The sections 2.7 above highlight the potential VR applications have for therapeutic and training systems with an interest in social applications.We have seen that the simulation of virtual audiences is a key feature and that various solutions already exist to simulate complex audience behaviours. We focused on the behaviour models, which offer several behaviours and interactions modelled via dimensions and representative variables. These models generate immersive virtual environments populated with believable virtual

agents. However, they do not necessarily have VR evaluations, which could lead to different perceptual results. Their implementations are not available, and none of the mentioned applications provides a generic architecture for the design of therapeutic or training applications. We featured two types of approaches, which both have issues regarding virtual audience behaviour control. WoZ and autonomous systems are complementary, but applications like the IVT-T provide high-level controls with autonomous scenarios [Delamarre et al., 2021], whilst WoZ applications benefit from live interactions and modification of the virtual environment from the experts like in *Breaking Bad Behaviours* [Lugrin et al., 2016].

This research project faces common issues related to VR applications' scalability, model implementation and technical requirement for end users. Consequently, this project has mandatory challenges to solve. Thus, our research focuses on designing an efficient behaviour model with rich communicative capacities evaluated in VR and its validation within VR applications by professionals regarding its acceptance and usability.

2.8 Conclusion

In this chapter, we set forth the research problem this thesis tackles regarding the design of a virtual audience, notably by its non-verbal behaviour and how it is perceived. We first introduced the concept of a virtual audience by providing a wide definition of all the notions related to virtual humans simulation and VR such as the feeling of presence. From this section, we expressed the challenges and key features to simulate virtual audiences in VR.

Another challenge studied in the second part is the simulation of expressive agents and how the VR users perceive their behaviour. We had a specific interest in virtual audience models and the different methods used to model non-verbal behaviour and how they are evaluated. Thus, we highlighted the need for further VR perception studies to confirm prior works.

Finally, in the last section, we pictured all these issues through multiple training and therapeutic systems implementing virtual audiences in VR that all suffer from limitations due to the audience controls. Therefore, the following chapter will present the technical approach used to design a virtual audience and provide an efficient animation pipeline which was the foundation for our research to build and study our behaviour model in VR.

Chapter 3

AMBIANCE Behaviour Model

3.1 Introduction

In chapter 2, we listed the requirements and features for virtual audience simulation in VR. We particularly focused on two virtual audience behaviour models from Chollet and Scherer, 2017 and Kang et al., 2016, which provide rich social signals according to a dimensional approach to emotion perception. We also described performance requirements for VR applications to simulate audiences without decreasing the user experience.

These models use valence and arousal as dimensions, which respectively represent the opinion and the engagement towards the speech and the user. Various user evaluations provide nonverbal behaviours according to valence or arousal. For instance, the spectators' gaze away frequency is linked with the arousal and the type of head movements and facial expressions to the valence. Together with these behaviours, the models provide more insight into the audience design and perception, e.g. the significance of each behaviour for user perception or the proportion of spectators with a congruent behaviour for the user to recognise the attitude.

In this chapter, we describe how we used these models to build ours. It relies on the same two dimensions but with the specificity that it models the audience's behaviour from a given attitude. The resulting model called AMBIANCE (Attitude Model defining the Behaviour of Individual AgeNts for Constructing audienCEs) aggregates VR user perception studies' results to match with a set of five attitudes, i.e. bored, indifferent, critical, interested and enthusiastic. This model's novelty lies in regrouping nonverbal behaviours under attitudes to ease the AMBIANCE integration and control in various VR applications. Finally, this chapter ends with a section describing the model implementation in a game engine and providing more insight into its operating principle.

3.2 The AMBIANCE Model

The work from Chollet and Scherer, 2017; Kang et al., 2016 evidenced that people can tell apart different audience attitudes, how to generate these attitudes and how the non-verbal behaviours are associated with them. Both models use the Valence and Arousal dimensions to associate diverse behaviours with a paired value of opinion and engagement. While the approach and the methodologies used differ from one study to another, they provide similar or complementary results. Thus, we decided to use a slightly different method and build a model by formalising the results from both Kang et al., 2016 and Chollet and Scherer, 2017.

3.2.1 Valence-Arousal Model

The AMBIANCE is a continuous two-dimensional Valence-Arousal model, which groups non-verbal behaviours related to the same attitude. We built behaviour rules to generate audience attitudes from valence-arousal pair values ranging from low (bored, impatient) to high (interested, enthusiastic). These rules use probabilities to gather a behaviour with an attitude from Kang et al., 2016's studies. Then by formalising the results and guidelines from Chollet and Scherer, 2017's statistical analysis, we linked the previous behaviours and attitudes to the corresponding rules. Thus, we designed the AMBIANCE's rules to be compatible with the most detailed results, i.e. to take into consideration all kinds of behaviours and their variations. Chollet and Scherer, 2017 online studies' results provide different parameters to model the generated attitude and measure the perceived opinion and engagement. They also give insight into how users perceive the entire audience, such as the proportion of agents needed to display an attitude and how frequently agents should display their behaviours. Thus, we grouped the behaviours for a given pair of valence and arousal, which results in a minimal and maximal frequency for each behaviour (arousal) and a condition on the proportion of agents displaying the same type of behaviour (valence). As in both models, we clustered the non-verbal behaviour in types which merge into four: firstly, postures which include the torso, the legs and the arms movements, secondly head movements which are non-verbal backchannels like nodding or shaking the head, thirdly the facial expressions and fourthly the gaze to determine whether or not virtual spectators look at the speaker.

Then, these clustered behaviours are expressed in the form of a set of user-customisable rules. A rule is a series of parameters describing a nonverbal behaviour that can be applied

to the audience's virtual agents. An audience's attitude is therefore specified through rules. Rules are divided into categories corresponding to model's nonverbal behaviours (posture, gaze, head movement and facial expression). A rule follows the following format:

$$rule_x(Type, Frequency, Proportion) \tag{3.1}$$

where x is the nonverbal behaviour category of the rule (eg. posture, gaze), $Type$ a predefined parameter characterising the nonverbal behaviour in the category, $Frequency$ how often the behaviour is displayed for each active agent, and $Proportion$ the number of agents in the audience which will be actively displaying the behaviour. An example of a rule for the facial expression category would be:

$$rule_{FacialExpression}(Smile, 0.5, 0.7) \tag{3.2}$$

This can be read as 70% of the agents smile 50% of a given period. Hence, the overall virtual audience's attitude is an aggregate of each nonverbal behaviour which are allocated to the virtual agents. The rest of the virtual audience's agents which are not affected by any of the rules can either keep their current nonverbal behaviours or can be assigned any. However, the model should prevent contradictory nonverbal behaviours to be associated. It can lead to ambiguous agent behaviour, *e.g.* to prevent agents to smile and shake their heads at the same time or to nod while gazing away from the user. These sets of rules can also include more than one rule per category and can for instance trigger three different rules for the posture which would allow more complexity and variations in the agents' behaviours.

3.2.2 The AMBIANCE as a Sum of Behaviours

In a nutshell, an attitude is the sum of applied behaviour rules, which when implemented combine any set of given rules into audience animations. Thus, the attitude is composed of two types of rules, the ones regarding the nonverbal behaviours (see Equation 3.3) and those to keep the audience's behaviour coherent with respect to the attitude to display (see Equation 3.4).

$$
\begin{aligned}
&rule_{FacialExpression}(FacialExpressionTypes, Frequency, Proportion)\\
&rule_{HeadMovement}(HeadMovementType, Frequency, Proportion)\\
&rule_{Posture}(PostureType, Frequency(*), Proportion)\\
&rule_{Gaze}(GazeDirection, Frequency, Proportion)
\end{aligned}
\tag{3.3}
$$

(*) At first, we considered the postures' frequency as fixed to simplify the model.

$$
\begin{aligned}
&rule_{Congruence}(Dimension, Level, Proportion)\\
&rule_{Uncanniness}([Type; ID], [Type; ID])
\end{aligned}
\tag{3.4}
$$

This second type of rule specifies the condition for an attitude to be valid and thus correctly perceived by the users. For instance, behaviours associated with a negative valence seem to be more vivid cues than positive ones. In consequence, the proportion of agents displaying negative or positive behaviour requires a check so that the user's perception is congruent with the intended one. Thus, in the equation, Dimension, Level, and Proportion respectively correspond to the valence or arousal dimensions, the dimension's intensity level ranging from very low to very high and the proportion of virtual agent in per cent displaying a congruent behaviour, i.e. corresponding to the given attitude. Another subtype of rules we called the "*uncanniness rules*" is responsible for preventing two behaviours from blending because they can lead to a weird or uncanny movement, e.g. nodding while rotating the neck to look at the window instead of the speaker. An *uncanniness rule* simply takes a type of non-verbal behaviour and a unique ID to single it out (see Equation 3.5).

$$
rule_{Uncanniness}([HeadMovement; Nod], [Gaze; Downward]) \tag{3.5}
$$

Some other parameters are available to design behaviour rules, such as the influence virtual agents' position or multi-modal backchannels have on user perception. These additional parameters were not included in the original model because they are either not significant or not included in past studies on audience perception, such as the multi-modal backchannel. The AMBIANCE did not contain sound in its rules as well, because none of the works studied evaluated it. As for the agents' reactions to other behaviours, Kang et al., 2016 provide some insight into how the agents should behave when a disruptive

event happens depending on the audience's attitude. However, we did not consider results for disruptive events since participants were given a context during the study, which might have influenced their answers to the questionnaires. Also, we developed the AMBIANCE to be a context-free model. Consequently, we decided not to include domain specific reaction behaviours in the model. For example, a training application for pre-service teacher such as *Breaking bad behaviour* requires students reactions such as laugh when a phone rings in the classroom. Nonetheless, Chapter 4 introduces our proposition for such interactions to be played out in our VR training application for instructors to trigger disruptive events.

In its final form, the model is expressed as follows (Equation 3.6):

$$\begin{aligned} Attitude = & \sum rule_x(Type, Frequency, Proportion) \\ & \wedge \sum rule_{Congruence}(Dimension, Level, Proportion) \\ & \wedge \sum rule_{Uncanniness}([Type; ID], [Type; ID]) \end{aligned} \tag{3.6}$$

Where X is the type of behaviour rule.

3.3 The AMBIANCE Implementation

The AMBIANCE model is implemented as an Unreal Engine Plugin. This game engine developed by Epic Games provides tons of development-ready assets and advanced features, such as a dynamic animation pipeline with inverse kinematic, facial expressions and skeleton controls. We chose to embed the AMBIANCE into a plugin to ease its integration in different projects so that only the plugin is needed to start using the model independently of the application. Hence, this plugin only contains the critical features to start simulating audiences with our model. In a nutshell, it regroups three modules: first, the AMBIANCE contains a *Manager* to drive the entire virtual audience and expose the main features to developers, then a character module that executes the behaviours' logic and finally, an *Animation module* responsible for displaying the behaviours when the *Character module* orders it. The *Manager module* is a unique instance in the simulation, whilst each agent populating the virtual audience has a *Character* and an *Animation module*. Additionally, to speed up the prototyping process, we designed the plugin to remain compatible with the engine's visual scripting tool called Blueprint. Appendix C provides

some examples of blueprint functions.

3.3.1 Plugin's Architecture

The plugin takes advantage of existing classes within the Unreal Engine 4 (UE4) to speed up the process of adding new virtual agents. The Character and the Animation Instance Class from the UE4 respectively provide basic movement controls, physics interactions, a skeleton component and a set of animation features to display complex animations. The skeleton is a hierarchy of bones and joints used to mimic humans' natural movements. Additionally, the plugin includes an example set-up for the UE4's mannequin, which is the reference for the controls and the animations. When animations are imported from external animations tools and designed for another character, they often need to be re-targeted to the UE4's mannequin first, i.e. to adapt the animation to the mannequin's skeleton. In doing so, it avoids dedicating the plugin to a specific character type.

3.3.2 Audience Manager

The *Audience Manager* explicitly provides methods with the same signature as in the aforementioned rules. For each behaviour, the *Manager* needs an animation type, a proportion of virtual agents which play it and a display frequency. Otherwise, the *Manager* can directly apply a behaviour to a single agent. Thus, the *Manager* can quickly change the audience's attitude by using a set of rules or adjusting a single agent's behaviour. Moreover, the *Manager* provides complementary features to target a specific population. Sometimes it is preferable to target a subgroup of agents with a distinct behaviour, for instance, when only the virtual agents gazing toward the speaker need to change their behaviour. The plugin is compatible with networked applications as well. It is compatible with a classical client-server architecture commonly used in multiplayer games. The plugin provides a simple host and joins services working on local networks, which lets the plugin work for multi-user applications. Hence this manager is meant to be connected with other modules to expose the model's parameters and to provide controls over the virtual audience, e.g. by clustering the rules into attitudes.

3.3.3 Character Module

Each virtual spectator in the audience receives the rules allocated by the Audience Manager and then computes and triggers the animations' blending for the targeted agent.

This module inherits from the UE4's character, which allows the combination of our system with all the advanced features available in the game engine, e.g. state machines, navigation, and character's senses modules. The character module manages the behaviour logically and computes the timing aspects of the various behaviours to trigger and blend them according to the rules. This module is responsible for the model's logic and regroups all parameters needed to apply the rules. Section 3.3.5 describes this logic in detail. Besides the model's logic, the character module provides the components to link the agent's skeleton to the model. These components can be combined with behaviour trees and state machines to extend the model, e.g. pathfinding, field of view simulation, or spatialized audio.

3.3.4 Animation Module

The animation module achieves the animation blending and computes the agent's movements according to the nonverbal behaviours the character module asks to display. The animation played out by an agent is dynamically constructed from one to several animations and may use rotations on characters' bones, each corresponding to a specific behaviour type. There are at least three animation layers to fully display an agent's behaviour: the posture, head, and gaze layers. The posture and head layers handle the body and head movements and mix these two layers to simultaneously display a blended animation, e.g., leaning backwards, arms crossed and shaking the head. The gaze layer is responsible for moving the neck in various directions. There is an optional layer which can control eye movements. Also, the head and gaze layers never blend to avoid incoherent head movements. Finally, facial expressions are handled separately through morph targets (shape keys) which warp the face of the virtual spectator so that it can display the desired facial expression.

Existing features from the game engines operate these blending animations by applying animations on specific sections of the skeleton's hierarchy. Additionally, it interpolates movements when transitioning between animations. Therefore, the animations blend according to the model's parameters, e.g. if a spectator's head parameter is valid, the animation module moves it.

The plugin can handle 3D models and animations from various free 3D character modelling tools such as Mixamo, or Autodesk Character Generator. The system can be easily enriched by adding new animations and head movements. It can also include new facial expressions by adding morph targets to characters' 3D models.

Therefore, each agent has an animation loop and a state machine, which handles the animation blending and the postures. Because each posture has a unique ID, the state machine associates one state with a posture. In doing so, transitions are applied to some postures to increase the movements' realism, e.g. with sitting animations when the agent stops walking, or animations avoiding limbs crossing each other.

As for the networking features, the animation module is the only one that should not replicate over the network because animation data would be excessively detrimental for the network bandwidth. Thus, each application's instance computes its animations, and the server replicates the model's parameters.

3.3.5 Operating principle

The system relies on a loop alternatively showing or hiding the animation-building process result.

Base animation loop: The animation phase has two parts (Figure 3.1), one awaiting and another playing the animation. Two independent timers handle these two states and the transitions between each one. An internal timer manager from the game engines computes the time elapsed for each timer. The base loop starts with a timer and waits for the period the agent is idling before showing any behaviour. When it reaches its end, it triggers the call to the next timer in the loop and so forth until the agent plays out all behaviours. Finally, the loop returns to its await state when this last timer completes itself. Appendices D.1 and D.2 provide UML sequence diagrams depicting this animation loop.

Every timer duration relies on a fixed value called the reference period. Each wait-and-play timer computation uses this parameter with the corresponding display frequency parameter's value provided in the manager module, i.e. the frequency from the behaviour rule. A fixed period can represent the pace at which we want to display an agent's given behaviour, e.g. gazing away from the user for 60% of the period.

To avoid the behaviours synchronising, we randomly delay the first execution loop. This time desynchronises the agent's behaviour starts and breaks a negative effect induced when every agent in the audience begins to display at the same time their respective animations.

The system does not rely on the update function UE4 provides since all animations are not continuously displayed. This update function is internally called every frame

and constantly evaluates animations. Consequently, it severely decreases the application's performance. Instead, we decided to concatenate function calls to avoid updating every frame and awake timer only when needed.

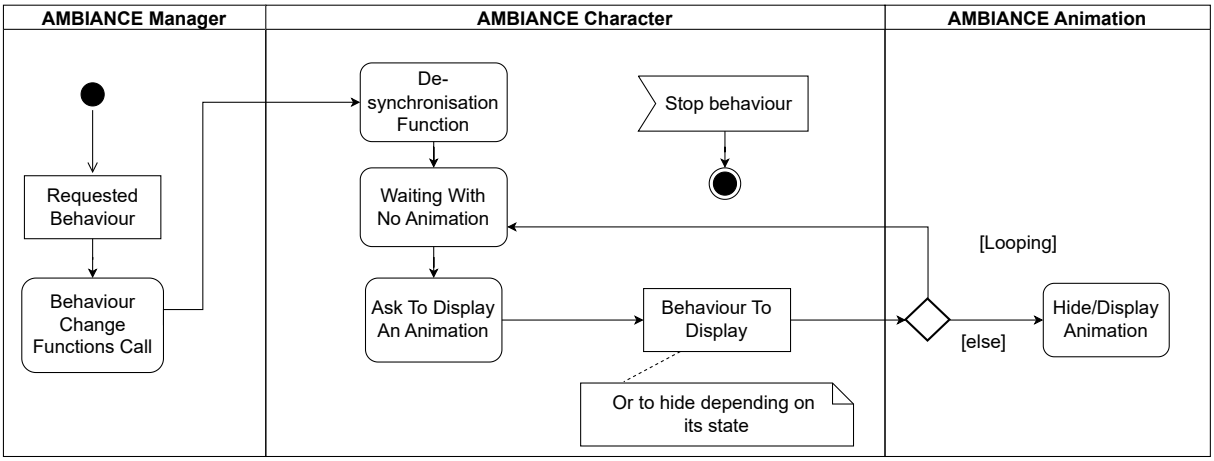


Figure 3.1 – Animation loop's execution flow. From the *AMBIANCE* manager to the different stages needed to follow the behaviour model and finally to the animation process.

Animation prioritisation: During this module's components development, we noticed that the generated animations do not blend well in some cases. When performed, animations could lead to physical oddities and overlapping behaviours that can break the immersion for the user, creating uncanny situations. As a solution, we decided to avoid the blending of certain animation types, e.g. the gaze direction and the head movement. Yet, the system allows the activation of incompatible animation but one at a time, e.g. one is triggered when the agent does not show any animation.

The solution to avoid incompatible movement is simple. The manager calls the first animation and becomes the prioritised one, i.e. it also signifies to other animations the priority order. In case of conflict between two incompatible animations, we cancel the activation of the lower-priority one. Due to asynchrony between timers, the module might never play out some behaviours. Therefore, to address this issue, a time threshold interrupts the prioritised animation, e.g. if an agent is gazing toward the window all the time, the threshold will stop this behaviour, letting the head node animation run.

3.4 Complementary Behaviours

The AMBIANCE model has been incrementally built and designed all along our research work. Thus, we extended the model according to the experiments carried out and included the remarks concerning the lack of sound, interactions between the user and the agents or the lack of posture variations which were too repetitious. This section reports complementary rules we built from our statistical analysis and users' feedback (see section 2.5). Since none of the models from Chollet and Scherer, 2017 and Kang et al., 2016 integrated backchannels and sound, we based our complementary behaviours on studies evaluating backchannels for ECA. Bevacqua, Pammi, et al., 2010 provide a list of backchannels together with the emotion they are related to when displayed in a conversation between human and ECA, e.g. a head nod and a vocal "mmhmm" are related to the agreement.

3.4.1 Multimodal Backchannels

Even if backchannels are often related to dyadic conversations, such interactions can still occur or at least increase the agents' engagement cue. We added rules that include the different elements defining a backchannel by extending the head movement rule. The rules for multimodal-backchannels follow this format:

$$rule_{backchannel}(HeadMovement, AudioCue, Delay) \tag{3.7}$$

where the *HeadMovement* is the behaviour displayed by the agent, the *AudioCue* is the sound corresponding to displayed head movement (nod, shake) and the *Delay* is the time between the head movement displayed and the audio cue start.

3.4.2 Interactions

So far, we have not mentioned user-agent interactions, even if Kang et al., 2016 provide insight into reactions to disruptive events. As we designed the first model's version to be free of context or any domain application, it does not contain such interaction. Therefore, the only user-agent interaction is the spectators' gaze toward the speaker. This behaviour is the only interaction which does not need to consider the application domain but still conveys a strong engagement signal for the speaker. By default, virtual agents' behaviours are continuously staring at the speaker and moving their necks accordingly, i.e. so that

if the speaker moves, all agents follow. Thus, the animation module updates the speaker location through time and computes the desired neck rotation. However, we anticipated the agent reaction feature and the possibility of using customised rules to create domain-specific behaviours and design scenarios. Because an agent's reaction depends on the situation appraisal, we integrated this type of behaviour in the custom rules to let domain experts decide which reaction better suits a given situation. To integrate this feature into the model, we used a rule formalism like the ones introduced in section 2. Thus, these custom rules depend on a given behaviour that can be application-specific, e.g. a student behaviour in the case of pre-service teacher training. It then takes a second behaviour as a reaction and finally to which spectators this reaction must be applied and in which delay. For instance, if a disruptive event occurs, the system will trigger after n seconds a reaction to it for a given number of agents depending on the current attitude or scenario.

$$\begin{aligned} rule_{Custom}(&CustomBehaviourDelegate,\\ &ReactionDelegate,\\ &TargetedSpectators,\\ &reactionDelay) \end{aligned} \tag{3.8}$$

We only used these customised rule in our main application described in Chapter 5.

3.5 Scalability Benchmark

Once we implemented the model, we were interested in the system's performance because low latency is critical for VR systems. It is a negative factor in simulator sickness, and it also considerably affects interaction [Lugrin et al., 2013]. Low latencies and jitter are also critical requirements for enabling collaborative applications with VR systems using virtual agents and embodied avatars [Latoschik et al., 2016]. Consequently, the evaluation focuses on measuring our system's impact on the frame rate and identifying the maximum threshold number of simultaneous agents in VR we can support without any animation and mesh optimisations. Consequently, we measured the system's impact on the frame rate and identified the maximum threshold number of simultaneous agents in VR we can support without any animation and mesh optimisations. This section presents our benchmarking method and the results and conclusions we draw from them.

The following results are divided in two parts. The first part shows the performances with low detailed agents (Figure 3.2) and the second with a more detailed agents (Figure 3.3), approximately four times more complex in terms of polygons count.

Figure 3.2 – Desktop and virtual reality screenshots of the first scalability benchmark. From left to right: 2 agents with a desktop GUI example, 28 agents in a VR overview, 1000 agents with desktop GUI.

Figure 3.3 – Benchmark environment view. Close view of 60 Mixamo's Remy agent.

3.5.1 Device

To perform the evaluation we used a laptop running with Windows 10 64 bits, Intel®Core i7-7700HQ processor (Quad core, 2.80 GHz, 8MB cache,8 GT/s) and NVIDIA®GeForce GTX 1070 Graphics Processing Unit (GPU) 8GB GDDR5. A HTC®Vive was used to carry out the VR evaluation. We choose to carry out the evaluation with a VR ready laptop we are commonly using to run our evaluations.

3.5.2 Environment Configuration

A simple VR scene was built without any superfluous post-processing or shadows. Thereby, it makes it easier to detect the bottleneck of our plugin. Thus, only a plane,

the agents and a default sky-box was rendered (Figures 3.2, 3.3). Several settings were selected, for instance VSync and UE4's frame smoothing were disabled because they can interfere with the frame rate in order to fit with the monitor's refresh rate. With regards to the VR head mounted display, it was at a fixed position in order to avoid rotations and thus fluctuations in the processor's draw calls. All measures are medians done on a fixed amount of time, i.e. 100 frames.

3.5.3 Results

Our results with the low detailed agents are visible on Figure 3.4 (blue curve). From the performances data we distinguish three thresholds : i) Up to 30 agents the average frame rate is high (> 90 Hz) ii) Up to 100 the frame rate decreases while remaining acceptable for VR usage ($]80, 90]$ Hz), iii) Above 200 agents, the system is no more suitable for VR use (< 30 Hz). We noticed that our system was CPU bounded. This is why we investigated which process was the most consuming for the CPU between the draw time and the time needed to update the audience, i.e. the time to update behaviours, meshes, and animations. It appears that the render time is not the bottleneck of our system but the cumulative time to update each agents in the audience.

The second evaluation does not face the same problem. In fact, while the agents' triangles are increasing fast, the number of agent does not (Figure 3.4 red curve). We are here facing the opposite problem with a draw time which is reducing the frame rate. With complex agents the system can handle around 60 agents before the draw time reduces the frame rate which is causing less fluidity in agents' movements for instance.

For ease of reading we are only displaying the frames per second count while the system is not yet bounded by the rendering. The step with 100 agents is also displayed in the second evaluation even if the system is already bounded by the rendering in order to compare the different impacts of the audience's update time and the draw time mainly affected by the number of characters to render. The details of the draw time and the update time are not displayed because they correspond to the frame rate given in the figure, depending on which one is the bottleneck. Additionally, the draw time is approximately half of the update time in our first evaluation before the draw time becomes more important, i.e. between 300 and 400 agents.

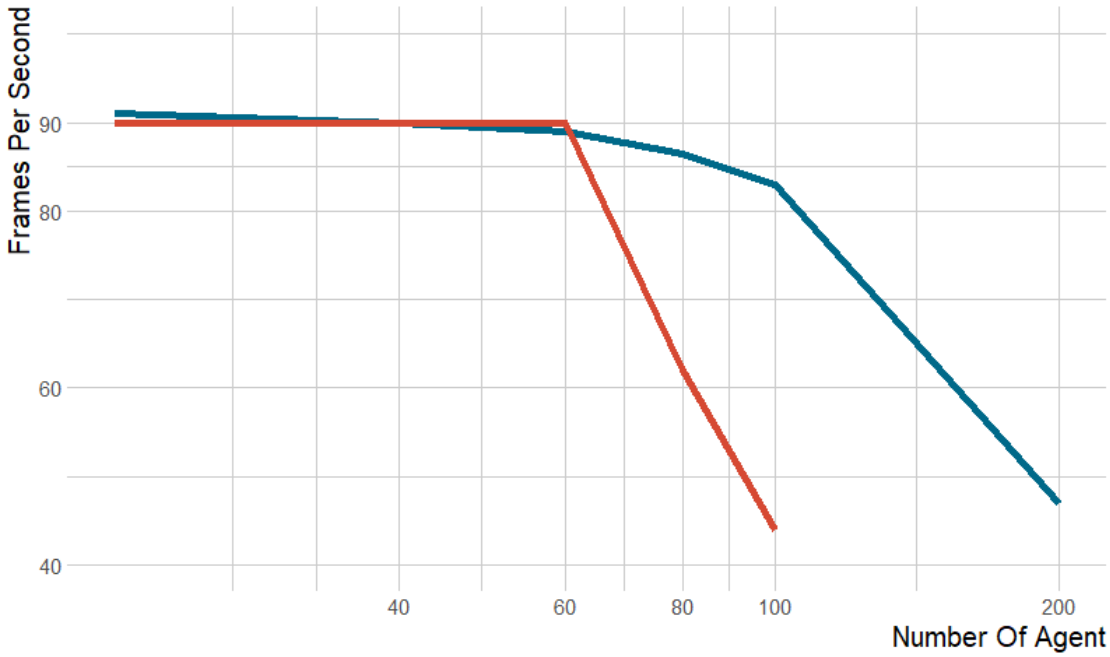


Figure 3.4 – Frame rate depending on the number of agent. In red the complex agent scalability data. In blue the lightweight agent scalability data.

3.6 Conclusion

In this chapter, we introduced the AMBIANCE model and its operating system. Previous works provide various statistical approaches which describe how users perceive audience non-verbal behaviours in different contexts. Previous works used to build the AMBIANCE draw their inspiration from cognitive psychology theories that model the user perception onto two continuous dimensions, arousal and valence. We have seen in chapter 2 that both have different approaches to evaluate their model, whether from video-recorded audiences and VR evaluation or the literature and crowdsourced evaluations. Therefore, the works from Chollet et al. and Kang et al. manipulate various nonverbal behaviours to enact different audience attitudes, such as postures, facial expressions and gaze. Their behaviour models also issue various audience simulation aspects, e.g. if agents should react to a disruptive event, how many agents must display the same attitude for the user to recognise it or how nonverbal behaviours interact together. However, these two models only provide a limited nonverbal behaviour roster and thus reduce the number of possible

attitudes to display. It is also necessary to add variability, some contrast between agents, and interactions to preserve the audience's believability if creating a broader range of attitudes. Besides these limitations, we proposed a formalisation expressed in a set of behaviour rules that vary depending on the targeted attitude to display, based on their perception studies' results and guidelines. The resulting model regroups three types of rules, one for the nonverbal behaviours, one for the overall behaviours congruence and one for the behaviour believability to avoid uncanny behaviours. The sum of all these rules gives rise to an audience attitude corresponding to what the users perceive.

Finally, we present the model implementation with insight into the different modules composing the AMBIANCE and its VR performances. Section 2 portrays the three main modules with their operating principle: the management modules drive the entire system and expose to the external system the model's parameter, and the character module and its animation module handle the model's logic for each agent and animates them with various animation techniques. The results from our scalability benchmark testify to the system's good VR performance allowing the simulation of virtual audiences ranging from 30 to 100 agents. Audiences of this size would suit VR applications which simulate professional meetings, classrooms or small conferences for public speaking training and even wider audiences like theatres. Therefore, this model and its implementation seem to yield a powerful tool for developing a system using virtual audiences. Nevertheless, the AMBIANCE is our interpretation and formalisation of two different behaviour models not yet evaluated in VR. Thus it requires a perception study to confirm VR users recognise the generated attitudes. The next chapter depicts the different user perception studies we ran to evaluate the model.

Chapter 4

AMBIANCE Perception Studies

4.1 Introduction

In the previous chapter, we introduced the AMBIANCE behaviour model and benchmarked the system in VR. Thus we demonstrated the model's ability to sustain at least 30 agents without decreasing the VR performances. The AMBIANCE relies on two user perception studies formalisation. These two studies came up with different approaches to evaluate user perception but offer complementary results regarding non-verbal behaviours perception according to the Valence and Arousal dimensions and the overall audience perception. This chapter will further detail the methods selected to evaluate the model and introduce our methodology and protocol for the AMBIANCE's evaluation. Finally, we present the results and propose guidelines for the design of virtual audiences in virtual reality applications.

Two questions arise when evaluating a new behaviour model in VR: one regarding nonverbal behaviour perception and the other regarding the generated attitudes perception. Even though we built the model from the literature, we interpreted statistical analysis and guidelines from two studies. Moreover, the model is domain-specific, so we should not evaluate the AMBIANCE for a given context. The perception study from Kang et al., 2016 does not provide enough information to create a model that would be consistent across multiple simulation contexts, e.g. for a classroom, a business meeting or for a play.

Thus, we chose to adhere to the protocol used by Chollet and Scherer, 2017: during the first phase, users had to design nonverbal behaviours according to the given pair value of valence and arousal. In the second experiment phase, another set of participants had to rate audiences' behaviours according to a pair value of valence and arousal. A dissimilarity in methodology has arisen from a shift in technology: Chollet and Scherer, 2017 used a crowd-sourced study, which was difficult to achieve in VR, so we have run a VR user perception study instead. In doing so, we can compare nonverbal behaviour perception and audiences' perceived attitudes to those obtained in the original crowd-sourced evaluation.

Comparisons can highlight dissimilarities in users' perceptions and allow users to provide guidelines for designing virtual audiences in VR. However, our hypotheses are slightly different, especially regarding facial expressions, for which we decided to change how we evaluate an agent's face at rest, whilst Chollet and Scherer, 2017 did not consider the absence of facial expressions.

4.2 First User Study: Nonverbal Behaviour Perception of Individual virtual Spectators

In this study, we investigated how nonverbal behaviours of individual agents are associated with the valence and arousal dimensions. We aimed to confirm whether the use of these behaviours for different audience attitudes corresponded with a valence-arousal pair. Hence we asked participants to design a virtual spectator's behaviour according to a given pair value of opinions and engagement toward the speech, e.g. a very low engagement and a negative opinion towards a speech.

4.2.1 Hypothesis

During this study, we aimed to confirm the following hypotheses, which have been validated for desktop-based simulation by earlier work by Chollet and Scherer, 2017. The aim was to provide a set of validated nonverbal behaviour for individual agents to build the audience model. In the following hypotheses, the independent variables (IV) are the levels of valence or arousal, and the dependent variables (DV) are the different nonverbal behaviours the users can select from a GUI (Figure 4.1). Both IVs can take 5 valence or arousal levels ranging from very low to very high. As for the DVs, we considered the same nonverbal behaviours as from the desktop study from Chollet and Scherer, 2017 Accordingly, we evaluated postures in terms of proximity (leaning forward or backward) and openness (arms crossed or hand behind the head). For the rest of the behaviours visible in Figure 4.1, we used the same behaviours as Chollet and Scherer, 2017.

Hypothesis 1 (**H1.1**), arousal and expressions: Higher arousal leads to more feedback, more facial expressions, more head movements, and more gaze directed at the speaker.

Hypothesis 2 (**H1.2**), valence and expressions: Smiles and nods are associated with positive valence, frowns and head shakes with negative valence, and eyebrow raises and a face at rest with neutral valence.

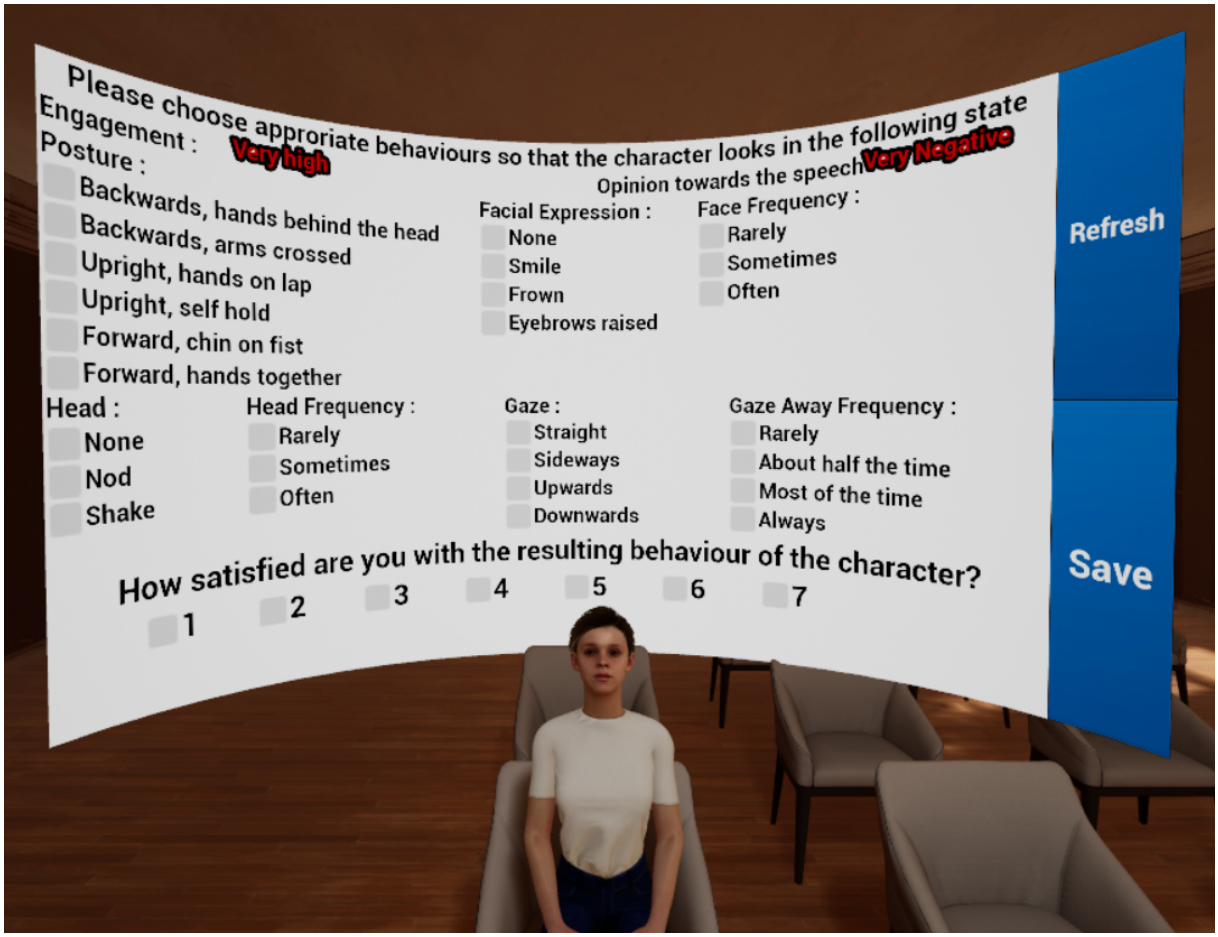


Figure 4.1 – Participant's view during the design of a spectator's attitude

Hypothesis 3 (**H1.3**), arousal and postures: Postures chosen for high arousal involve leaning closer to the speaker than postures chosen for lower arousal.

Hypothesis 4 (**H1.4**), valence and postures: Relaxed postures lead to a more positive valence compared with more closed postures.

4.2.2 Method

For this study, 20 people participated, 4 women and 16 men, aged from 18 to 28, 13 of which are students and 7 are in the workforce (see appendix F.1 for the consent form provided). Participants had to select behaviours for all the possible pairs of valence and arousal. Each value of valence and arousal has 5 levels: respectively very negative, negative, neutral, positive and very positive for the valence and very low, low, medium,

high and very high for the arousal. Participants had to select the different behaviours on a GUI directly in VR (Figure 4.1). In doing so, the users could directly see the changes in behaviour without taking or removing the head-mounted display. When satisfied by the resulting behaviour, the participants could validate the answer and continue to the next pair of valence and arousal values. These pair values were randomised to avoid any order effect. The agents used were either male or female. The agent's gender was balanced in between each answer to avoid effects. The virtual environment was a simple room with chairs for the agent and a desk behind the user. Once the participant had finished specifying an agent's behaviour, they were asked to rate how satisfied they were with the result using a 7-point satisfaction scale. A poor rating would have led to the behaviour being discarded from the model. However, none of the participants gave a rating under 4 for any of the agents. Each session started with a short training phase where the users were able to become familiar with the virtual environment and the GUI. There was no time limit set for designing each agent. At the end of the session, participants were interviewed about their overall experience. During a semi-guided interview participant were questioned regarding the behaviour believability, if they were able to design agent behaviours for each pair and were also asked to give some feedback about the VR setup usability and their feeling in terms of VR sickness.

To carry out this study we used an HTC®Vive Pro running on a stationary computer running with Windows 10 64 bits, Intel®Core i7-7700k processor (8 cores, 4.20GHz, 11MB cache,11 GT/s) and NVIDIA®GeForce GTX 1080Ti Graphics Processing Unit (GPU) 16GB GDDR5. The simulation runs on average with 90 frames per second on this computer.

4.2.3 Results

For H1.1, H1.3 and H1.4 we use non-binary factors and ordinal numerical variables. The postures which were described by openness and the proximity transformed into numerical variables where a high proximity value represents a forward posture and low represents a backward posture (backward is 1 and forward is 3), and a high openness value represents open and a low value represents closed (arms crossed and self-hold: 1, arms behind the head: 3, the rest: 2). We used the exact same transformation as in Chollet and Scherer, 2017. We also transformed the frequencies into an ordinal variable similarly to the model. Where no behaviour is 0, *rarely* is 0.25, *sometimes* and *about half the time* are 0.5, *often* and *most of the time* are 0.75, and *Always* is 1. In doing so, we can easily

Table 4.1 – Wilcoxon's Pairwise Tests, numbers are p-values from each tests, the levels of arousal are on both sides of the table with the Spearman's effect size for the arousal and the four variables.

Gaze Away Frequency					
Arousal	High	Low	Medium	Very High	Spearman's effect size (ρ)
Low	<0.01	-	-	-	
Medium	<0.01	<0.01	-	-	-0.6
Very High	0.462	<0.01	<0.01	-	
Very Low	<0.01	0.054	<0.01	<0.01	
Facial Expressions Frequency					
Arousal	High	Low	Medium	Very High	Spearman's effect size (ρ)
Low	0.012	-	-	-	
Medium	0.345	1.0	-	-	0.2
Very High	1.0	0.003	0.689	-	
Very Low	0.040	1.0	1.0	0.016	
Head Movements Frequency					
Arousal	High	Low	Medium	Very High	Spearman's effect size (ρ)
Low	<0.01	-	-	-	
Medium	0.042	1.0	-	-	0.3
Very High	1.0	<0.01	<0.01	-	
Very Low	<0.01	1.0	0.560	<0.01	
Proximity					
Arousal	High	Low	Medium	Very High	Spearman's effect size (ρ)
Low	<0.01	-	-	-	
Medium	<0.01	<0.01	-	-	0.7
Very High	1.0	<0.01	<0.01	-	
Very Low	<0.01	0.079	<0.01	<0.01	

compare our results to those obtained from the literature. It is also necessary because we used a within-subjects design and non-binary nominal factors which prevent us from using Pearson's Chi-squared test or non-parametric Cohen's test for independence. Hence, because the distribution is not normal, we ran a non-parametric Friedman test instead of a repeated-measures ANOVA test.

4.2.4 Arousal and Expressions

For H1.1, we set the arousal as the IV and conducted tests with the face, head and gaze-away frequencies as DV. Pairwise Wilcoxon's tests using Bonferroni's adjustment method have been used on each feature of the IV. For all three DVs, the arousal has a significant effect on the behaviours' frequencies (gaze: $\chi^2 = 192, df = 4$, p-value < 0.05, facial expressions: $\chi^2 = 17.7$, $df = 4$, p-value < 0.05), head movements: $\chi^2 = 47.6$, $df = 4$, p-value < 0.05). The pairwise tests (Table 4.1) show which levels are significant compared to the others. With regard to the facial expressions, low arousal leads to less frequent facial expressions and high arousal to more frequent facial expressions. The participants could not significantly differentiate two levels of arousal which were not markedly different, for instance, *Very Low* and *Low*. The Figure 4.2 gives a representation of these differences

while showing which levels are significant to each other. However, it is important to underline that the effect size for the facial expression frequency and the arousal is small (Spearman's $\rho = 0.2$). Thus we can only slightly agree with these conclusions.

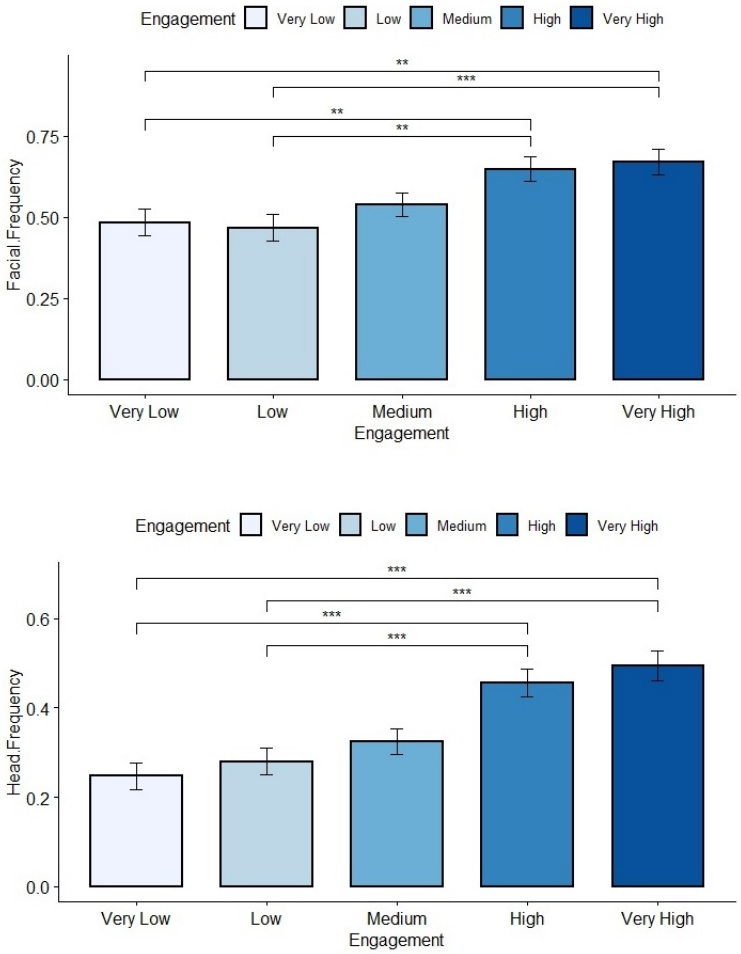


Figure 4.2 – Distribution plots of the facial expressions and head movements frequency depending on the engagement (Arousal) levels. Significant results with the *Medium Engagement* level was removed for clarity purpose, see Table 4.1 for details.

The results for the head movement frequency are much alike. Low arousal leads to less frequent head movements, and high arousal leads to more frequent head movements. The compared frequencies are only significant for opposed levels of arousal, *e.g. Low* and *High* arousal (see Figure 4.2). For these two first results, the behaviour frequency for medium arousal is not significantly different from all other levels. These mild results were probably due to the size of the frequency scale, which only had three levels. The following

tests had 4 levels, which allowed the users to design behaviour with significantly different behaviour frequencies. Therefore, the gaze-away frequency is significant for all levels of arousal except between a *High* and a *Very High* level and a *Low* and a *Very Low* level of arousal. This means that a higher arousal level leads to a less frequent gaze away while a low level of arousal leads to a more frequent gaze away from the user (see Figure 4.3). Overall, lower arousal leads to less frequent expressions and high arousal leads to more frequent expressions.

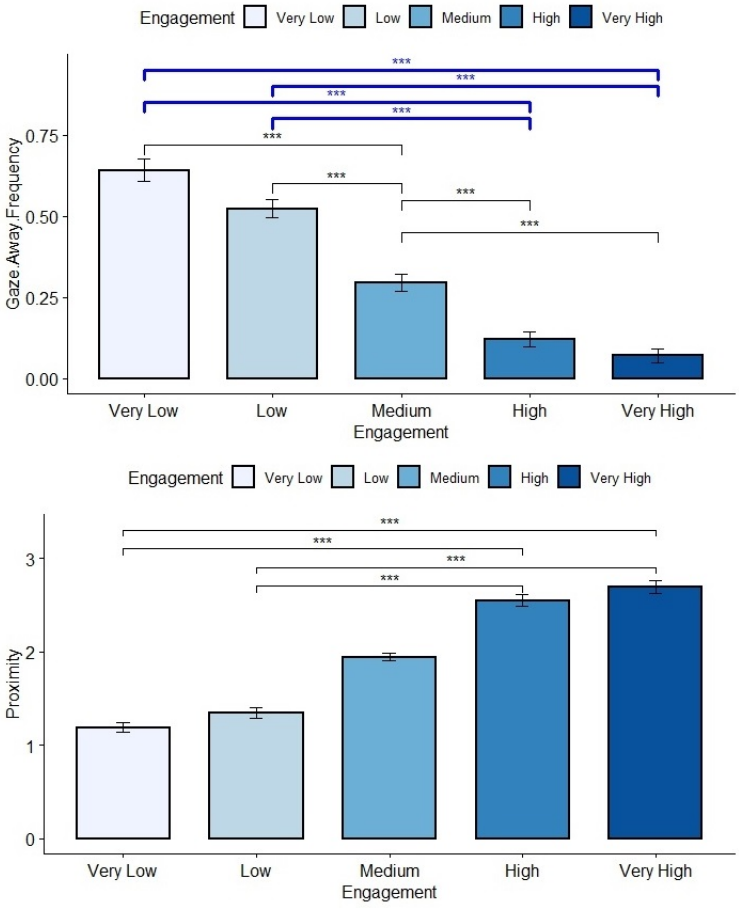


Figure 4.3 – Distribution plot of the gaze away frequency and the posture proximity depending on the engagement (Arousal) levels. Significant results with the *Medium Engagement* level was removed for clarity purpose. Main significant results from the pairwise tests related to the gaze are shown in blue above the bar charts, see Table 4.1 for details.

Arousal and Postures

For H1.3, we set the arousal as the IV and the proximity as the ordinal DV. Results are significant as well ($\chi^2 = 221, df = 4, p - value < 0.05$); higher arousal leads to closer posture proximity, while lower arousal leads to a further posture. The pairwise tests show that the proximity is mainly significant for markedly different levels of arousal: for instance for a *High* and a *Low* level of arousal. The Table 4.1 and the Figure 4.3 resume these results.

Valence and Postures

For H1.4, we set the valence as the IV and relaxation as the DV. Surprisingly, this test is not significant so we failed to reject the null hypothesis. In our case, the relaxation or the openness of the proposed postures does not seem to be related to the valence. This result differs from the findings in the desktop-based audience research by Chollet and Scherer, 2017.

Valence and Expressions

Finally, for H1.2, we set the valence as the IV and conducted tests with the facial expression and head movement categories as the nominal DVs. Here, both the IV and DVs have more than two levels and the study was a within-subjects design, so we used a multinomial logistic regression. If we do not transform our DV into ordinal data, it is because we are interested in getting the influence of each behaviour type per level of valence, unlike the previous tests in which we were comparing mean frequencies per subject or the average proximity values for the posture. Hence we used a face without facial expressions for the agent as the reference event to determine all odds ratios for the IV. The regression model is expressed as:

$$g_j \ = \ \beta_0 \ + \beta_j * x_i \tag{4.1}$$

where x_i represents the different levels of valence, j is the behaviour parameter selected by the participants, and β_0 is the intercept parameter. Results regarding the head movements indicate that shaking the head is associated with a negative valence, and nodding with positive valence (with an odds ratio within the confidence interval CI95%). Head shake is significantly negative and head nod is significantly positive. Details on the tested

Table 4.2 – Multinomial Logistic Regression significance table.

H1.2 Head Mov.	coeff.	std. error	$z$value	Pr($>\|z\|$)
Nod:(Intercept)	-1.9	0.3	-6.3	< 0.05
Opinion Positive	2.9	0.4	7.7	< 0.01
Opinion Very Positive	3.4	0.4	8.5	< 0.01
Shake:(Intercept)	-4.5	1.0	-4.4	< 0.05
Opinion Negative	5.7	1.0	5.5	< 0.01
Opinion Very Negative	5.9	1.0	5.6	< 0.01
H1.2 Facial Exp.	coeff.	std. error	$z$value	Pr($>\|z\|$)
Frown:(Intercept)	-1.9	0.3	-6.0	< 0.01
Opinion Negative	3.7	0.4	8.5	< 0.01
Opinion Very Negative	3.8	0.4	8.6	< 0.01
Smile:(Intercept)	-2.2	0.4	-6.1	< 0.01
Opinion Positive	4.8	0.5	8.8	< 0.01
Opinion Very Positive	6.0	0.8	7.5	< 0.01

features are visible in Table 4.2 and Figure 4.4 shows the distribution for each behaviours depending on the levels of valence.

$$g_{shaking} = -4.5 + 5.7 * x_{negative} - 9.8 * x_{positive} + 5.9 * x_{veryNegative} - 9.6 * x_{veryPositive} \quad (4.2)$$

$$g_{nodding} = -1.9 + 0.9 * x_{negative} + 2.9 * x_{positive} + 0.8 * x_{veryNegative} + 3.3 * x_{veryPositive} \quad (4.3)$$

To study the relationship between the different facial expressions and the valence, we used the same formula (Equation 4.1) where j represents the facial expression selected by the participants. Results indicate that smiling is significantly associated with positive valence, and frowning with negative valence (with an odds ratio within the confidence interval CI95%). Almost none of the participants chose eyebrows raised, and this is why we do not propose an analysis for it. The Figure 4.4 highlights the lack of selections of eyebrows raised.

$$g_{smile} = -2.8 - 0.4 * x_{negative} + 4.8 * x_{positive} + 0.9 * x_{veryNegative} + 6.0 * x_{veryPositive} \quad (4.4)$$

$$g_{frown} = -1.9 + 3.6 * x_{negative} + 2.0 * x_{positive} + 3.84 * x_{veryNegative} + 2.2 * x_{veryPositive} \quad (4.5)$$

These results partially confirm the findings from Chollet and Scherer, 2017 with an exception to the *eyebrows raised* behaviour with which we have no available data. We can also add that the default face displayed by the virtual agents was significantly preferred

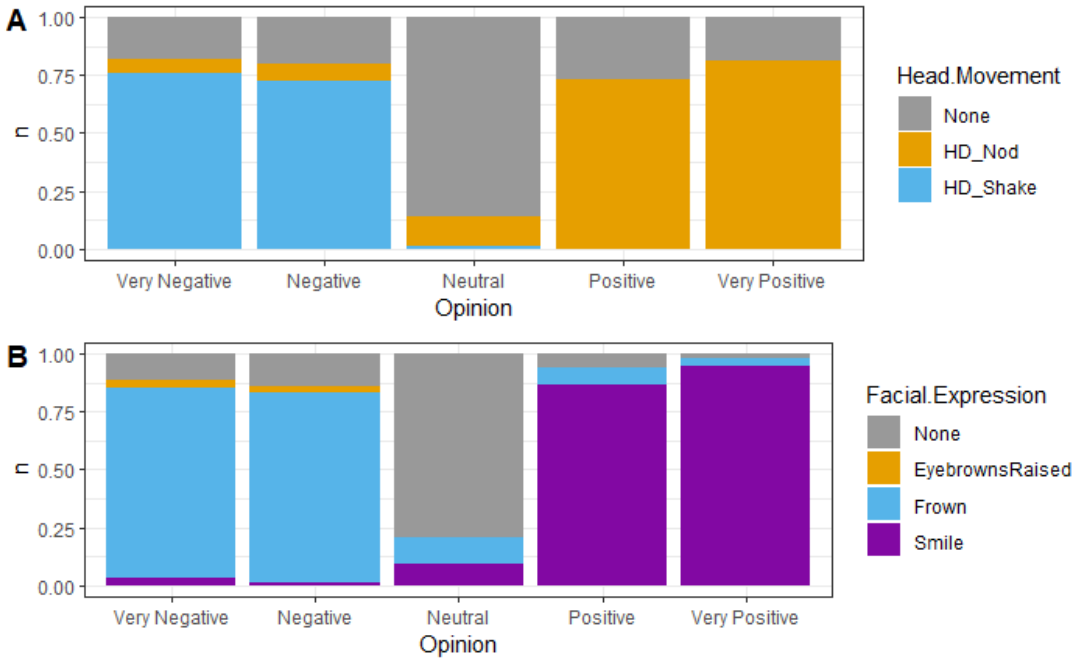


Figure 4.4 – Distribution of behaviours per state levels for the investigated hypotheses. The five bars in each sub-figure correspond to the five possible values of valence, very negative, negative, neutral, positive and very positive. (A) is related to the head movements types and (B) to the facial expressions types.

to the other facial expressions; this confirms our hypothesis on the association between a neutral valence and a face at rest.

4.3 Second User Study: Virtual Audience Attitudes Perception Evaluation

This study consisted of an evaluation aimed to validate the perceived audience attitudes generated by our model. Based on the nonverbal behaviour rules using values of valence and arousal and the results from our first user evaluation, we designed different virtual audiences. We investigated whether the audience attitudes generated with these rules could be identified by the users in terms of valence and arousal. The relationships between the nonverbal behaviours and the valence or the arousal are mainly from Chollet and Scherer, 2017's studies in which they provide the proportion of agents with a certain behaviour needed to let the users recognize the audience's attitude. The section below

describes how we used the literature and our first study to build the virtual audiences used in this study.

4.3.1 Hypothesis:

Our hypotheses for this study are that VR users can significantly perceive different attitudes generated with our system in terms of valence and arousal in VR. This means the model we use to generate the VA attitudes can be used to create the virtual agents' behaviours which allow the users to perceive the targeted attitude.

Hypothesis 1 **(H2.1)**: The higher the positive valence, the higher the positive perceived opinion is for the participant (respectively for negative valence and opinion).

Hypothesis 2 **(H2.2)**: The lower the arousal, the lower the perceived engagement is for the participant (respectively for high arousal and perceived engagement).

In the above two hypotheses, the terms *high valence* and *low arousal* express the intensity of the displayed audience's attitude. Therefore, in this study, a higher proportion of virtual spectators displaying behaviours matching the targeted attitude populate the virtual audience. For instance, an attitude with very positive valence and very low arousal corresponds to an audience where more than 60% of the agents display positive behaviour and where more than 30% of them are highly engaged [Chollet and Scherer, 2017]. The behaviours from the first study are then used to generate the virtual audience. For this example, what we call an agent exhibiting a positive behaviour is an agent displaying the nonverbal behaviours corresponding to a positive valence (head nod and smile). The engagement for this same agent is also based on the behaviour frequencies previously shown in our first study where a highly engaged agent would frequently nod and smile while leaning forward. We designed the audience bearing in mind that, according to Chollet and Scherer, 2017, the more frequent a behaviour is, the stronger it is perceived. Table 4.3 gives an example of what parameters we use to populate our virtual audiences for different attitudes. Finally, for the attitude displaying a neutral valence and medium arousal, we used a mix of positive, negative and neutral agents as well as a mix of agents with a low, medium or high engagement like in Chollet and Scherer, 2017.

Table 4.3 – Example of a rules set for an Enthusiastic and critical attitudes. The frequency parameter used are the same as in our user study.

Enthusiastic				
	Arousal	Very High		
	Valence	Very Positive		
Rules set		Type	Frequency	Proportion
	$Rule_{Posture}$	Lean Forward chin on fist	Always (1.0)	20%
	$Rule_{Posture}$	Upright hands on laps	Always (1.0)	20%
	$Rule_{Posture}$	Forward hands together	Always (1.0)	20%
	$Rule_{FacialExpression}$	Smile	Most of the time (0.75)	60%
	$Rule_{HeadMovement}$	Nod	Most of the time (0.75)	50%
	$Rule_{gaze}$	Sideways	Rarely (0.25)	10%
Critical				
	Arousal	Medium/High		
	Valence	Negative		
	$Rule_{Posture}$	Backward arms crossed	Always (1.0)	20%
	$Rule_{Posture}$	Upright self hold	Always (1.0)	20%
	$Rule_{FacialExpression}$	Frown	Often (0.75)	60%
	$Rule_{HeadMovement}$	Shake	Sometimes (0.5)	40%
	$Rule_{gaze}$	Downward	Sometimes (0.5)	20%
	$Rule_{gaze}$	Sideways	Sometimes (0.5)	20%

4.3.2 Method

For this study, we kept the same virtual environment as that we had for the first evaluation. We used 10 different characters from Adobe®Mixamo, 5 females and 5 males (Figure 4.5). All these virtual spectators provided implementation of facial expressions. They were driven by our system according to the nonverbal behaviours previously identified during the individual agent nonverbal behaviour study. Figure 4.6 provides examples of both a critical and an interested audience.

Figure 4.5 – Participant's view during the audience perception evaluation.

Figure 4.6 – Example of an annoyed **(A)** and an interested **(B)** virtual audiences.

38 people participated in the evaluation: 9 women and 29 men aged from 19 to 28. None of them participated in the first study and all were students. Participants had to rate their perceived audience's opinion and engagement on two 5-point scales representing the value of valence and the level of arousal (*i.e.* from 1 to 5: very negative, negative, neutral, positive and very positive valence and very low, low, medium, high and very high for the arousal). The same audiences were shown to participants but in randomised orders to avoid any order effects. We designed these audiences to correspond to 5 different types of attitudes:

Attitude 1, (**A1**): very negative valence and very low arousal;

Attitude 2, (**A2**): very negative valence and very high arousal;

Attitude 3, (**A3**): neutral valence and medium arousal;

Attitude 4, (**A4**): very positive valence and very low arousal;

Attitude 5, (**A5**): very positive valence and very high arousal.

Each session started with a short training phase where the users were able to become familiar with the virtual environment and the GUI. There was no time limit set for answering. At the end of the session, users were interviewed about their overall experience. The same hardware was used in this evaluation.

4.3.3 Results

The evaluation followed a within-subjects design, and the distribution was not normal, so we ran a non-parametric Friedman test for H2.1 and H2.2. Concerning H2.1, we set the generated attitudes defined above (A_i) as the IV and conducted a test with the perceived opinion value as the DV. For H2.2, we also set the generated attitudes defined above (A_i) as the IV and conducted a test with the perceived engagement value as the DV. Pairwise Wilcoxon's tests using the Bonferroni adjustment method have been used for each modality (Table 4.4).

For both the valence and the arousal values, we found a significant effect on the perceived attitudes (valence: $\chi^2 = 43.7, df = 4$, p-value < 0.01, arousal: $\chi^2 = 26.4, df = 4$, p-value < 0.01). In the pairwise test for H2.1, the 3 different values of valence are correctly associated with the audience attitudes of the participants (negative, neutral and positive). Moreover, the perceived values of valence are only significant when comparing attitudes from markedly different levels of valence: there are no differences between attitudes with the same value of valence (See Figure 4.7). However, the neutral audience ($A3$) is not significantly perceived as neutral by the participants but slightly positive. We believe this

Table 4.4 – Wilcoxon's Pairwise Tests, numbers are p-values from each tests, the attitudes are on both sides of the table with the Spearman's effect size for the valence and then the arousal.

Valence					
Attitudes	Very Low & Very Negative	Very High & Very Negative	Medium & Neutral	Very Low & Very Positive	Spearman's effect size (ρ)
Very High & Very Negative	1.0	-	-	-	
Medium and Neutral	<0.01	<0.01	-	-	0.5
Very Low and Very Positive	<0.01	<0.01	1.0	-	
Very High and Very Positive	<0.01	<0.01	1.0	1.0	
Arousal					
Attitudes	Very Low & Very Negative	Very High & Very Negative	Medium & Neutral	Very Low & Very Positive	Spearman's effect size (ρ)
Very High & Very Negative	1.0	-	-	-	
Medium and Neutral	1.0	1.0	-	-	0.2
Very Low and Very Positive	<0.01	<0.01	<0.01	-	
Very High and Very Positive	1.0	1.0	1.0	<0.01	

is because we mixed different types of nonverbal behaviours which were not only neutral.

For H2.2, the pairwise test shows that participants cannot significantly identify all levels of arousal. The results exhibit a significant difference in the users' perception between low and high arousal when the valence is positive (A4 and A5, Figure 4.7). However, in A1, A2 and A3, there are no significant differences in terms of perceived arousal. Participants cannot significantly differentiate two different levels of arousal when the valence is negative. The same for A3 which is supposed to be perceived as moderately engaged was perceived as highly engaged. We believe it is also due to the use of negative nonverbal behaviours to generate the attitude. All three attitudes we designed with negative nonverbal behaviours were perceived with the same high level of arousal (Figure 4.7). Further investigations can be done to test if negative nonverbal behaviours also influence perceived arousal. Thus, H2.2 is only partially validated, and we cannot completely reject the null hypothesis. Figure 4.8 reports an example of attitudes the users cannot significantly recognise.

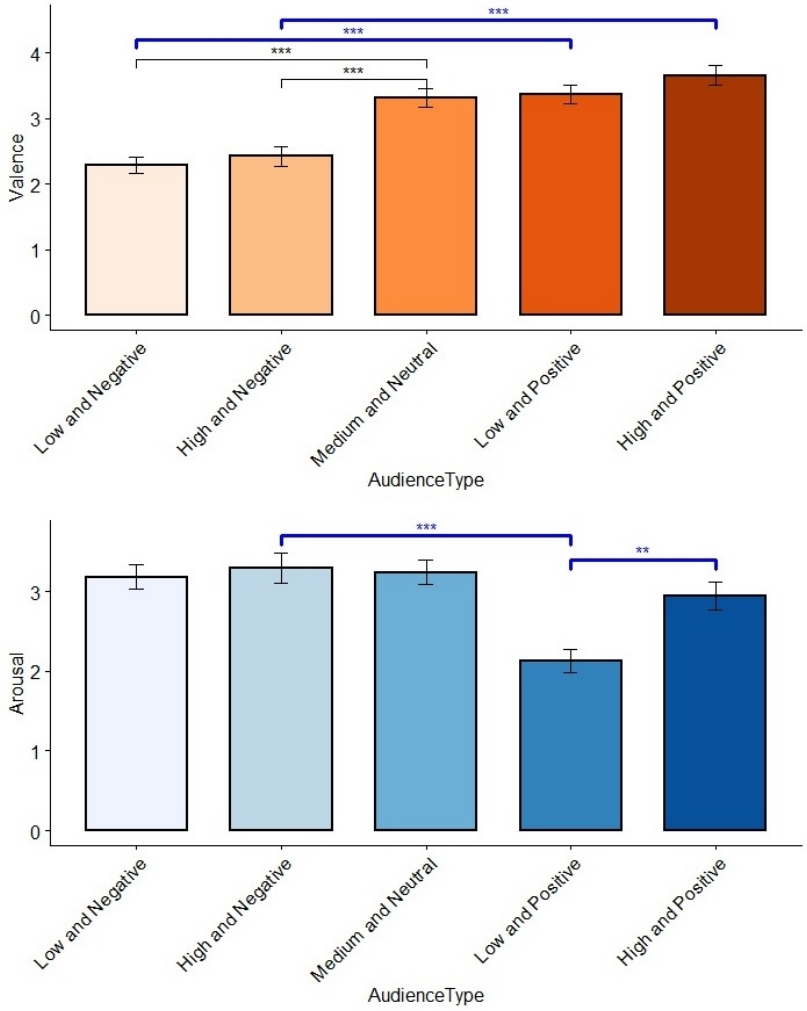


Figure 4.7 – Distribution plot of the selected valence and arousal depending on the attitudes. Main significant results from the pairwise tests for our hypotheses are shown in blue above the bars charts.

4.4 Recommendations for Audience Simulation in Virtual Reality

Concerning our results and considering the existing literature, we can propose recommendations and guidelines for virtual audience design and attitude generation. These guidelines focus on nonverbal behaviours, but the following section 4.5 tackles other aspects, such as sound or backchannels.

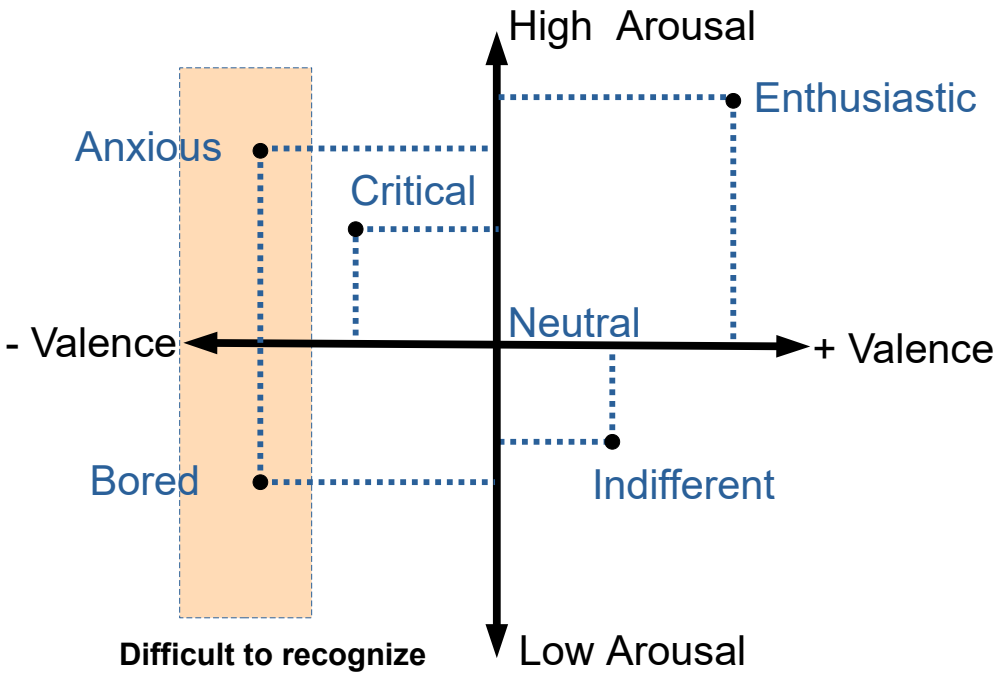


Figure 4.8 – Mapping of virtual audiences' attitudes on the models' dimensions using the categorisation from Kang et al., 2016. On the horizontal axes the valence and on the vertical one the arousal.

Engagement towards the speech (Arousal): the audience's engagement is significantly related to the gaze, the frequency of movements and the posture's proximity. As for facial expressions, we would advise alternating between the targeted facial expression and a face at rest to let the users perceive and compare those differences instead of displaying them continuously. Kang et al., 2016 already offered such guidelines for facial expressions.

Opinion towards the speech (Valence): The opinion is significantly related to the nonverbal behaviour type. Consequently, we would advise using distinct head movements and facial expressions to help users differentiate them. Moreover, if VR users are close to the audience, they might be able to perceive subtle facial expressions and head movements but significantly less if users are distant from it.

Recommendation for negative opinion and engagement: according to our results, users may have difficulties differentiating levels of engagement (Arousal) when the virtual audience has a negative opinion (Valence). Consequently, we would recommend using a lower proportion of negative nonverbal behaviours with regard to the percentage of positive nonverbal behaviour, such as frowning or shaking the head and crossing the arms. Chollet and Scherer, 2017 also reported that fewer negative behaviours are required to let the users correctly recognise a negative attitude compared to positive ones.

4.5 Limitations

Our evaluations have shown that despite not being domain-specific, we can successfully create different audience attitudes significantly identified by VR users, such as indifferent, critical, and enthusiastic (Figure 4.9). We mostly validated our hypotheses on virtual audiences' attitude perception and confirmed its efficiency in virtual reality, except for virtual audiences displaying low engagement and negative opinions as per the model. Audiences with low engagement and negative opinions are not identified correctly by VR users, which might prevent the perception of bored or anxious attitudes. The reasons for this shift in perception are probably due to the greater intensity of negative signals and, as reported by Kang et al., 2016, the interaction between negative valence and low arousal.

Our results also raise questions about some behaviours from the model and their perceptions. Posture relaxation, deemed significant for the perception of valence in the model, has not been confirmed as such in our first study. However, users still significantly distinguish a positive audience from a negative one in our second study when we used posture relaxation. It probably comes from the fact that VR users recognise more easily negative behaviour than those related to arousal.

The interviews with the participants gave us more insight into their understanding and perception of the audience's attitudes. Some of them mentioned that facial expressions were strong signals, which corresponds to the results from our first evaluation. Participants also mentioned they use a known context to make their decision, e.g. a lecture or a professional meeting. We gave no context before the experiment, but a precise context might help the users to associate nonverbal behaviours with a known audience attitude. Lastly, the most recurrent comment was the lack of sounds from the audience when it was playing some behaviour. None of the studies we used to build our model defined sound or backchannels for the virtual audience. Therefore, we chose not to add sound to our

Figure 4.9 – Example of nonverbal behaviours used to display 4 different attitudes, (A) shows this behaviours on one agent and (B) for the entire virtual audience.

evaluations. However, other studies concerning virtual agent behaviour, such as research on conversational agents may fill in the gap between the nonverbal behaviours and the associated multi-modal backchannels [Barmaki and Hughes, 2018; Kistler et al., 2012;

Laslett and Smith, 2002; Poppe et al., 2010].

Based on this feedback, we can see that some of the limitations in the audience's attitude perception may be solved when using the system in a specific context. Because more behaviour diversity and sound may also help to improve perception, the features introduced in Chapter 3.4 can improve the model. These behaviours were not part of the evaluations we present here to preserve the integrity of the relationships we tested. However, behaviours such as chatting or leaving the room warrant exploration in future perception studies.

4.6 Conclusion

In this chapter, we presented our perception studies aiming for evaluating the users' perceived attitudes generated with our model. The first evaluation provides insight into non-verbal behaviour perception according to the model's dimensions, valence and arousal. This study partially reproduces the result from Chollet and Scherer, 2017 in VR, except for the eyebrows, for which we could not run a test and the posture's openness. In our second perception study, we evaluated audience attitude user perception and confirmed the users' ability to distinguish opposed levels of valence and arousal except for the low arousal perception, which might interact with negative valence appraisal. Kang et al., 2016 already reported a potential interaction between low arousal and negative valence. However, adding some audio to the environment and domain-specific behaviour to root the users in a given context could solve this issue.

From our point of view, the solution to extend the model was to deploy the AMBIANCE in a domain specific application to evaluate it with experts. Thus the following Chapter 5 introduces a public speaking training application for the university curriculum deployed for bachelor students from the University of Würzburg. This new application helps us study new context-specific behaviours and backchannels with lecturers and students and evaluate the system's acceptance and usability. In fact, as explained in Chapter 2, virtual audience control can be difficult and can produce a workload for instructors depending on the method used, i.e. Wizard of Oz or fully autonomous. This public speaking system helps us investigate how a hybrid approach can ease audience control and still suit the instructors' needs. VR GUIs, similar to Figure 4.1, offer the possibility to fine-tune the audience or to provide high-level controls directly influencing the audience's attitude.

Chapter 5

Model deployment and evaluations

5.1 Introduction

We have seen in the previous chapter that the AMBIANCE can generate different audience attitudes. Yet, these user evaluations do not investigate how to control such virtual audiences. In existing systems, audience control is an issue when experts drive a simulation whilst they supervise a training or therapeutic session. The Breaking Bad Behaviour application we took as a reference for our work relies on experts in classroom management to manage a training simulation and author the classroom's behaviour to follow a pedagogical plan. In this application, instructors drive each virtual agent's behaviour and adapt it at runtime accordingly to the trainee's actions. In this system, the supervisor has to adapt each virtual agent's behaviour to fit with the ongoing pedagogical scenario while listening to the trainee teacher and taking notes for the post-training briefing. In this case, the experts can be overwhelmed by the controls and the pre-service teacher supervision, which significantly increases the workload. Moreover, preparing a pedagogical plan or a therapeutic strategy with such a system might be complex for computer science neophytes Mouw et al., 2020.

Delamarre et al., 2021 propose another classroom management training system to drive the virtual audience behaviour according to scenarios designed by domain experts. As a result, experts script the virtual audience's behaviour, and thus it has no scenario flexibility and might suffer from simulation repetitiveness. Still, it provides a high-level authoring tool for users without knowledge of scripting languages. However, virtual audience systems such as Breaking Bad Behaviour relying on a tutor-in-the-loop could benefit from higher-level user control to manipulate the audiences. Such autonomous features could profit VR training systems, regarding the trade-off between a fully autonomous simulation and a Wizard of Oz system where each spectator is individually controlled. For instance, when replacing tutor expertise with a self-sufficient component is not desirable, e.g. VR therapy and training could need real-time adjustments and temporary fine

control of the environment.

In Chapter 3, we evidenced how virtual audiences' ability to modify their behaviour is beneficial to convey emotions VR users can perceive. These emotions emerge from non-verbal behaviours and multiple social signals, such as backchannels or interactions between agents and users [Chollet and Scherer, 2017; Kang et al., 2013; Pertaub et al., 2002]. Therefore, VR social applications can benefit from attitude control to become suitable for teaching and therapeutic environments. Additionally, VR exposure therapy applications use virtual audience as fear stimulus due to the ability system have to fine-tune the audience's behaviour to elicit various fear response [Owens and Beidel, 2015]. However, as explained in section 2.7, a dynamically controlled environment is mostly unfeasible or unsafe during an in vivo simulation and virtual training simulations require controllable environments to supply instructors with training scenarios and plausible environments. Therefore, fine control of the audience behaviour is paramount for rooting the user in the virtual scene and providing training and therapeutic adaptive environments.

Our proposal consists of a novel scenario control tool that aims at solving specific requirements for training or therapeutic VR systems. Our approach is to use a behaviour model to create pedagogical scenarios relying on the affect it arouses in the users. Unlike regular training scenarios, this approach does not rely on a sequence of actions and choices that makes the scenario branch but instead focuses on the affective experience. Thus, during the presentation, the audience's attitude changes modulate the students' affects. This section presents two contributions: we first describe how we used a user-centred development process to develop a VR training system for bachelor students and then how we solved the trade-off between a fully autonomous and a Wizard of Oz system. In doing so, we extended the AMBIANCE with non-verbal behaviours, backchannel, and affective cues based on the instructor's feedback. Finally, we present a high-level control interface helping instructors to design pedagogical narratives via a high-level application programming interface (API). Our system and its development process provide insights into the successful integration of VR-based formative educational tools into a university curriculum training application.

We introduce and discuss the STAGE system (Speaking To an Audience in a digGital Environment), a high-level control system built around a state-of-the-art virtual audience simulation. The STAGE allows leveraging the potential of co-presence in finely controlled and tutor-led training for public speaking through the creation of pedagogical scenarios. These scenarios rely on events that encompass affective phenomena rather than organising

events changing the course of a training scenario.

5.2 Public Speaking VR Training Application

5.2.1 Development Context

We developed the STAGE for *Scientific Writing and Presentation* seminars for post-graduate and undergraduates at the University of Würzburg in Germany. We used the system for two semesters: during the first semester, volunteers participated in a preliminary study that helped us develop the first STAGE prototype, and in the second one, all students practised in VR to prepare for their final presentation. The application was designed for the ATMCS module (actual trend in mensch und computer), an HCI course in which students learn how to prepare a scientific presentation and review a paper. Students have to choose an article from the last CHI conference, present it, and provide a review. During the exam, they have 10 minutes of presentation, then a 5-minute long review and finally 10 minutes of questions from the audience. Before the oral presentation, students have a series of lectures on public speaking skills, slides preparation for scientific presentations and scientific paper review.

Before we deployed the STAGE into the seminar, the students had no compulsory training and were mainly preparing for their exams with the lectures. According to the lecturers, only a minority of students contacted the professors to get feedback on their presentations. Therefore, we proposed the STAGE as a VR training tool to let the students practise their public speaking skills, especially those which can be challenging to learn with online presentations, e.g. how to react to the audience's behaviours or how to use the space on stage. Thus the STAGE provides a learning tool that can expose students to situations they may experience during real-life presentations.

The STAGE was then designed to fit this seminar and provides a safe learning environment for the students and a flexible educational tool. We designed the training sessions to supply the students with a practise session in front of a virtual audience with the professors supervising them. On the one hand, to help during the presentation and on the other hand, to give a personalised review of the student's slides and presentation quality. Thus, we used the seminar marking to provide a virtual environment that lets the supervisors evaluate the students with the different criteria that are usually used for the seminar (see appendix E.1).

In a desire to focus the system's development on the needs of the different users (i.e. the instructors and students), we first targeted the critical functionalities making it possible to provide a functional virtual training environment. Then we iteratively added different software improvements providing better control of the environment and the best user experience.

5.2.2 Development Methodology

To provide the most suitable system to the instructors, we followed a user-centred development driven by the lecturers in charge of the seminar. Figure 5.1 shows the development process we followed.

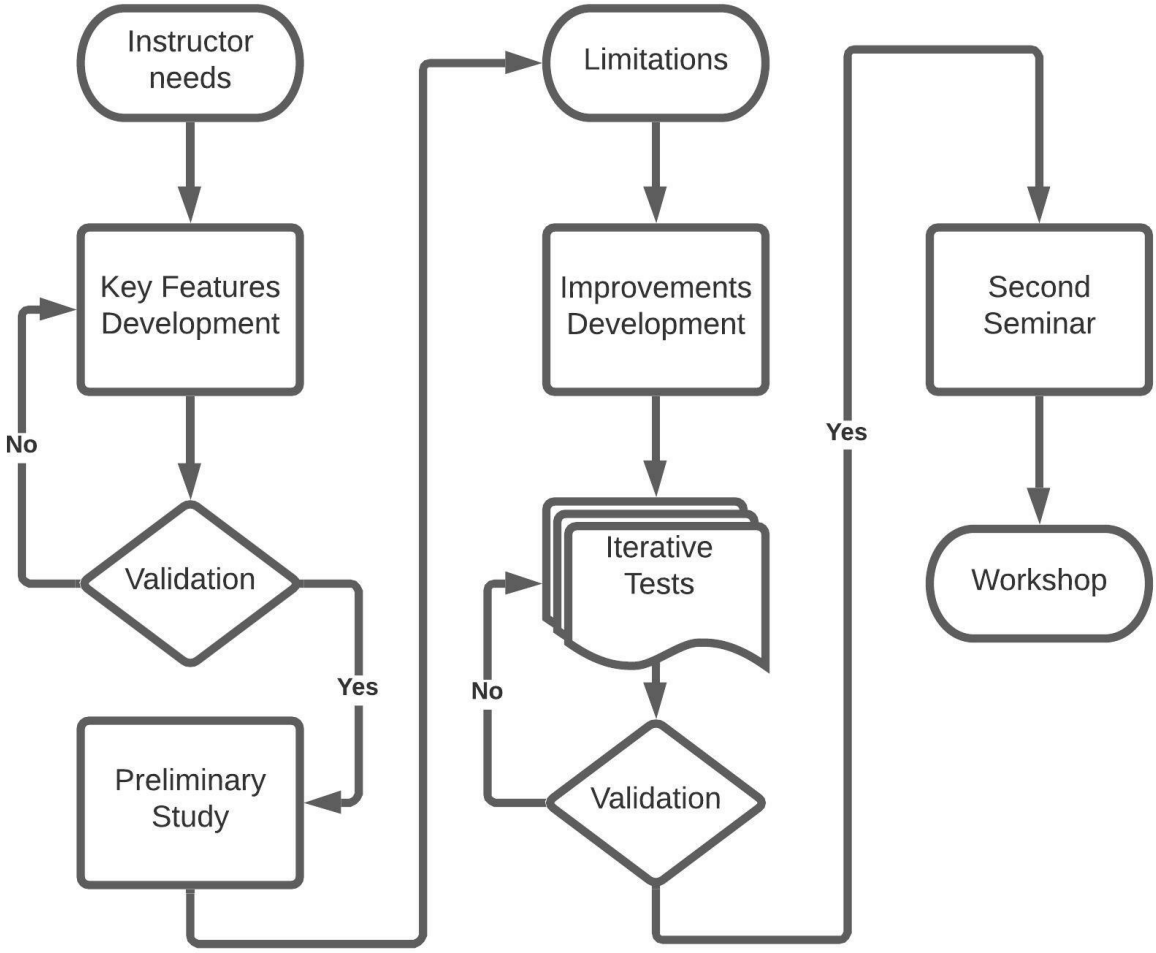


Figure 5.1 – The methodology used for the design of the system.

Our first milestone was the critical feature identification required by the instructors according to their pedagogical needs. Thus, after analysing the instructor's and the student's tasks, we listed the features required for the seminar to happen in VR. The instructors needed to be able to listen to the presentation while watching the slides and the student's movements. As for the students, they needed to be able to display their slides in VR, control them with a remote controller, and have feedback on the current state of their presentation, i.e. current slide shown and remaining time. On top of these features, the system requires a plausible and believable virtual environment populated with a controllable virtual audience to expose the user to various public speaking situations.

Moreover, the system had to allow the application to be used in mixed reality (MR) with a projected virtual audience with Kinect-based speaker tracking to accommodate students uncomfortable in VR and provide a more natural conference-like situation.

With these key features in mind, we developed a prototype with an iterative process where instructors test and validate each implemented feature. After the prototype was functional and validated, we ran an eight-week-long preliminary study in which undergraduate student volunteers participated in training sessions. The training session was structured as follows: the students sent their slides beforehand for test purposes, and then on the training day, they had a training session in which they could test their slides in both VR and MR. After this training, the students had to choose between VR and MR and stand ready for the presentation (Figure 5.2). The presentation was 10 minutes long, with questions from the instructor at the end. Additionally, a semi-structured interview and a briefing between the students and the instructor were following it. In this discussion, the instructor gives feedback about the slides' quality, the presentation content, and the public speaking skills.

Figure 5.2 – The three possible roles during the seminar: (a) Embodied Teacher with controls, (b) the virtual reality speaker, and (c) the mixed reality speaker.

Thanks to this preliminary study, we gathered the first students' impressions of the system. Instructors also provided a list of additional requirements from their system uses during this seminar, namely regarding the controls and the cognitive load when it comes to following the presentation and handling the virtual audience. Thus, a second iterative process started with PhD student volunteers and instructors to test each improvement requested. Students were rehearsing their presentations for incoming research meetings and were able to provide further feedback for each iteration. Some others experimented with specific system features, such as the slide controls or the training instructions. PhD students also participated in Monkey testing to anticipate and avoid usability issues during

the presentations. As for the improvements made to the AMBIANCE, instructors asked for new attitudes and behaviours and a new control interface to widen the possibilities for designing pedagogical scenarios.

After the instructors validated the second prototype, a second seminar used the training system to let students practise VR before their final exam. In parallel, we organised a workshop with lecturers to get more insight into its possible use in subsequent lectures and seminars.

5.3 Application Architecture

The prototype relies on the instructors' pedagogical needs, which were namely: listening to the student's presentation, watching the slides and seeing the students' movements provided that the simulation takes place in a believable conference environment.

We developed the prototype with *Unreal Engine 4* [1] [Epic Games, 2022b] as a VR application for students and as a desktop application for the instructors (Figure 5.3). This application follows a client-server architecture where the client side is responsible for the student's controls and the server for the virtual audience attitude and the instructor's controls. This network architecture allows instructors to attend the presentation remotely, e.g. during the first seminar, students and instructors were in two different rooms but connected via the university network.

1. Unreal Engine, https://www.unrealengine.com/en-US/, [Accessed March 29, 2022]

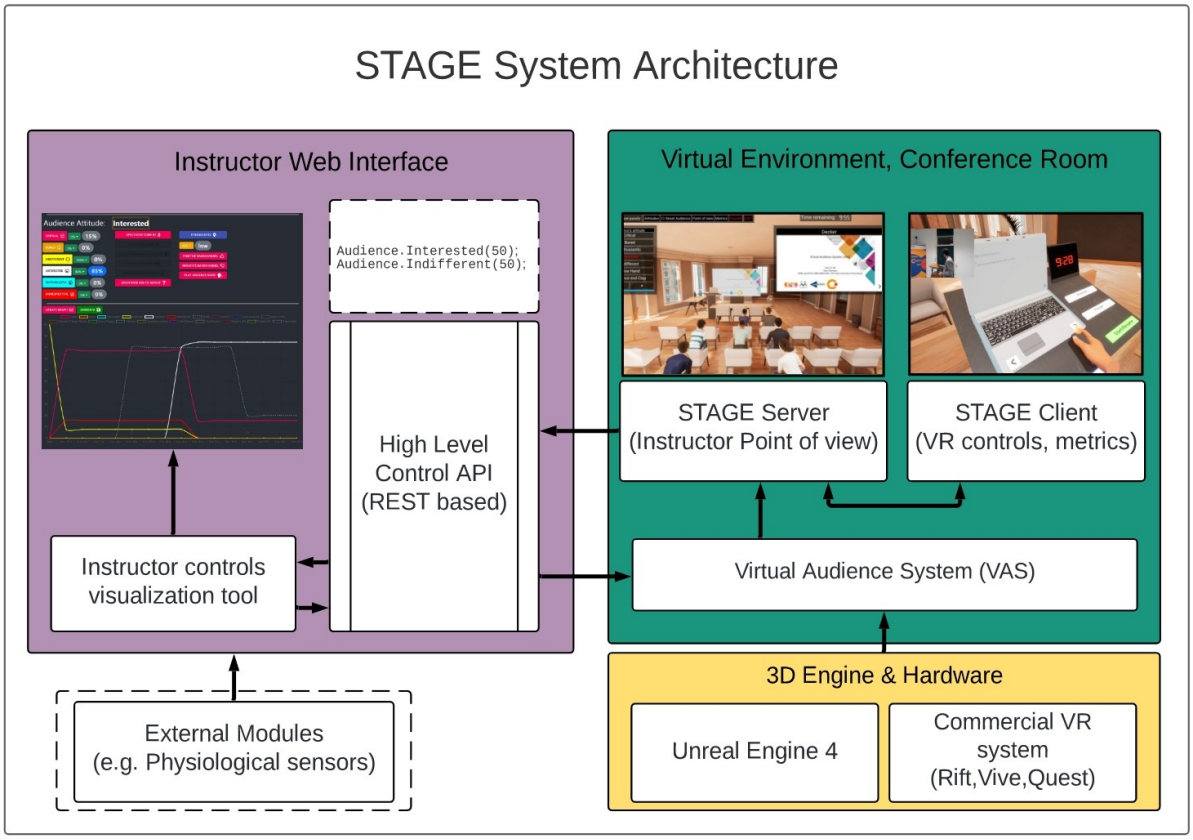


Figure 5.3 – Speaking To an Audience in a diGital virtual Environment (STAGE) system architecture.

To meet the student's requirements, we created a virtual environment allowing them to control their slides. Thus, we used Decker [2] [Latoschik & Team (Uni Wuerzburg) et al., 2022], an open-source slide creation tool based on the Markdown language, interpreted in HTML by a web browser. Seven German universities and lecturers at the University of Würzburg already use this tool. Lecturers were already using Decker for the seminar, and students had presentation templates to become familiar with the system. Thus, we created a natural VR interaction metaphor with the slides. As in Figure 5.2, we implemented a virtual remote slide presenter with a laser pointer, appearing in the user's virtual hands and controlled with the VR controller buttons or thumb-sticks.

Therefore, students can use the controllers to interact with slides by clicking on the virtual screen and highlighting elements of their slides with the laser. Also, the presenter

2. Decker sources repository, University of Würzburg https://gitlab2.informatik.uni-wuerzburg.de/decker/decker,[Accessed March 29, 2022]

can press the controllers' buttons to move to the next or previous slides. Two web browsers display the slides on a large panel representing the projected slides and on a laptop, which allows the user to have feedback on the current slide while facing the virtual audience. A timer displays the remaining time as soon as they start the presentation by pressing a button next to the virtual laptop (Figure 5.4) to keep track of the time spent during the presentation. Finally, a panel can show presentation-quality metrics that were possible to visualise and export for the students, such as the percentage of time looking at the audience, the time on each slide, which agents the user looks at the most, or the time talking. Besides the slides and laser pointer interactions, the user can embody an avatar composed of two virtual hands holding the laser controllers, a head-mounted display, and transparent footprints on the ground to locate the user. The head-mounted display is not visible from the student presenter's point of view. The hands are animated and move according to the capacitive sensors of the VR controllers, which provide the student's thumb and index location.

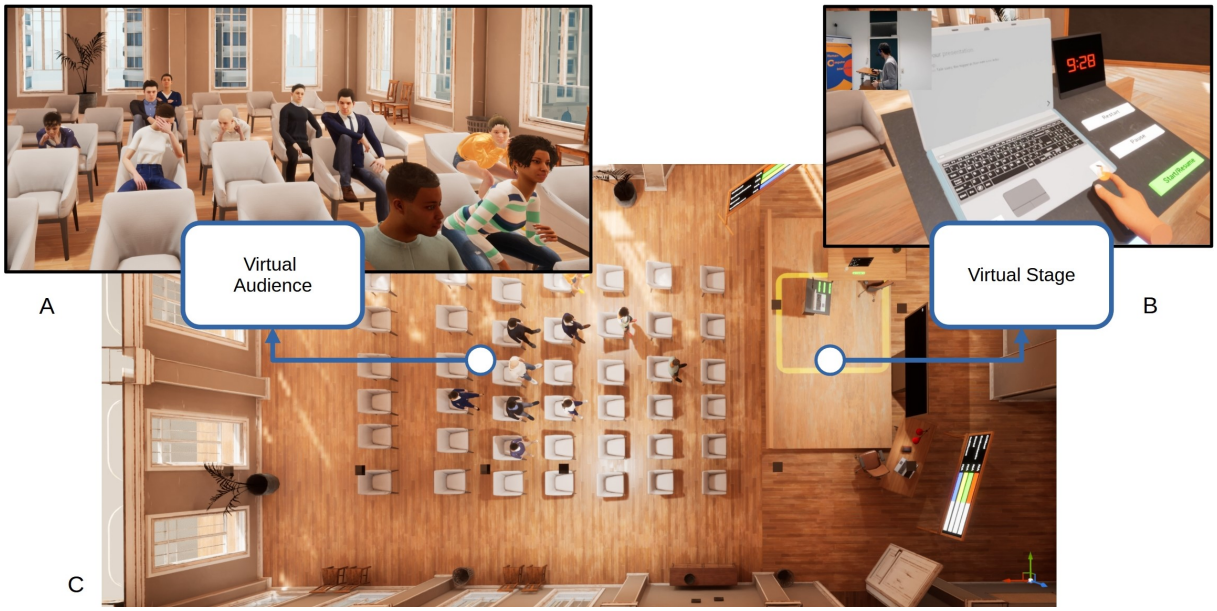


Figure 5.4 – System Overview, from a top-down perspective: With (a) the virtual audience, (b) the virtual stage with the laptop and the student menu, (c) the virtual conference room from a top-down perspective.

To improve the menu buttons' usability, we created a press interaction with the user's hands, which triggers a visual cue and a hand animation when the student's hand gets closer to a button to encourage the user to press it with the index. Then when the student

interacts with the button, it triggers a visual effect and a smooth vibration in the controller to notify the student of the ongoing interaction. Finally, to increase the interactivity with the environment, the user can grab some objects in the virtual environment.

The STAGE can also use scanned avatars (Figure 5.5), allowing the users to embody their photo-scanned avatar using inverse kinematics to partially track the body movements based on the head and the controllers' location. This feature has not been used during the seminars for multiple reasons, the scanning pipeline takes time and might require redoing the process in case of a malfunction, and we were not sure all the students would agree to be scanned. So to keep the same protocol inbetween students, we avoided using scanned avatars.

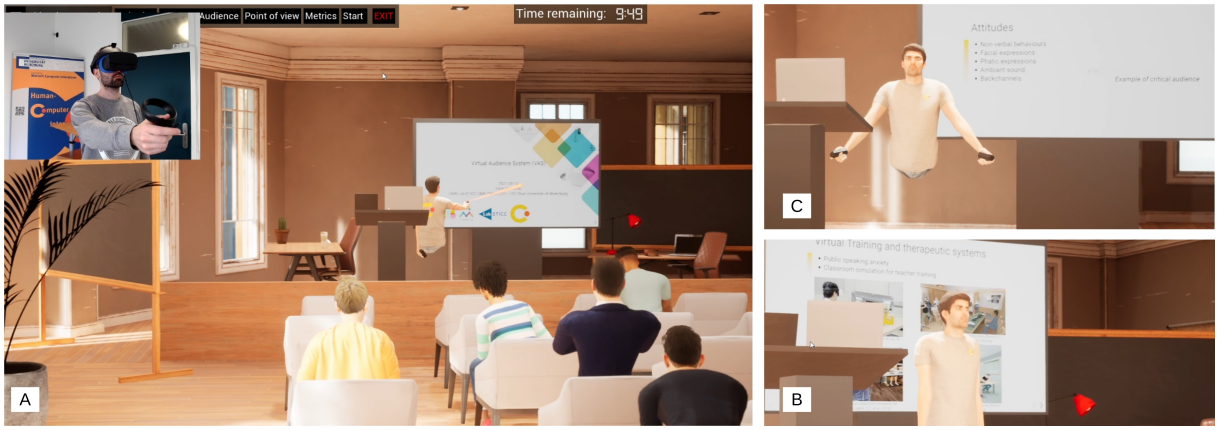


Figure 5.5 – Example of scanned avatar embodiment for the speaker (a) with the point of view from virtual spectators (b) and (c).

The audience was also a critical prerequisite for the instructors. As stated in section 5.1, VR training simulation's usefulness lies in their ability to expose users to particular situations while controlling the degree of exposure, in our case, the virtual audience and the virtual spectators' behaviour populating it. In the prototype, the AMBIANCE plugin described in Chapter 3 generated the virtual audience, which made it possible to display four audience attitudes, i.e. bored, enthusiastic, indifferent, and critical. Then, through our iterative development process, we extended the model with new audience attitudes, context-related behaviours and social interactions such as backchannels. Lecturers also added a small set of behavioural cues to support the pedagogical plans, e.g. spectator leaving or coming into the room.

Finally, we added a GUI to extend the instructor desktop application to let lecturers

have high-level control of the virtual audience's attitude (Figure 5.6). This GUI also allows the instructors to use a live question system with which they can embody a virtual spectator to raise a hand and talk through a microphone. It also shows the slides and has a camera system to get different points of view of the virtual environment, e.g. from the back of the room, from the front row or from the stage. After compiling the instructors' feedback for the preliminary study, we extended the desktop application with a web graphical user interface providing a high-level control API to control the virtual audience at run-time and pre-scripting pedagogical scenarios. A visualisation tool accompanies this second web GUI to keep track of the ongoing narrative.

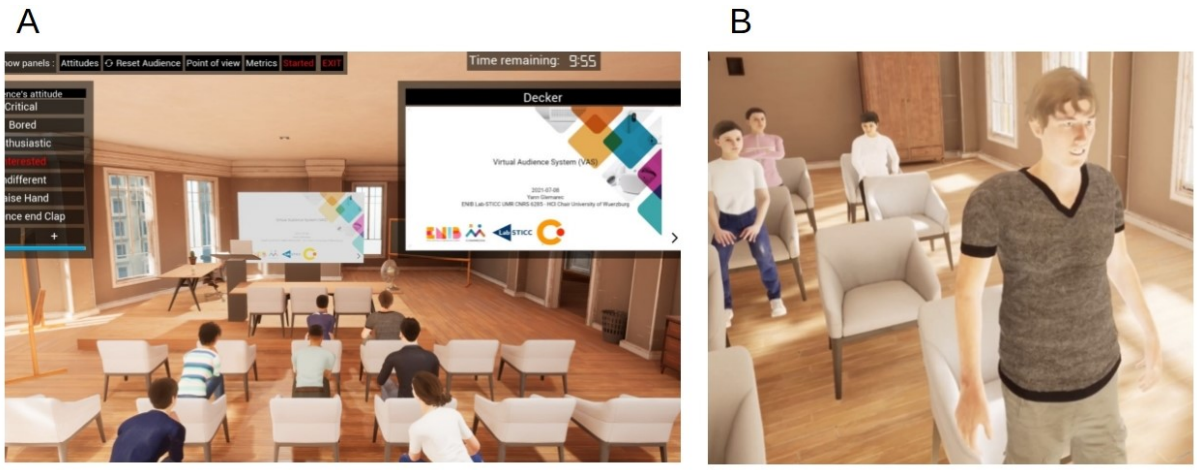


Figure 5.6 – Instructor Graphical User Interface with controls and slides preview (a) and embodied virtual spectator by the instructor for the end questions (b).

A training session was structured as follows: first, the students sent their slides made with Decker a bit in advance to test them beforehand, and then they had a training session in which they could test their slides in VR and MR. After this training, the student had to choose between VR and MR and stand for the presentation. The presentation was 10 minutes long, with questions from the instructor at the end. A semi-structured qualitative interview and a briefing between the student and the instructor followed the presentation to give feedback about the slides' quality, the presentation content, and the public speaking skills.

The first preliminary results led us to develop better interaction techniques, extend the VAS, and add a high-level GUI allowing us to launch previously established scenarios while providing graphical feedback on the audience's state. Consequently, all the devel-

opments described in the following sections were made based on successive iterations. Following these developments, a second seminar took place as well as a workshop with university professors and lecturers who tried out the system. Thus, the following sections introduce the AMBIANCE's extensions added to compensate for the limitations identified during the user study and the behavioural cues requested by the instructors to let them create more suitable audiences for the scenarios (section 5.4.1). Then we describe the instructor interface allowing the design of pedagogical plans with a high-level API. It aims at granting fine control of the virtual audience. This GUI provides a scenario and the current audience's state visualisation (see section 5.6). Finally, all the feedback from students and lecturers who participated in the second seminar and workshop are given and discussed to provide guidelines on the development of a similar training system for the university curriculum (see section 5.7.2).

5.4 The STAGE System Implementation

From the instructor's point of view, the STAGE is a teaching aid by which the student can experience simulated scientific talk. Such a simulation implies a believable virtual audience in terms of reactions toward the presentation. The challenge is to provide an audience whose behaviour communicates a perceptible attitude toward the speaker. This phenomenon has the effect of arousing positive or negative affects. The aim is to supply the student with an environment that will provide the best possible experience of a scientific talk.

5.4.1 Virtual Audience Implementation

As introduced in section 5.3, we integrated the AMBIANCE plugin into the stage architecture. Thus, the rule-based model helped us adjust the virtual agents' behaviours to the desired audience attitude.

We modelled the attitudes into objects containing the rules and triggering them when needed. These attitude objects encapsulate the associated behaviours and expose straightforward controls to the experts. It avoids directly using the behaviour rules whilst they can change the proportions of agents displaying the targeted attitude. Moreover, we integrated the reactions to disruptive events described in Chapter 3.3.5 and the custom behaviours needed by the lecturers. As for the posture modifications, the rules were not

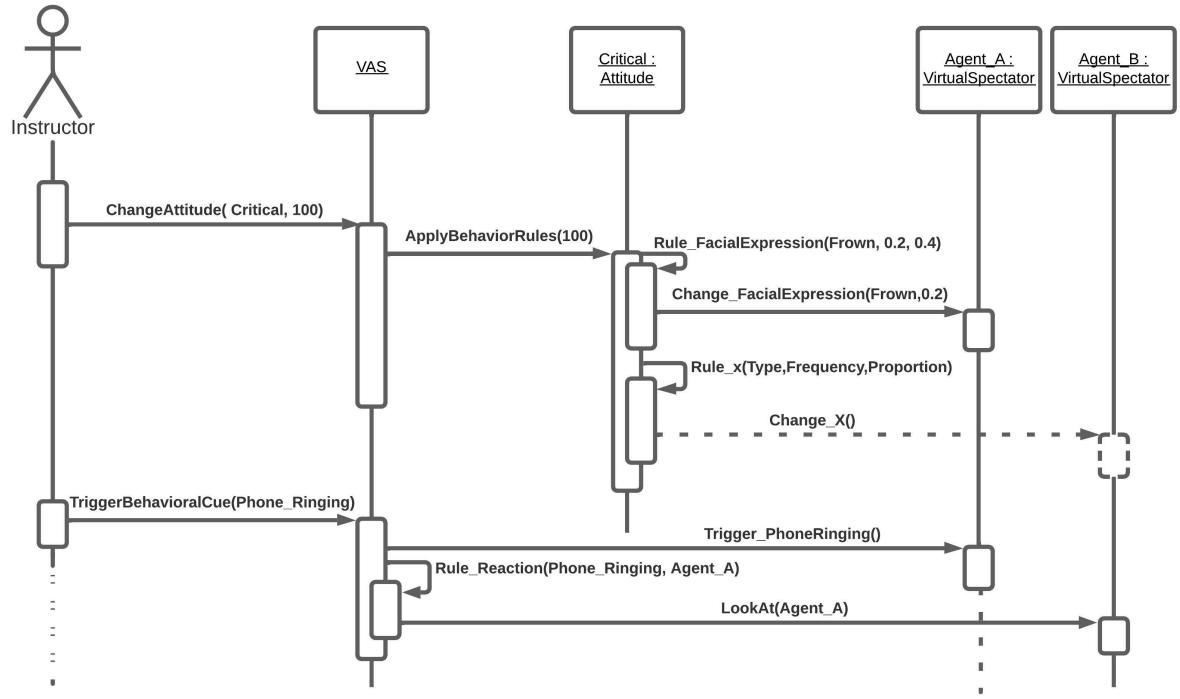


Figure 5.7 – Sequence diagram describing an attitude change into a critical one, with an example behavioural cue triggered with its reaction.

changing the displayed animations, but the instructor had to request a new posture to see changes. Alternatively, we decided to consider different posture types to cluster animations and assemble them into posture pools responsible for changing the postures through time, in a random order, in a pre-establish order or based on a probabilistic basis. Overall the new operating principle is very similar to the previous one but provides higher-level controls, Figure 5.7.

5.4.2 Behavioural Cues

The AMBIANCE does not include backchannels or behavioural cues because it was not evaluated in a specific context and is only relying on non-verbal behaviours. Thus, the STAGE had to extend the first implementation with context-specific behaviours. Based on the feedback from the first seminar, the instructors gave us a list of behaviours they needed in their pedagogical scenarios to design the audience's attitudes. They created two scenarios using multiple attitudes to let the students experiment with different types of audiences and various presentation phases they could encounter in real life. For instance,

the audience is interested at the beginning of the talk but then gets bored and indifferent to the speech, and finally, the audience becomes critical because the spectators do not appreciate the presentation. Instructors punctuated these scenarios with contextual affective cues used as disruptive or supportive events during presentations, such as a spectator coming into the room during the speech, someone yawning loudly, or playing a supportive backchannel. The two scenarios were very different, the first was meant to be supportive and less stressful as possible, whilst the second one was meant to be challenging, including frequent disruptive behaviours and a majority of negative attitudes displayed. The purpose of such scenarios is to let the students face the different situations they cannot experience training individually and provide personalized feedback from the lecturers.

Therefore, to provide plausible scenarios to students and let them face these public speaking situations, we added to the model several behaviours. In a nutshell, we first extended the number of postures and variations of the existing model implementation to discard the looping behaviours. Students often mentioned these looping animations to be noticeable during the development phase and interviews. These posture variations are changes in some body parts, like crossing the legs differently or resting on the opposite hand. Then, because the instructors felt limited by the four original attitudes from the model, we created new ones with the same rule-based system. One guideline for building new attitudes was to use attitude-related behaviours to let the users distinguish the difference when two attitudes are close in terms of perceived valence and arousal, e.g. to differentiate a bored audience from an indifferent one. Thus, we extended the new model with two new attitudes, interested and disrespectful. The model defines the interested one with a high level of arousal and positive valence, i.e. equivalent to positive opinion and engagement toward the speaker. Thus, based on the model's rules, this attitude triggers frequent nodding and smiles with virtual agents leaning forward and mostly staring at the speaker or the slides. As for the specific behaviours related to the interested attitudes, we added two evocative behaviours according to the instructor. The supplemental behaviours were "taking notes" and "leaning sideways" to look at the slides when someone obstructs the agent's sight. By contrast, the disrespectful attitude displays less frequent head movements and facial expressions while the virtual agents lean backwards. In this case, the specific behaviours were agents texting, chatting together, or putting their arms behind the head. The new set of attitudes now includes around seventy postures, four different head movements, and four facial expressions. The STAGE also includes specific behaviours such as yawning for the bored attitude, texting for the disrespectful attitude,

or taking notes for the interested one. We made these new animations from motion capture data which were then applied to the different virtual agents.

Besides these new behaviours, to support the scenarios and improve the overall audience believability, we added some affective cues and social interactions made with motion capture as well, e.g. a phone ringing. Along with the domain-specific behaviours like yawning or texting, we added contextual behaviours such as spectators asking to repeat with German voice lines depending on the virtual agent's gender or a phone ringing followed by apologies from the virtual agents. As for the social interactions, we added reactions to these events and behaviours based on a proxemic awareness of what is happening around them. For instance, when a phone rings, the surrounding agents will look at the virtual spectator trying to switch off its phone. It works the same with spectators coming in late, virtual agents close to the newcomer stare at it because they are distracted. Thus, to create these reactions, we added new features, such as pathfinding from the game engine and audio components to play 3D sounds.

The affective cues are periodic, either triggered manually by the instructor or automatically by the narrative. These reactions are conditioned and can rely on pre-established rules such as the distance between virtual agents, the current attitude displayed, or some user metrics. We implemented these utility-based rules from audiences' accounts in which the instructors precisely described what type of behaviour happens when they occur and in which circumstances. Consequently, all reactions rely on heuristics from the instructors' experience of public speaking but change according to the current audience's attitude. For instance, when the virtual spectator's phone rings, some agents can look at it and frown if there are interested but might not react if bored or disrespectful (see Figure 5.8).

To improve the interactions between the student presenting and the virtual audience, we added backchannels to increase the virtual audience's engagement toward the talk. However, in this situation, the audience is not involved in a conversation but only in a listening situation. Consequently, we added supportive and negative multimodal backchannels that do not involve analyzing the talk. The supportive ones notify the user that the spectator better understands what she or she is saying, i.e. the agent nods and emits a long "mmh". As for the negative ones, spectators were notifying a misunderstanding, with the spectator frowning and emitting a specific negative or even rude multimodal backchannel. The students who recorded the voice lines recorded a specific one from the German language that can be compared to the long "what" in English. Also, we based the moment when to trigger these backchannels on advice from the lecturers. A utility-

Figure 5.8 – Example of audience reaction: when a new virtual spectator enters the conference room.

based function uses the user metrics gathered during the presentation to decide when the backchannel should be supportive or not. So, for instance, when the speaker has a regular pace, and the student is often looking at the audience, it is more likely that the audience will trigger a supportive backchannel, while if the student always looks at the notes and needs to go back to previous slides, it is even more likely that a negative backchannel is triggered. Finally, to avoid long absences of noise coming from the virtual audience, we added some noisy behaviours which do not affect any pedagogical plan or attitudes, e.g. a virtual spectator picking up a pen or others who cough. PhD students evaluated each of these modifications during their training through structured interviews, before instructors approved them.

5.5 Performances and Scalability

In the STAGE the virtual audience is composed of 13 different virtual characters from *Adobe Mixamo*[3] [Adobe Systems, 2022]. During the seminar, the students were given an *Oculus®Rift S* with a constant 80 Hz refresh rate bound to the hardware. However, to provide the most plausible environment, such a virtual conference room should be able to issue larger crowds. Thus, we evaluated the STAGE performances and

3. Mixamo character and animation library, https://www.mixamo.com/#/ Accessed February 12 2022

Figure 5.9 – The three virtual audiences were used to run the scalability performance evaluation: (a) the 13 agents used during the seminar, (b) 26 agents, and (c) 36 agents which correspond to a full conference room.

ran a scalability benchmark. We first measured the AMBIANCE performances within the seminar setup over 5 minutes, and as expected, the behaviour model implementation does not considerably impact the performance ($\mu = 1.68$ ms, $\sigma = 0.2$). To perform the evaluation we used a computer running with Windows 10 64 bits, Intel®Core™i7-9700K central processing unit (CPU) at 3.60GHz, and NVIDIA®GeForce RTX 2080 SUPER Graphics Processing Unit (GPU) 8GB GDDR5. An Oculus®Rift S was used to carry out the VR evaluation.

Nonetheless, a load test shows that the STAGE is quickly GPU-bounded due to the virtual agent rendering (Table 5.1). Further investigations have shown that shaders used to render some agents' hair had a significant impact on the frame rate. Translucency is a common issue for VR rendering and is detrimental to performance. Concerning the number of agents and the resulting frame rate, the system can handle 19 virtual agents from Mixamo. Although, if we double the number of virtual agents, the system runs already under 45 frames per second which is not suitable enough for VR uses (Figure 5.9).

Table 5.1 – STAGE Scalability Performances measured over a 2-minute long period.

Number of Virtual Agents	Number of frames per seconds (FPS)	Game Thread (ms)	Rendering Thread (ms)	GPU Thread (ms)
13	83	$\mu = 11, \sigma = 0,2$	$\mu = 6.8, \sigma = 0.3$	$\mu = 12, \sigma = 0.2$
26	41	$\mu = 24, \sigma = 1.1$	$\mu = 12, \sigma = 0.9$	$\mu = 25, \sigma = 0.3$
36 (full room)	23	$\mu = 32, \sigma = 3.4$	$\mu = 24, \sigma = 1.4$	$\mu = 42, \sigma = 5.0$

Even if our system already uses different levels of detail for the virtual agents' meshes and props in the environment, these measures testify to a need for a virtual audience

Figure 5.10 – Three different types of virtual agents used within the STAGE: (a) the characters used during the seminar from Mixamo®, (b) the Meta-Humans from Epic®Games, and (c) Photo-scanned avatars from the University of Würzburg.

with a mix of highly detailed characters and less detailed ones. For instance, blending photo-scanned avatars, Meta-Humans from Epic®Games, Mixamo characters, and others with simplified mesh structures could solve such limitations. The most detailed virtual agents should be close to the speaker for the facial expressions to be visible so that the further the agent is, the less it is. But the STAGE system can already use these different virtual agents on the condition that the animations are compatible with all agents' builds (Figure 5.10).

5.6 STAGE Control Interface

After we evaluated the first prototype, instructors reported limitations due to the complexity of manually and continuously controlling the virtual audience, even though the instructor only had to change the overall attitude or trigger specific behaviour accordingly to the presentation. Therefore, we proposed to design a web interface in which they could monitor pedagogical narratives. In doing so, they only have to listen to the presentation and monitor the ongoing pedagogical scenario or eventually adopt the virtual audience's attitude if the current training session does not fit with the pre-established narrative.

Therefore, we designed a graphical user interface to enable us to control the virtual audience by changing the attitude and triggering specific behaviour like backchannel or context-related behaviours and by controlling and monitoring the ongoing narrative. Thus, we created a web GUI based on the REACT framework using a REST API to communicate with the STAGE application. Concerning the scenario design, we implemented a high-level control API to directly control the STAGE and change the audience's attitude or trigger disruptive events.

5.6.1 Audience Controls

The instructors required a high-level control interface and to keep fine control over the audience to adapt it. The application has a central component responsible for providing a simple control interface with the behaviour model and the different rules. Thanks to the attitude model, instructors can directly change the attitudes in percentages without taking care of each agent individually. This component provides direct access to the AMBIANCE manager. Thus, we developed a high-level API to drive the virtual audience with simple instructions. Figure 5.11 shows how these instructions can quickly adapt the audience to the presentation. However, the system cannot always follow the instructions and tries to get as close as possible to it. For instance, two agents speaking together can only be displayed under certain conditions, they have to be next to each other and close enough.

Such controls can elicit a heavy workload, so the GUI had to ease its use. Consequently, based on the instructor's audience feedback, we linked the different behavioural cues to each attitude. It enables us to dynamically adapt the displayed buttons of each domain-specific behaviour to activate behaviours associated with the current attitude. These dynamic buttons nudge the instructors to only manipulate attitude-related behaviours. Aside from the audience controls, the interface provides controls over the virtual environment, such as the student's timer, a reset of the slide, or a logging system for the instructor to add information within the visualisation tool, e.g. the speaker perceived stress.

5.6.2 Virtual Audience Control API

The aforementioned high-level API also provides the instructors with a simple scenario editing tool. In addition to these controls, a state machine lets the instructors use successive states containing the high-level instructions adapting the VA attitude and be-

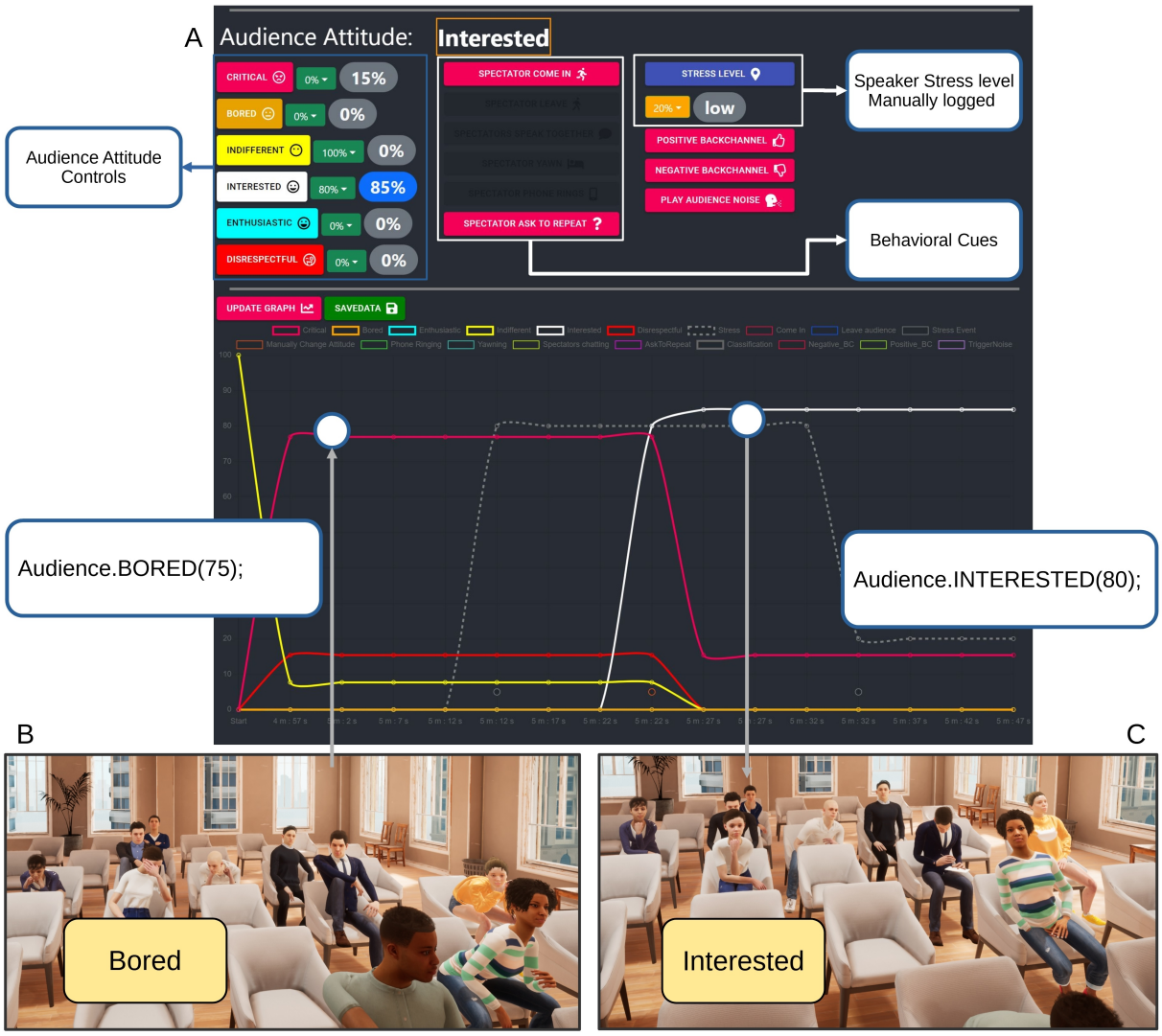


Figure 5.11 – Example of manual change of the virtual audience attitude: with (a) the initially bored audience, (b) the instructor web interface with the controls, and (c) the resulting interested audience obtained with a high-level instruction.

haviour cues. In doing so, the audience's attitude can be timed or conditioned to events, user metrics, or even external tools which can communicate through the REST API, such as physiological sensors.

Instructors have tested physiological-driven scenarios in which the virtual audience's attitude changes according to the student's data obtained through an *Empatica E4* [4] wristband [Empatica, 2022]. However, we never used this pedagogical scenario during the seminar to avoid unfairness between students since none of these scenarios has been evaluated. Moreover, the stress classifier was a prototype as well. For the same reason, all the scenarios were linear during the seminar and were not branching to create alternative ones, so all the students had the same pedagogical scenario. Nonetheless, such features could provide significant pedagogical help in terms of stress monitoring. Some students might suffer from fear of public speaking and could benefit from training sessions taking into account their stress where the audience can change its attitude to lower the students' stress. Moreover, branching scenarios could lead to adaptive narratives and personalised sessions. We could already use such narratives with the control API we provide by conditioning the audience behaviour on user behaviours or physiological data (Snippet 5.1).

```
OnStartNarrative(){
    this.Audience.Interested(60);
    this.Audience.Enthusiastic(40);
}

NarrativeLoop(float deltatime){
   if(this.Intructor.SpeakerEstimatedStressLevel>75){
   {
        // try to calm down speaker if too stressed
        this.Audience.Interested(60);
        this.Audience.Enthusiastic(40);
   }else{
        if(this.Speaker.TimeLookingAtSlides>50){
            this.Audience.Bored(60);
            this.Audience.Indifferent(30);
        }
```

4. Empatica E4 wristband technical page, https://www.empatica.com/en-eu/research/e4/ Accessed February 12, 2022

```
17          else {
18              this.Audience.Interested(70);
19              this.Audience.Indifferent(30);
20          }
21  }
22
23  void OnEndNarrative(){
24      if(this.Speacker.TimeLookingAtSlides>50) {
25          this.Audience.Applaude(90);
26      }
27      else {
28          this.Audience.Leave(90); // delay in seconds
29      }
30  }
```

Listing 5.1 – Example of a simple training plan using the Virtual Audience Control API in Javascript for training on maintaining the visual contact with the audience.

Once more, to ease the use of the STAGE, the virtual audience affective modulation automatically begins when the student press the *Start* button on the virtual laptop menu. Thus, it helps the instructors to focus on the starting presentation without over-monitoring the training settings.

5.6.3 Visualisation Module

Despite being autonomous, the scenarios and the audience's affective modulations still need to be monitored by the instructor to be adapted to the presentation. To do so, we used the control API to log each attitude and behavioural cue change in the ongoing scenario and draw it on a continuously updated graph. This graph draws the attitudes with different curves based on the percentage of affected agents. The visual representation illustrates the behavioural cues or events by coloured circles (Figure 5.11). Finally, a CSV file saves these data for further analysis or replay purposes.

Instructors could use such training data for post-training briefing or replay the presentation with the students to provide a formative evaluation. Moreover, it could feed a performance analysis system supplied with the simulation data, the user metrics or physiological data to provide a qualitative report about the presentation. For example,

students could see when they were stressed and on what slide they were at that time, but they could also know how much time they spent per slide or when they needed to look at their notes.

5.7 Preliminary User Study

In this section, we provide feedback from the 16 students who participated in the seminars gathered from questionnaires and semi-structured interviews (see appendix F.2 for the consent form used and appendix G for the questionnaire). We exclude the PhD students who participated in the iterative tests. We also report a review of the STAGE made by the lecturers in charge of the seminar and three researchers in computer science.

5.7.1 Methods

During the two seminars, we had the opportunity to gather the students' feedback about different aspects of the STAGE, namely, the system acceptance, its usability and the virtual audience believability.

The first seminar was exploratory. Thus, we mainly focused on semi-structured interviews since we needed specific details that can be harder to get with a questionnaire. For instance, the slide interactions system or the simulation controls with the instructors were brand new and required multiple iterations. The second seminar focused on the system's usability and acceptance evaluation rather than its development. Thus, we put the main items from the first set of interviews into questionnaires. Hence, we added questionnaires focusing on the acceptance and usability of the system. On top of these questionnaires, we added a public speaking anxiety scale (see appendix H for the PSAS questionnaire's items) [Bartholomay and Houlihan, 2016] to measure the students' public speaking anxiety (PSA) and compare it to their self-estimated stress and performances. In both seminars, students had a short training in VR to get used to the controls and the virtual environment, then they had the presentation, followed by the debriefing with the instructor, and finally, they had the questionnaires and a semi-structured interview. We chose not to give any questionnaires before the presentation to let the students keep their focus on the training since it was part of a lecture and not only a user study.

Regarding the workshop, all participants had the opportunity to play both roles, i.e. the student role in VR and the instructor role. They were first doing a presentation of

their own, for instance, a lecture or a scientific presentation, and then evaluating their colleague's presentation. They provided feedback on all the different aspects of the system. The discussion following the trial was two-part, with a conversation between the participants and smaller guided discussions based on various aspects we wanted to explore, such as the controls usability, the virtual audience behaviour believability, or the system acceptance for further uses in other seminars.

5.7.2 Results

Regarding these three lines of research, we got promising results. All students agreed in our questionnaires and interview that the STAGE could help improve their presentation skills and also agreed on the usefulness of such a VR training system at the university, *e.g. "[it could help] to get more confident with the presentation itself"*, *"I noticed where I had problems in finding words"*, *"Especially in times of Covid, it makes practising easy"*. Regarding the feeling of engagement during the presentation the results are mixed and 50% of them still believe a real audience is much more engaging for a training session. However, they all agreed that using the STAGE for a practise session is "funnier" compare to what they usually do. In fact, 50% of the students declared practising their presentation alone, the others prepare notes or ask other students to help them. Some comments highlight the reason why students agreed on it and it is probably due to the narratives instructors designed, *e.g. "I feel like it can be helpful to practise with distraction sounds, although I was very surprised when I first noticed"*, *"It felt almost like a real experience and it helped me a lot during the presentation because I could notice how the audience was behaving and I could adapt a bit the way of presenting."*. Concerning the system usability, a frequent comment is a difficulty to read some figures or slides on the small laptop's screen especially when the colour contrast is weak.

For the PSA scores we obtained with a questionnaire, we were able to first identify students who stressed about their presentation and who might also suffer from PSA while we were interested to know if there was a possible correlation. These preliminary results seem to show a correlation between the public speaking anxiety score and the reported stress during the presentation. We ran a correlation test on the students' PSA and the self-estimated stress from the second seminar, but we removed the students who had issues with their slides or with the VR application which might have induced some additional stress, e.g. video not playing in the slides or tracking issues that implied a restart of the VR device. Since we have a small sample with ties and a distribution which does

not follow a normal one we used Kendall's Tau correlation test and adjusted the p-value when ties occurred. Thus, the PSA and the self-estimated stress seem positively correlated (Kendall's $\tau_c = 0.796, p - value = 0.048$). However, these results are preliminary and only include 8 students from the second seminar. Moreover, it is worth mentioning that none of the students declared being stressed but at least stressed "as normal" which corresponds to the middle value in the provided scale. Moreover, it does not seem that the self-estimated performance is linked to the anxiety level measured with the questionnaire.

Finally, regarding the virtual environment, all students agreed it was a believable environment. Concerning the virtual audience, the behaviours and the virtual agents' reactions were considered believable for a conference, *e.g. "In comparison to a real audience at a conference or a similar event the virtual audience was probably very realistic"*, *"Looked a little bored at the end, I think this could also be in reality"*. As for the impact on the presentation, 70% stated the audience behaviour impacted their own behaviour, *e.g. "I felt shortly distracted when a phone in the audience rang"*, *"I looked more towards the audience and pointed out details."*, *"It made me feel a bit unsure about how my presentation was going when people were leaving the room. A ringing phone also made me lose focus for a bit."*. However, only 50% declared adapting their presentation to the virtual audience, *e.g "I tried to refer to them directly for example as "all of you", which I probably would not have done if I was talking to just one person"*. Eventually, almost all students could recognise the audience attitude displayed and remember after the presentation when a specific attitude was displayed according to our questionnaires. They all remembered that the audience started interested and then became bored, only one student did not remember any specific attitudes.

We believe it is worth mentioning that a student got a very high score of public speaking anxiety (75/85) and stated not having paid attention to the virtual audience at all. Moreover, the student declared being stressed by the fact that real persons were listening to the presentation, i.e. the instructor. This student also stated to be disturbed by all the noises coming from the virtual audience. Knowing that the PSA seems to be considered a subgroup of social anxiety disorders in the literature [Blöte et al., 2009], it might be interesting when using such systems to detect students who might suffer from it. Adapting the scenario to them and thus providing a less stressful training session could be a solution, either with specific scenarios or with dynamic ones adjusting the virtual audience's attitude while measuring the anxiety with physiological sensors. Yet, such a hypothesis would need further investigation.

5.7.3 Virtual Audience Believability

The comments from the workshop's participants regarding the virtual audience believability seem similar to the students: the different attitudes are noticeable, and the different behavioural cues are even more noticeable. However, the virtual audience needs a better audio system with more sounds. It seems to have a lack of *"ambient noise"*. For instance, when a virtual agent changes its posture, its chair should sometimes creak. The room in which the seminar was running might also play a role. The experimentation room produces an echo when the students talk whilst the computer's fans were covering the sounds from the environment played out by the HMD's speakers. Students reported this issue as well, despite the sounds being spatialised, *"One noise, I could not identify what it was supposed to be. The noise being directly in your ear makes it seem a bit unrealistic"*. A solution for this would be to use headphones that do not cover the student's voice or speakers in the seminar room which would play the sound coming from the virtual audience.

A proposition from a lecturer was to improve the scenarios with agents displaying a certain *"personality"*, meaning that instead of letting the model freely change the virtual agent's behaviour, it would take into account its past behaviours. The virtual agents could avoid displaying an opposite attitude or only display specific behaviour, e.g. an agent with a disrespectful attitude would not suddenly become interested.

Concerning the feeling of social presence, participants from the workshop proposed to add some VR interaction with a human embodying an avatar before the beginning of the presentation. For instance, the training session in VR could be held with the student and instructors embodying their avatars, who explain the controls and directly show how to use them in the virtual environment. Such rich interaction between co-located agents and embodied avatars seems to increase the feeling of co-presence and the possibility of interaction with the virtual environment [Latoschik et al., 2019].

5.7.4 Scenario Controller

As for the control interface, the visualisation graph and the user metrics logs seem to be of great interest when looking at the students' performances afterwards or using it as a replay tool. Hence, instructors could cross the narrative and the metrics to provide even more personalised feedback. Such metrics visualisation would also be a first step for the system to be used alone by the students without the instructors. For instance, it would

allow the students to watch their presentation with a quantitative assessment of their performance, like the time spent per slide and how long they looked at their notes.

Nonetheless, there are some areas of improvement in terms of usability: reading the graph while trying to stay focused on the presentation is too complicated as well as reading the current percentage of agents displaying a specific attitude, e.g. a pie chart might have been more suitable to read the audience attitude. Also, we automated almost all the GUI so the instructors could focus on the presentation and not on starting the scenario or manually changing the attitude.

Concerning the scenarios, the high-level API seems promising. It provides high-level instructions to design simple pedagogical scenarios by modulating the virtual audience's affective cues. Still, a graphical representation of the state machine would ease the design of scenarios, at least to see the following state and the transition. Moreover, instructors could use this scenario controller to author the virtual audience without being bound to a specific system with high-level controls and direct behaviour changes, provided they have some knowledge of computer science.

5.7.5 Integration in University Curriculum

Participant all agreed on the potential the STAGE represent for an ecological environment for a formative evaluation. Such VR training systems like the Breaking Bad Behaviour system are used to practise classroom management skills through successive training sessions either by using the system or by watching peers practising [Lugrin et al., 2016]. Lecturer participants in the workshop recommended letting the students practise more than once in VR. For instance, according to the lecturers, students who do not remember their slides keep reading them and look less at the audience. Thus, having multiple training sessions with a specific focus could improve the training process, e.g. a first session could just be dedicated to the slides without VR, while the following could use the STAGE to focus on public speaking skills. In addition to repeated training sessions, peer review sessions where students help each other improve their presentations might help.

Instructors could also use the STAGE during hybrid sessions in which other students could join the presentation and embody a virtual spectator. This feature already exists in the STAGE but would need further controls allowing the spectators to have partial control over the avatar behaviour or at least to participate in the overall audience attitude, similar to online conferences in which attendees can use emojis to interact or share their mood. Such features echo the aforementioned recommendation for adding social interactions with

humans to increase the feeling of co-presence.

5.8 Deployment in a Therapeutic Practise

Following our user study at the University of Würzburg, we had the opportunity to collaborate with a speech therapy practise in the city of Augsburg. We were able to offer the STAGE as a therapeutic tool for public speaking exercises in virtual reality. The therapists from Logopädie-Plus wanted to conduct a user study with patients suffering from fear of public speaking to test the effectiveness of virtual reality audience simulations for therapy exercises. Thus, after testing the application to grasp its potential, the therapists provided us with a typical therapy scenario in the form of a list of events with a timeline. Then we implemented it in the Scenario-Controller so that our tool could trigger all the exercise's steps while respecting the given timeline. Particular attention was given to the design of the virtual environment so that patients would start the session in a replica of the practise (Figure 5.12). Patients began the sessions by receiving instructions on the study and the use of the headset and then moved into a booth to be isolated when they started the simulation, they were in a virtual replica of the booth to acclimatise to the tool.

Figure 5.12 – Silent booth from the Logopedic-Plus practise (A) and the STAGE's virtual environment (B).

For this study, the scenario's timeline was seven minutes long. It begins with an interested audience which progressively becomes disrespectful and bored until all spectators talk to each other without giving any attention to the speaker, i.e. the patient who is practising. The scenario is scattered with disruptive events to disturb the patient, e.g. a

spectator comes in late, another leaves, a phone rings, and at the scenario's end a group of agents stop next to the virtual door's room and speak loudly.

This collaboration allows us to test the acceptance and usability of our application for expert users and patients. Since we embedded the entire starting process in a batch file and automated the scenario start, therapists had no difficulties using the STAGE when conducting an exercise. If the starting process was straightforward, therapists reported having issues with the hardware twice, which seems to be a critical limitation since it takes a lot of time to start over. A crucial advantage for the therapists was that we especially designed the scenarios for them, thus easy to use and suit their needs. From a speech therapist's perspective, the STAGE is a suitable element of therapy for the future. However, such an application must be stable. Moreover, if therapists could adapt and design scenarios themselves to better adapt the session to their patients, they would use it spontaneously and individually.

5.9 Conclusion and Guidelines

This chapter describes how we used user-centred developments for the STAGE, a scientific presentation VR training application. Then we introduced the scenarios driving the system relying on users' affects aroused by the virtual audience's attitude. The scenario controller provides a high-level API for controlling the audience's attitude and the behavioural cues required to design such scenarios. Thus, we insisted on this method to manage the application because it could partially solve the compromise to find between a fully autonomous system, where instructors cannot adapt their scenario during the training session, and a Wizard-of-Oz system in which the instructors have to manually author each virtual agent.

The results from the preliminary user study we ran during two seminars seem promising. All participants agreed on the potential pedagogical interest the STAGE has for university seminars and concurred on the audience behaviour believability. Yet, the system may benefit from more sounds and audio feedback from the VA to improve the users' feeling of immersion.

Concerning the STAGE's control interface, the workshop we held with professors and lecturers from the university highlighted possible improvements. The visualisation tool can improve and ease the monitoring of the training, whilst the current visualisation tool already has some value for post-training feedback and the high-level controls it provides.

The current audience's attitude has to be easy to read, and the interface should only display relevant information. All the same for the current scenario's state, which is hidden in the main graph, whilst a simple state machine graph could solve this issue.

The seminar could provide repeated training sessions and become a hybrid system for formative evaluation where other students could join the session to embody virtual spectators and participate in the presentation with non-verbal behaviour controls for the embodied avatars. Such modification could lead to peer review training sessions where students assist each other on the condition the STAGE provides a qualitative data visualisation tool from the user metrics in case an instructor would not attend the presentation.

The STAGE could now profit from a longitudinal study regarding the learning outcomes it provides. Previous studies underlined the need for further research regarding what contributes to the success of VR public speaking training systems [Poeschl, 2017], even though recent studies show good user acceptance of such training systems [Palmas et al., 2019]. The new AMBIANCE model should also be evaluated in terms of perceived attitudes even though the instructors validated the audience behaviours and students seem able to recognise the current attitude. Otherwise, the resulting attitude might be biased, and the students' and instructors' perceptions might differ. Though the pedagogical scenarios seem to impact students, it would be interesting to further test adaptive narratives based on user metrics or physiological data. Such interactions may better suit students suffering from PSA by adjusting the audience's attitude to elicit a positive affect and decrease the anxiety induced by the virtual audience.

With these preliminary results in mind, we can provide guidelines regarding the control of virtual audiences in the context of a VR training system.

Firstly, audience behaviour controls have an essential role to play in the system usability for the instructors. Use high-level behaviour controls along with an evaluated behaviour model to guarantee the users would distinguish the displayed behaviour and avoid the instructors focusing on editing the virtual agents' behaviour.

Secondly, we identify the scenarios as a great tool to improve the training sessions. The design of plausible storytelling is essential to root the users in the training context. Like in role-playing games, instructors can author the ongoing narrative. In our case, it can even be based on users' metrics to provide specific interest exercises.

Thirdly instructors should use such training systems multiple times to let users get used to it and the VR devices. Then they can let trainees face exercises focusing on specific skills. In doing so, the trainee can improve from one session to another, similar to

therapeutic systems.

The following chapter 6 discusses the aforementioned limitations and potential improvements for the STAGE and the AMBIANCE. Finally, we will develop our ongoing work regarding how we use these two applications in other application domains and how we plan to push further the design of hybrid authoring tools for the STAGE.

Chapter 6

Conclusion

6.1 Summary of Contributions

The goal of this research was to propose a new virtual audience behaviour model evaluated for VR applications. Existing models and applications provide different approaches to creating virtual audiences and evaluating the user perception, but they barely evaluate it directly in VR, e.g. they rely on experts' knowledge instead or emotion models. In our case, we based our model on two-dimensional behaviour models which rely on valence and arousal. These two models from Chollet and Scherer, 2017 and Kang et al., 2016 provide various non-verbal behaviours to create multiple types of audiences. From these works, we developed the AMBIANCE model, which also relies on the valence and the arousal dimensions. In order to confirm the model's ability to generate different audience attitudes, we first implemented the model in a popular game engine plugin that we benchmarked to validate the system's performances, and then we ran perception studies in VR.

At first, we studied the relationship between non-verbal behaviours and the model's dimensions. Our strategy was to study behaviours associated with the model's dimensions by construction. Thus, we asked participants to build an agent's behaviour according to a given pair of valence and arousal by associating non-verbal behaviours in VR. The results confirmed our hypotheses and partially reproduce those from Chollet et al. regarding the non-verbal behaviours' relationship with opinion and the frequency with engagement, e.g. head nod is significantly perceived as a positive behaviour, whilst the more an agent gazes away, the more likely it will be perceived as disengaged. Afterwards, we evaluated the user perception of the audiences in VR by exposing the participant to audiences with different attitudes. This perception study allowed us to determine whether VR users recognise the different attitudes. The study's results confirmed the AMBIANCE's ability to generate multiple attitudes, such as enthusiastic, indifferent or critical.

Once the model was evaluated, we integrated AMBIANCE into a new VR application for the students from the University of Wurzburg to prepare their scientific presentations.

The application called STAGE provides a virtual environment where students can present their slides in front of a virtual audience controlled by the model. The STAGE was the opportunity to evaluate the audience behaviour model in an academic context and testify to its potential usage in VR training applications. We successfully used the STAGE for two semesters while we iteratively developed and improved the system to fit the users' needs. During this iterative process, we assessed the system's usability and acceptance, as well as the students' PSA induced by the virtual audiences. We also developed a GUI for the lecturers to control the audience's behaviour and the simulation's scenarios through high-level controls, which does not require expertise in programming. Similar applications providing controls seem to struggle with it, since a too-detailed model can overwhelm the instructor and hinder the session supervision.

Finally, after using the STAGE for a year, we gathered valuable feedback from VR users and instructors regarding the system's acceptance, usability and potential improvements. For instance, the model would benefit from domain-specific behaviours to reproduce real-life situations. Moreover, the control interface seems useful for the instructors but needs to be simple and would need an editing tool for the scenarios, e.g. a graphical editor with a library of nodes could be used to create timelines without programming knowledge. Additionally, we collaborated with therapists and ran a preliminary study which confirmed the STAGE potential for therapeutic uses.

6.2 Future Works And Open Research Questions

The following paragraphs describe the future works for the AMBIANCE and the STAGE. We also present our new lines of research regarding our work on audience behaviour modelling and high-level scenario controls development.

6.2.1 AMBIANCE model's Responsiveness and Believability

In our last study (Chapter 5), we used the AMBIANCE model with high-level controls and custom rules to support training scenarios, but the simulation was mainly relying on scripted scenarios to lower the instructors' workload and avoid unfairness between students. The controls allowed the lecturers to directly adapt user-agent interactions to the simulation, but in doing so, they reported that they sometimes lost focus on the student's presentation.

Therefore, a series of developments are planned for the model to become responsive. Adding new authoring features to the manual controls to improve the model's autonomy would ease the audience supervision. Using physiological and environmental data could enhance the model to display more adaptive behaviours and spontaneous reactions to the user's actions. Some public speaking training systems already integrate behavioural metrics and physiological data in their behaviour model to drive the virtual audience accordingly [Chollet et al., 2022; Palmas et al., 2021]. In these works authors used such data to evaluate the user's presentation quality, e.g. the system can collect different behavioural metrics such as the user's gaze direction, voice or position.

The audience should be more believable and responsive to the VR user's actions if we integrate this new kind of autonomous reaction into the model. Ultimately, it should reduce the supervisor's workload. For instance, disruptive events from the existing scenarios could trigger autonomous reactions in the audience according to specific criteria, such as the current audience's attitude, agents' personalities or the current scenario's state. Therefore, It would be crucial to consult application domain experts to establish these new evaluation criteria. This approach is very close to appraisal theories, which rely on such criteria to evaluate events and elicit emotions accordingly. Consequently, a new series of perception studies would be necessary to validate the new behaviours and reactions.

6.2.2 STAGE Instructors' Scenarios and Controls

The STAGE has been proven usable for research purposes and can be the groundwork for further investigation regarding audience behaviour models, innovative pedagogical tools and simulation controls. Therefore, we are currently working on an open-source version of the model and a free demo version of the STAGE. However, the application requires a few modifications to meet expectations, such as those identified during our user studies. For instance, improving the scenario controller to ease the creation of new ones by adding a graphical tool and providing a collection of actions for the supervisors to create their scenario timelines are critical features to becoming a commercial tool.

Besides its potential to become a commercial application, the STAGE's virtual audience is easy to modify thanks to the model parameters and the number of features it already provides, e.g. multi-user, VR interactions, eye-tracking and a simple control API. Therefore, we are able to study various assumptions on audience user perception, such as gender bias, gaze avoidance, and environmental influence.

Additionally, we have the opportunity to evaluate the STAGE efficiency as a training or

therapeutic application. During our work with the bachelor students from the University of Würzburg, we reported a significant PSA for a majority of students. Thus, we are able to evaluate the STAGE's ability to lower this anxiety level through repeated practise sessions as well as the students' presentation skills. Such a study would require consecutive training sessions and a long-term follow-up to detect a significant reduction in PSA in the participating students. For instance, Premkumar et al., 2021 reported a significant decrease in PSA with two VR exposure sessions with a one-month follow-up with their VR public speaking application. In a similar fashion, our collaboration with therapists is the occasion to study if the STAGE is a suitable platform for professionals by evaluating the application with patients.

Nonetheless, designing social skill training or therapeutic systems, including interactive virtual agents, presents several challenges. An important challenge the STAGE partially addresses is the control of a virtual audience to follow a training plan whilst allowing it to react to the user's behaviours and interactions. Even if the STAGE can simulate an audience without a human-in-the-loop, it still requires someone to script the timelines and scenarios. Interactive storytelling methods are good candidates to mitigate this issue. For this reason, we plan to investigate how these techniques can benefit the STAGE to obtain more realistic and variable scenarios and audiences.

Search-based algorithms can be used to search the space of possible behaviours and reactions to be applied to the current event. For example, in Lugrin and Cavazza, 2006, the *Death Kitchen* application relies on an ontology of possible actions for objects. In our case, the ontology could be based on the audience's attitude and the agents' current behaviour. The attitude can be represented by different dimensions like the levels of Valence and Arousal and the behaviours by different states. Such an ontology could condition the triggered interactions or reactions, e.g. is holding an object, is standing, or gazing toward the user. Alternatively, providing causally coherent narrative experiences can be issued by logic and rule-based perspectives [Bosser et al., 2010; Martens et al., 2013] or more classical plan-based perspectives [Cavazza et al., 2002; Young, 1999]. Existing rule-based systems like *Ceptre* [Martens, 2015] or *Celf* [Schack-Nielsen and Schürmann, 2008] can be used for the emergence of system behaviours in interactive storytelling. For instance, linear logic seems promising to design interactive narrative scenarios due to the benefit it provides regarding the findings of deadlocks and flows in the scenario when accompanied by a proof tool [Dang et al., 2011]. As such, various systems have provided interactive experiences where virtual agents are controlled by a narrative engine seeking to balance

authorial intent, the impact of user intervention, and narrative coherence. Previous work has, for instance, investigated how causally coherent stories can unfold from virtual agent interactions [Cavazza et al., 2002], how to control the unfolding of the story based on high-level narrative goals [Cheong and Young, 2014, Lindsay et al., 2017,Porteous et al., 2010], means to take into account user intervention, including affective input [Gilroy et al., 2013], or how to patch up the narratives in systems where user intervention may break the causal coherence [Riedl and Stern, 2006]. Another benefit of using an interactive storytelling approach for educational systems lies in the fact that stories have the ability to influence attitudes and behavioural intentions of people [Dettori and Paiva, 2009].

6.3 Concluding Remarks

This thesis has extensively investigated the issues related to the simulation of audiences in virtual reality (VR) and validated the audience behaviour model developed during this research work through multiple evaluations and benchmarks. The behaviour model development process led us across all aspects of audience simulation, from the animation pipeline, the performances in virtual reality, the attitude generation, the user perception, and the audience controls for instructors to finally end with the model deployment in a professional application. This research explores the limitations reported by previous studies to provide a novel method for the design of virtual audiences in virtual reality. Thus, we conducted two perception studies which showed the model's ability to display various attitudes in VR. Then, we conducted a series of experiments with lecturers and bachelor students to evaluate the system's controls, usability and acceptance, leading to the development of our scenario controller for instructors and model improvements regarding the instructions to create context-specific rules. Finally, we concluded from our research that virtual audience behaviour models should be directly evaluated in VR because user perception seems to be slightly different compared to less immersive systems and that hybrid applications blending high-level controls and autonomous features seem to be a good compromise for instructors.

Appendix A

Publications

This research project resulted in 6 peer-reviewed publications from journals and conferences. The exhaustive list below contains two journals in Frontiers in VR, a full paper at a national conference (Mensch und computer) and three posters or short papers. A Late-Breaking Work at CHI (2019) depicts the initial project idea. The model's development and the VR benchmark have been presented to VRST in a first poster, which supported a paper published at a national conference (MUC 2020) summarising our development process and the model's operating principle. The perception studies led to a publication in the Frontiers in VR journal (2021). Finally, we published a short paper and a journal regarding the STAGE application, its implementation and evaluation. The short paper (VRST) recounts the first semester we used the STAGE, whilst the journal paper summarises the entire year of development, the method used, and the evaluations from the students and lecturers who used the VR side and the instructor side with its control application. This last publication highlights the benefit of using high-level controls to author the audience's behaviour and thus generate behaviours that support a pedagogical plan.

1. Glemarec Y, Lugrin J-L, Bosser A-G, Buche C and Latoschik ME(2022) Controlling the Stage: A High-Level Control System for Virtual Audiences in Virtual Reality. Front. Virtual Real. 3:876433. doi: 10.3389/frvir.2022.876433
2. Glemarec, Y., Lugrin, J. L., Bosser, A. G., Buche, C., & Latoschik, M. E. (2021, December). Conference Talk Training With a Virtual Audience System. In Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (pp. 1-3).
3. Glemarec Y, Lugrin J-L, Bosser A-G, Collins Jackson A, Buche C and Latoschik ME (2021) Indifferent or Enthusiastic ? Virtual Audiences Animation and Perception in Virtual Reality. Front. Virtual Real. 2:666232. doi:10.3389/frvir.2021.666232
4. Glemarec, Y., Lugrin, J. L., Bosser, A. G., Cagniat, P., Buche, C., & Latoschik, M.

(2020, September). Pushing Out the Classroom Walls: A Scalability Benchmark for a Virtual Audience Behaviour Model in Virtual Reality. In Mensch und Computer.

5. Glemarec, Y., Bosser, A. G., Buche, C., Lugrin, J. L., Landeck, M., Latoschik, M. E., & Chollet, M. (2019, November). A Scalability Benchmark for a Virtual Audience Perception Model in Virtual Reality. In 25th ACM Symposium on Virtual Reality Software and Technology (pp. 1-1).
6. Lugrin, J. L., Bosser, A. G., Latoschik, M. E., Chollet, M., Glemarec, Y., & Lugrin, B. (2019, May). Towards narrative-driven atmosphere for virtual classrooms. In Extended abstracts of the 2019 CHI conference on human factors in computing systems (pp. 1-6).

Open-source model : in addition to this research work, we also published the AMBIANCE model as an open-source plugin compatible with Unreal Engine. The plugin supports the blueprint system provided by the engine, which is a graphical programming language. This tool allows users unfamiliar with programming to use all the audience generation features we provide in the AMBIANCE. Moreover, the plugin contains various demos, one for each model feature and multiple blueprint examples for setting up a virtual agent with the plugin and its animations using the Unreal Engine retargeting system. We also plan to make the STAGE system available as a free-to-download binary application containing the VR student presentation application and the instructors' scenario controller.

To obtain access to the sources, you can ask one of the project's members to grant you access to the repository. When registered, you can receive a dedicated fork of the project on the HCI Group's GitLab. Visit https://hci.uni-wuerzburg.de/projects/virtual-audiences/ to contact the project team.

Appendix B

Acronyms

— **AMBIANCE** Attitude Model defining the Behaviour of Individual AgeNts for Constructing audienCEs

API Application Programming Interface

ATMCS Aktual Trend in Mensch und Computer

CHI Conference on Human Factors in Computing Systems

COMMEDIA COgnition, Models and Machines for Engaging Digital Interactive Applications

CPU Central Processing Unit

CSV Comma-Separated values

DV Dependent Variable

ECA Embodied Conversational Agent

ENIB École Nationale d'Ingénieurs de Brest

FPS Frame Per Second

GPU Graphics Processing Unit

GUI Graphical User Interface

HCI Human Computer Interaction

HMD Head Mounted Display

IPIP-NEO International Personality Item Pool - Neuroticism Extraversion Openness to experience

IV Independent Variable

IVA Intelligent Virtual Agent

MR Mixed Reality

PSA Public Speaking Anxiety

PSAS	Public Speaking Anxiety Scale
SAM	Self-Assessment Manikin
SIP	Semantic Interpretation Principal
STAGE	Speaking To an Audience in a digGital Environment
UE4	Unreal Engine 4
VR	Virtual Reality
WOZ	Wizard of OZ

Appendix C

Blueprints

The AMBIANCE plugin provides all the features as Unreal Engine's Blueprints. The engine introduces the blueprints in its documentation as such: *"The Blueprint Visual Scripting system in Unreal Engine is a complete gameplay scripting system based on the concept of using a node-based interface to create gameplay elements from within Unreal Editor. As with many common scripting languages, it is used to define object-oriented classes or objects in the engine."* [Epic Games, 2022a]. Figure C.1 presents a few methods and parameters from the AMBIANCE manager the user can call, overload or override with blueprints without modifying the plugin's scripts. For instance, the user can access the list of character instances compatible with the model, and change their behaviour or trigger a specific reaction to an event.

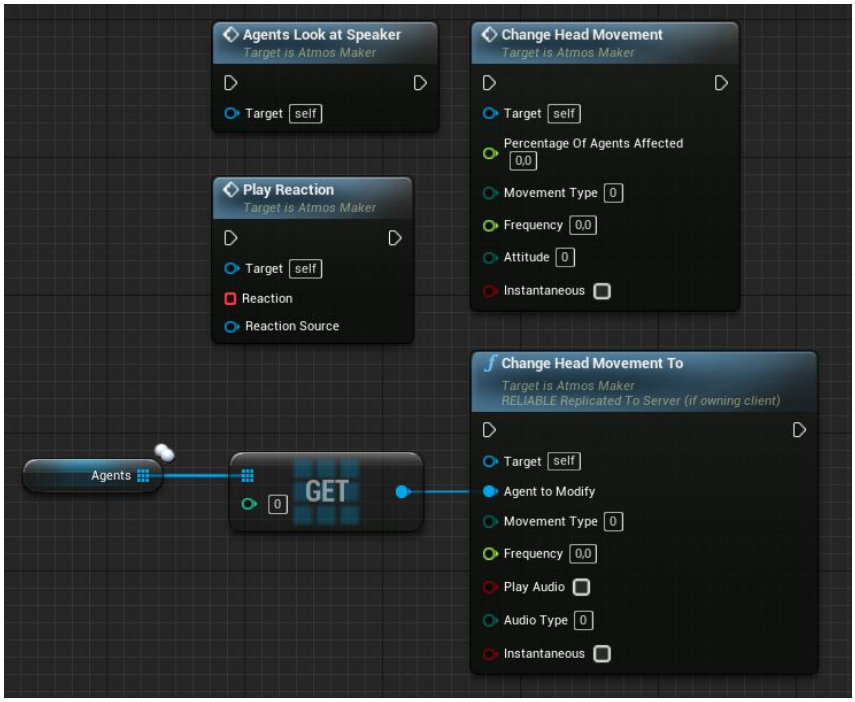


Figure C.1 – Example of blueprint methods accessible from the plugin's manager.

Appendix D

UML Diagrams

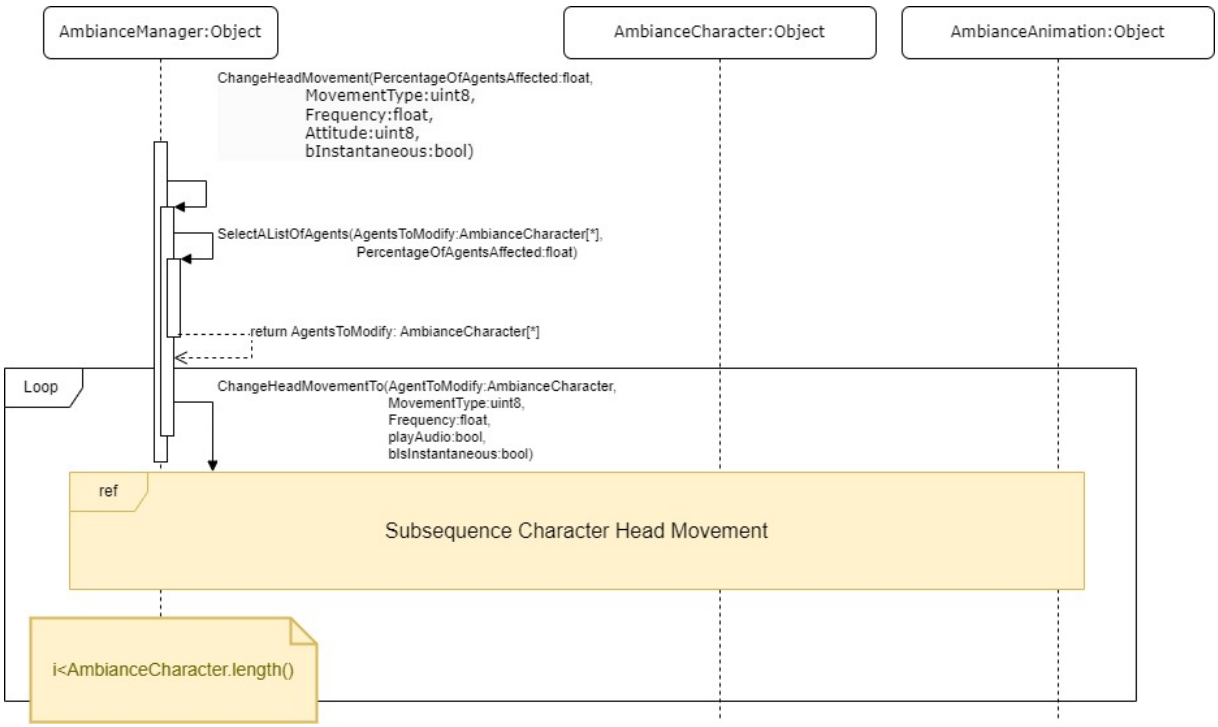


Figure D.1 – Sequence Diagram : AMBIANCE manager changing the audience's head movements.

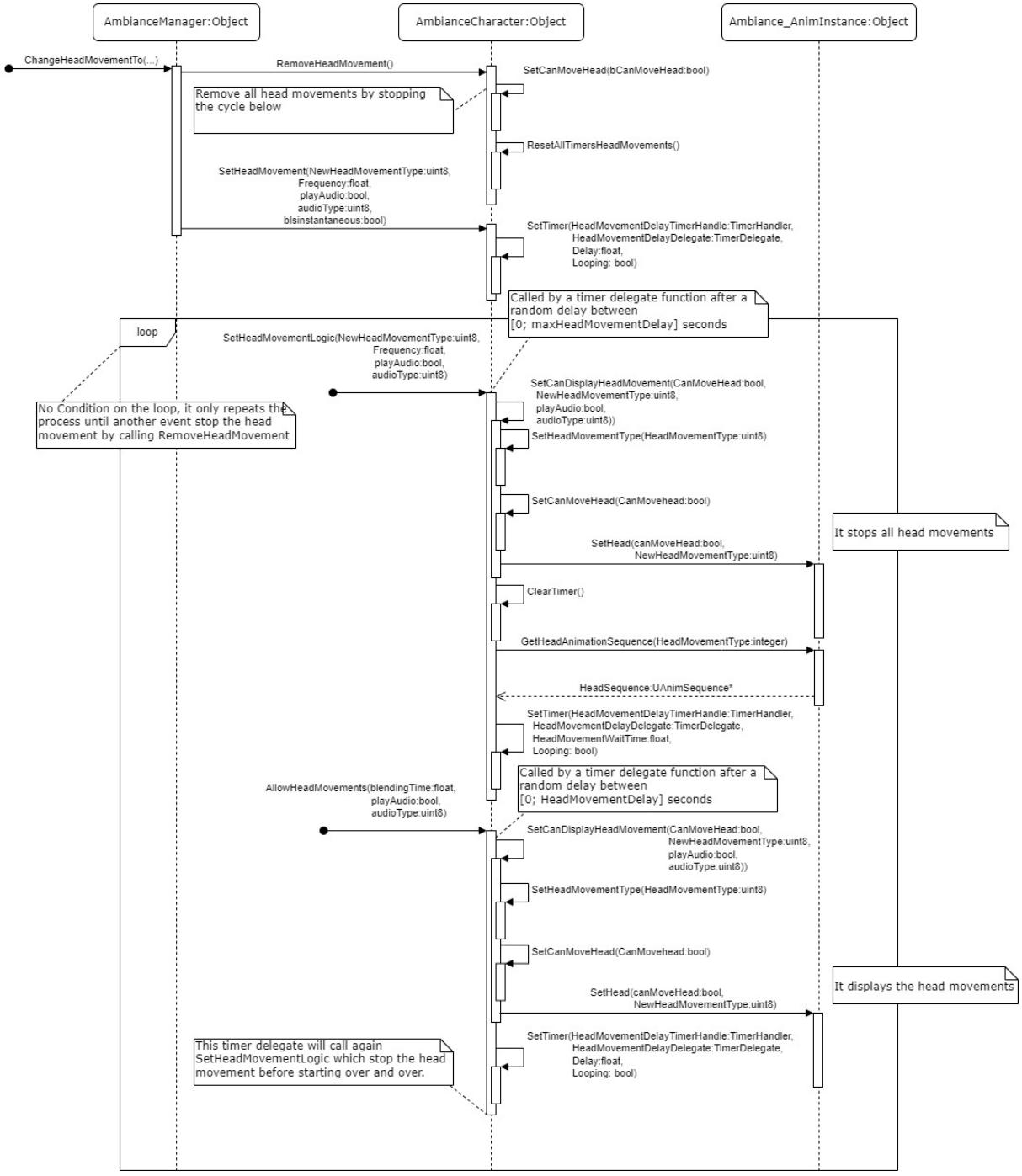


Figure D.2 – Sequence Diagram : Subsequence Character Head Movement logic

Appendix E

Seminar Marking

E.1 Marking that was used to design the system

ATMCS Seminar Marking	WS 22 - 17.02.23
Student Evaluated	
Primary Evaluator	Dr. Jean-luc Lugrin
Secondary Evaluator	

TALK EVALUATION						
	Qualities	**Details**	**Done** (1 = Yes, 0 = no)	**Comments**	**Points**	**Percentage of Point**
1	**Be On time Presentation**	*On time- too short/long ? (15 min - per min offset -20%) +/- 1 min is ok*	1		-25	0 %
2	**Be On time Review**	*On time- too short/long ? (5 min - per min offset -20%)*	1		-10	0 %
3	**Be Correct**	*Explanation was free from fault/misexplanation - misunderstand the paper?, Correctly answers to questionsCorrectly explain Previous and related work*	1		-10	0 %
4	**Be Kini (Concise)**	*Cover on the essentials Brief in form but comprehensive: all essential parts of research have been discussed - Questions, Methodology, Contributions, Future.*	1		-10	0 %
5	**Be Unambiguous**	*No Ambiguity, Not different interpretations possible*	1		-10	0 %
6	**Be Talking**	*Not reading notes or slides*	1		-10	0 %
7	**Be Simple**	*(Simplification and synthesis) easy to understand research question, motivation, methodology, results, contributions, and future work, Appropriate use of pictures, use analogy, metaphor*	1		-10	0 %
8	**Be Consistent**	*Consistent presentation style and concept definition*	1		-10	0 %
9	**Be Progressive**	*Big picture first, increasing complexity, example*	1		-10	0 %
10	**Be Structured**	*Answer basic audience question (Who, What, How, Why, What the future)*	1		-10	0 %
11	**Be Logical**	*Logical explanation, transitions, no contradiction*	1		-10	0 %
12	**Be Readable**	*slide content: no info overload - no long sentence or tables or complex figure, large font and Free of Spelling/Grammar mistake*	1		-30	0 %
13	**Be Audible**	*Speak clearly and loudly,Energetic and Enthusiastic, Eye Contact, Polite,Be Comprehensible*	1		-10	0 %
14	**Be Energetic**	*Speak clearly and loudly,Energetic and Enthusiastic, Eye Contact, Polite,Be Comprehensible*	1		-10	0 %
15	**Be Pedagogue**	*Easy to understand the essence, explain and re-explain concept, repeat object, method and results*	1		-10	0 %
16	**Be definition and giving concepts**	*explaining all main concepts, clear definition*	1		-10	0 %
17	**Be Concluding**	*Summary of what , how, why and main result + figure workTake aways slide, or recap on what, why, how and main result*	1		-10	0 %
18	**Be Using Minimalist Design**	*Slide should not contain information which is irrelevant or rarely needed (not chart junk, decoration,..)*	1		-10	0 %
19	**Be Explaining types of main contribution**	*Contribution type identified and benefits clearly discussed*	1		-10	0 %
20	**Be Explaining videos**	*Commented the video or why/what has been displayed*	1		-10	0 %
21	**Be Explaining Experiment tasks and measurements**	*The Tasks associated to the results presented have been clearly explain*	1		-10	0 %
22	**Be Explaining Experiment design**	*Between, within, factorial?*	1		-10	0 %
23	**Be Explaining Interaction Technique(s)**	*How to the user will use them*	1		-10	0 %
24	**Be Explaining Interaction Technique(s) Design and implementation**	*How to the developer programmed them*	1		-10	0 %
25	**Be Explaining System**	*How to the Software was build*	1		-10	0 %
26	**Be Starting with nice introduction**	*Take always slide, or recap on what, why, how and main result*	1		-10	0 %
27	**Be Serious**	*No too many visual funny images, not taking presentation seriously, acting casual, no visual artwork unrelated to topic*	1		-10	0 %
28	**Be Answering Audience**	*Correctly answers question*	1		-10	0 %
29	**Be Answering self**	*Do not always ask audience question, speaker should be answering him/herself*	1		-10	0 %
30	**Be Summarising and correctly Citing Previous work**	*previous work discusses? correct citation?*	1		-10	0 %

31	**Be Using the slide space**	*too much of the slides are not in used (large white space)*	1		-10	0 %
32	**Dont use Any Artwork**	*should not use decoration on slide, instead of relevant visual information*	1		-10	0 %
33	**Be Prepared (Be not hesitating)**	*dont hesitate too much, wrong text, slides, order, or missing slide*	1		-10	0 %
34	**Dont be too fast**	*too often to fast on critical information*	1		-10	0 %
35	**Be Finishing The Complete Talk in Time**	*Did not have time to explain all slides*	1		-10	0 %
36	**Dont have long quote or sentence on slide**	*dont have fully sentence, or long quote extracted from paper questionnaire*	1		-10	0 %
37	**Bonus**	*Presentation in English*	0		10	0 %
38	**Other Good qualities**	*Innovative Metaphors, Figures, demos?*				
39	**Other Bad qualities**	*Specific issues? (specifiy negative points)*				

	Mininum Points	**0**
	Maximum Points	**50**
	Total Points to Remove	**0,00**
Talk Mark		**50,00**

REVIEW EVALUATION

	Qualities	Details	Done	Comments	Points	Percentage of Point
	Correct Size	*max 1500 words (about 3 pages) - min 1000 words (about 2 page)*	1		-25	0 %
	Correct Conent	*Synthesis is correct, main contributions types correctly identified and discussed, main reviewer criticism are correct*	1		-25	0 %
	Complete	*Evaluate Originality, Correctness, Complexity, Clarity, Replicability, Relevance and Impact*	1		-20	0 %
	Balanced	*Discuss negative/positive aspects*	1		-10	0 %
	Clear	*Argumentation is clear and makes sense, sentences are understandable*	1		-10	0 %
	Concise	*Focus on main problems, explained them quickly*	1		-10	0 %
	Accurate	*Report clear examples of problems and acts from paper*	1		-10	0 %
	Justified	*Citing previous work or other papers, books to backup criticism or issues found*	1		-10	0 %
	Objective	*With evidences, unbiased, judgment based on paper quality, not on reviewer dont report personal feelings, opinion without any evidence*	1		-10	0 %
	Respectful	*Polite tone& Formal English*	1		-10	0 %
	Constructive	*Propose solutions to improve quality*	1		-10	0 %
	Proofread	*no spelling/grammar mistakes*	1		-10	0 %
	Approved	*Evaluators agrees with review decision*	1		-10	0 %
	Other good aspects					
	Other bad aspects					

	Mininum Points	**0**
	Maximum Points	**50**
	Total Points to Remove	**0**
Review Mark		**50,00**

	Overall Comments and Final Grade				
	Primary Evaluator's Talk Comments ***Indicative content:*** *Best & Worst Parts?* *What was wrong?* *Main Unclear or Missing Parts?* *Interesting and Enjoyable Presentation?* *Main Research Contributions, Motivations and Limitations Understood?*		**Total Mark**		**100**
	Secondary Evaluator's Talk Comments **Indicative content:** *Best & Worst Parts?* *What was wrong?* *Main Unclear or Missing Parts?* *Interesting and Enjoyable Presentation?* *Main Research Contributions, Motivations and Limitations Understood?*		**Final Grade**		**1,0**
	Primary Evaluator's Review Comments **Indicative content:** *Best & Worst Parts?* *What was wrong?* *Main Unclear or Missing Parts?* *Interesting and Enjoyable Review?* *Main Research Contributions & Motivation & Limitations Understood?*				

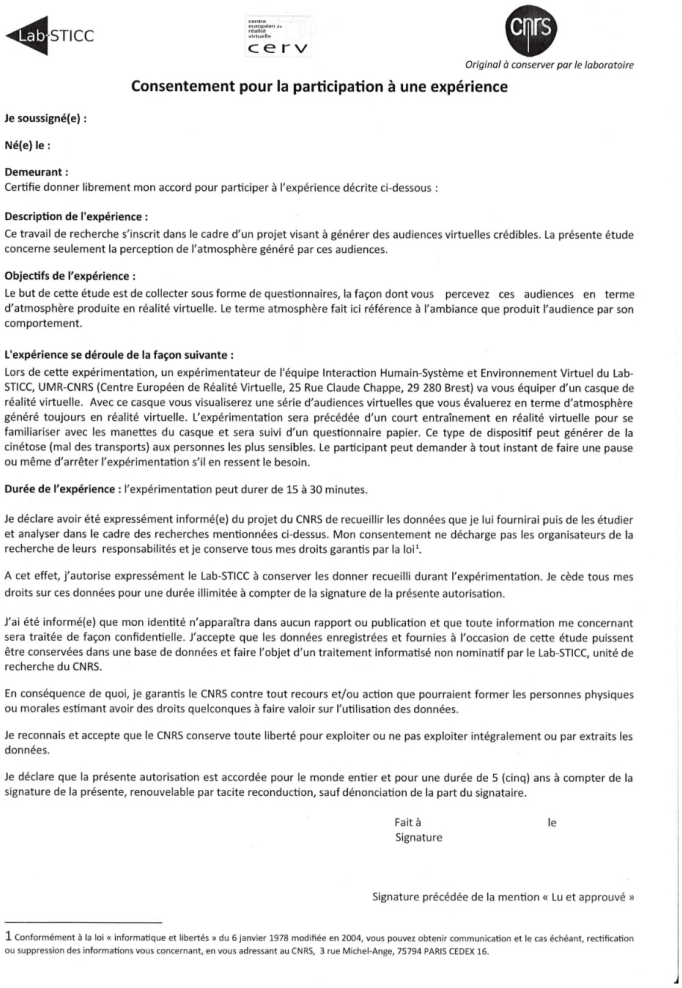
pointgrades

Points	Grade	Verbal Grade
0	5	nicht ausreichend
50	4	ausreichend
55	3,7	ausreichend
60	3,3	befriedigend
65	3	befriedigend
70	2,7	befriedigend
75	2,3	gut
80	2	gut
85	1,7	gut
90	1,3	sehr gut
95	1	sehr gut

Appendix F

Consent Forms

F.1 Consent Form Perception Study



Lab-STICC

cerv

Original à conserver par le laboratoire

Consentement pour la participation à une expérience

Je soussigné(e) :

Né(e) le :

Demeurant :
Certifie donner librement mon accord pour participer à l'expérience décrite ci-dessous :

Description de l'expérience :
Ce travail de recherche s'inscrit dans le cadre d'un projet visant à générer des audiences virtuelles crédibles. La présente étude concerne seulement la perception de l'atmosphère généré par ces audiences.

Objectifs de l'expérience :
Le but de cette étude est de collecter sous forme de questionnaires, la façon dont vous percevez ces audiences en terme d'atmosphère produite en réalité virtuelle. Le terme atmosphère fait ici référence à l'ambiance que produit l'audience par son comportement.

L'expérience se déroule de la façon suivante :
Lors de cette expérimentation, un expérimentateur de l'équipe Interaction Humain-Système et Environnement Virtuel du Lab-STICC, UMR-CNRS (Centre Européen de Réalité Virtuelle, 25 Rue Claude Chappe, 29 280 Brest) va vous équiper d'un casque de réalité virtuelle. Avec ce casque vous visualiserez une série d'audiences virtuelles que vous évaluerez en terme d'atmosphère généré toujours en réalité virtuelle. L'expérimentation sera précédée d'un court entraînement en réalité virtuelle pour se familiariser avec les manettes du casque et sera suivi d'un questionnaire papier. Ce type de dispositif peut générer de la cinétose (mal des transports) aux personnes les plus sensibles. Le participant peut demander à tout instant de faire une pause ou même d'arrêter l'expérimentation s'il en ressent le besoin.

Durée de l'expérience : l'expérimentation peut durer de 15 à 30 minutes.

Je déclare avoir été expressément informé(e) du projet du CNRS de recueillir les données que je lui fournirai puis de les étudier et analyser dans le cadre des recherches mentionnées ci-dessus. Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités et je conserve tous mes droits garantis par la loi[1].

A cet effet, j'autorise expressément le Lab-STICC à conserver les donner recueilli durant l'expérimentation. Je cède tous mes droits sur ces données pour une durée illimitée à compter de la signature de la présente autorisation.

J'ai été informé(e) que mon identité n'apparaîtra dans aucun rapport ou publication et que toute information me concernant sera traitée de façon confidentielle. J'accepte que les données enregistrées et fournies à l'occasion de cette étude puissent être conservées dans une base de données et faire l'objet d'un traitement informatisé non nominatif par le Lab-STICC, unité de recherche du CNRS.

En conséquence de quoi, je garantis le CNRS contre tout recours et/ou action que pourraient former les personnes physiques ou morales estimant avoir des droits quelconques à faire valoir sur l'utilisation des données.

Je reconnais et accepte que le CNRS conserve toute liberté pour exploiter ou ne pas exploiter intégralement ou par extraits les données.

Je déclare que la présente autorisation est accordée pour le monde entier et pour une durée de 5 (cinq) ans à compter de la signature de la présente, renouvelable par tacite reconduction, sauf dénonciation de la part du signataire.

Fait à le
Signature

Signature précédée de la mention « Lu et approuvé »

1 Conformément à la loi « informatique et libertés » du 6 janvier 1978 modifiée en 2004, vous pouvez obtenir communication et le cas échéant, rectification ou suppression des informations vous concernant, en vous adressant au CNRS, 3 rue Michel-Ange, 75794 PARIS CEDEX 16.

F.2 Consent Form STAGE Evaluation

Yann Glémarec
Versuchsleiter

Julius-Maximilians-Universität Würzburg
Lehrstuhl für Mensch-Computer-Interaktion
Am Hubland
D-97074 Würzburg

Raum 01.005, Gebäude M1, Campus Süd

Telefon: +336 72 62 79 59
E-Mail: yann.glemarec@uni-wuerzburg.de
Internet: www.hci.uni-wuerzburg.de

Würzburg, 11, January 2022

Virtual audience system, user study for learning outcome and stress mitigation.

1. Ablauf der Studie

Das folgende Experiment findet in einer virtuellen Umgebung statt. Die Versuchsleitung wird Ihnen helfen, das dafür benötigte Head-Mounted Display (HMD) aufzusetzen. Anschließend haben Sie 10 Minuten Zeit um ein CHI-Paper Ihrer Wahl zu präsentieren. Die Präsentationsfolien, die vorab im Decker Format erstellt worden sein müssen, werden im Rahmen der Virtual Reality (VR) Simulation genutzt. Im Anschluss an die Präsentation gibt es eine ca. 10-minütige Frage- und Antwortrunde bezüglich Ihrer Wahrnehmung des Systems. Am Ende des Experiments erhalten Sie Feedback von Dr. Jean-Luc Lugrin zur Qualität der Präsentation. Das gesamte Experiment wird ca. 45 Minuten dauern.

Folgende Daten werden im Rahmen des Experiments gesammelt: Geschlecht, Alter, Vorerfahrung mit VR Systemen, objektive Messungen während der Präsentation (z. B. die Zeit, die pro Präsentationsfolie benötigt wird) und bestimmte physiologische Daten wie den elektrodermalen Leitwert oder Ihre Temperatur.

Es werden geeignete Vorkehrungen getroffen, damit Unbefugte keinen Zugriff auf die Aufzeichnungen erhalten, Die Aufzeichnungen werden ausschließlich von projektbezogenen Mitarbeitern für die Auswertung des Experiments verwendet. Bei Fragen wenden Sie sich bitte an den Experimentator.

2. Freiwilligkeit und Anonymität

Die Teilnahme an der Studie ist freiwillig. Sie können jederzeit und ohne Angabe von Gründen die Teilnahme an dieser Studie beenden, ohne dass Ihnen daraus Nachteile entstehen. Die im Rahmen dieser Studie erhobenen, oben beschriebenen Daten und persönlichen Mitteilungen werden vertraulich behandelt. So unterliegen diejenigen ProjektmitarbeiterInnen, die durch direkten Kontakt mit Ihnen über personenbezogene Daten verfügen, der Schweigepflicht. Des Weiteren wird die Veröffentlichung der Ergebnisse der Studie in anonymisierter Form erfolgen, d. h. ohne dass Ihre Daten Ihrer Person zugeordnet werden können.

4. Datenschutz

Die Erhebung und Verarbeitung Ihrer oben beschriebenen persönlichen Daten erfolgt pseudonymisiert am Lehrstuhl für Informatik 9 unter Verwendung einer Nummer und ohne Angabe Ihres Namens. Es existiert eine Kodierliste auf Papier, die Ihren Namen mit der Nummer verbindet. Die Kodierliste ist nur den Versuchsleitern und dem Projektleiter zugänglich; das heißt, nur diese Personen können die erhobenen Daten mit meinem Namen in Verbindung bringen. Die Kodierliste wird in einem abschließbaren Schrank aufbewahrt und nach Abschluss der Datenerhebung, spätestens aber am 9. August 2021, vernichtet. Ihre Daten sind dann anonymisiert. Damit ist es niemandem mehr möglich, die erhobenen Daten mit Ihrem Namen in Verbindung zu bringen. Die anonymisierten Daten werden mindestens 10 Jahre gespeichert. Solange die Kodierliste existiert, können Sie die Löschung aller von Ihnen erhobenen Daten verlangen. Ist die Kodierliste aber erst einmal gelöscht, können wir Ihren Datensatz nicht mehr identifizieren. Deshalb können wir Ihrem Verlangen nach Löschung Ihrer Daten nur solange nachkommen, wie die Kodierliste existiert.

4. Sicherheitshinweise

Folgende Vorerkrankungen stellen Ausschlusskriterien für das folgende Experiment dar. Bitte wenden Sie sich an den Versuchsleiter sollte einer oder mehrere der Punkte auf Sie zutreffen (ohne den konkreten Punkt zu nennen).

- Sie haben Abnormitäten in Bezug auf Ihre binokularen Sehfähigkeiten (extrem starkes Schielen, extrem starke Sehschwäche, Einschränkungen in der räumlichen Wahrnehmung oder andere).
- Sie leiden an einer Herzkrankheit oder anderen schweren Krankheiten.
- Sie leiden an starkem Schwindel, Krämpfen, epileptischen Krämpfen oder Blackouts, die durch Blitzlichter oder Muster ausgelöst werden können, z.B. wenn Sie Fernsehen schauen, Videospiele spielen oder bei Aufenthalt in einer virtuellen Realität.
- Sie hatten in der Vergangenheit ein oder mehrmals starke oder epileptische Krämpfe, Bewusstseinsverlust oder ein anderes Symptom welches mit einem epileptischen Zustand in Verbindung gebracht werden könnte.
- Sie leiden unter Gleichgewichtsstörungen.
- Sie leiden unter sozialer Phobie oder pathologischer Angst vor öffentlichen Auftritten

Bitte beachten Sie zusätzlich folgende Punkte.

- Brechen Sie den Versuch sofort ab wenn Sie eines der folgenden Symptome feststellen: Müdigkeit, Benommenheit, überstarker Speichelfluss, exzessives Schwitzen, Schwindel, Übelkeit, Desorientierung, beeinträchtigte Auge-Hand Koordination, beeinträchtigte Balance, überstrapazierte Augen, verschwommenes Sehen, Doppelsehen oder andere visuelle Anomalitäten, Beschwerden oder Schmerzen im Kopf oder den Augen, unabsichtliche Bewegungen, Augen- oder Muskelzucken oder Krämpfe.
- Diese Symptome können bis Stunden nach der Erfahrung in der virtuellen Realität bestehen bleiben oder sich noch verstärken. Im Falle, dass eines oder mehrere der genannten Symptome auftreten, fahren Sie kein Auto, bedienen Sie keine Maschinen oder führen Sie keine visuell oder physisch anspruchsvollen Aufgaben, die einen funktionierenden Gleichgewichtssinn oder eine funktionierende Auge-Hand Koordination voraussetzen (z.B. Sport oder Fahrradfahren), bis Sie sich vollständig von den Symptomen erholt haben.
- Suchen Sie einen Arzt auf, wenn Sie schwere oder anhaltende Symptome aufweisen.

5. Einwilligungserklärung

Ich (Name des Teilnehmers /der Teilnehmerin in Blockschrift)

_____

bin schriftlich über die Studie und den Versuchsablauf aufgeklärt worden. Ich willige ein, an der Studie „Virtual audience system, user study for learning outcome and stress mitigation." teilzunehmen. Ich bin über den Ablauf der Studie informiert worden. Sofern ich Fragen zu dieser vorgesehenen Studie hatte, wurden sie vom/von der VersuchsleiterIn vollständig und zu meiner Zufriedenheit beantwortet.

Mit der beschriebenen Erhebung und Verarbeitung der Daten bin ich einverstanden. Die Aufzeichnung und Auswertung dieser Daten erfolgt pseudonymisiert im Lehrstuhl für Informatik 9, unter Verwendung einer Nummer und ohne Angabe meines Namens. Es existiert eine Kodierliste auf Papier, die meinen Namen mit dieser Nummer verbindet. Diese Kodierliste ist nur den Versuchsleitern und dem Projektleiter zugänglich, das heißt, nur diese Personen können die erhobenen Daten mit meinem Namen in Verbindung bringen. Nach Abschluss der Datenerhebung, spätestens am 9. August 2021, wird die Kodierliste gelöscht. Meine Daten sind dann anonymisiert. Damit ist es niemandem mehr möglich, die erhobenen Daten mit meinem Namen in Verbindung zu bringen. Mir ist bekannt, dass ich mein Einverständnis zur Aufbewahrung bzw. Speicherung dieser Daten widerrufen kann, ohne dass mir daraus Nachteile entstehen. Ich bin darüber informiert worden, dass ich jederzeit eine Löschung all meiner Daten verlangen kann. Wenn allerdings die Kodierliste bereits gelöscht ist, kann mein Datensatz nicht mehr identifiziert und also auch nicht mehr gelöscht werden. Meine Daten sind dann anonymisiert. Ich bin einverstanden, dass meine anonymisierten Daten zu Forschungszwecken weiter verwendet werden können und mindestens 10 Jahre gespeichert bleiben.

Des weiteren versichere ich, dass ich die Sicherheitshinweise gelesen habe und dass kein Ausschlusskriterium auf mich zutrifft.

Sollten behandlungsbedürftige Auffälligkeiten in der Testdiagnostik erkannt werden, bin ich damit einverstanden, dass mir diese mitgeteilt werden, so dass ich diese ggf. weiter abklären lassen kann. Ich wurde darüber informiert, dass die Information über auffällige Befunde u.U. mit versicherungsrechtlichen Konsequenzen verbunden sein kann.

Ich hatte genügend Zeit für eine Entscheidung und bin bereit, an der o.g. Studie teilzunehmen. Ich weiß, dass die Teilnahme an der Studie freiwillig ist und ich die Teilnahme jederzeit ohne Angaben von Gründen beenden kann. Ich weiß, dass ich in diesem Fall Anspruch auf Versuchspersonenstunden für die bis dahin erbrachten Stunden habe. Eine Ausfertigung der Teilnehmerinformation über die Untersuchung und eine Ausfertigung der Einwilligungserklärung habe ich erhalten. Die Teilnehmerinformation ist Teil dieser Einwilligungserklärung.

Name des Teilnehmers in Druckschrift:	Name des Versuchsleiters in Druckschrift:
_____	_____
Ort, Datum & Unterschrift des Teilnehmers:	Ort, Datum & Unterschrift des Versuchsleiters:
_____	_____

Appendix G

QUESTIONNAIRE

G.1 Demographic Questions

■ Please fill in the following text boxes:
- First Name
- Last Name
- Age
- Occupation

■ Please select your gender:
- Female
- Male

■ What is your prevailing hand?
- Right Hand
- Left Hand

G.2 Performance

■ How would you rate the quality of your presentation ?
- Very Bad -Bad - Neither Good nor Bad -Good -Very Good

■ How do you think the virtual audience would rate your presentation quality?
- Very Bad -Bad - Neither Good nor Bad -Good -Very Good
(This question is mandatory)

■ Would you say the audience was reacting to your presentation?
- Yes, they were reacting to my presentation or my behaviour,
- No, they were not reacting to my presentation or my behaviour,
- Other: (free answer)

→ For the following three questions, participants had to place their answers on a timeline or to select *never* on a radio button.

■ During your presentation when was the audience looking interested ?

- At the beginning
- At the end
- All the time
- Never
- Other:

(This question is mandatory)

■ During your presentation, when was the audience looking bored?

- At the beginning
- At the end
- All the time
- Never
- Other:

■ During your presentation, when was the audience looking enthusiastic?

- At the beginning
- At the end
- All the time
- Never
- Other:

(This question is mandatory)

■ During your presentation, when was the audience looking critical?

- At the beginning
- At the end
- All the time
- Never
- Other:

■ Did you adapt your presentation to make the virtual audience more interested? If yes, why and how? You can add more details into the comment section.

- Yes I did adapted my presentation or my behaviour,
- No I did not adapted my presentation or my behaviour,
- I have not seen any specific audience behaviours,

Please enter your comment here:

(This question is mandatory)

■ Would you say that the audience's behaviour impacted your behaviour, feeling or emotion? If yes, what impact and when?

- Yes, I changed my behaviour during the presentation
- No, I did not changed my behaviour during the presentation
- I did not put much attention to the audience

Please enter your comment here:

G.3 Presentation

■ How many times did you practiced your presentation before the study ?

(free answer)

■ Do you consider yourself as having experience doing public presentations ?

- yes
- no

■ Using one of the five propositions below, how would you describe your stress before the presentation?

- Not stressed at all
- As normal
- Slightly stressed
- Very Stressed
- Extremely Stressed

■ Would you say that the simulation is less stressful than a real audience to practice a presentation?

- It was less stressful than practicing in front of a real audience
- It was more stressful than practicing in front of a real audience
- There is no difference, it is as stressful as usual

Other:

■ How do you practice for a presentation? Please select one of the propositions bellow and if none is fitting you, please add your answer to 'other'

- I do not practice
- I prepare notes before
- I practice alone
- I ask friends to listen to me

Other:

G.4 The Simulation

■ Based on your own experience, would you say this simulator could be used as a practice system for public presentations at the university?

- Yes, it could be useful
- No, it won't be useful

Please enter your comment here:

■ Would you say this simulator could improve your presentation skills?

- Yes, it could help me to improve my presentation skills
- No, it won't help me to improve my presentation skills

Please enter your comment here:

■ Would you say this simulator is more or less engaging than a normal practice session without virtual reality?

- Yes, it is more engaging
- No, it is less engaging

■ Would you say that using this simulator for a practice session is funnier than for a practice session without virtual reality?

- Yes, it is funnier to practice with this system
- No, it is not funnier to practice with this system
- It is even less fun to practice with it

Please enter your comment here:

G.5 The Audience behaviour

■ Would you say the reaction from the virtual audience and the virtual spectators was believable compared to a real audience?

- Yes
- No, please add some details in the comment box:

Please enter your comment here:

■ Would you say the behaviours of the virtual audience and the virtual spectators were believable compared to a real audience?

- Yes

- No, please add some details in the comment box:

Please enter your comment here:

■ Would you say the sounds from the virtual environment were believable compared to a lecture room or meeting room?

- Yes

- No, please add some details in the comment box:

Please enter your comment here:

■ If you have any remarks and comment to do about the system or the study please write them down into the frame below:

Appendix H

Public Speaking Anxiety Scale

The PSAS is a 17-item self-report measure with responses measured in a Likert-format with score ranging from 1 "not at all" to 5 "extremely." Scores on this scale can range from 17 to 85. There are five items on this scale that are reverse coded.

1. Giving a speech is terrifying
2. I am afraid that I will be at a loss for words while speaking
3. I am nervous that I will embarrass myself in front of the audience
4. If I make a mistake in my speech, I am unable to re-focus
5. I am worried that my audience will think I am a bad speaker
6. I am focused on what I am saying during my speech*
7. I am confident when I give a speech*
8. I feel satisfied after giving a speech*
9. My hands shake when I give a speech
10. I feel sick before speaking in front of a group
11. I feel tense before giving a speech
12. I fidget before speaking
13. My heart pounds when I give a speech
14. I sweat during my speech
15. My voice trembles when I give a speech
16. I feel relaxed while giving a speech*
17. I do not have problems making eye contact with my audience*

Note. 1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely.

∗ *Reverse-coded.*

Bibliography

Adobe Systems, I., (2022), *Animated 3d characters library*. https://www.mixamo.com/#/ (accessed: 28.03.2022)

Allbeck, J., & Badler, N. I., (2001), Consistent communication with control, *Center for Human Modeling and Simulation*, 85.

Allwood, J., Nivre, J., & Ahlsén, E., (1992), On the semantics and pragmatics of linguistic feedback, *Journal of semantics*, *9 1*, 1–26, https://doi.org/https://doi.org/10.1093/jos/9.1.1

American Psychiatric Association, A., Association, A. P. et al., (2013), *Diagnostic and statistical manual of mental disorders: dsm-5* (Vol. 10), Washington, DC: American psychiatric association.

Anderson, K., André, E., Baur, T., Bernardini, S., Chollet, M., Chryssafidou, E., Damian, I., Ennis, C., Egges, A., Gebhard, P., et al., (2013), The tardis framework: intelligent virtual agents for social coaching in job interviews, *International Conference on Advances in Computer Entertainment Technology*, 476–491.

Anderson, P., Rothbaum, B. O., & Hodges, L. F., (2003), Virtual reality exposure in the treatment of social anxiety, *Cognitive and Behavioral Practice*, *10 3*, 240–247, https://doi.org/https://doi.org/10.1016/S1077-7229(03)80036-6

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., & Calamaras, M. R., (2013), Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial., *Journal of consulting and clinical psychology*, *81 5*, 751, https://doi.org/https://psycnet.apa.org/doi/10.1037/a0033559

Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O., (2005), Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure, *Depression and Anxiety*, *22 3*, 156–158, https://doi.org/https://doi.org/10.1002/da.20090

Argelaguet, F., Hoyet, L., Trico, M., & Lecuyer, A., (2016), The role of interaction in virtual embodiment: effects of the virtual hand representation, *2016 IEEE Virtual Reality (VR)*, 3–10, https://doi.org/10.1109/VR.2016.7504682

Argyle, M., & Dean, J., (1965), Eye-contact, distance and affiliation, *Sociometry*, 289–304.

Arsenault, D., (2005), Dark waters: spotlight on immersion, *GAMEON-NA International Conference*, 50–52.

Arthur, K. W., Booth, K. S., & Ware, C., (1993), Evaluating 3d task performance for fish tank virtual worlds, *ACM Transactions on Information Systems*, *11 3*, 239–265.

Aylett, R. S., (2004), Agents and affect: why embodied agents need affective systems, *Hellenic Conference on Artificial Intelligence*, 496–504.

Barmaki, R., & Hughes, C., (2018), Gesturing and embodiment in teaching: investigating the nonverbal behavior of teachers in a virtual rehearsal environment., *Proceedings of the AAAI Conference on Artificial Intelligence*, *32 1*, https://ojs.aaai.org/index.php/AAAI/article/view/11394

Bartholomay, E. M., & Houlihan, D. D., (2016), Public speaking anxiety scale: preliminary psychometric data and scale validation, *Personality and Individual Differences*, *94*, 211–215.

Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., & Scherer, S., (2013), Cicero-towards a multimodal virtual audience platform for public speaking training, *International workshop on intelligent virtual agents*, 116–128, https://doi.org/https://doi.org/10.1007/978-3-642-40415-3_10

Becker, C., Prendinger, H., Ishizuka, M., & Wachsmuth, I., (2005), Evaluating affective feedback of the 3d agent max in a competitive cards game, *International Conference on Affective Computing and Intelligent Interaction*, 466–473.

Bevacqua, E., De Sevin, E., Hyniewska, S. J., & Pelachaud, C., (2012), A listener model: introducing personality traits, *Journal on Multimodal User Interfaces*, *6 1*, 27–38.

Bevacqua, E., Hyniewska, S. J., & Pelachaud, C., (2010), Positive influence of smile backchannels in ecas, *International Workshop on Interacting with ECAs as Virtual Characters*, 13.

Bevacqua, E., Pammi, S., Hyniewska, S. J., Schröder, M., & Pelachaud, C., (2010), Multimodal backchannels for embodied conversational agents, *International Conference on Intelligent Virtual Agents*, 194–200.

Biocca, F., (1997), The cyborg's dilemma: progressive embodiment in virtual environments, *Journal of computer-mediated communication*, *3 2*, JCMC324.

Biocca, F., Harms, C., & Burgoon, J. K., (2003), Toward a more robust theory and measure of social presence: review and suggested criteria, *Presence: Teleoperators and Virtual Environments*, *12 5*, 456–480, https://doi.org/10.1162/105474603322761270

Blanke, O., (2012), Multisensory brain mechanisms of bodily self-consciousness, *Nature Reviews Neuroscience*, *13 8*, 556–571.

Blanke, O., & Metzinger, T., (2009), Full-body illusions and minimal phenomenal selfhood, *Trends in cognitive sciences*, *13 1*, 7–13.

Blöte, A. W., Kint, M. J., Miers, A. C., & Westenberg, P. M., (2009), The relation between public speaking anxiety and social anxiety: a review, *Journal of anxiety disorders*, *23 3*, 305–313.

Bosser, A.-G., Cavazza, M., & Champagnat, R., (2010), Linear logic for non-linear storytelling [Subject to restrictions, author can archive publisher's version/PDF. ; 19th European Conference on Artificial Intelligence, ECAI 2010 ; Conference date: 16-08-2010 Through 20-08-2010], *Frontiers in artificial intelligence and applications*, 713–718, https://doi.org/10.3233/978-1-60750-606-5-713

Botvinick, M., & Cohen, J., (1998), Rubber hands 'feel'touch that eyes see, *Nature*, *391 6669*, 756–756.

Bouvier, E., & Guilloteau, P., (1996), Crowd simulation in immersive space management. *Virtual environments and scientific visualization'96* (pp. 104–110), Springer, https://doi.org/https://doi.org/10.1007/978-3-7091-7488-3_11

Bowman, D., Kruijff, E., LaViola Jr, J. J., & Poupyrev, I. P., (2005), *3d user interfaces: theory and practice*, Addison-Wesley.

Braun, A., Musse, S., de Oliveira, L., & Bodmann, B., (2003), Modeling individual behaviors in crowd simulation, *Proceedings 11th IEEE International Workshop on Program Comprehension*, 143–148, https://doi.org/10.1109/CASA.2003.1199317

Buche, C., & Bigot, N. L., (2018), Revam: a virtual reality application for inducing body size perception modifications, *2018 International Conference on Cyberworlds (CW)*, 229–236, https://doi.org/10.1109/CW.2018.00049

Cafaro, A., Vilhjálmsson, H. H., Bickmore, T., Heylen, D., Jóhannsdóttir, K. R., & Valgarðsson, G. S., (2012), First impressions: users' judgments of virtual agents' personality and interpersonal attitude in first encounters, *International conference on intelligent virtual agents*, 67–80.

Cavazza, M., Charles, F., & Mead, S., (2002), Character-based interactive storytelling, *IEEE Intelligent Systems*, *17 4*, 17–24, https://doi.org/10.1109/MIS.2002.1024747

Chenney, S., (2004), Flow tiles, *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 233–242, https://doi.org/https://doi.org/10.1145/1028523.1028553

Cheong, Y.-G., & Young, R. M., (2014), Suspenser: a story generation system for suspense, *IEEE Transactions on Computational Intelligence and AI in Games*, *7 1*, 39–52, https://doi.org/https://doi.org/10.1109/TCIAIG.2014.2323894

Chollet, M., Marsella, S., & Scherer, S., (2022), Training public speaking with virtual social interactions: effectiveness of real-time feedback and delayed feedback, *J. Multimodal User Interfaces*, *16 1*, 17–29, https://doi.org/10.1007/s12193-021-00371-1

Chollet, M., Ochs, M., & Pelachaud, C., (2014), From non-verbal signals sequence mining to bayesian networks for interpersonal attitudes expression, *International conference on intelligent virtual agents*, 120–133, https://doi.org/https://doi.org/10.1007/978-3-319-09767-1_15

Chollet, M., & Scherer, S., (2017), Perception of virtual audiences, *IEEE computer graphics and applications*, *37 4*, 50–59, https://doi.org/https://doi.org/10.1109/MCG.2017.3271465

Chollet, M., Sratou, G., Shapiro, A., Morency, L.-P., & Scherer, S., (2014), An interactive virtual audience platform for public speaking training, *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, 1657–1658, https://doi.org/10.5555/2615731.2616111

Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A., & Scherer, S., (2015), Exploring feedback strategies to improve public speaking: an interactive virtual audience framework, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1143–1154, https://doi.org/https://doi.org/10.1145/2750858.2806060

Clower, D. M., & Boussaoud, D., (2000), Selective use of perceptual recalibration versus visuomotor skill acquisition, *Journal of Neurophysiology*, *84 5*, 2703–2708.

Cummings, J. J., & Bailenson, J. N., (2016), How immersive is enough? a meta-analysis of the effect of immersive technology on user presence, *Media Psychology*, *19 2*, 272–309.

Dang, K. D., Hoffmann, S., Champagnat, R., & Spierling, U., (2011), How authors benefit from linear logic in the authoring process of interactive storyworlds, *Interactive Storytelling: Fourth International Conference on Interactive Digital Storytelling, ICIDS 2011, Vancouver, Canada, November 28 – 1 December, 2011. Proceedings*, 249–260, https://doi.org/10.1007/978-3-642-25289-1_27

De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., & De Carolis, B., (2003), From greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent, *International journal of human-computer studies*, *59 1-2*, 81–118.

Delamarre, A., Shernoff, E., Buche, C., Frazier, S., Gabbard, J., & Lisetti, C., (2021), The interactive virtual training for teachers (ivt-t) to practice classroom behavior management, *International Journal of Human-Computer Studies*, *152*, 102646, https://doi.org/https://doi.org/10.1016/j.ijhcs.2021.102646

Delamarre, A., (2020), *Interactive virtual training: implementation for early career teachers to practice classroom behavior management* (Doctoral dissertation), FIU Electronic Theses; Dissertations. 4595., https://doi.org/https://dx.doi.org/10.25148/etd.FIDC009190

Dermouche, S., & Pelachaud, C., (2019), Generative model of agent's behaviors in human-agent interaction, *2019 International Conference on Multimodal Interaction*, 375–384.

Dettori, G., & Paiva, A., (2009), Narrative learning in technology-enhanced environments, *In* N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: principles and products* (pp. 55–69), Springer Netherlands, https://doi.org/10.1007/978-1-4020-9827-7_4

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., et al., (2014), Simsensei kiosk: a virtual human interviewer for healthcare decision support, *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 1061–1068.

Ekman, P., (1999), Basic emotions, *Handbook of cognition and emotion*, *98 45-60*, 16, https://doi.org/https://doi.org/10.1002/0470013494.ch3

Ekman, P., & Friesen, W. V., (1975), *Unmasking the face: a guide to recognizing emotions from facial clues* (Vol. 10), Ishk.

Empatica, I., (2022), *Empatica e4 wristband.* https://www.empatica.com/research/e4/ (accessed: 28.03.2022)

Epic Games, I., (2022a), *Introduction to blueprints.* https://docs.unrealengine.com/4.27/en-US/ProgrammingAndScripting/Blueprints/GettingStarted/#:~:text=The%20Blueprint%20Visual%20Scripting%20system,or%20objects%20in%20the%20engine. (accessed: 28.03.2022)

Epic Games, I., (2022b), *Unreal engine.* https://www.unrealengine.com/en-US/ (accessed: 28.03.2022)

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C., (2007), The world of emotions is not two-dimensional, *Psychological science*, *18 12*, 1050–1057, https://doi.org/https://doi.org/10.1111/j.1467-9280.2007.02024.x

Fukuda, M., Huang, H.-H., Ohta, N., & Kuwabara, K., (2017), Proposal of a parameterized atmosphere generation model in a virtual classroom, *Proceedings of the 5th International Conference on Human Agent Interaction*, 11–16, https://doi.org/10.1145/3125739.3125776

Gallego, M. J., Emmelkamp, P. M. G., van der Kooij, M., & Mees, H., (2011), The effects of a dutch version of an internet-based treatment program for fear of public speaking: a controlled study, *International journal of clinical and health psychology*, *11 3*, 459–472.

Gamito, P., Oliveira, J., Santos, P., Morais, D., Saraiva, T., Pombal, M., & Mota, B., (2008), Presence, immersion and cybersickness assessment through a test anxiety virtual environment, *Annual Review of CyberTherapy and Telemedicine*, *6*, 83–90.

Garrod, S., & Pickering, M. J., (2004), Why is conversation so easy?, *Trends in cognitive sciences*, *8 1*, 8–11.

Gilroy, S. W., Porteous, J., Charles, F., Cavazza, M., Soreq, E., Raz, G., Ikar, L., Or-Borichov, A., Ben-Arie, U., Klovatch, I., & Hendler, T., (2013), A brain-computer interface to a plan-based narrative, *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, 1997–2005.

Goffman, E., (1963), *Behavior in public places: notes on the social organization of gatherings*, Free Press of Glencoe.

Grandjean, D., Sander, D., & Scherer, K. R., (2008), Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization, *Consciousness and cognition*, *17 2*, 484–495, https://doi.org/https://doi.org/10.1016/j.concog.2008.03.019

Graziano, M. S., & Botvinick, M. M., (2002), How the brain represents the body: insights from neurophysiology and psychology, *Common mechanisms in perception and action: Attention and performance XIX*, *19*, 136–157.

Gunes, H., & Pantic, M., (2010), Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners, *International conference on intelligent virtual agents*, 371–377, https://doi.org/https://doi.org/10.1007/978-3-642-15892-6_39

Gunes, H., Schuller, B., Pantic, M., & Cowie, R., (2011), Emotion representation, analysis and synthesis in continuous space: a survey, *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 827–834, https://doi.org/https://doi.org/10.1109/FG.2011.5771357

Hale, K. S., & Stanney, K. M., (2006), Effects of low stereo acuity on performance, presence and sickness within a virtual environment, *Applied Ergonomics*, *37 3*, 329–339.

Harrigan, J. A., (n.d.), Proxemics, kinesics, and gaze. *The new handbook of methods in nonverbal behavior research*, Oxford University Press.

Harris, S. R., Kemmerling, R. L., & North, M. M., (2002), Brief virtual reality therapy for public speaking anxiety, *Cyberpsychology & Behavior*, *5 6*, 543–550.

Hayes, A. T., Hardin, S. E., & Hughes, C. E., (2013), Virtual, augmented and mixed reality. systems and applications: 5th international conference, vamr 2013, held as part of hci international 2013, las vegas, nv, usa, july 21-26, 2013, proceedings, part ii, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-642-39420-1_16

Hendrix, C., & Barfield, W., (1996), The sense of presence within auditory virtual environments, *Presence: Teleoperators & Virtual Environments*, *5 3*, 290–301, https://doi.org/https://doi.org/10.1162/pres.1996.5.3.290

Heudin, J., (2004), Evolutionary virtual agent, *Proceedings. IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2004)*, 93–98, https://doi.org/10.1109/IAT.2004.1342929

Hofmann, S. G., (2004), Cognitive mediation of treatment change in social phobia., *Journal of consulting and clinical psychology*, *72 3*, 392.

Huang, H.-H., Fukuda, M., & Nishida, T., (2019), Development of a platform for rnn driven multimodal interaction with embodied conversational agents, *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 200–202, https://doi.org/10.1145/3308532.3329448

Iribarnegaray, L., (2021), " c'est beaucoup plus anxiogene que de rendre une copie ": la peur de l'oral, une angoisse française, *Le Monde*.

Janowski, K., & André, E., (2019), What if i speak now?: a decision-theoretic approach to personality-based turn-taking., *AAMAS*, 1051–1059.

Jicol, C., Wan, C. H., Doling, B., Illingworth, C. H., Yoon, J., Headey, C., Lutteroth, C., Proulx, M. J., Petrini, K., & O'Neill, E., (2021), Effects of emotion and agency on presence in virtual reality, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13.

Kahlon, S., Lindner, P., & Nordgreen, T., (2019), 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*, *13 1*, 1–10, https://doi.org/https://doi.org/10.1186/s13034-019-0307-y

Kang, N., Brinkman, W.-P., van Riemsdijk, M. B., & Neerincx, M., (2016), The design of virtual audiences: noticeable and recognizable behavioral styles, *Computers in Human Behavior*, *55*, 680–694, https://doi.org/https://doi.org/10.1016/j.chb.2015.10.008

Kang, N., Brinkman, W.-P., van Riemsdijk, M. B., & Neerincx, M. A., (2013), An expressive virtual audiencewith flexible behavioral styles, *IEEE Transactions on Affective Computing*, *4 4*, 326–340, https://doi.org/10.1109/TAFFC.2013.2297104

Kelly, O., Matheson, K., Martinez, A., Merali, Z., & Anisman, H., (2007), Psychosocial stress evoked by a virtual audience: relation to neuroendocrine activity, *CyberPsychology & Behavior*, *10 5*, 655–662.

Kilteni, K., Groten, R., & Slater, M., (2012), The sense of embodiment in virtual reality, *Presence: Teleoperators and Virtual Environments*, *21 4*, 373–387.

Kistler, F., Endrass, B., Damian, I., Dang, C. T., & André, E., (2012), Natural interaction with culturally adaptive virtual characters, *Journal on Multimodal User Interfaces*, *6 1*, 39–47.

Klinger, E., Bouchard, S., Légeron, P., Roy, S., Lauer, F., Chemin, I., & Nugues, P., (2005), Virtual reality therapy versus cognitive behavior therapy for social phobia: a preliminary controlled study, *Cyberpsychology & behavior*, *8 1*, 76–88.

Knight, M. M., & Arns, L. L., (2006), The relationship among age and other factors on incidence of cybersickness in immersive environment users, *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, 162–162.

Krämer, N. C., Lucas, G., Schmitt, L., & Gratch, J., (2018), Social snacking with a virtual agent–on the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities, *International Journal of Human-Computer Studies*, *109*, 112–121.

Kringlen, E., Torgersen, S., & Cramer, V., (2001), A norwegian psychiatric epidemiological study, *American journal of psychiatry*, *158 7*, 1091–1098.

Kucherenko, T., Hasegawa, D., Henter, G. E., Kaneko, N., & Kjellström, H., (2019), Analyzing input and output representations for speech-driven gesture generation, *Pro-*

ceedings of the 19th ACM International Conference on Intelligent Virtual Agents, 97–104.

Kucherenko, T., Jonell, P., van Waveren, S., Henter, G. E., Alexandersson, S., Leite, I., & Kjellström, H., (2020), Gesticulator: a framework for semantically-aware speech-driven gesture generation, *Proceedings of the 2020 International Conference on Multimodal Interaction*, 242–250.

Laslett, R., & Smith, C., (2002), *Effective classroom management: a teacher's guide*, Routledge.

Latoschik, M. E., Kern, F., Stauffert, J.-P., Bartl, A., Botsch, M., & Lugrin, J.-L., (2019), Not alone here?! scalability and user experience of embodied ambient crowds in distributed social virtual reality, *IEEE transactions on visualization and computer graphics*, *25 5*, 2134–2144, https://doi.org/10.1109/TVCG.2019.2899250

Latoschik, M. E., Lugrin, J.-L., Habel, M., Roth, D., Seufert, C., & Grafe, S., (2016), Breaking bad behavior: immersive training of class room management, *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology*, 317–318, http://dl.acm.org/authorize?N40662

Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., & Botsch, M., (2017), The effect of avatar realism in immersive social virtual realities, *Proceedings of the 23rd ACM symposium on virtual reality software and technology*, 1–10.

Latoschik & Team (Uni Wuerzburg), M., Tramberend (Beuth Hochschule Berlin), H., & Botsch (TU Dortmund), M., (2022), *Decker*. https://gitlab2.informatik.uni-wuerzburg.de/decker/decker (accessed: 28.03.2022)

LaViola Jr, J. J., (2000), A discussion of cybersickness in virtual environments, *ACM Sigchi Bulletin*, *32 1*, 47–56.

Lee, C., Bonebrake, S., Bowman, D. A., & Höllerer, T., (2010), The role of latency in the validity of ar simulation, *2010 IEEE Virtual Reality Conference (VR)*, 11–18.

Lee, K. M., (2004), Presence, explicated, *Communication theory*, *14 1*, 27–50, https://doi.org/https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lee, S., & Son, Y.-J., (2008), Integrated human decision making model under belief-desire-intention framework for crowd simulation, *2008 Winter Simulation Conference*, 886–894, https://doi.org/10.1109/WSC.2008.4736153

Lester, J. C., Voerman, J. L., Towns, S. G., & Callaway, C. B., (1997), Cosmo: a life-like animated pedagogical agent with deictic believability.

Lhommet, M., & Marsella, S. C., (2015), Expressing emotion through posture and gesture. *The oxford handbook of affective computing* (pp. 273–285), Oxford University Press, https://doi.org/https://psycnet.apa.org/doi/10.1093/oxfordhb/9780199942237.013.039

Lindner, P., Dagöö, J., Hamilton, W., Miloff, A., Andersson, G., Schill, A., & Carlbring, P., (2021), Virtual reality exposure therapy for public speaking anxiety in routine care: a single-subject effectiveness trial, *Cognitive Behaviour Therapy*, *50 1*, 67–87.

Lindsay, A., Read, J., Ferreira, J. F., Hayton, T., Porteous, J., & Gregory, P., (2017), Framer: planning models from natural language action descriptions, *In* L. Barbulescu, J. Frank, Mausam, & S. F. Smith (Eds.), *Proceedings of the twenty-seventh international conference on automated planning and scheduling, ICAPS 2017, pittsburgh, pennsylvania, usa, june 18-23, 2017* (pp. 434–442), AAAI Press, https://aaai.org/ocs/index.php/ICAPS/ICAPS17/paper/view/15735

Lisetti, C., Amini, R., Yasavur, U., & Rishe, N., (2013), I can help you change! an empathic virtual agent delivers behavior change health interventions, *ACM Transactions on Management Information Systems (TMIS)*, *4 4*, 1–28, https://doi.org/https://doi.org/10.1007/s11423-020-09819-9

Lugrin, J.-L., & Cavazza, M., (2006), Ai-based world behaviour for emergent narratives, *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, 25–es, https://doi.org/10.1145/1178823.1178853

Lugrin, J.-L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., & Grafe, S., (2016), Breaking bad behaviours: a new tool for learning classroom management using virtual reality, *Frontiers in ICT*, *3*, 26, https://doi.org/https://doi.org/10.3389/fict.2016.00026

Lugrin, J.-L., Wiebusch, D., Latoschik, M. E., & Strehler, A., (2013), Usability benchmarks for motion tracking systems, *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, 49–58.

Lugrin, J.-L., Wiedemann, M., Bieberstein, D., & Latoschik, M. E., (2015), Influence of avatar realism on stressful situation in vr, *2015 IEEE Virtual Reality (VR)*, 227–228.

Martens, C., (2015), Ceptre: A language for modeling generative interactive systems, *In* A. Jhala & N. R. Sturtevant (Eds.), *Proceedings of the eleventh AAAI conference on artificial intelligence and interactive digital entertainment, AIIDE 2015, november*

14-18, 2015, university of california, santa cruz, ca, USA (pp. 51–57), AAAI Press, http://www.aaai.org/ocs/index.php/AIIDE/AIIDE15/paper/view/11536

Martens, C., Bosser, A.-G., Ferreira, J. F., & Cavazza, M., (2013), Linear logic programming for narrative generation, *In* P. Cabalar & T. C. Son (Eds.), *Logic programming and nonmonotonic reasoning* (pp. 427–432), Springer Berlin Heidelberg.

Matthews, S. L., Uribe-Quevedo, A., & Theodorou, A., (2020), Rendering optimizations for virtual reality using eye-tracking, *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*, 398–405, https://doi.org/10.1109/SVR51698.2020.00066

McCauley, M. E., & Sharkey, T. J., (1992), Cybersickness: perception of self-motion in virtual environments, *Presence: Teleoperators & Virtual Environments*, *1 3*, 311–318.

McMahan, A., (2013), Immersion, engagement, and presence: a method for analyzing 3-d video games. *The video game theory reader* (pp. 67–86), Routledge.

Meehan, M., Insko, B., Whitton, M., & Brooks Jr, F. P., (2002), Physiological measures of presence in stressful virtual environments, *Acm transactions on graphics (tog)*, *21 3*, 645–652.

Mehrabian, A., (1972), *Nonverbal communication*, Aldine-Atherton, https://books.google.fr/books?id=Gc5-AAAAMAAJ

Mehrabian, A., (1996), Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament, *Current Psychology*, *14 4*, 261–292, https://doi.org/https://doi.org/10.1007/BF02686918

Metzinger, T., (2009), Why are out-of-body experiences interesting for philosophers?: the theoretical relevance of obe research, *cortex*, *45 2*, 256–258.

Mignault, A., & Chaudhuri, A., (2003), The many faces of a neutral face: head tilt and perception of dominance and emotion, *Journal of nonverbal behavior*, *27 2*, 111–132, https://doi.org/https://doi.org/10.1023/A:1023914509763

Ministere de l'Education Nationale et de la Jeunesse, (2020), *Epreuve orale dite 'grand oral' de la classe de terminale de la voie générale á compter de la session 2021 de l'examen du baccalauréat.* https://www.education.gouv.fr/bo/20/Special2/MENE2002780N.htm (accessed: 15/11/2022)

Minsky, M., (1980), Telepresence.

Morency, L.-P., Christoudias, C. M., & Darrell, T., (2006), Recognizing gaze aversion gestures in embodied conversational discourse, *Proceedings of the 8th International*

Conference on Multimodal Interfaces, 287–294, https://doi.org/10.1145/1180995.1181051

Morency, L.-P., Sidner, C., Lee, C., & Darrell, T., (2005), Contextual recognition of head gestures, *Proceedings of the 7th international conference on Multimodal interfaces*, 18–24.

Moss, J. D., & Muth, E. R., (2011), Characteristics of head-mounted displays and their effects on simulator sickness, *Human factors*, *53 3*, 308–319.

Mouw, J. M., Fokkens-Bruinsma, M., & Verheij, G.-J., (2020), Using virtual reality to promote pre-service teachers' classroom management skills and teacher resilience: a qualitative evaluation, *Proceedings of the 6th International Conference on Higher Education Advances (HEAd'20)*, 325–332.

Narayan, M., Waugh, L., Zhang, X., Bafna, P., & Bowman, D., (2005), Quantifying the benefits of immersion for collaboration in virtual environments, *Proceedings of the ACM symposium on Virtual reality software and technology*, 78–81.

Niederberger, C., & Gross, M., (2003), Hierarchical and heterogenous reactive agents for real-time applications, *Computer Graphics Forum*, *22 3*, 323–331, https://doi.org/ https://doi.org/10.1111/1467-8659.00679

Niewiadomski, R., Bevacqua, E., Mancini, M., & Pelachaud, C., (2009), Greta: an interactive expressive eca system, *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, 1399–1400.

Niewiadomski, R., Demeure, V., & Pelachaud, C., (2010), Warmth, competence, believability and virtual agents, *International Conference on Intelligent Virtual Agents*, 272–285.

Nilsson, N., Nordahl, R., & Serafin, S., (2016), Immersion revisited: a review of existing definitions of immersion and their relation to different theories of presence, *Human technology*, *12 2*, 108.

North, M. M., North, S. M., & Coble, J. R., (1998), Virtual reality therapy: an effective treatment for the fear of public speaking, *International Journal of Virtual Reality*, *3 3*, 1–6.

Norvig, S., & Russel, P., (2002), *Artificial intelligence. a modern approach*, Upper Saddle River, NJ, USA: Prentice Hall.

Ochs, M., Niewiadomski, R., Pelachaud, C., & Sadek, D., (2005), Intelligent expressions of emotions, *International Conference on Affective Computing and Intelligent Interaction*, 707–714, https://doi.org/https://doi.org/10.1007/11573548_91

Ochs, M., Pelachaud, C., & Sadek, D., (2008), An empathic virtual dialog agent to improve human-machine interaction, *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, 89–96, https://doi.org/10.5555/1402383.1402401

Ochs, M., Ravenet, B., & Pelachaud, C., (2013), A crowdsourcing toolbox for a user-perception based design of social virtual actors, *Computers are Social Actors Workshop (CASA)*.

Oculus VR LLC, (2017), *Oculus best practices*. Retrieved October 21, 2022, from https://scontent.oculuscdn.com/v/t64.5771-25/12482206_237917063479780_486464407014998016_n.pdf?_nc_cat=105&ccb=1-7&_nc_sid=489e6e&_nc_ohc=4dMhYyH-pyEAX_wdh4V&_nc_ht=scontent.oculuscdn.com&oh=00_AT9IJXLrd8sU3rm5BRvbLwAHiLLsMw&oe=63578E92

Oman, C. M., (1990), Motion sickness: a synthesis and evaluation of the sensory conflict theory, *Canadian journal of physiology and pharmacology*, *68 2*, 294–303.

Ortony, A., Clore, G. L., & Collins, A., (1988), The cognitive structure of emotions cambridge, *UK: Cambridge University Press9*, https://doi.org/https://doi.org/10.1017/CBO9780511571299

Owens, M. E., & Beidel, D. C., (2015), Can virtual reality effectively elicit distress associated with social anxiety disorder?, *Journal of Psychopathology and Behavioral Assessment*, *37 2*, 296–305.

Palmas, F., Cichor, J., Plecher, D. A., & Klinker, G., (2019), Acceptance and effectiveness of a virtual reality public speaking training, *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 363–371.

Palmas, F., Reinelt, R., Cichor, J. E., Plecher, D. A., & Klinker, G., (2021), Virtual reality public speaking training: experimental evaluation of direct feedback technology acceptance, *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 463–472, https://doi.org/10.1109/VR50410.2021.00070

Palmer, M. T., (1995), Interpersonal communication and virtual reality: mediating interpersonal relationships, *Communication in the age of virtual reality*, 277–299.

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M., (2013), Putting yourself in the skin of a black avatar reduces implicit racial bias, *Consciousness and cognition*, *22 3*, 779–787.

Pelechano Gómez, N., O'Brien, K., Silverman, B. G., & Badler, N., (2005), Crowd simulation incorporating agent psychological models, roles and communication, *First International Workshop on Crowd Simulation*, http://hdl.handle.net/2117/10144

Pertaub, D.-P., Slater, M., & Barker, C., (2002), An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience, *Presence: Teleoperators and Virtual Environments*, *11 1*, 68–78, https://doi.org/10.1162/105474602317343668

Poeschl, S., (2017), Virtual reality training for public speaking—a quest-vr framework validation, *Frontiers in ICT*, *4*, 13.

Poeschl, S., Wall, K., & Doering, N., (2013), Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence, *2013 IEEE Virtual Reality (VR)*, 129–130, https://doi.org/https://doi.org/10.1109/VR.2013.6549396

Poggi, I., (2007), Mind, hands, face and body, *A goal and belief view of multimodal communication. Weidler, Berlin*, https://doi.org/http://dx.doi.org/10.1515/9783110261318.627.

Poggi, I., Pelachaud, C., Rosis, F. d., Carofiglio, V., & Carolis, B. D., (2005), Greta. a believable embodied conversational agent. *Multimodal intelligent information presentation* (pp. 3–25), Springer.

Poppe, R., Truong, K. P., Reidsma, D., & Heylen, D., (2010), Backchannel strategies for artificial listeners, *Proceedings of the 10th International Conference on Intelligent Virtual Agents*, 146–158.

Porteous, J., Cavazza, M., & Charles, F., (2010), Applying planning to interactive storytelling: narrative control using state constraints, *ACM Transactions on Intelligent Systems and Technology (TIST)*, *1 2*, 1–21, https://doi.org/https://doi.org/10.1145/1869397.1869399

Potdevin, D., Clavel, C., & Sabouret, N., (2021), Virtual intimacy in human-embodied conversational agent interactions: the influence of multimodality on its perception, *Journal on Multimodal User Interfaces*, *15 1*, 25–43, https://doi.org/https://doi.org/10.1007/s12193-020-00337-9

Prada, R., & Paiva, A., (2005), Believable groups of synthetic characters, *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, 37–43, https://doi.org/10.1145/1082473.1082479

Premkumar, P., Heym, N., Brown, D. J., Battersby, S., Sumich, A., Huntington, B., Daly, R., & Zysk, E., (2021), The effectiveness of self-guided virtual-reality exposure therapy for public-speaking anxiety, *Frontiers in Psychiatry*, *12*, https://doi.org/10.3389/fpsyt.2021.694610

Prepin, K., Ochs, M., & Pelachaud, C., (2012), Mutual stance building in dyad of virtual agents: smile alignment and synchronisation, *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, 938–943.

Reason, J. T., & Brand, J. J., (1975), *Motion sickness.*, Academic press.

Riccio, G. E., & Stoffregen, T. A., (1991), An ecological theory of motion sickness and postural instability, *Ecological psychology*, *3 3*, 195–240.

Riedl, M. O., & Stern, A., (2006), Believable agents and intelligent story adaptation for interactive storytelling, *In* S. Göbel, R. Malkewitz, & I. Iurgel (Eds.), *Technologies for interactive digital storytelling and entertainment* (pp. 1–12), Springer Berlin Heidelberg.

Roseman, I. J., (1991), Appraisal determinants of discrete emotions, *Cognition and Emotion*, *5 3*, 161–200, https://doi.org/10.1080/02699939108411034

Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., & Price, L., (2000), A controlled study of virtual reality exposure therapy for the fear of flying., *Journal of Consulting and Clinical Psychology*, *68 6*, 1020–1026, https://doi.org/10.1037/0022-006x.68.6.1020

Russell, J. A., (1980), A circumplex model of affect., *Journal of personality and social psychology*, *39 6*, 1161, https://doi.org/https://psycnet.apa.org/doi/10.1037/h0077714

Sander, D., Grandjean, D., & Scherer, K. R., (2005), A systems approach to appraisal mechanisms in emotion [Emotion and Brain], *Neural Networks*, *18 4*, 317–352, https://doi.org/https://doi.org/10.1016/j.neunet.2005.03.001

Savicki, V., & Kelley, M., (2000), Computer mediated communication: gender and group composition, *CyberPsychology & Behavior*, *3 5*, 817–826.

Schack-Nielsen, A., & Schürmann, C., (2008), Celf - A logical framework for deductive and concurrent systems (system description), *In* A. Armando, P. Baumgartner, & G. Dowek (Eds.), *Automated reasoning, 4th international joint conference, IJCAR 2008, sydney, australia, august 12-15, 2008, proceedings* (pp. 320–326), Springer, https://doi.org/10.1007/978-3-540-71070-7_28

Scherer, K. R., Schorr, A., & Johnstone, T., (2001), *Appraisal processes in emotion: theory, methods, research*, Oxford University Press.

Schubert, T., Friedmann, F., & Regenbrecht, H., (2001), The experience of presence: factor analytic insights, *Presence: Teleoperators & Virtual Environments*, *10 3*, 266–281.

Serino, A., Alsmith, A., Costantini, M., Mandrigin, A., Tajadura-Jimenez, A., & Lopez, C., (2013), Bodily ownership and self-location: components of bodily self-consciousness, *Consciousness and cognition*, *22 4*, 1239–1252.

Sheridan, T. B. et al., (1992), Musings on telepresence and virtual presence., *Presence Teleoperators Virtual Environ.*, *1 1*, 120–125.

Shernoff, E. S., Von Schalscha, K., Gabbard, J. L., Delmarre, A., Frazier, S. L., Buche, C., & Lisetti, C., (2020), Evaluating the usability and instructional design quality of interactive virtual training for teachers (ivt-t), *Educational Technology Research and Development*, 1–28.

Short, J., Williams, E., & Christie, B., (1976), *The social psychology of telecommunications*, Toronto; London; New York: Wiley.

Slater, M. et al., (1999), Measuring presence: a response to the witmer and singer presence questionnaire, *Presence: teleoperators and virtual environments*, *8 5*, 560–565.

Slater, M., (2003), A note on presence terminology, *Presence connect*, *3 3*, 1–5.

Slater, M., (2009), Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments, *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364 1535*, 3549–3557.

Slater, M., Lotto, B., Arnold, M. M., & Sanchez-Vives, M. V., (2009), How we experience immersive virtual environments: the concept of presence and its measurement, *Anuario de psicología*, *40 2*, 193–210, https://doi.org/http://hdl.handle.net/2445/49643

Slater, M., Pérez Marcos, D., Ehrsson, H., & Sanchez-Vives, M. V., (2009), Inducing illusory ownership of a virtual body, *Frontiers in neuroscience*, 29.

Slater, M., Pertaub, D.-P., Barker, C., & Clark, D. M., (2006), An experimental study on fear of public speaking using a virtual environment, *CyberPsychology & Behavior*, *9 5*, 627–633.

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R., (2000), Small-group behavior in a virtual and real environment: a comparative study, *Presence*, *9 1*, 37–51.

Slater, M., Spanlang, B., & Corominas, D., (2010), Simulating virtual environments within virtual environments as the basis for a psychophysics of presence, *ACM transac-*

tions on graphics (TOG), *29 4*, 1–9, https://doi.org/https://doi.org/10.1145/1778765.1778829

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O., (2010), First person experience of body transfer in virtual reality, *PloS one*, *5 5*, e10564.

Slater, M., Steed, A., McCarthy, J., & Maringelli, F., (1998), The influence of body movement on subjective presence in virtual environments, *Human factors*, *40 3*, 469–477.

Stanney, K. M., Kennedy, R. S., & Drexler, J. M., (1997), Cybersickness is not simulator sickness, *Proceedings of the Human Factors and Ergonomics Society annual meeting*, *41 2*, 1138–1142.

Steed, A., Pan, Y., Zisch, F., & Steptoe, W., (2016), The impact of a self-avatar on cognitive load in immersive virtual reality, *2016 IEEE virtual reality (VR)*, 67–76.

Stein, D. J., Lim, C. C., Roest, A. M., De Jonge, P., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Benjet, C., Bromet, E. J., Bruffaerts, R., et al., (2017), The cross-national epidemiology of social anxiety disorder: data from the world mental health survey initiative, *BMC medicine*, *15 1*, 1–21.

Stein, M. B., Torgrud, L. J., & Walker, J. R., (2000), Social phobia symptoms, subtypes, and severity: findings from a community survey, *Archives of general psychiatry*, *57 11*, 1046–1052.

Stone III, W. B., (2017), *Psychometric evaluation of the simulator sickness questionnaire as a measure of cybersickness* (Doctoral dissertation), Iowa State University.

Straightlabs GmbH & Co.KG, (2023), *Speech trainer*. https://vr-speech.com/en/ (accessed: 23.01.2023)

Swartout, W. R., Gratch, J., Hill Jr, R. W., Hovy, E., Marsella, S., Rickel, J., & Traum, D., (2006), Toward virtual humans, *AI Magazine*, *27 2*, 96–96, https://doi.org/https://doi.org/10.1609/aimag.v27i2.1883

Thompson, P. A., & Marchant, E. W., (1995), A computer model for the evacuation of large building populations, *Fire safety journal*, *24 2*, 131–148, https://doi.org/https://doi.org/10.1016/0379-7112(95)00019-P

Ting-Toomey, S., & Dorjee, T., (2018), *Communicating across cultures*, Guilford Publications.

Treisman, M., (1977), Motion sickness: an evolutionary hypothesis, *Science*, *197 4302*, 493–495.

Tsakiris, M., Prabhu, G., & Haggard, P., (2006), Having a body versus moving your body: how agency structures body-ownership, *Consciousness and cognition*, *15 2*, 423–432.

Ulicny, B., & Thalmann, D., (2001), Crowd simulation for interactive virtual environments and vr training systems. *Computer animation and simulation 2001* (pp. 163–170), Springer, https://doi.org/https://doi.org/10.1007/978-3-7091-6240-8_15

Ulicny, B., & Thalmann, D., (2002), Towards interactive real-time crowd behavior simulation, *Computer Graphics Forum*, *21 4*, 767–775, https://doi.org/https://doi.org/10.1111/1467-8659.00634

Väljamäe, E., Larsson, P., Västfjäll, D., & Kleiner, M., (2004), Auditory presence, individualized head-related transfer functions, and illusory ego-motion in virtual environments, *in in Proc. of Seventh Annual Workshop Presence 2004*, https://doi.org/https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

van Waterschoot, J., Bruijnes, M., Flokstra, J., Reidsma, D., Davison, D., Theune, M., & Heylen, D., (2018), Flipper 2.0: a pragmatic dialogue engine for embodied conversational agents, *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 43–50.

Varner, D., Scott, D., Micheletti, J., & Aicella, G., (1998), Umsc small unit leader nonlethal trainer, *Proc. ITEC*, *98*.

Virre, E., (1996), Virtual reality and the vestibular apparatus, *IEEE engineering in medicine and biology magazine*, *15 2*, 41–43.

VirtualSpeech, L., (2022), *Virtualspeech*. https://virtualspeech.com/ (accessed: 28.03.2022)

VRSpeaking, LLC, (2022), *Ovation*. https://www.ovationvr.com/ (accessed: 28.03.2022)

Wallach, H. S., Safir, M. P., & Bar-Zvi, M., (2009), Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial, *Behavior modification*, *33 3*, 314–338, https://doi.org/https://doi.org/10.1177/0145445509331926

Wang, I., & Ruiz, J., (2021), Examining the use of nonverbal communication in virtual agents, *International Journal of Human–Computer Interaction*, *37 17*, 1648–1673, https://doi.org/10.1080/10447318.2021.1898851

Weech, S., Kenny, S., & Barnett-Cowan, M., (2019), Presence and cybersickness in virtual reality are negatively related: a review, *Frontiers in Psychology*, *10*, https://doi.org/10.3389/fpsyg.2019.00158

Williford, J. S., Hodges, L. F., North, M. M., & North, S. M., (1993), Relative effectiveness of virtual environment desensitization and imaginal desensitization in the

treatment of acrophobia, *Proceedings of Graphics Interface '93*, 162–162, http://graphicsinterface.org/wp-content/uploads/gi1993-21.pdf

Witmer, B. G., & Singer, M. J., (1998), Measuring presence in virtual environments: a presence questionnaire, *Presence*, *7 3*, 225–240.

Wittchen, H.-U., & Fehm, L., (2003), Epidemiology and natural course of social fears and social phobia, *Acta Psychiatrica Scandinavica*, *108*, 4–18.

Yee, N., & Bailenson, J., (2007), The proteus effect: the effect of transformed self-representation on behavior, *Human communication research*, *33 3*, 271–290.

Yngve, V. H., (1970), On getting a word in edgewise, *Chicago Linguistics Society, 6th Meeting, 1970*, 567–578.

Young, R. M., (1999), Notes on the use of plan structures in the creation of interactive plot, *AAAI fall symposium on narrative intelligence*, 164–167.

Zimmons, P., & Panter, A., (2003), The influence of rendering quality on presence and task performance in a virtual environment, *IEEE Virtual Reality, 2003. Proceedings.*, 293–294.

COLLEGE MATHSTIC
DOCTORAL BRETAGNE
BRETAGNE OCEANE

ÉCOLE NATIONALE D'INGÉNIEURS DE BREST

Titre : Simulation et perception d'audiences en réalité virtuelle.

Mot clés : Réalité virtuelle, Agent Virtuel, Audience virtuelle, Comportement non-verbal

Résumé :

De nombreuses applications d'entraînement ou de thérapie simulent des audiences en réalité virtuelle (RV) pour fournir des environnements sûrs et écologiques. Cependant, la simulation d'audience en RV présente de nombreux défis liés à la création d'attitudes à partir du comportement non-verbal des personnages qui la composent. De plus, en RV le nombre de personnages, aussi appelé agents, leurs animations ou leur réalisme peuvent aussi créer des problèmes de performances. Les modèles de comportement utilisés dans ces systèmes ne sont pas directement évalués en RV et se basent sur des études en ligne ou l'avis d'experts du domaine d'application. Aussi, la différence de technologie et la subjectivité de la perception utilisateur pourraient influencer les résultats de ces évaluations. Nous proposons donc un modèle de comportement d'audience évalué en RV qui génère les comportements non-verbaux de ses membres à partir d'une attitude donnée, cela dans le but d'améliorer la qualité des audiences et faciliter leur utilisation dans des scénarios pédagogiques et thérapeutiques. Nous présentons une série d'évaluations des performances et de la perception utilisateur en RV visant à valider la capacité du système à simuler différents types d'attitudes (ennuyé, intéressé ou critique) tout en préservant une expérience de RV optimale et en offrant une application de contrôle de haut niveau facilitant le changement d'attitude en temps réel, notamment pour la création de scénarios de formation. Enfin nous validons la faisabilité du déploiement de ce modèle dans des applications d'entraînement et de thérapie par l'exposition utilisées par des professionnels.

Title: Audience simulation and perception in virtual reality.

Keywords: Virtual reality, Virtual Agent,Virtual audience, non-verbal behaviour

Abstract: Many training or exposure therapy applications simulate audiences in virtual reality (VR) to provide safe and ecological environments. However, audience simulation in VR presents many challenges related to creating attitudes from the agents' non-verbal behaviour. Furthermore, in VR, the animations and the realism of the characters, also called agents, can also create performance problems. The behaviour models used in these systems are not directly evaluated in VR and rely on online studies or the application domain experts' knowledge. Thus, the difference in technology and the user perception subjectivity could influence evaluations' results. Therefore, we propose an audience behaviour model evaluated in VR that generates the non-verbal behaviours of its members from a given attitude to improve the audiences' quality and facilitate their use in educational and therapeutic scenarios. We present a series of performance and user perception evaluations in VR aiming at validating the system's ability to simulate different types of attitudes (bored, interested or critical) while preserving an optimal VR experience and providing a high-level control application facilitating seamless attitude change in real-time, notably for the creation of training scenarios. Finally, we validate the deployment feasibility of this model in training and exposure therapy applications used by professionals.