Simulating Interactions with Virtual Characters for the Treatment of Social Phobia

THÈSE N^{O} 4466 (2009)

PRÉSENTÉE LE 29 JUILLET 2009

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE RÉALITÉ VIRTUELLE
PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Helena GRILLON

acceptée sur proposition du jury:

Prof. C. Petitpierre, président du jury
Prof. D. Thalmann, directeur de thèse
Prof. O. Blanke, rapporteur
Dr E. Klinger, rapporteur
Prof. P. Sharkey, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2009

CONTENTS

Remerciements

Je tiens à remercier le professeur Daniel Thalmann, mon directeur de thèse, pour m'avoir donné l'opportunité de faire ce doctorat. Merci également aux membres de mon jury, docteur Evelyne Klinger, professeur Paul Sharkey, professeur Olaf Blanke et professeur Claude Petitpierre.

Merci au docteur Françoise Riquier et à son équipe de stagiaires, Diana Felli, Caroline Nicod et Johann Caffaro, pour leur aide et leur collaboration dans le domaine de la psychologie et de la psychiatrie.

Merci à toute l'équipe du VRLab, passée et présente, pour les discussions intéressantes, les bons moments passés ensemble et la bonne ambiance de travail. En particulier, je souhaite remercier Barbara Yersin, Jonathan Maïm et Damien Maupu pour leur bonne humeur et leur amitié, et Achille Peternier pour avoir relu une grande partie de ce texte. Je souhaite également remercier Ronan Boulic pour ses bons conseils et sa disponibilité et bien sûr Mireille, pour son travail de design de qualité.

Un grand merci à Bruno Herbelin et Julien Pettré pour leurs conseils avisés et nos collaborations. Merci à Josiane Bottarelli pour sa gentillesse et sa bonne humeur.

Merci aussi à toutes les personnes que j'ai pu rencontrer en conférences et avec qui j'ai pu avoir des discussions extrêmement intéressantes et constructives.

Merci à Pierre-Alban, Daniela, Lili, Paul, Johanna, Véronique et Amparo, membres de ma famille et amis, pour leur soutien inconditionnel.

Enfin, un merci tout particulier à Benoît Le Callennec pour son soutien, sa disponibilité, ses conseils et son aide dans les moments de doute comme dans les moments heureux.

Résumé

La Réalité Virtuelle (RV) est devenue un outil pratique pour les thérapeutes dans le traitement des phobies. Elle permet de simuler des situations difficiles à reproduire dans la vie réelle, et de les répéter autant que voulu. De plus, elle permet d'avoir un contrôle de la situation. La simulation peut être arrêtée si le patient ne peut pas la gérer ou être affinée afin d'être graduelle. La Thérapie par Exposition à la Réalité Virtuelle (TERV) s'est montrée efficace dans le contexte de différentes phobies telles que l'acrophobie. Les phobies sociales, par contre, sont beaucoup plus difficiles à gérer. En tant qu'humains, nous sommes experts en représentations et comportements humains. Cela rend beaucoup plus difficile d'obtenir des résultats crédibles et immersifs.

Dans cette thèse, nous décrivons une série d'outils et d'applications que nous avons développés, pour la TERV des phobies sociales et de l'agoraphobie avec foules. Nous décrivons comment nous créons différents scénarios pour la TERV de la phobie sociale. Nous exposons également un programme que nous avons créé, permettant des interactions élaborées entre un utilisateur et des personnages virtuels. Plus particulièrement, nous avons conçu et implémenté un programme permettant à nos personnages virtuels de changer de comportement en fonction du comportement visuel de l'utilisateur. Il permet de les rendre intéressés lorsqu'ils sont regardés et distraits lorsqu'ils sont évités.

Nous décrivons ensuite un modèle que nous avons implémenté permettant aux personnages d'une foule d'avoir des comportements visuels. Ceci consiste en une méthode qui détecte automatiquement où et quand chaque personnage doit regarder, ainsi qu'en un moteur de cinématique inverse dédié afin que les personnages puissent satisfaire les contraintes définies par les points d'intérêt préalablement détectés. Ils peuvent ainsi effectuer des mouvements de regard de façon naturelle. Nous décrivons ensuite une architecture que nous avons développée qui combine le travail que nous avons effectué dans le domaine de la phobie sociale et ce modèle d'attention visuelle. Nous avons utilisé notre modèle d'attention visuelle afin de permettre aux personnages dans la foule de se regarder. Nous avons également utilisé la capture oculaire et de la capture de mouvements afin de définir où un utilisateur regarde dans un environnement CAVE. Les personnages répondent alors en regardant l'utilisateur, en regardant là où l'utilisateur regarde, ou en regardant d'autres personnages dans la foule. Nous avons ainsi obtenu un environnement immersif et interactif pour la TERV dans le domaine de l'agoraphobie.

Dans la troisième partie de cette thèse, nous décrivons diverses expériences que nous avons faites afin de valider nos divers outils et applications. Notre première étude consiste en l'utilisation de la RV dans un HMD. Nous y utilisons également un système de capture de mouvements oculaires afin d'analyser les comportements d'évitement du contact visuel avant et après la thérapie. Nous discutons ensuite l'utilisation de cette capture oculaire en tant qu'outil afin d'aider à l'évaluation et au diagnostic de la phobie sociale. Comme les comportements d'évitement du contact visuel sont fréquents chez les personnes souffrant de telles phobies, la capture oculaire peut s'avérer être un outil précieux. Nous décrivons une expérience dans laquelle nous avons testé la capture oculaire en tant qu'outil de diagnostic et d'évaluation. Nous décrivons ensuite une expérience que nous avons menée afin d'évaluer le potentiel de notre application permettant des interactions entre utilisateur et personnages virtuels. Pour finir, nous décrivons l'expérience que nous avons conduite afin d'évaluer l'application que nous avons développée dans le contexte de l'agoraphobie avec foules.
Mots clés : Thérapie d'Exposition par Réalité Virtuelle, Phobie Sociale, Agoraphobie avec foules, Comportements d'attention, Eye-tracking, Interactions Humain - Virtuel

Abstract

Virtual Reality (VR) has nowadays become a very useful tool for therapists in the treatment of phobias. Indeed, it allows the simulation of scenarios which are difficult to reproduce in real life. It also allows for a situation to be repeated as much as one wants. Moreover, it allows for a complete control over the situation. The simulation can be stopped if the patient cannot handle it. It can also be tweaked for gradual exposure. Virtual Reality Exposure Therapy (VRET) has proven to be efficient in the context of phobias such as acrophobia or the fear of flying. Social phobia, however, are much harder to deal with. Indeed, as humans, we are experts in human representations and behaviors; it makes it much harder to obtain credible and immersive environments.

In this thesis, we describe a set of tools and applications which we have developed to be used in VRET of social phobia and agoraphobia with crowds. We first describe how we create different scenarios for VRET of social phobia. We then expose the application we have developed which allows for elaborate interactions between a user and virtual characters. In particular, we have designed and implemented a software which allows for virtual characters to change behavior depending on the user's eye contact behavior. It allows them to seem interested when being looked at and distracted when not.

We then describe the model we have implemented to simulate gaze attention behaviors for crowds of virtual characters. This consists of a method that automatically detects where and when each virtual character in a crowd should look. Secondly, it consists of a dedicated gaze Inverse Kinematics (IK) solver in order for the virtual characters to satisfy the constraints defined by the automatically detected points to be looked at. This allows for the characters to perform the looking motion in a natural and human like way. We then describe the architecture we have developed to combine the work we have done in the domain of social phobia and this model of attention behaviors for crowd characters. We thus use our model of looking behaviors to allow for crowd characters to look at each other. We also use eye-tracking and optical motion capture to determine where a user is looking in a CAVE environment. The virtual characters then respond by either looking at the user, looking at what the user is looking at, or looking at other characters in the crowd. We thus obtain an immersive and interactive environment for VRET in the domain of agoraphobia with crowds.

The third part of this thesis describes various experiments we have conducted in order to validate our applications. Our first study consists of using VR in a head-mounted display (HMD) for the treatment of social phobia. In this study, we also use eye-tracking in order to analyze eye contact avoidance behaviors before and after therapy. We then discuss the use of eye-tracking as a tool to help assess and diagnose social phobia. Since eye contact avoidance behaviors are frequent in people suffering from such phobias, eye-tracking can certainly be a helpful tool. We describe an experiment in which we tested eye-tracking as a diagnosis and assessment tool on a phobic population and on a control group. We also describe an experiment to evaluate the potential of our proposed interaction loop in the context of social phobia. Finally, we describe the experiment we have conducted to evaluate our application in the context of agoraphobia with crowds.

Keywords: Virtual Reality Exposure Therapy, Social Phobia, Agoraphobia with Crowds, Attention Behaviors, Eye-tracking, Human - Virtual Human Interaction

CHAPTER 1

Introduction

1.1 Context and Motivations

In the domain of treatment of social anxiety disorders, one of the principles used in Cognitive Behavioral Therapy (CBT) is exposure to anxiety provoking situations [Wolpe, 1969; Marks, 1987]. This consists of gradually exposing the patient to the fearful situation in order to reduce anxiety and encourage confrontation behavior. Traditional exposure therapies use desensitization techniques which require for the patient to either imagine the fearful situation, or expose themselves to it in-vivo. Nowadays, exposure to Virtual Reality (VR) presents itself as an alternative to standard in-vivo exposures.

Indeed, obstacles to the traditional techniques are numerous. For example, some patients are unable to imagine the anxiety provoking situation. It can also be difficult for the therapist to know exactly what the patient is imagining. There can be a strong rejection from the patient towards in-vivo exposures. Moreover, these are difficult to control or repeat. Finally, it may be difficult to respect the patient's privacy in in-vivo situations.

VR techniques can be used to overcome difficulties which are inherent to the traditional mode of treatment of anxiety disorders. On one hand, they allow controlled exposure of the patient to complex, dynamic and interactive stimuli in 3D. On the other hand, they allow the evaluation and treatment of the patient's cognitive, behavioral and functional performances. These exposures take place in the therapist's office and therefore respect patient intimacy [North et al., 1998]. VR offers anxiety provoking scenarios which are difficult to access and are not easily available in real life. As an example, it would be extremely difficult for a therapist to fill his/her office with spiders in order to treat a patient suffering from arachnophobia. Equally, it would be extremely expensive and time consuming to repeatedly take a patient on an airplane in order to treat him/her against fear of flying. VR also allows the repetition of exposures without limitations. For example, a job interview is an accessible but exceptional situation. It would be difficult to have to go through a job interview every week, as a habituation exercise.

Different factors contribute to the efficacy of VR exposure. The main ones are:

- The patient has to feel present in the Virtual Environment (VE) and have the possibility to experiment it in a subjective way.
- The stimuli delivered by the VE have to evoke emotions such as anxiety.
- The learned behaviors and the changes in the way of thinking have to be generalized to the real situations.

Many studies have already been conducted in the domain of Virtual Reality Exposure Therapy (VRET) and have proven to be efficient in the case of phobias such as acrophobia or fear of flying. However, when dealing with social phobia or agoraphobia with crowds, it is much harder to obtain good results. Indeed, as human beings, we are experts in human representations and behaviors. This makes it much harder to create credible virtual characters. Moreover, to help treat social phobia with VR, there is a real need to endow the virtual characters with possibilities to interact and even with a form of empathy. A user should be able to feel that the virtual character understands him/her and that it reacts to what he/she is doing or saying. Also, a user should be able to understand the virtual character's reactions and associate them with those that a real person could have.

In this thesis, we describe the tools and applications that we have created and developed in order to allow for this interaction between a user and virtual characters. We have focused on the non-verbal interactions and more specifically on gaze and eye contact. As discussed by Cassel and Thorisson, non-verbal feedback is the most important of all feedbacks, at least concerning virtual characters [Cassell and Thorisson, 1999]. Moreover, the first study we conducted, presented in Chapter 6.2 together with literature have motivated us to focus on eye contact. It is one of the most important non-verbal behaviors (if not the most) and is often avoided by people suffering from social phobia. We thus create interactions between users and virtual characters based on gaze. We also describe a model we have developed to simulate gaze attention behaviors for virtual characters, applicable to social phobia (a single or few characters) and agoraphobia with crowds (hundreds of characters) in order for those interactions to seem natural and human like.

1.2 VRET for Social Phobia versus Agoraphobia

In the first part of this thesis, we present work conducted in the context of social phobia. We describe the various tools and applications we have developed for the VRET of social phobia. Social phobia, also known as social anxiety disorder or social anxiety, is a diagnosis within psychiatry and other mental health professions which refers to an excessive anxiety in social situations. These cause considerable distress and impair people's ability to function in areas of daily life. The diagnosis can be of a specific disorder (when only some specific situations are feared) or a generalized disorder. The main characteristic of social phobia is the underlying fear of being negatively evaluated or judged by others [American Psychiatric Association, 1994]. Patients suffering from social phobia fear of being embarrassed or humiliated. Though these fears may be recognized by the patient as excessive or unreasonable, it can be extremely difficult to overcome them. Phobic patients completely or partially avoid certain situations by minimizing visual contact, verbal communication or physical presence. They rely on avoidance behaviors to calm down their fears. Social phobia is an anxiety disorder which concerns up to 13.3% of the population [Heimberg, 1995] and has an important social impact on the way people live their lives. The aim of the tools and applications we have developed is to assist therapists in the treatment of social phobia. Exposure to virtual scenarios can replace in-vivo exposure in many contexts and can be much easier to use.

The second part of this thesis is focused on agoraphobia with crowds. Agoraphobia is an abnormal and persistent fear of public places or open areas. It is an anxiety provoked by being in places or situations from which escape is difficult. It can also be a fear of being in a place where it is impossible to obtain help in case of panic attack. Different forms of this anxiety translate in a fear of open spaces, traveling, or crowds [American Psychiatric Association, 1994]. It is this last specific type of agoraphobia that we work on in this thesis. Agoraphobia is an anxiety disorder which occurs twice as much in women than in men and concerns 2% to 4% of the population. Here as well, the aim of the application we have developed is to provide therapists with a tool which can be used in the treatment of social phobia. The benefit which can be found in using VRET for the treatment of agoraphobia with crowds is even more conspicuous since it is a situation very difficult to regulate in vivo. Our application allows for gradual exposure to virtual crowds in a very immersive context.

Finally, in the last part of this thesis, we describe the various experiments we have conducted in order to validate our various tools and applications in the context of social phobia and agoraphobia with crowds.

1.3 Contributions

In the first part of this thesis, we desribe the various tools and applications we have developed to be used in VRET to treat social phobia and more specifically, fear of public speaking. Previous studies on social phobia together with our first experiments in the domain, have raised our interest towards eye contact analysis and simulation. Eye contact is a crucial part of our interactions with others, and it is a known to be often avoided by people suffering from social phobia. In this context, we first describe our methods to create various VRET scenarios and to endow our virtual characters with various behaviors. We then describe a behavioral model we have developed for characters to seem aware of the presence of the user or the phobic patient and to interact with him/her.

In the second part of this thesis, we focus on VRET of agoraphobia with crowds. In this context, we describe a gaze attention model we have developed for virtual crowd characters. In this model, we automatically detect where and when each character should look and adapt their motion in order for them to perform the gaze motions in a smooth and natural way.

Finally, we extend both this attention model and the interaction possibilities proposed in the context of social phobia to agoraphobia with crowds in an immersive CAVE environment. We thus propose a very immersive simulation in which a user or patient can interact with crowd characters using gaze.

In the last part of this thesis, we describe the various experiments we have conducted in order to validate our methods and applications. In the first one, we test a VRET protocol together with eye contact behavior analysis. In our second experiment, we test eye-tracking as a diagnosis and assessment tool for abnormal gaze behaviors known to be present in social phobics. Our third experiment tests the interaction model we propose in the context of social phobia. Finally, our last experiment tests the combination of our interaction and attention models in an immersive CAVE environment, in the context of agoraphobia with crowds.

1.4 Preliminaries

Since we have worked on the treatment of social phobia with VR, some terms in this text belong to the therapeutic language. These may not be familiar to all readers. We have thus regrouped a number of definitions and concepts in the Glossary (Appendix D), at the end of this document.

All abbreviations used in this thesis are listed at the end of this document in Appendix C. Each expression replaced by an abbreviation will be introduced as complete words followed by the abbreviation in parentheses, the first time it occurs in the text. Further references to these expressions will be done by abbreviation only.

Concerning mathematical notations and conventions, this document uses the following. Scalar numbers are denoted by small letters such as s. Vectors and quaternions are indicated by small, boldface letters such as $\boldsymbol{v}$ and $\boldsymbol{q}$. Matrices are expressed as capital, boldface letters such as $\mathbf{M}$. Additional notations are introduced when necessary.

Finally, all images in this document which are not owned by the VRLab are courtesy of their respective authors.

1.5 Organization of this Document

In Chapter 2 we discuss the state of the art in the domains of VRET for the treatment of phobias, and more specifically of social and agoraphobia. We also discuss the use of eye-tracking as a diagnosis and assessment tool and as mode of interaction. Finally, we also go over models of visual attention developed for virtual humans and motion editing techniques.

This thesis is then separated into three distinctive parts.

Chapter 3 is dedicated to the various tools and applications we have developed in the context of social phobia. More specifically, we discuss the various scenarios we have setup in this context. We describe the use of scripting to guide the interactions between users and virtual characters. Finally, we describe how we use this together with eye-tracking to modify virtual character behavior with respect to human eye contact behavior.

Then, Chapters 4 and 5 discuss the applications we have developed in the context of agoraphobia with crowds. More specifically, in Chapter 4 we present an architecture which allows the addition of attention behaviors to crowds of virtual characters. This model automatically detects where and when each character should look. It equally defines a form of how to look. The set of rules used in this architecture are derived from biomechanics and occulometric measures. In Chapter 5 we describe a real-time version of the attention model presented in the previous chapter. This model combines visual attention for crowd characters and interactions with the user. This architecture allows for a user to interact with crowd characters using eye-tracking. Thus, depending on where a user is looking, in a CAVE environment, the characters composing the crowd can either look at the user, look at other characters, or look at where the user is looking.

Finally, Chapters 6 and 7 are dedicated to our experimental validations. In Chapter 6, we describe three different experiments conducted in the context of social phobia. In the first one, we describe the experimental protocol we have used on 8 social phobic patients and the results obtained through VRET using our application. We also discuss the patients' results obtained through various rating scales and eye-tracking. In the second experiment, we describe how eye-tracking can be used as assessment and diagnosis tool for VRET. We present the study we have conducted in order to demonstrate the efficacy of this tool. We equally present the experimental protocol used over 5 social phobic subjects and 5 control subjects and present the results obtained through eye-tracking measures. The third experiment describes a study we have conducted in order to validate our character behavior modification model. This study was conducted on a cohort of 12 healthy subjects. We first describe our experimental protocol and then present the subjects' results using various rating scales. We also describe a case study we have conducted over a young girl suffering from Asperger syndrome. In Chapter 7, we present an experiment on the immersive capacities of the architecture we have developed in the context of agoraphobia on 30 healthy subjects. We discuss the results obtained by subjective ratings as well as eye-tracking and discuss how this model can be used in VRET of agoraphobia.

Finally, we conclude this thesis in Chapter 8.

CHAPTER 2

State of the Art

2.1 Virtual Reality Exposure Therapy

It has now been almost 20 years that VR technologies have been experimented and evaluated in order to treat phobias. The objective of VRET sessions is to desensitize the phobic patient by exposing him/her to the anxiety provoking stimulus.

Many studies have been conducted regarding the use of VR in the treatment of phobias. Traditionally, during in-vivo exposure, patients are asked to evaluate which 8 situations are the most anxiety provoking for them. Then, they are gradually exposed to the fearful situations and guided by the therapist. This is done in order to habituate the patient to the anxiety provoking situation. In VRET, the same principle is used, but the patients are exposed to computer generated scenes and scenarios instead of the in-vivo situation.

The first and most frequent type of phobia which has been tackled by VRET is acrophobia, or the fear of heights. One of the pioneer studies using VRET was conducted by Rothbaum et al. [Rothbaum et al., 1995]. They designed their environment using Wavefront software (now belonging to Autodesk [Autodesk, 2009]) which they then integrated into a VR application called The Simple Virtual Environment Library [Verlinden et al., 1993]. This simple application allowed the loading of the scene composed of hierarchically grouped objects and the creation of events for walkthroughs (e.g. pressing a key on the keyboard to move forward in the VE). They presented a case study in which a patient was gradually exposed to increasingly anxiety provoking situations through a Head Mounted Display (HMD). Their results indicated a significant habituation from the patient regarding both anxiety symptoms and avoidance of anxiety provoking situations. A second study was conducted by Hodges et al. using the same software and hardware [Hodges et al., 1995]. They created three types of

Figure 2.1: Various environments available at virtually better [VirtuallyBetter, 2009]. *Left:* Fear of driving. *Middle:* Virtual Iraq. *Right:* Panic attacks.

environments, namely an elevator, a series of balconies and a series of bridges, which they then tested over half of 20 acrophobic students; the other half being in a wait-list control group. Their results showed significant differences between subjects having been exposed to VRET and those of the control group. Emmelkamp et al. then compared in-vivo exposure with VRET using low-cost VR hardware and software [Emmelkamp et al., 2001, 2002]. They designed two situations: a diving board and an elevator using Superscape VRT 5.0 software, a commonly used VR modeling and visualization toolkit [superscape, 2009]. Here as well, they used an HMD to visualize the scene in 3D. Their results showed that VRET allowed the reduction of the levels of anxiety and avoidance. They equally concluded that VRET proved to be as efficient as the in-vivo exposure.

Even though all these setups only worked at a frame rate of 10 Frames Per Second FPS, the efficacy of VRET in the treatment of acrophobia was from then on well established.

Following these pioneer developments and studies, VRET has been experimented on many other types of phobias. Several early applications have been developed and tested (under the form of case studies) for the treatment of fear of flying [Rothbaum et al., 1996; North and Coble, 1997; Wiederhold et al., 1998]. These were later followed by more complex environments and complete studies [Rothbaum et al., 2000, 2002; Wiederhold et al., 2002; Botella et al., 2004b]. Other applications have been developed and tested in the treatment of arachnophobia [Carlin et al., 1997; Garcia Palacios et al., 2001; Bouchard et al., 2006]. Post Traumatic Stress Disorder (PTSD) has also been worked on by several researchers; for veterans of the war in Vietnam [Rothbaum et al., 1999, 2001], for people suffering from PTSD following the World Trade Center events on September 11th 2001 [Difede and Hoffman, 2002; Difede et al., 2002], and for soldiers coming back from the wars in Iraq and Afghanistan [Rizzo et al., 2005; Spira et al., 2006; Reger et al., 2009].

Other developments and studies have also been done in the context of claustrophobia, fear of driving, fear of many kinds of animals, attention deficits, pain distraction, eating disorders and many more.

What we are more interested in is the work which has been done in the domain social phobia and agoraphobia with crowds. By extension, this implies work which has been done in creating empathy between virtual characters and human beings.

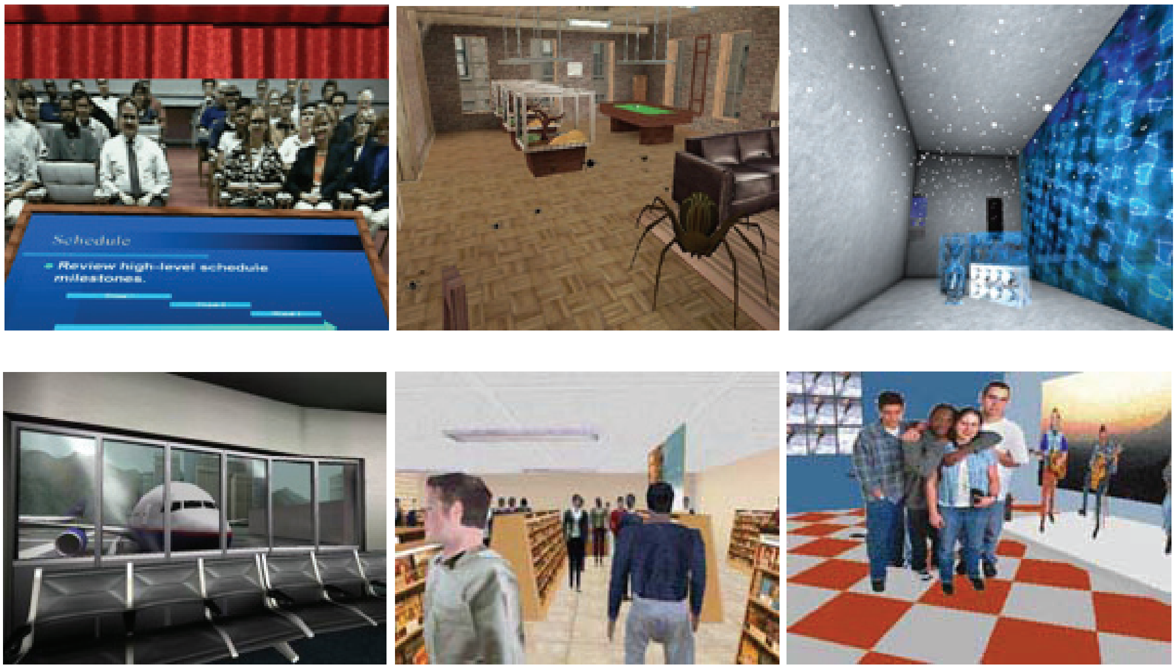

Figure 2.2: Various scenarios available at vrphobia [vrphobia, 2009]. *Top left:* Fear of public speaking. *Top middle:* Arachnophobia. *Top right:* Pain distraction. *Bottom left:* Fear of flying. *Bottom middle:* Panic attack. *Bottom right:* Social phobia.

2.1.1 VRET For Social Phobia

The first application to be used in VRET of fear of public speaking was proposed by North et al. [North et al., 1998]. Their scenes were created using VREAM Virtual Reality Development Software Package and Libraries [VREAM, 2009]. They created a model of an auditorium which could contain up to 100 characters. The software was designed to allow for the characters to enter the auditorium one by one and then by groups of five until it was filled. They equally used pre-recorded audio to simulate various audience responses such as clapping, laughing, or encouragements. A loudspeaker was plugged in during the sessions, allowing for users to hear the echo from their voice. They tested this application on 8 subjects suffering from fear of public speaking. The treatment consisted in 5 weekly sessions of $10 - 15$ minutes each: the patient, standing behind the wooden podium had to talk to the auditorium. During the session, the therapist could vary the number of people in the audience and their attitudes by alternating between different pre-recorded video sequences. A control group of 8 other subjects was exposed to a trivial virtual reality scene. The control subjects were asked to manage their phobia by using visualization techniques or self- exposure. The authors explain that the patients from the VRET group experimented various physical and emotional impacts (heart rate acceleration, damp hands, loss of balance, etc.) similar to those felt during the in vivo exposures. This study showed that VRET was able to reduce patients' anxiety when facing a public which was not the case for the control group subjects.

Slater et al. used the DIVE (Distributed Interactive Virtual Environment) developed by the Swedish Institute of Computer Science [SICS, 2009] to create a public speaking simulation [Slater et al., 1999]. They created a model of a seminar room in which were seated

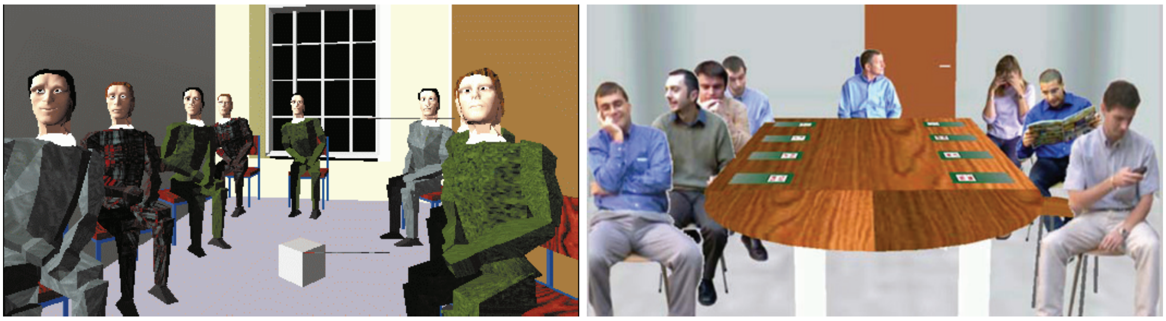

Figure 2.3: *Left:* A negative audience greeting the speaker [Slater et al., 1999]. *Right:* Public speaking exercise in front of an audience [Klinger et al., 2005].

8 characters. These characters presented random autonomous behaviors such as twitches, blinks and nods. In order to make them move, they used TCL (Tool Command Language) scripts attached to the various character body parts in the DIVE database. In addition, they simulated eye contact and various facial expressions for these characters. An example of their model is depicted on the left of Figure 2.3. The authors tested their setup on 10 students in 3 different public speaking situations. In the first, they had to do a talk in front of an either very hostile or very friendly audience. In the second, they had to repeat their talk in front of the audience they had not faced during the first situation. Finally, in the third, they all started off with a very hostile audience, which then became friendly. The results obtained by this study show that an audience with higher interest reduces anxiety due to public speaking. This study was then extended to 40 subjects [Pertaub et al., 2001, 2002] and confirmed the results to the first one. Finally, the same team of researchers pursued these studies by comparing behaviors and responses between 16 phobic subjects and 20 non phobic ones [Slater et al., 2004]. The subjects were all asked to do a public speaking exercise in an empty room and in front of a 5-people audience. Their results showed that non-phobic people reacted in the same way in both situations. However, phobic people felt increased anxiety in front of the audience as compared to the empty room.

Harris et al. proposed a scenario depicting an auditorium [Harris et al., 2002]. In their system, the environment starts off by being empty. It then gradually fills up with characters having either positive or negative attitudes. The characters could either applaud, to encourage the user, or talk between themselves and laugh, thus not pay attention to what the user was saying. Moreover, they allowed the characters to talk louder an louder in order to amplify discomfort. They then tested their system on 8 students suffering from fear of public speaking. Their results were compared to those of a control group of 6 people. The authors report that the difference between the two sets of students relied in their confidence in public speaking situations but not in their answers to the different anxiety scales.

Herbelin et al. used symbolic representations of characters, depicted on the left of Figure 2.4, to create a virtual audience of people [Herbelin et al., 2002]. The authors used snapshots of eyes which they placed in the environment in order to create their audience. The snapshots were placed in concentric circles around the user. Their system then allowed the definition of the number of circles, the number of snapshots and the distance to the user in order to change induced anxiety. They asked 10 voluntary participants to give a speech to

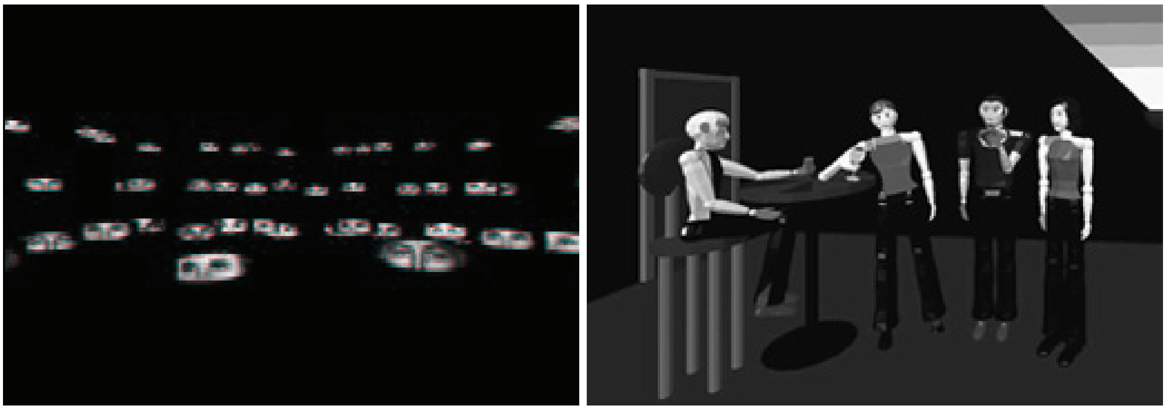

Figure 2.4: *Left:* An audience consisting of picture of eyes [Herbelin et al., 2002]. *Right:* Disinterested characters in bar environment [James et al., 2003].

the virtual audience in an HMD. From their results, the authors concluded that VRET had a high potential in replacing in-vivo situations.

Anderson et al. proposed a VR scenario consisting in a classroom with a virtual desk on which text was written [Anderson et al., 2003]. This text could be used as notes for a public speaking exercise. They used video inlay to insert an audience of 5 people in the virtual classroom. The authors then tested their application on two social phobic patients exposed to the scenario through an HMD. Their results proved this treatment to be efficient.

James et al. proposed a two-fold scenario to evaluate the potential of generating social anxiety in VEs [James et al., 2003]. The first part consisted in an underground train in which the characters expressed neutral behaviors. The second, depicted on the right of Figure 2.4 was a bar in which the characters seemed very disinterested. They simulated those behaviors with gaze and pre-recorded sentences as in [Slater et al., 1999]. They then asked 10 subjects to evaluate different scenarios. Their results showed that a socially demanding VE is more anxiety provoking for a phobic subject than a non-socially demanding one.

Within the VEPSY Updated project [vepsy, 2009], Klinger et al. proposed a system to tackle various factors of social phobia, namely, performance, scrutiny, intimacy and assertion [Klinger et al., 2002; Roy et al., 2003]. To this end they created four different environments depicting a meeting room, outside a coffee shop, an appartment, and a walkthrough going from an appartment to a shop, including a lift. These were done in 3dsMax [Autodesk, 2009] and integrated into an interactive environment using Virtools Dev 2.0 [Dassault Systèmes, 2009], providing the rendering engine as well as a behavioral engine to define the various possible interactions between the user and the virtual characters or objects in the environment (e.g. pull a chair with the mouse). For the characters, they used 3D sprites. They used simple billboards on which were projected pictures of real people involved in daily situations. Finally, they used background sound and pre-recorded sentences for the dialogues with the patient. The authors then tested this in a defined evaluation protocol on 10 phobic patients. Klinger et al. then further experimented this same protocol on 36 patients diagnosed with social phobia [Klinger et al., 2005]. The results of both these studies showed that VRET and traditional CBT both were clinically valid and that the difference in results between the two was trivial. An example of their system is illustrated on the right of Figure 2.3.

Various methods have also been developed for the treatment of social phobia. For example, some researchers have proposed the use of Internet to expose patients to videos [Botella et al., 2000, 2004a]. Others have proposed to use a combination of virtual environments with live video [Lee et al., 2002b].

Similar to many of the existing applications for the treatment of social phobia, we propose an architecture which allows the use of various environments such as a bar, a cafeteria or an auditorium. Our architecture equally allows the definition the number of characters to be present in each scene. In addition, we use recorded background sounds in order for our environments to be more immersive as well as pre-recorded sentences to allow our characters to talk. Our application can be used on a simple computer screen, can be projected on a large back-projection screen or even in an HMD. This latter is used in combination with a head tracker in order for the images to change with respect to the user's head movements.

2.1.2 VRET For Agoraphobia

The term "agoraphobia" is now taken to include fears not only of open spaces but also of related aspects such as the presence of crowds and the difficulty of immediate easy escape to a safe place (usually home). The lack of an immediately available exit is one of the key features of many of these agoraphobic situations [WHO, 1993].

North et al. proposed a series of scenarios to be used in VRET treatment of agoraphobia [North et al., 1996]. These scenarios, however, did not consider crowds but only open spaces. To create the scenarios, they used the same software setup as in their application for social phobia [North et al., 1998]. These depicted a balcony, an empty room, a dark barn with or without a black cat, a covered bridge, a lift, a canyon and hot air balloons. They tested their scenarios from least to most anxiety provoking on a population of 30 individuals. They then compared their results with those of a control group of 30 people. These showed significant improvement in the patients having followed the VRET. The anxiety towards the feared situations decreased in the case of the patients having been through VRET whereas it stayed the same for the control group subjects.

Moore et al. proposed a series of environments to be used by people suffering from panic attacks and agoraphobia [Moore et al., 2002]. These environments were developed by Giuseppe Riva's team and represented a lift, a supermarket, a town square with and without people, and a beach. The characters were represented as billboards on which were applied real human textures. The exploration of the environments was done through HMD and navigation was obtained by using a joystick. They tested those environments on 9 healthy people by recording their physiological responses during exposure. They measured the skin conductance and temperature, heart rate and breathing rate. Their results showed that there was a physiological arousal during exposure. The aim of the authors with this study was to propose a baseline in studying physiological measures on agoraphobic patients.

Vincelli et al. proposed a set of 4 environments to be used with agoraphobic patients [Vincelli et al., 2003]. These environments were also developed by Giuseppe Riva. They represented a lift, a supermarket, a metro car, and a large square. The software actually consisted in one application, the Virtual Environments for Panic Disorders (VEPD), depicting a four-zone environment developed with the Superscape VRT 5.6 toolkit [superscape, 2009]. The

Figure 2.5: *Left:* Virtual supermarket for the treatment of panic attacks and agoraphobia [Villa Martin et al., 2007]. *Right:* Metro station for the exposure to panic attacks and agoraphobia [Botella et al., 2004c].

application allowed the definition of the zone to be explored, the length of exposure and the number of characters to be present in the scene (from none to a crowd). The authors then tested these environments on 12 agoraphobic patients randomly divided into three groups. The first group followed VRET with an HMD as part of the therapy during 8 sessions, the second followed traditional CBT during 12 sessions and the third was on a waiting list. Their results showed that both VRET and CBT could significantly reduce the number of panic attacks. Moreover, this was the case for fewer sessions in VRET than in traditional CBT. The authors therefore suggest that VRET could actually be more efficient.

Botella et al. proposed a set of 5 different environments to be used in VRET of agoraphobia [Botella et al., 2004c]. These depicted a room, a bus, a tunnel, a supermarket, and a metro station. The supermarket and metro station are depicted in Figure 2.5. The environments were developed within the VEPSY Updated project [vepsy, 2009] using Virtools Dev 2.0 [Dassault Systèmes, 2009]. They also simulated various sensations such as increased heart beat, being out of breath, and blurry vision by modifying images and sound. Their effect was to bring the patients back to the symptoms they may feel during a panic attack. They also allowed their situations to be modulated for gradual exposure. For example, they could vary the number of characters, the trip length or the type of conversation to be held. The authors then tested this setup on 37 patients [Botella et al., 2007] divided into three groups, those following VRET, those following in-vivo exposure and those on a waiting list. Their results supported the efficacy of VRET in the treatment of panic attacks and agoraphobia. The same team further tested their environments in the treatment of a 26-year old woman suffering from panic attacks and agoraphobia [Villa Martin et al., 2007]. Her results showed a decrease in anxiety following the VRET sessions which was maintained 12 months after.

Finally, Peñate et al. proposed a set of 7 virtual environments depicting a square and street, an airport building and an airplane, a bank office, a lift and an underground car park, a beach, a highway, and a cableway [Peñate et al., 2007]. These were developed using OpenGL and were based on a Torque engine from Garage Games [GarageGames, 2009]. Two projectors were used in order to send two different images on screen and allow for the environments to be visualized in stereo using polarized glasses. Finally, the environments were projected on a 2.5 x 2.5 meters screen allowing increased immersion as compared to a normal computer screen. The authors tested their application on 28 patients suffering from chronic agoraphobia. 15 of them followed a combined CBT and VRET treatment and 13 followed a traditional CBT treatment. Their results showed a slight but significant improvement of most patients at follow up (3 months later). Moreover, this was amplified in the case of those having followed the combined VRET and CBT treatment.

On a different note, somewhat between social phobia and agoraphobia, is the creation of empathy between virtual characters and human beings. In this context, Slater et al. have proposed a virtual replica of the Stanley Milgram obedience experiment [Slater et al., 2006]. They asked 34 participants to give a series of word association memory tests to a virtual character. When the character gave an incorrect answer, they were asked to administer her an electric shock. Their results showed that even though the character and the shocks were obviously not real, the participants tended to respond to the situation as if it were real.

A group of researchers conducted a study on paranoid thinking in VR [Valmaggia et al., 2007; Freeman et al., 2008]. Their simulation consisted of a metro ride containing 20 characters exhibiting neutral behaviors, displayed in a CAVE. In the former study, 21 subjects with high risk of psychosis were tested. Their results showed that most participants reported some paranoid experiences but found the environment to be mainly neutral or positive. In the latter study, the authors tested 200 subjects from the general population in an HMD. The majority found the characters to be neutral or positive but a substantial minority reported paranoid concerns. The authors concluded that paranoia could be understood in terms of cognitive factors and that VR will probably lead to rapid advance in its understanding.

In the application we have developed, we have principally focused on its use for agoraphobia with crowds. We have allowed our system to work in a highly immersive CAVE environment as opposed to a computer screen or an HMD. We have also focused on the simulation of character gaze in order to be able to increase the feeling of discomfort during immersion as well as give the possiblity of interaction between virtual characters and the user. Our system allows the definition the number of characters present in the scene, the proportion of characters which will be interested in the user or other characters, what the characters will be looking at and to what extent.

2.1.3 Eye-tracking in VRET

Eye-tracking has already been used in quite a number of domains, such as neurosciences, psychology, industrial engineering, marketing, and computer science [Duchowski, 2002]. The domains we are interested in, here, are on the one hand, psychology and on the other hand, computer sciences and in particular, virtual reality. More specifically, we first discuss the use of eye-tracking for interaction. We then describe the work which has been done with regards to the use of eye-tracking for the diagnosis and assessment of psychiatric illnesses.

2.1.3.1 Eye-tracking for Interaction

Concerning the use of eye tracking for object selection and movement, Hutchinson described one of the first applications [Hutchinson et al., 1989] . The author proposed the use of approximate locations of point of regard (POR) on screen to select menu objects. The computer then executed the command associated to this object.

Jacob proposed to use the eye as a replacement to the mouse [Jacob, 1990]. He proposed an intelligent gaze-based informational display. In this system, a text window would unscroll in order to give information about items selected by gaze. In this work, the author equally identified one of the main problems in using the eyes as a pointing and selection tool: the difficulty to know whether the eyes are scanning or selecting. This is also known as the Midas touch problem. In order to sidestep this, he proposed to use dwell time for selection.

In their paper, Starker and Bolt presented an information display system [Starker and Bolt, 1990]. In their system, a user equipped with an eye-tracker could control navigation in a 3D environment by gaze. These 3D environments equally contained characters which would change behavior when looked at. More specifically, they would start blushing or speaking. They used synthesized speech to interactively describe the objects which were being looked at on screen. Here, dwell times were used to zoom into the environment. Their setup was in front of a monitor screen and the users had to use a chin-rest in order to avoid head movements, which were not monitored.

Colombo et al. proposed a system coupling eye- and head-tracking [Colombo et al., 1995]. They monitored the various types of possible movements to trigger different types of events. Smooth gaze shifts were assimilated to image scanning, head movements were used to drag objects on screen, and “eye pointing”, if long enough, was identified as a mouse-click. They tested their method on a virtual museum application, where a user could explore the museum environment and select desired information on the various paintings.

Cassel and Thórisson conducted an experiment in which they tested different types of conversational agents [Cassell and Thorisson, 1999]. In a first phase, the agent gave content feedback only. In the second, it gave content and envelope feedback (non-verbal behaviors related to conversation such as gaze or tapping of the fingers), and in the third, content and emotional feedback. Their aim was to confirm their hypothesis, that envelope feedback was much more important than any other feedback. In their study, the subject was eye-tracked in order for the conversational agents, which consisted in simple 2D cartoon characters, to be able to respond with respect to where the subject was looking.

More recently, Tanriverdi and Jacob proposed an interactive system in VR [Tanriverdi and Jacob, 2000]. They used eye-tracking to select objects in a 3D environment. They then compared the efficacy of using eyes instead of hands for object selection. They concluded that the use of eyes for selection was more efficient than hands for far away objects but not for close ones. However, they also concluded that subjects had more difficulty remembering interaction locations when using their eyes than when using their hands.

In most of the above mentioned work, there was a requirement for the user’s head to be static because it was not tracked. Zhu and Ji developed an eye-tracking system which did not require a static head [Zhu and Ji, 2004]. Moreover, their system did not require calibration. One of the tests they did to evaluate their system, was to use eye-tracking for region selection and magnification, which was achieved by blinking thrice.

Finally, Wang et al. developed a system in which they used eye-tracking in order to change the behavior of a software agent in tutoring exercises [Wang et al., 2006]. When eye movement fell under a certain threshold and/or when the pupil size was smaller than a given threshold, indicating loss of interest, the software agent reacted by showing anger or by alerting the subject. On the other hand, if both these values were above the given thresholds, the agent looked happy. They equally used the gaze point area as a topic selector. They used their system in front of a computer screen and did not track the head.

We propose to use eye-tracking to define character behavior during exposure. More specifically, our method consists in using coupled eye- and head-tracking data to determine whether a character is being looked at by a user or not during a public speaking exercise. Our application then allows the modification of character behavior with respect to these data. The character thus remains attentive and interested while being looked at and starts to become distracted when eye contact is avoided.

2.1.3.2 Eye-tracking for Diagnosis and Assessment

Some studies have already been conducted on the use of eye-tracking systems to diagnose certain illnesses and phobias. These systems have also been used for the treatment of various ocular and mental dysfunctions.

Horley et al. used an eye-tracker to evaluate how phobic people processed interpersonal stimuli [Horley et al., 2003]. They used a retinal and corneal reflection eye-tracker recording at a frequency of 50 Hz. Their aim was to track phobic subjects' eyes while they were being exposed to various facial expressions. More specifically, they used photographs of a same person with a neutral face, a happy face and a sad face. They then recorded fixation points of 15 phobic and 15 non phobic subjects while showing them the different photos. Their aim was to empirically verify that social phobic subjects really did demonstrate eye contact avoidance. Their results showed that there was a clear avoidance of facial features, and especially the eyes, in the phobic subjects. The authors therefore suggested that the avoidance of salient facial features was an important marker of social phobia. The same team repeated and extended this study to 22 social phobic subjects and the same number of control subjects [Horley et al., 2004]. In addition to the analysis of facial feature avoidance, they analyzed hyperscanning behaviors, which consists of increased scanpath lengths. The results they obtained from this study were consistent with the previous one and with theories emphasizing the role of information processing biases in people suffering from social phobia.

Ramloll et al. described a gaze contingent therapeutic environment used to foster social attention in autistic children [Ramloll et al., 2004]. They used an ISCAN desk-mounted eye-tracking system [Iscan, 2009] based on pupil and corneal reflection. The authors proposed an automatic calibration technique as it is extremely difficult to ask an autistic child to respond to gaze direction cues. They then presented an application which consisted of interactions with a virtual character. The application allowed the rewarding of a user with vestibular effects and oral congratulations when typical gaze behaviors were performed. This was done by monitoring the time it took the user to select a character's face (with the eyes) and the duration of fixation periods when being asked to look at a virtual character in the eyes, look at two characters interacting, and follow characters' gaze directions. They then tested their application and used it as a rehabilitation technique for 24 to 54 months old autistic children suffering from attention deficits. The aim was to entice them to meet typical gaze behaviors.

Lange et al. used an ASL 501 head-mounted eye-tracking system [ASL, 2009] in order to determine differences in gaze behaviors between 16 spider phobics and 16 non spider phobics [Lange et al., 2004]. They started by putting the participants in a control condition during which they were asked to look at a TV video in absence of the feared stimulus. They then tested two different experimental conditions. In the first, they put the feared stimulus, a live tarantula, right next to the safety area (the only exit to the room). In the second, they introduced this feared stimulus away from the safety area. The phobic subjects reduced their viewing of the video and increased their viewing of the feared stimulus and of the safety area. Moreover, they made more eye motions across the room in which the experiment was taking place. The authors concluded that their results were in line with previous studies and that phobic people scanned the environment as part of safety behavior.

Certain types of ocular movements are symptomatic of specific illnesses or phobias. As an example, tasks requiring saccadic eye movements can identify cognitive deficit in children suffering from schizophrenia [Ross et al., 2005]. In this study, the authors tested Delayed Oculomotor Response (DOR) on 187 children aged 5 to 16 years old. 45 of them had childhood-onset schizophrenia, 64 had a first-degree relative with schizophrenia, and 84 of them were typically developing children. Their results showed that the children with childhood-onset schizophrenia demonstrated impaired response inhibition and spatial accuracy compared to the two other groups of children. However, there were no notable differences between these two latter groups.

Smith used an ASL series 5000 eye-tracking device [ASL, 2009] combined with magnetic head tracking to test the presence of hypervigilance and avoidance strategies in 46 undergraduate subjects [Smith, 2005]. The author hypothesized that socially anxious people would demonstrate these vigilance-avoidance behaviors when exposed to disgust-faces (threatening stimuli) but not when exposed to happy-faces (non-threatening stimuli). High socially anxious subjects and low socially anxious ones did not differ in attention behaviors towards the non-threatening stimuli. They did not differ in initial attention towards the threatening stimuli either. However, high socially anxious subjects took more time to disengage from this threatening stimuli. This study concluded that high socially anxious individuals have a tendency to present delayed disengagement from social threat.

Herbelin described the use of eye-tracking to determine fixation periods on areas of interest, as well as detect salient facial feature avoidance [Herbelin, 2005]. The author developed a visualization tool to help therapists analyze their patients' gaze behaviors. The author also described a gaze map color coding method, depicted on the left of Figure 2.6. This method allows the determination of which parts of a virtual character are being looked at or even if the character is being looked at at all, and this, even when the characters are in movement. The author conducted a case study using both these tools. He concluded that eye-tracking could be very beneficial as it allows for therapists to confront the phobic subjects to their gaze avoidance behaviors.

Finally, Mühlberger et al. developed the CyberSession software allowing them to manipulate and navigate in a VE [Mühlberger et al., 2008]. They designed a virtual lift which would stop at 60 floors of a building. At each floor level, the lift doors would open and the user would face three different scenarios. In the first one, the user would face a happy character and an angry one. This can be seen in Figure 2.6, right. In the second, the user would face a happy character only and in the third one, an angry character only. The authors used electrooculogram based eye-tracking combined with head tracking in order to allow subjects to be free of movement. They tracked the eye movements of 26 students testing the

Figure 2.6: *Left:* Illustration of a gaze map color coding method [Herbelin, 2005]. *Right:* Trial with a happy and an angry person [Mühlberger et al., 2008].

lift environment, half of which were told they would have to do a speech after the experiment. Their results indicated that the participants initially attended more to the happy people than to the angry ones, and this was amplified in the case of those who thought would have to give a talk after the experiment. However, they also had a tendency to completely avoid the characters, should they be happy or angry.

Similarly to Herbelin [Herbelin, 2005], we propose to use eye-tracking to analyze various aspects of gaze behaviors during virtual public speaking exercises. Our method consists of projecting various environments containing a central character and other characters in the background on a large back projection screen. Users are then asked to do a public speaking exercise while their eye is being tracked using an ISCAN eye-tracking device [Iscan, 2009] combined with an Ascension head-tracker to allow freedom of movement [Ascension, 2009]. During exposure, we record the user's eye movements and then analyze several factors, namely, fixation periods and areas, eye scan velocity, eye blink frequency and duration. We thus propose to use eye-tracking as an assessment and diagnosis tool for the presence of gaze avoidance behaviors and by extension, of social phobia itself.

2.2 Models of Visual Attention

The synthesis of human vision and perception is a complex problem which has been tackled in many different ways. Models of synthetic vision based on memory have been developed for character navigation. For example, Kuffner and Latombe proposed a method which combines a low-level path planner, a path-following controller, and cyclic motion capture data to generate character animations [Kuffner Jr and Latombe, 1999]. They use graphics rendering hardware to simulate character visual perception together with synthetic vision and memory update rules. They thus simulate autonomous animated character vision in real-time and in a dynamic environment. An example of their system is depicted in Figure 2.7, left.

Similarly, Peters and O'Sullivan proposed a memory model base on stage theory [Peters and O'Sullivan, 2002]. This consists of sensory memory, short-term memory and long-term

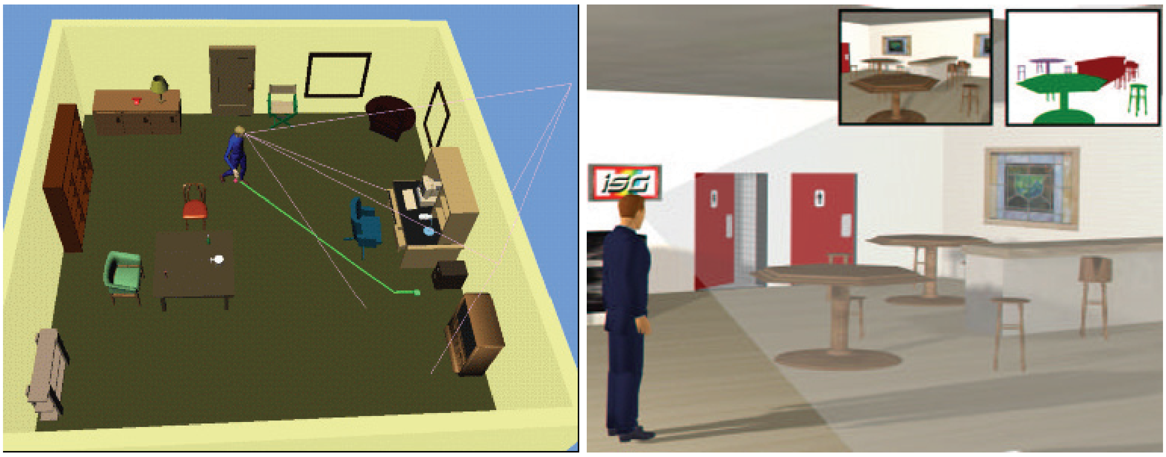

Figure 2.7: Different models of vision and perception. *Left:* [Kuffner Jr and Latombe, 1999]. *Right:* [Peters and O'Sullivan, 2002].

memory. The items attended to in the environment are stored in these three types of memory depending on their importance and the number of times they have been attended to. In order to simulate vision, they use false-coloring and encoded scene description information. Rendering hardware is then used to render the scene from the character's perspective. An example of their system is illustrated on the right of Figure 2.7. In both these methods, the authors simulated vision but not the actual human looking behavior.

Much work has equally been conducted in the simulation of visual attention and gaze in Embodied Conversational Agents (ECA). A model of perception was introduced by Hill which allows the creation of plausible virtual helicopter pilots for military simulations [Hill, 2000]. In this model, the objects in the pilot's environment are grouped according to various criteria such as type. The virtual pilot then receives information from objects in the environment such as location or velocity. The virtual pilot then decides on which object to attend to depending on these sensory data. Kim et al. further expanded the approach by using a benefit and cost function to determine when a character should look at an object [Kim et al., 2005]. They equally added aural perception to their model.

Gillies created behavioral animations through the simulation of vision and attention [Gillies, 2001]. In this model, object features are used to determine where attention should be set. Object features are modeled under the form of complex reasons for which an object should be attended to, such as artistic appreciation. Actors are thus endowed with interests and pay attention to objects relative to those interests.

Chopra Khullar and Badler proposed a computational framework for generating visual attending behaviors in an embodied simulated human agent [Chopra Khullar and Badler, 2001]. Their implementation associates a set of motor activities, such as walking or reaching, and cognitive actions, such as monitoring or visual search, with predefined patterns of looking behavior. In their model, a user first enters actions into their system as tasks on a queue. A task queue manager then coordinates the motor and cognitive activities which are requested and creates the appropriate attention behavior. An arbitrary process determines where an agent looks by selecting from three levels of behavior: endogenous, exogenous, and idling. The different types of behaviors have different weights. As an example, the task

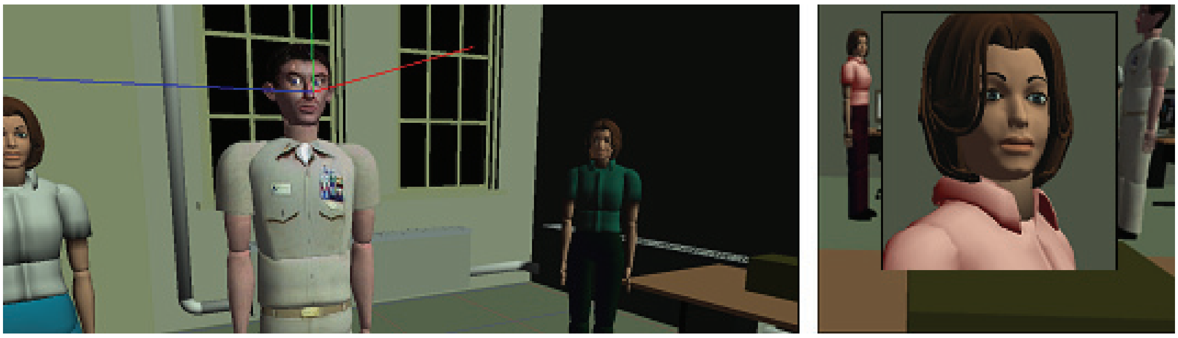

Figure 2.8: Mixture of bottom-up and top-down attention behaviors [Gu and Badler, 2006].

related eye behaviors have the highest precedence. They also use uncertainty thresholds to determine when an object should be looked at again.

Peters et al. proposed a model of attention and interest using gaze behaviors for speaking and listening agents [Peters et al., 2005]. They segment characters into eye, head, and body regions. They then retrieve direction and locomotion information for the character from an object database. The eyes, and regions oriented towards the viewer then receive higher weighting. They use these results as an attention level metric stored in the short-term memory system. In parallel, they use texts which contain what the speaker will say and the meaning to convey. They then use a finite state machine for each ECA in order to determine the gaze behaviors to perform in relation to the text.

Similarly, Gu and Badler proposed a visual attention model which integrated both bottom-up and top-down filters, and combined 2D snapshots of a scene with 3D structural information [Gu and Badler, 2006]. Their attention model affects eye motor control by specifying gaze direction, degree of eye aperture and size of the pupil, depending on luminance. Their ECA first attend to the top-down cues (a speaker for instance) and may have their attention diverted by bottom-up cues (a falling object for example). This is depicted in Figure 2.8.

Lance and Marsella proposed a model allowing emotionally expressive head and body movements during gaze shifts [Lance and Marsella, 2007]. They recorded real actors performing such movements together with neutral gaze shifts. They then extracted the parameters which defined the difference between neutral and expressive gaze shifts. Finally, they used a Gaze Warping Transformation (GWT), which consisted in temporal scaling and spatial transformation parameters to describe the manner of an emotionally expressive gaze shift. All these types of models give very convincing results but are designed for ECA and are not applicable to crowds.

In their paper, Itti et al. discuss a neurobiological model of visual attention [Itti et al., 1998] which they later on applied to virtual characters [Itti et al., 2003]. They use neurophysiologic data from the primate brain to create their visual processing model. They then use recordings of freely behaving Rhesus monkeys in order to derive their eye/head movement model. The input video is processed with various filters in order to obtain feature maps which contribute to the creation of a unique saliency map. The maximum of this map is then used to point the model's attention. This result then drives the eye/head movement controller and animates the virtual character.

Peters and O'Sullivan equally proposed a model based on the saliency maps previously discussed [Peters and O'Sullivan, 2003]. They actually combined it with the stage theory model of memory presented in [Peters and O'Sullivan, 2002].

Courty et al. also proposed a model based on saliency maps [Courty et al., 2003]. They modeled the human perception process by using a saliency map based on geometric and depth information. In order to do this, they combined a spatial frequencies feature map with a depth feature map. They then applied this to a virtual character in order for it to perceive its environment in a biologically plausible way.

On a different note, Peters and Itti conducted an experiment in which they tracked subjects' gazes while they played computer games [Peters and Itti, 2006]. They tested various heuristics to predict where the users would direct their attention. They compared outlier-based heuristics and local heuristics. Their results showed that heuristics which detect outliers from the global distribution of visual features were better predictors than the local ones. They concluded that bottom-up image analysis could predict an important part of human gaze targets in the case of video games.

Yu and Terzopoulos proposed a decision network framework to simulate how people make decisions on what to attend to and on how to react [Yu and Terzopoulos, 2007]. Their virtual characters are endowed with an intention generator, based on internal attributes and memory. They receive perceptual data by querying the environment. This data comes under the form of position, speed and orientation. They then decide on what to attend to depending on their current intention and on possible abrupt visual onsets. Finally, they endow their virtual characters with a memory system which allows them to remain consistent in their behaviors and adapt to changes in the environment. This approach equally aims at animating single characters or small groups of characters, but not large amounts of them, such as would be seen in virtual crowds.

The approach we propose for character attention behavior synthesis resides on the automatic detection of interest points based on bottom-up visual attention. Our method uses character trajectories from pre-existing crowd animations to automatically determine the interest points in a dynamic environment. Since it relies on trajectories only, it is generic, and can be used with any kind of crowd animation engine. Moreover, it allows the generation of attention behaviors for large crowds of characters. In a second step, we propose an alleviated version of our method, directly integrated in a crowd engine. In this method, we determine the interest points from user position, user's interest position, and character positions. It also allows the generation of attention behaviors for large crowds of characters in real-time.

2.3 Motion Editing

Motion editing, i.e. the modification of character movements, is also a vast domain which has been worked on in profusion. A large category of methods relies on the skillful manipulation of motion clips from a motion capture database by blending [Kovar and Gleicher, 2003] or by defining motion graphs [Kovar et al., 2002a; Arikan and Forsyth, 2002; Lee et al., 2002a]. Due to the many possible configurations in attention behaviors, this would require a very dense database in our case.

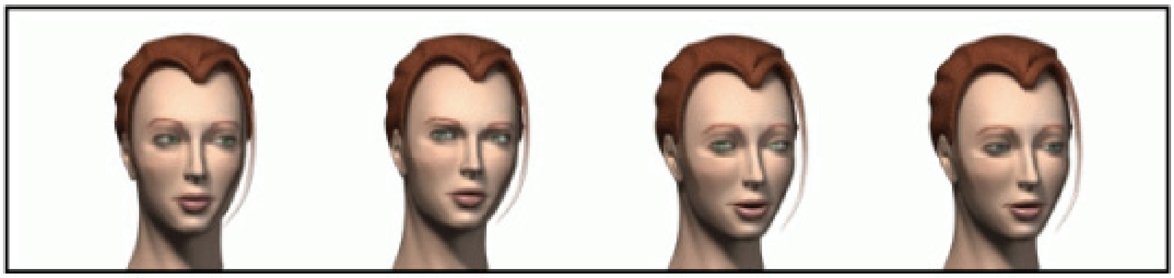

Figure 2.9: Samples of an animated face with eye movements [Lee et al., 2002c].

Several other methods used Jacobian based Inverse Kinematics (IK) solvers to edit motions. For example, Choi and Ko discussed a method for online retargeting [Choi and Ko, 2000]. Le Callennec and Boulic introduced the notion of prioritized constraints to solve possible conflicts between user-defined constraints [Le Callennec and Boulic, 2004]. While these methods are generic enough to possibly use any kind of constraints, the use of Jacobian inversion causes prohibitive computational costs that are not compatible with our framework.

Another category of methods, in which our approach resides, uses analytic IK. For example, Badler et al. proposed one of the pioneer methods in this domain [Badler et al., 1985]. They describe a human movement simulator which would execute motion descriptions. This simulator could perform rotations, twists, facings, motion paths, shapes, contact relationships, and goal-directed positions. These goal-directed positions consist in placing an end-effector; the different joints then revolve around their respective degrees of freedom (DOF) in order to attain a convenient posture in order to satisfy the effector position constraint, as long as these remain within the joint limit specifications.

Tolani et al. proposed a set of inverse kinematics algorithms to animate an anthropomorphic leg or arm [Tolani et al., 2000]. They proposed a combination of analytical and numerical methods to solve position, orientation, and aiming constraints. Their system equally allows for the user to interactively explore the set of possible solutions.

Lee proposed a method to edit a pre-existing animation to satisfy a set of user defined constraints for human like figures [Lee and Shin, 1999]. Their method allows them to adjust the posture or configuration of an articulated figure at each frame of an animation. They introduce a hierarchical motion representation which allows them to adaptively manipulate a motion in order to satisfy a set of constraints and also allows the edition of an arbitrary portion of the motion through direct manipulation. Kovar et al. then extended this method to remove footskate (foot sliding effects) from motion captured animations [Kovar et al., 2002b]. They propose an online algorithm which allows the enforcement of footplant constraints in an existing motion.

Shin et al. proposed a method to map motion capture data on a character of different size from the performer while maintaining the important aspects of the captured motions [Shin et al., 2001]. They use Kalman filters in order to remove noise from the motion captured data. They also propose a set of rules to dynamically assign varying importance to a set of tasks. Finally, they propose a dedicated IK solver which solves these constraints in real-time.

Kulpa et al. designed a motion representation which is independent from character morphology and containing the constraints intrinsically linked to the motion, such as the foot-

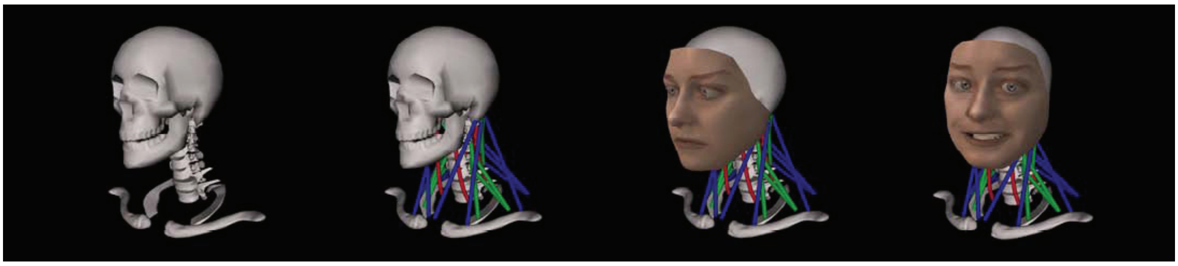

Figure 2.10: Biomechanical system for natural head-neck movements [Lee and Terzopoulos, 2006].

plants [Kulpa et al., 2005]. This method allows them to share a motion between several characters having different morphologies. They equally adapted a hierarchical Cyclic Coordinate Descent (CCD) algorithm taking advantage of this representation to deal with spacetime constraints for the characters. All these analytic methods are dedicated to the positioning of end-effectors. In this thesis, we are interested in controlling the final orientation of the eyes, head, and spinal joints over time instead.

Somewhat between models of human vision and motion editing, Lee et al. proposed an eye movement model which they based on empirical models of saccades and statistical models of eye-tracking data [Lee et al., 2002c]. Their approach consisted in using the spatio-temporal trajectories of saccadic eye movements to synthesize the kinematic characteristics of the eye. They first analyzed a sequence of eye-tracking images in order to extract the spatio-temporal trajectories of the eye. With this, they derived two statistical models of the saccades which occur, for talking mode and listening mode. Their model reflects saccade magnitude, direction, duration, velocity, and inter-saccadic intervals. Figure 2.9 illustrates their method with a couple of examples.

Lee and Terzopoulos proposed a head-neck model based on biomechanics which emulates the anatomy of the human neck [Lee and Terzopoulos, 2006]. They also presented a neuromuscular control model in order to animate the head and the neck. Their method allows the simulation of head pose and movement, but also the stiffness of the head-neck multibody system, as shown in Figure 2.10. Even though such a method gives stunning results, its computational times are prohibitive for crowd animation.

Instead, we present an extremely fast analytic dedicated gaze IK solver to handle the orientation of both eyes, the head, and the spine. Given a pre-existing animation, our solver computes the displacement maps for characters to satisfy predefined gaze constraints. These displacements are smoothly propagated onto the original motion to ensure that the final results are continuous. Our method deals with dynamic constraints and manages both the spatial and temporal distribution of the displacements.

CHAPTER 3

VRET For Social Phobia

Much work has already been conducted with regards to the use of VR as therapeutic tool for the treatment of phobias. Many types of phobias have thus been tackled such as acrophobia, arachnophobia, and claustrophobia. It is quite simple to simulate heights, spiders, or closed spaces in VEs. On the other hand, it is much harder to simulate humans. Indeed, as human beings, we are experts in human being representation and movement. For this reason, much less work has been undergone in social phobia. VRET for social phobia, and more specifically, for fear of public speaking is only about a decade old.

In this chapter, we first present the various scenarios we have created, including environments, characters and scenario scripting. We then present a method we have developed in order to obtain interaction between the user and a character using eye-tracking data. We then discuss how this has motivated us to further work on character gaze behaviors.

3.1 Preliminaries

The first part of this thesis is based on previous work which has been undergone in the Virtual Reality Laboratory (VRLab, EPFL). More specifically, the platform which we use has been developed by Herbelin during his PhD thesis [Herbelin, 2005].

This framework allows for the real-time animation of a small group of characters. They are endowed with gaze control and facial animation. The platform works with Python scripts. These scripts allow interactive control of characters in order to make them talk for example; this actually consists of playing a pre-recorded sentence and animating the character's face and eyes accordingly.

These scenarios can either be viewed on monitor, in an HMD or on a large back-projection screen. While using the HMD, we can track the user's head in order to modify the images with regards to head rotation for enhanced immersion. While using the back-projection screen, the user can be equipped with a coupled eye- and head- tracking device in order to determine where the user is looking on screen. The combination of the two allows freedom of movement in front of the screen.

Moreover, we use an eye-tracking data visualization tool. It is based on a gaze-map chromatic gradient coding. This allows the representation of the eye-tracked points on the virtual character even if it is dynamic. It therefore serves as assessment tool to analyze recorded eye-tracking data and illustrate possible eye contact avoidance behaviors. More details on this method can be found in [Herbelin, 2005] and [Herbelin et al., 2007].

3.2 Creating Scenarios For VRET

Using this existing platform, we have created various different scenarios for VRET of social phobia. Using environments created under Autodesk 3DSMax [Autodesk, 2009] as well as characters modeled with the same 3D software, we have set up different scenes. Also, by creating various textures for a same character template, we have been able to diversify our characters without having to design them all.

3.2.1 Environments

The main scenes we have created and worked with are depicted in Figure 3.1. They were all designed to exercise public speaking in the various situational domains as described by Holt et al., namely, formal speaking or interaction, informal speaking or interaction, assertion and observation by others [Holt et al., 1992].

The first one is an office environment, depicted on the top left and top right of Figure 3.1. We have created two different scenarios with this environment; the first is an interview with the boss of a company. We have further diversified this by letting the boss be either a man or a woman. The second scenario takes place in another room of the same environment and consists of sitting in front of 5 people from the company and having to do a speech. These serve as exercise in the formal speaking or interaction situational domain.

The second environment, depicted on the middle left of Figure 3.1 is a bar. The scenario we have created consists of being seated facing a person in a bar. The user would have to imagine that this person is a new friend or a new colleague. Here as well, this character can be either a man or a woman. Other characters are seated at different tables in the bar. This serves as exercise in the informal speaking or interaction situational domain.

Our third environment is a cafeteria, depicted on the middle right of Figure 3.1. This scenario is actually very similar to the one in the bar. Here as well, the user is seated facing a person, man or woman. Some social phobic people are unable to eat in front of others. We thus added food on a plate in front of the user. This scenario can serve as exercise in the observation by others situational domain.

Figure 3.1: Various VEs used for public speaking exercises. *Top Left:* Job interview simulation. *Top Right:* In an office, facing five people. *Middle Left:* Meeting in a bar. *Middle Right:* Meeting in a cafeteria. *Bottom Left:* Speech in front of an auditorium. *Bottom Right:* Sitting at the back of an auditorium.

Figure 3.2: Various designed characters.

The last environment we have created is an auditorium, illustrated on the bottom left and bottom right of Figure 3.1. We have also created various scenarios with this setup. In the first, the user is standing on the platform and has to do a speech or presentation in front of a jury of 5 characters. This can serve either in the formal speaking or the assertion situational domains. In the second scenario, the user is also standing on the platform, but has to do a speech or presentation in front of approximately 20 characters. Finally, in our last scenario, the user is seated at the back of the auditorium and has to ask questions to the character on the platform, presenting something. Here as well, approximately 20 other characters are seated in the auditorium. Moreover, these characters can turn around and look at the user "seated" behind them. All these scenes are loaded into the viewer using xml scripts. These scenarios can can serve in the observation by others or assertion situational domains.

3.2.2 Characters

We have designed various characters with multiple textures in order to artificially enlarge our character set. Some of these characters' faces are depicted in Figure 3.2. Figure 3.3 illustrates how we can obtain a variety of characters using multiple textures on a single mesh. All our characters have a skeleton structure composed of 76 bones in order to animate them. To create these characters, we have gone through various stages: modeling (to create the mesh), texturing (to give it a face and clothes) and skinning (to be able to move the mesh and animate the characters using bones).

The characters to be loaded in the scene are defined in the initialization xml script. In this initialization script, we also define their position and orientation in the environment.

3.2.3 Sound

Our scenarios have been enhanced with various sounds. We have used background sounds in order to create an atmosphere to our various scenes. For example, in the auditorium, background sound consists of ventilation, people coughing, rustling of papers, etc. In the bar, it consists of people talking and clinking glasses. Moreover, we have recorded various sentences to be "spoken" by our characters. This allows the avoidance of breaks in presence during exposure. Instead of having the therapist or experimenter talk to the user, the characters can ask the user to begin, continue or end their presentation. These sentences are triggered by the therapist or experimenter using different keys.

Figure 3.3: Multiple textures on a single mesh to obtain variety in character representations.

3.2.4 Scripts

As previously mentioned, the existing architecture developed in VRLab allowed the use of Python scripts in order to give instructions as to how the scenarios should unfold. We have thus created different scripts for the various scenarios in order to instruct the characters on what they should do. A simple view of the general animation loop is shown in Algorithm 3.1:

Before launching the script containing the general animation loop, an initialization script is executed in order to define the characters which are present in the scene and their original posture. Now that we have explained how we setup our various characters in various scenarios, we discuss in more detail how we animate them and assign them different behaviors.

3.3 Eye-Tracking For Interaction

In order to use VR as a tool in CBT, and more specifically, in the treatment of phobias or other social related problems, patients need to feel present in the VEs and feel the presence of its characters. When it comes to characters, in order for the sense of presence to be increased, both their representation and behavior have to be worked on in parallel, otherwise, sense of presence decreases [Garau et al., 2003]. This is also known as the “uncanny valley” effect, which has first been presented in the field of robotics [Mori, 1970]. This need is especially true in the case of social phobia, since patients need to feel as if they were in a real situation; they need to experience the anxiety they would experience in-vivo and feel that they are really facing a character that understands their presence. In other words, since we are already quite good in virtual human representations, what is needed is an increase in realism in their behaviors. The difficulty is that as humans, we are experts in human behavior. It makes it much harder to create credible characters. Much research has already been conducted in this aim, such as speech synthesis, character modeling or animation.

Algorithm 3.1: Character Animation Loop

```
begin
    Wait between 1 and 2 seconds
    i = Randomly select an agent
    j = Last selected agent
    while i = j
    do
        i = Randomly select an agent
    Deactivate previous agent action
    Pick new agent action
    if Keypress 1 then
        Start sound start
        Animate agent face
    if Keypress 2 then
        Start sound continue
        Animate agent face
    if Keypress 3 then
        Start sound end
        Animate agent face
end
```

Following the two first experiments we have conducted, which are described in detail in Chapter 6 and in which we analyzed eye contact behaviors in phobic subjects and healthy subjects, we have been motivated to further work on this factor. Moreover, eye contact is a very important part of communication and is known to be often avoided by people suffering from social phobia. We thus propose to use gaze and eye contact to drive interactions between a user and virtual characters. We suggest to further enhance character believability by making them responsive to eye contact. To do so, we introduce an application, based on real-time eye- and head-tracking, which allows for virtual characters to be aware of when they are looked at. They can then respond in many different ways. To illustrate this, we chose to make them seem attentive and interested when looked at, and bored and distracted when avoided by eye contact. This choice was motivated by the fact that eye contact avoidance is a known feature present in people suffering from social phobia [American Psychiatric Association, 1994]. People suffering from social phobia are very affected by the way other people perceive them. By increasing virtual character awareness and giving them the possibility to look at the real subject (or avoid looking at him/her), it is possible to greatly amplify immersion and therefore, the induced anxiety and the training ability of VRET.

The algorithm we use in order to determine whether a character is being looked at or not is presented at the beginning of this chapter and further described in [Herbelin et al., 2007]. It is a texture-based color picking technique which works on dynamic meshes. Even if the character is moving on screen, it is still possible to know whether it is being looked at or not. The user is equipped with an eye-tracking device coupled with a head-tracker. We can thus determine where the user is looking and more specifically, if he/she is looking at a character or avoiding it. We then used Python scripts in order to, amongst other things, assign animations to the characters on screen. Algorithm 3.2 summarizes the different steps of the animation loop. We use two animation pools: one containing attentive behaviors and one containing distracted ones. Each pool contains 10 different animations. Then, if the eye-tracked coordinates are within the character bounds, an animation is randomly chosen from the attentive pool and applied to the character. On the other hand, if the eye-tracked coordinates are outside the character bounds, an animation is randomly chosen from the distracted pool. A random time lapse is defined between the assignment of each animation. The character thus does not become attentive or distracted immediately. Whenever an animation is to be assigned to a character, we verify whether it is being looked at or not. Depending on this, the chosen animation will be of type attentive or distracted.

Algorithm 3.2: Character behavior selection

```
Data: input point of regard on screen (POR)
Result: picked character
begin
    Wait between 1 and 2 seconds
    i = Randomly select an agent
    j = Last selected agent
    while i = j
    do
        i = Randomly select an agent
    Deactivate previous agent action
    if agent = central character then
        Retrieve eye-tracking coordinates
        Check if in character bounds
        if In character bounds then
            Pick agent action in attentive pool
        else
            Pick agent action in distracted pool
    else
        Pick new agent action
end
```

3.4 Results

Figure 3.1 depicts the results obtained from the combination of our environments with our characters. As mentioned in Section 3.2, we have created various scenarios by positioning the user and the characters in different positions in the different environments. Regarding the use of eye-tracking for interaction, Figure 3.4 illustrates the difference in behavior between

Figure 3.4: Variations in character behavior depending on eye-tracking coordinates. *Left:* Attentive. *Right:* Distracted.

the attentive and distracted versions of a character animation. On the left of the image, we can see the attentive version of the character, whilst the right part of the image depicts the distracted one.

Chapter 6 describes the experimental results we have obtained using the various scenarios described in this chapter. It presents an experiment we have conducted on a social phobic population using these environments, characters and scenarios. It also describes an experiment we have conducted to evaluate the potential of our method to change character behavior with respect to user eye contact behavior.

3.5 Gaze Behaviors to Enhance Character Realism

The experiments we have conducted in order to validate our scenarios in the context of social phobia have underlined the importance of gaze behaviors for both phobic subjects and virtual characters. First, gaze avoidance behaviors are known symptoms in people suffering from social phobia [American Psychiatric Association, 1994]. Moreover, several studies (including our own) have further confirmed this. Eye contact is also an essential part of non-verbal communication. This has thus motivated us to further enhance character realism by providing them with gaze attention behaviors. Additionally, we wanted to be able to create scenarios for both the social phobia and the agoraphobia with crowds contexts. We thus chose to develop an architecture which could be used both on a single character and on a crowd of characters. This architecture was developed in two phases: an offline and an online one. By offline, we mean that the method is a pre- or post- process one. Conversely, by online, we mean that the method is applied during the simulation, in real-time. The first phase of our architecture was developed in order to provide characters with gaze attention behaviors. We then adapted it in order to integrate it into an existing crowd simulation engine, making it online. We equally developed a system similar to that explained in the previous section to allow users to interact with virtual characters using gaze.

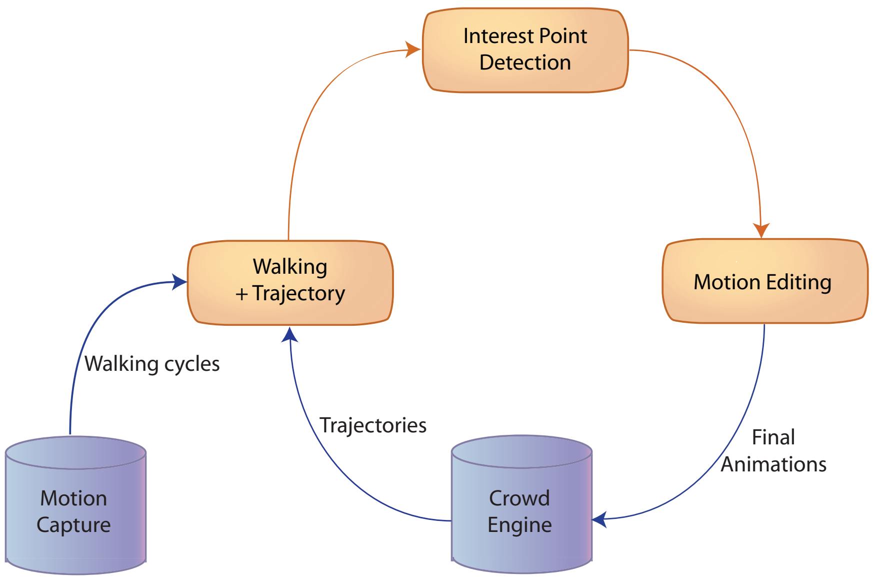

Figure 3.5: Overview of the offline architecture allowing the addition of gaze attention behaviors to virtual characters.

3.5.1 Gaze Behaviors for Crowd Characters

As mentioned, our first model to add gaze attention behaviors to crowd characters is an offline one. The overall outline of this architecture is depicted in Figure 3.5. We use the character trajectories from an existing crowd simulation, created using a crowd simulation engine developed in lab and described in [Maïm et al., 2009]. The trajectories simply consist in the position and orientation of each character at each time step. We then combine these trajectories with motion captured walking cycles in order to reconstruct the character walking motions. We adapt these walking cycles in order to eliminate footskating effects due to the changes in speed of the moving characters. Using the character and object trajectories, we then automatically detect where each character should look and when by assigning a score to each other character and object in the scene. This score is determined by a scoring function composed of various subscores, each associated to a particular feature such as distance or orientation. Finally, we adapt the character motions in order for them to perform the gaze behaviors. We compute the displacement maps to be applied to each recruited joint in order to perform the gaze motion. We then smoothly propagate these displacements over time in order to obtain a natural movement. The complete character animations are then re-injected into the crowd simulation engine in order to obtain the final results. This architecture and all of its components are discussed in detail in Chapter 4.

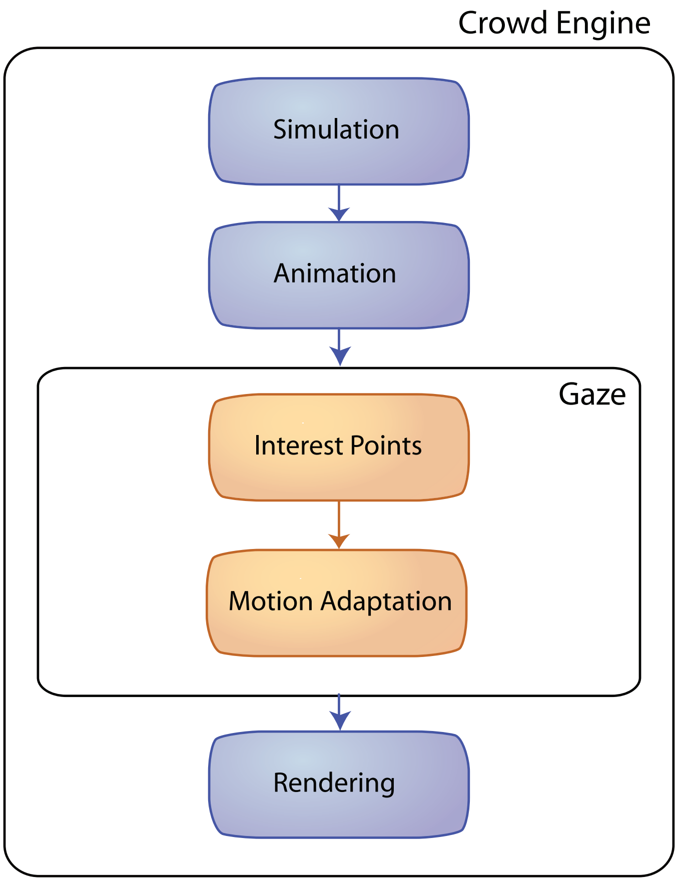

Figure 3.6: Overview of the online architecture allowing the addition of gaze attention behaviors to virtual characters.

3.5.2 Interaction with Virtual Crowds

The second model we propose is an online adaptation of our first one. As depicted in Figure 3.6, we integrate our architecture directly into the crowd simulation engine and adapt it in order for it to feature online properties. We simplify the interest point detection algorithm by pre-defining the points of interest the characters should look at instead of computing them. These points of interest can either be a user, something the user is looking at or any other randomly chosen character or object in the scene. Similarly to the offline method, we then adapt the original character motions in order for them to perform the gaze behaviors. However, the method we use in this online architecture is slightly different in order for it to work online. Moreover, we further implement the possibility to interact with the characters using gaze, similarly to the work presented in the previous section. By tracking a user's gaze, we can detect where he/she is looking and more specifically, at which character he/she is gazing at. We thus allow the virtual characters to react to gaze by looking back at the user or by paying attention to what the user is looking at. The detailed description of this architecture is presented in Chapter 5.

3.6 Discussion

Filtering of eye-tracking data. We could improve the efficacy of the algorithm we have developed which allows for interaction between user and virtual character using gaze. Indeed, our current framework decides on which animation to assign to a character depending on isolated eye-tracking data. This could be improved by taking into account the overall eye contact behavior and by filtering out small, saccadic movements. When we talk to people, we do not stare at them. Usually, we mostly look at them but our eyes wander around from time to time (when we are thinking of what to say next for example). Data filtering would thus allow to take into account fixation areas and to determine gaze behavior in a more general, overall way.

Closed Loop Potential. We trust that our gaze interaction algorithm can give the possibility to invert the vicious circle in which people suffering from social phobia find themselves; they fear a situation and therefore do not expose themselves to it which leads to fearing it even more. In the present application, a user can induce a positive feedback loop by looking at the character when talking to it and thus enter a virtuous circle as opposed to the vicious one described above. This positive feedback loop can enforce confidence and motivate the patient to reproduce these gaze behaviors in real conditions.

3.7 Conclusion

In this chapter, we have presented the various scenarios we have developed and used in the context of VRET of social phobia. We have also presented a method to change character behavior with respect to user eye contact behavior using an eye-tracking device.

The various scenarios we have proposed have allowed us to simulate different situations which are frequently avoided by people suffering from social phobia. The experiments we have conducted using these systems have given very promising results, which are discussed in Chapter 6. We believe that this type of exposure has a high potential as habituation exercises. We also believe that there is a high potential in such a feedback loop as exposed in this chapter, using eye-tracking to drive character behavior.

Finally, we have discussed how this application has motivated us to further work on the simulation of gaze attention behaviors for virtual characters. We have briefly outlined a first architecture which allows the addition of these gaze attention behaviors to an existing crowd simulation in an offline way. We have then described the main features of our second architecture, which allows the addition of the gaze attention behaviors in an online way and allows for a user to interact with the virtual crowd characters using gaze. In the two following chapters, we describe each of these architectures and their various components in detail.

CHAPTER 4

Simulating Visual Attention for Crowds

The experiments we have conducted in the context of social phobia, that will be discussed in detail in Chapter 6, have underlined the crucial importance of eye contact in human communication. In order to obtain believable and natural looking crowd characters, we thus decided to endow them with attention behaviors. We developed an architecture applicable to both social phobia and agoraphobia with crowds, in the sense that it can be applied to a single or small number of characters but also to a large crowd of characters. This architecture allows for characters to behave naturally in terms of gaze behaviors and to perform looking motions in a smooth and natural way.

When we walk in town, we pay attention to our surroundings by looking in different directions. We look at other people, objects, or even at nothing in particular. As mentioned, we believe that an important aspect which can greatly enhance crowd animation realism is for the characters to be aware of their environment and of other characters. This has partly been achieved with navigation and path planning algorithms. Our aim, however, is to obtain more advanced behaviors than those which navigation alone can provide. This raises the common problem of mandatory trade-off between rich, realistic behaviors and computational costs. Individual character animation may provide realistic results but is computationally expensive. Conversely, global crowd behavior design is much faster but results in a loss of character individuality. To add gaze attention behaviors to crowds, we are confronted with two issues. The first one is to detect the points of interest for characters to look at. The second one is to edit the character motions for them to perform the gaze behavior. This has to be done very rapidly in order to animate a large number of characters. In this chapter, we propose a two-fold method which meets all these requirements.

Our first step consists of a per-character automatic interest point detection algorithm based on *bottom-up* attention behaviors. When we attend to objects or people, it is either as *active* or as *passive* attention behavior [James, 1983]. The term bottom-up is used to describe passive or involuntary, stimulus-driven attention. For example, if a very tall person comes running towards us, our attention will be captured by this person. Our algorithm automatically detects *where* and *when* each character should look. It is based on a scoring method which is a weighted sum of elementary scores. These are determined by elementary functions using parameters such as distance or orientation.

Our second step is a very fast dedicated IK solver to satisfy these automatically detected gaze constraints. Our solver determines how the character motions are edited both spatially and temporally. It computes the displacement maps to satisfy the constraints and smoothly propagates the motion adjustments with adequate timing for the eyes, head, and spine in order for the final motion to be fluid and continuous.

4.1 System Overview

Our system works as an extra animation layer added to an existing crowd animation. The idea is to enhance this animation by providing its characters with gaze attention behaviors. For clarity purposes, we use the term *character* to refer to the individual for which we are generating the gaze behavior. Similarly, we use the term *entity* to refer to either a character or an object that can possibly attract attention. This can be a walking character or a poster on a wall for example. Finally, *interest points* are the locations in space which attract attention. Our method generates attention behaviors solely from the entities' trajectories. Thus, it is generic, and can potentially be used with any type of crowd animation engine.

We define a trajectory $\mathbf{T}_i(t)$ for an *entity* E as:

$$\mathbf{T}_i(t) = [\mathbf{p}_i(t), r_i(t)] \tag{4.1}$$

where i is the entity's identification, $\mathbf{p}_i(t) \in \mathbb{R}^3$ is its position at time t, and $r_i(t) \in \mathbb{R}$ its forward orientation (direction) in the horizontal plane at time t.

Since our method aims at enhancing crowd realism, we must deal with a large number of characters. It is therefore unthinkable for a user to define all the points in space which should be attended to by each character. This would be extremely time consuming. For this reason, it is mandatory to automatically detect the interest points. One of the key features of our method resides in this automatic detection over time. Based on the character and object input trajectories, it takes into account both the spatial and the temporal aspects of the gaze behaviors: the *where* to look and the *when* to look. Moreover, it induces a form of *how* to look. Finally, the detected interest points form a set of gaze constraints $\mathbf{L}$ to be satisfied.

Our method also consists of a robust and very fast dedicated IK solver. Given a pre-existing motion, we compute the displacement maps $\mathbf{m}(t_i)$ that adjust the motion postures in order to satisfy the set of constraints $\mathbf{L}$. We then propagate these displacements over time in order for the eyes, head, and torso to be desynchronized. By desynchronized, we mean that the movement is initiated by the eyes. The head and the remainder of the spine then follow and the eyes partially re-center with respect to the head. Indeed, when we look at something, our eyes move before the head, which itself moves before the spine joints. The motion of each character is thus adapted for him/her to attend to the automatically defined interest points in a smooth and natural way.

However, since our method uses character trajectories only in order to remain generic, our first step is to reconstruct the character walking motions from those trajectories and motion captured walking cycles.

4.2 Walking Motion Reconstruction

As basis to the motion, we use a walking cycle which was recorded by motion capture. The motion we use is of a person walking on a treadmill at 3 km/h. The trajectories obtained from the crowd simulation engine are recorded at a sample rate of approximately 30 fps but this amount varies slightly from one frame display to another. In order to solve this problem for the reconstruction of the walking motion, we first proceed to a temporal normalization. Moreover, during the simulation, the characters do not always walk around at the same speed. Their pace varies throughout the simulation, in particular during collision avoidance. To remedy to this problem, we thus proceed to a speed resolution.

4.2.1 Temporal normalization

In order to obtain the trajectories at exactly 30 fps, we first calculate the slope of the straight line going from the first position in time to the next one for each of the coordinates (x, y, z). We do the same for each of the orientation coordinates. By using the equation of the straight line, we can then determine which would be the value of the second coordinate at a specific time t (every 30^{th} of a second in our case).

$$x_i = a(t_i - t_{i-1}) + x_{i-1}$$

x_i being the interpolated value to be found, a, the slope, t_i, the next non-normalized time value, t_{i-1}, the previous normalized time value, and x_{i-1}, the value at the previous normalized time. We then repeat this process between each pair of consecutive points. The first point being, in each case, the one which has just been calculated.

4.2.2 Speed resolution

The trajectories that we output from the crowd simulator vary in time in the matter of speed. As an example, when a collision is to be avoided between two pedestrians, one can slow down and the other speed up. Since we use a base motion of someone walking at 3 km/h, the movement of the legs does not match with the displacement to be done. In order to solve this, we have to determine, at each time step, at which speed the character is moving from the trajectory values. Then, as we know the distance which is traveled within one walking cycle at 3 km/h, we can determine the distance which should be traveled at the calculated speed. Finally, we recover the body posture corresponding to the traveled distance in the base walking motion. In this way, we greatly reduce foot sliding effects. In order to eliminate them completely, we implemented the method proposed by Treuille et al. to enforce footplants [Treuille et al., 2007]. This first consists of labeling the footplant in the original walking cycle. Then, whenever a footplant is detected during motion reconstruction, we adjust the root joint displacement in order for the character to remain in its current position. Finally, once the footplant is deactivated, we propagate the cumulated adjustment on the next few frames in order for the character to "catch up" on the distance it has not traveled during footplant and avoid a sudden jump forward when applying the following root joint position.

4.3 Automatic Interest Point Detection

The first main step in our method consists of automatically detecting the interest points from the entity trajectories. We define an interest point IP as an entity which should be attended to by a given character. More formally, IP is defined as:

$$IP(t) = [\mathbf{p}_t, t_a, t_d, [t_b, t_e]] \tag{4.2}$$

where $\mathbf{p}_t \in \mathbb{R}^3$ is the interest point's position in space at time t. t_a is its *activation duration*, t_d its *deactivation duration*, and $[t_b, t_e]$ represents the interest points lifespan. Indeed, when we look at something or someone, we do not perform our looking motion instantaneously [Grossman et al., 1988]. For this reason, we introduce an activation duration t_a. Its purpose is to define the amount of time it will take for the looking motion to be executed. Conversely, the deactivation duration t_d defines the amount of time for a character to look away from an interest point. These parameters are further discussed in Section 4.4.2. It is to be noted that in the case where an interest point is replaced by another, the deactivation is skipped and replaced by the activation to go from the first interest point to the second one.

Another important factor in attention behaviors is that we do not look indefinitely at objects or people. We can either lose interest in something or find something else more interesting to look at. As depicted in Figure 4.1, this is regulated by $[t_b, t_e]$ in our method, which is the duration for which an entity *is* an interest point.

Algorithm 4.1: Elementary Scores Computation

```
Data: input trajectories
Result: elementary scores
begin
    for all frame f do
        for all character C do
            for all entity E do
                compute elementary scores
end
```

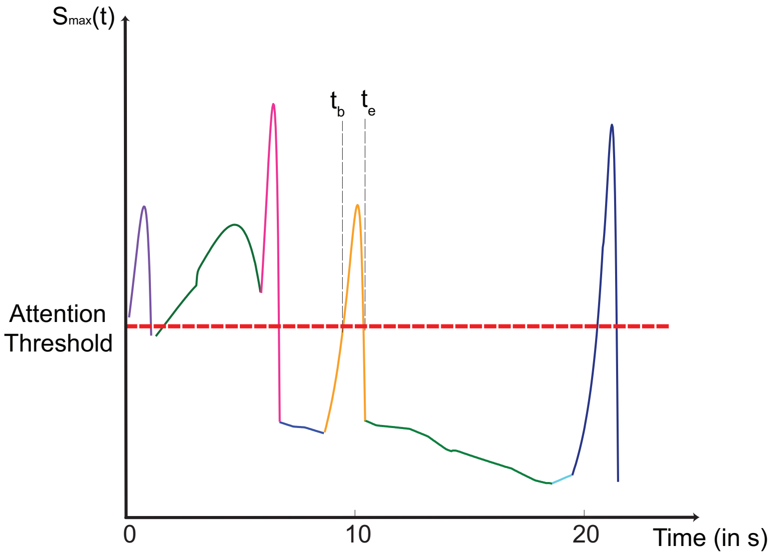

Figure 4.1: Overall maximum scores $S_{max}(t)$ for a character C. The different colors represent different interest points. t_b and t_e respectively represent the beginning and the end of a gaze constraint.

For each character C and at each time t, we define the level of interest other entities have by assigning them a score $S(t)$. The entity E which obtains the highest score $S_{max}(t)$ becomes the interest point that should be attended to by C at time t as long as it fulfills two conditions. $S_{max}(t)$ first has to be above an *attention threshold*. This defines the percentage of time C will be attentive to other entities. Second, E should obtain $S_{max}(t)$ for a minimal amount of time $[t_b, t_e]$ which we have empirically set to $1/3s$. Indeed, minimum gaze durations have been ranged from 80 to 150 milliseconds in literature [Optican, 1985], the smallest being in the case of highly skilled tasks. Since we are not in the context of highly skilled tasks, we based our minimum gaze duration on the upper bound. We doubled this number in order to take into account the time it takes to attain the interest point. $S(t)$ is then computed through a scoring function which contains various components. Algorithm 4.1 summarizes the computational course of the elementary scores associated to each component. Following is their description.

Previous studies in the domains of psychology, neurosciences, and vision such as [Neisser, 1967] explain that human attention is captured by substantial differences in one or more simple visual attributes. Simple visual attributes are features such as color, orientation, size, and motion [Wolfe and Horowitz, 2004]. Additionally, Yantis and Jonides pointed out that abrupt visual onsets equally attract human attention [Yantis and Jonides, 1990]. A person or object entering the field of view, i.e. movement in the periphery, is an abrupt visual onset [Egeth and Yantis, 1997]. These studies have motivated our choice of four different criteria:

- ***Proximity:*** this can be assimilated to size; closer objects or people seem larger and attract attention more easily than those far away. Moreover, those which are closer occlude those which are further away.

- ***Relative speed:*** a person will be more prone to set his/her attention on something moving at a very different speed than his/her own velocity.
- ***Relative orientation:*** we are more attentive to objects coming towards us than moving away from us. This is also related to size since something coming towards us seems to become larger.
- ***Periphery:*** we are very sensitive to movements occurring in the peripheral vision. More specifically, to objects or people entering the field of vision.

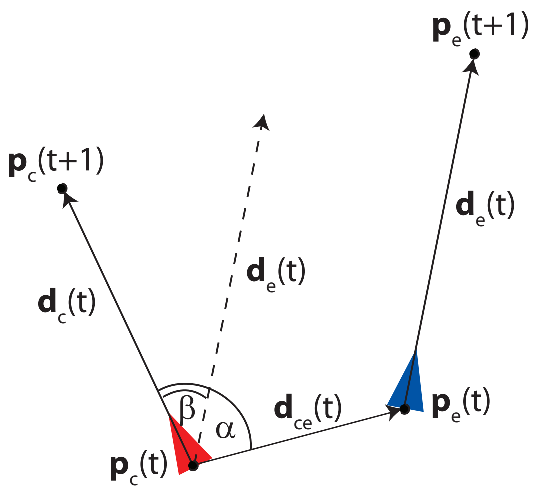

Figure 4.2: Schematic representation of the parameters used for the elementary scoring. $\mathbf{p}_c(t)$ is the character position at time t, $\mathbf{p}_e(t)$ is the entity position at time t, α is the entity orientation in the character's field of view, and β is the angle between the character and the entity forward directions.

In order to decide where a given character will look at a given time, we evaluate all entities in terms of the above mentioned criteria. As depicted in Figure 4.2, we first evaluate a set of parameters for each of these entities, namely the distance $d_{ce}(t)$, the relative speed $rs(t)$ defined by forward differentiation as $||\mathbf{d}_e(t) - \mathbf{d}_c(t)||$, which is actually the difference in distance traveled by the character and the entity in one frame, the orientation in the field of view $\alpha(t)$, and the relative direction $\beta(t)$. Similarly to Sung et al. [Sung et al., 2004], we combine these simple parameters to create more complex scoring functions that evaluate each of the criteria we have chosen: S_p for proximity, S_s for speed, S_o for orientation, and S_{pe} for periphery. These are described in detail in Sections 4.3.1 to 4.3.4. We thus assign a set of elementary scores to each entity. At this point, we can define, for each parameter individually, which entity is the most interesting for each character at each frame. However, these criteria need to be evaluated as a whole for them to have a meaning. To this end, we define a final scoring function. Moreover, in order to obtain variety in the gaze behaviors of the characters, we want to bring in subtle changes in the importance of each parameter. The subscores can therefore be weighted to have more or less influence on the overall score. These weights are assigned randomly by the application on a per-character basis. Our overall scoring function is thus defined as:

$$S(t) = I_E(\omega_p S_p(t) + \omega_s S_s(t) + \omega_o S_o(t) + \omega_{pe} S_{pe}(t)) \tag{4.3}$$

where I_E is the *impact factor* of E. I_E can be used to give more impact to certain entities. For example, a user may want objects to be more important than characters. Once the best overall scores $S_{max}(t)$ (best $S(t)$) have been computed, we define the attention threshold A which determines the minimum score value for an attention behavior to be activated. This cannot be defined as an absolute threshold since the overall score values can greatly vary. For example, overall scores tend to be much higher in dense areas than in sparser ones. We thus compute A as the *(100-a)*th percentile of $S_{max}(t)$. a being randomly generated by the application and different for each character. The character will therefore only pay attention to interest points a percent of the time. The corresponding interest points will be the ones that have a higher score than A. This is depicted in Figure 4.1.

Our method automatically generates attention shifts since we calculate the interest points at each frame and for each character. However, if the interest point stays the same for a very long time, this generates unlikely behaviors. For example, if two characters are walking side by side, their respective scores for each other may be very high due to their proximity. In this case, they will keep on staring at each other, producing unrealistic attention behaviors. We therefore define a threshold duration d_l. If an interest point lasts for more than d_l, the entity of next highest interest is chosen as new interest point (as long as its score remains above the attention threshold). We arbitrarily set d_l to a maximum value of 4 seconds for the attention behavior not to last indefinitely. This also simulates interesting emergent behaviors. In the example given above, two characters walking side by side will oscillate between looking at each other and looking at another entity or back in front of them. They will thus give the impression of characters talking together.

The following sections explain the different elementary functions we implemented in more detail. All algebraic notations in these sections are represented in Figure 4.2.

4.3.1 Proximity

The proximity parameter evaluates the distance between a character C and all other entities E. Given $d_{ce}(t)$, the distance between C and E at time t, and $\alpha(t)$ the orientation of E in C's field of view at time t, our proximity score is computed as:

$$S_p(t) = \exp\left(\frac{-(0.5(d_m - d_{ce}(t)) + (\frac{d_m}{2} - 1))^2}{2}\right) \tag{4.4}$$

where d_m is the maximal distance value beyond which C will stop looking. This value may be modified by the user; he/she may wish to set d_m to a small value for narrow and densely populated areas. Conversely, he/she may wish to set it to a higher value for large open spaces. Our computation allows for entities situated $2-3$ meters away from the character to obtain the highest proximity scores. Indeed, we believe that entities closer than this will already have been attended to and thus should lose their interest potential.

4.3.2 Speed

For speed, we follow the same principle as for the proximity parameter. It is calculated as:

$$S_s(t) = \omega_{sw}||\mathbf{d}_e(t) - \mathbf{d}_c(t)|| \tag{4.5}$$

where $||\mathbf{d}_e(t) - \mathbf{d}_c(t)||$ is the relative speed, corresponding to the difference in distance traveled by C and E in one frame, and ω_{sw} is a weighting factor to bring the elementary speed scores to vary in the same range as the proximity scores.

4.3.3 Orientation

Similarly, our orientation score is computed as:

$$S_o(t) = (\pi - \alpha(t))\beta(t) \tag{4.6}$$

The larger the angle α, the more opposite the direction of E is in comparison to the direction of C. We want to give more importance to the entities coming towards C. We therefore weight the orientation score in order for the entities in the central vision to be favored as opposed to the entities in the peripheral vision.

4.3.4 Periphery

The last criterion is the periphery. This actually works just as the orientation. Calculations are the same, however, we give more importance to the entities in the periphery. The smaller the angle, the closer the direction of E is in comparison to the direction of C. Entities entering the field of vision will thus have small angle values. The periphery score is therefore calculated as:

$$S_{pe}(t) = \begin{cases} 0 & \text{if } \beta(t) > \beta_m \\ \omega_{pw}\alpha(t)(\pi - \beta(t)) & \text{otherwise} \end{cases} \tag{4.7}$$

where β_m is the maximum angle between the forward directions of C and E. Here as well, we weight the score with a weighting parameter ω_{pw} for the score range to be similar to that of the other criteria. We thus obtain all our subscores.

It is important to note that we further improve our algorithm by pruning a number of computations. Indeed, brute force computation proved to be counterproductive. For example, all entities farther than a certain distance from C need not be computed. We therefore prune the scores computation for each entity E. First, we use the maximum distance d_m. All entities farther than this from C are automatically discarded from further computation. Out of this subset of entities, we prune the process once more by considering only those in C's field of view. All following computations are done solely on this remaining subset of entities. We thus greatly reduce computational costs.

4.4 Automatic Motion Adaptation for Gaze

In the previous section, we explained how we define the interest points over an animation for each character. In the present section, we explain how we adapt the initial motions to obtain the desired gaze attention behaviors. Algorithm 4.2 shows the course of action at each timestep once the elementary scores have been computed.

Algorithm 4.2: Simulation Loop

```
Data: elementary scores
Result: motion adjustments
begin
    for all frame f do
        for all character C do
            for all entity E do
                compute overall score
            select maximum score
            check attention threshold
            if maximum score above threshold then
                set gaze constraint
                check temporal contribution
            edit motion
end
```

Each of the interest points we have calculated for a character C can be considered as a gaze constraint l_i in a set of gaze constraints **L**. C's motion thus has to be adjusted to meet these constraints. Since the interest points can be dynamic (in the case where they are moving entities), we have to compute the joint displacements to be applied to the base motion at each frame. As this is done on a per-frame basis, the overall performance of our system critically depends on the performance of our IK solver. To this end, we propose a robust and very fast dedicated gaze IK solver.

The skeletons we use are composed of 86 joints. Our method adjusts 10 of them: 5 spinal cord joints, 2 cervical joints, 1 head joint and 2 eye joints in order for the characters to align their gaze to the interest points. The eyes are *swing* joints and have 2 DOF. All others are ball and socket joints that have 3 DOF. This amounts to 28 DOF in all. By considering only this subset of the full skeleton, we greatly reduce the complexity of our algorithm. This allows us to have very small computational times and thus to animate a large number of characters.

Our method consists of two distinct phases. The first one computes the displacement map to be applied in order to satisfy the current gaze constraint. We name this *spatial resolution*. At each timestep, if there is an active constraint, we launch an iterative loop starting with the bottom of the kinematic chain (lumbar vertebrae) and ending with the top of the kinematic chain (the eyes). At each iteration, we calculate the total remaining rotation to be done by the average eyes position (global eye) to satisfy the constraint and determine the ratio of this rotation to be applied to the current joint. The remaining rotation to be done by each eye joint is then computed in order for them to converge on the interest point. Moreover, for interest points in the $30°$ composing the central foveal area, only the eye joints are recruited. For the $15°$ farther on each side composing the central vision area, only the eye, head, and cervical joints are recruited. Small movements therefore do not recruit the heavier joints. The second component is the *temporal* propagation of the displacement map over an automatically defined number of frames. This number is different if considering the eye joints, the head and cervical joints, or the joints composing the remainder of the spine. In this way, we allow for the lighter joints to move more rapidly than the others. The eyes therefore converge on the interest point well before any of the other joints attain their final posture.

4.4.1 Spatial Resolution

The purpose of the spatial resolution is to find a *displacement map* $\mathbf{m}(t)$ that modifies the initial motion in order to satisfy a given gaze constraint l_i defined by our automatic interest point detection algorithm. Similarly to Lee and Shin [Lee and Shin, 1999], we consider the initial motion as a set of independent character postures. We adjust each of these postures individually to satisfy l_i. To determine the displacement which should be applied to each of the recruited joints to satisfy l_i, we first calculate the 3D rotation $\mathbf{q}_l \in \mathbb{S}^3$, that aligns the global eye orientation to the position of l_i. Let $\mathbf{M}_{wt}$ be the rigid transformation matrix that transforms a point $\mathbf{p}$ in a local coordinate frame to its world position $\mathbf{x}_{wt}$ at time t. Let $\mathbf{l}_{wt}$ be the interest point position expressed in world coordinates. The vector $\mathbf{v}_{lt}$ going from the global eye to the interest point in the global eye frame is defined as:

$$\mathbf{v}_{lt} = \mathbf{R}_{wt}^T(\mathbf{l}_{wt} - \mathbf{x}_{wt}) \tag{4.8}$$

where $\mathbf{R}_{wt}$ is the rotational part of $\mathbf{M}_{wt}$ and $\mathbf{x}_{wt}$ is the global eye position in the world coordinate frame. Let $\mathbf{d}_{lt}$ be the initial looking direction expressed in the global eye frame. The complete rotation $\mathbf{q}_{lt}$, in local coordinates, is thus the shortest rotation to go from $\mathbf{d}_{lt}$ to $\mathbf{v}_{lt}$.

However, the eyes are not the only joints to adjust. To reach a natural posture, we dispatch this rotation to the other recruited joints. To determine the contribution c_i of each joint to the complete rotation $\mathbf{q}_l$, we take inspiration from the work of Boulic et al. [Boulic et al., 2004]. The authors proposed a set of formulas to define the angle proportions to be assigned to the spine joints depending on the type of rotation to be done (pitch, yaw or roll). We use the formula they propose for the linearly increasing rotation distribution around the vertical axis. In our model, the rotations around the sagittal and frontal axes are very small in comparison to those in the vertical axis. We therefore keep the same formula for all types of rotations:

$$c_i = (-(i-n))(\frac{2}{n(n-1)}) \quad i = 1...9 \tag{4.9}$$

where n is the total number of joints through which to iterate and i is the joint index with $i = 1$ the lowest lumbar vertebra and $i = 9$ the global eye. At each step, c_i determines the percentage of *remaining* rotation to be assigned to joint i. The total rotation to be done by each joint for the character to satisfy the constraint may then be calculated by spherical linear interpolation (slerp) using these contribution values. To reach the final character posture, we compute the remaining rotation for each eye to converge on the interest point.

4.4.2 Temporal Resolution

The speed at which we look at thing or objects varies depending on what we look at. For example, we will look rapidly at a lightning bolt. On the contrary, when we look at people walking in the street, our looking behavior is much slower. To reproduce this, we dynamically determine the activation duration based on the best overall scores $S_{max}(t)$. A point of high interest triggers a rapid movement and one of low interest a slower one. In other words, the activation duration t_a for a character C to satisfy a gaze constraint l_i is computed with the $S_{max}(t)$ at time t_b associated to that constraint. Given the maximum possible score S_{MAX}, t_a is then computed as:

$$t_a = \frac{\alpha S_{MAX}}{v_m S_{max}(t)} \tag{4.10}$$

where α is the angle of the total rotation which would have to be done by the head to satisfy l_i, expressed in radians and v_m is the maximum possible head velocity. The choice of value for v_m is motivated by a study conducted by Grossman et al. [Grossman et al., 1988]. The authors experimented on the maximum head velocity during vigorous voluntary yaw rotations. They obtained a median maximal velocity v_m of 4π radian per second. However, we hardly use our maximal head velocity in everyday life. We therefore set it to 2π radian per second in our model. t_a defines the number of frames it will take the head and cervical joints to satisfy the constraint. We double this value to obtain the number of frames in which the remainder of the spine will satisfy the constraint and halve it to obtain the number of frames in which the eyes will converge. This allows the lighter joints to move faster than the heavier ones. Finally, since a motion can take as long as one wants, there is no particular time threshold under which it should be done. We therefore arbitrarily set an upper bound value for t_a at 2 seconds in order for the motion not to be unnaturally long. In this way, if $S_{max}(t)$ is very high, the turning motion will be done fast. Conversely, if $S_{max}(t)$ is low, the turning motion will be done in a larger number of frames. Moreover, the eyes converge on the interest point faster than the head and cervical joints, which in their turn will satisfy the constraint before the remainder of the spine.

If the gaze behavior is deactivated, either because the character has been looking at a point of interest for too long or if there are no more interest points above the attention threshold, the character will look back in front of him. The duration of the deactivation t_d is randomly generated within an adequate range.

When we look around, our movements are not performed at a linear velocity. They start with an acceleration or ease-in phase, reach a peak velocity, and end with a deceleration or ease-out phase [Lee and Terzopoulos, 2006]. Moreover, they are desynchronized in time. The eyes move before the head, which itself moves before the spinal joints. To reproduce this, we further weight our rotation contributions c_i with a temporal propagation function $f_P(t)$ which follows a Gauss error function curve:

$$f_P(t) = erf(n/2) = \frac{2}{\sqrt{\pi}} \int_{-n/2}^{n/2} e^{-t^2} \tag{4.11}$$

where n is the total number of frames over which the gaze motion will be done. This computation is done with the different activation time values for the three sets of joints (eyes, head and torso). As depicted in Figure 4.3, we thus obtain a slight delay in the movement initiation between these three sets of joints.

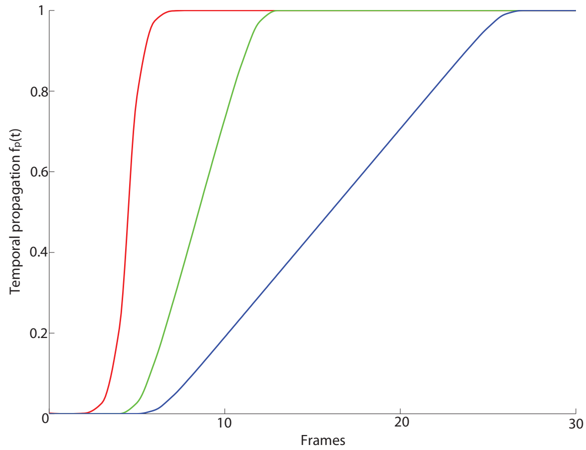

Figure 4.3: Desynchronization between the eyes, head, and torso. The eyes start moving before the head and satisfy the constraint first. The head and cervicals start moving and satisfy the constraint before the remainder of the spine.

Our final movement therefore allows the eyes to converge on the interest point and then partially re-center with respect to the head as the remainder of the joints move to satisfy the constraint. Indeed, they only partially re-center as a whole portion of the rotation is done by the eyes only. In our examples, most characters are in movement and the majority of the gaze constraints are associated to other characters in movement. These constraints are thus *dynamic*. We therefore recompute the displacement map to satisfy the constraint at each timestep. We can assume that the constraint's position from one frame to the next does not change much. We therefore recompute the rotation to be done at each frame but maintain the total contribution $f_P(t)c_i$ to apply which we calculated before the initiation of the gaze motion. However, we reset the contributions to 0 if the gaze constraint changes, i.e., if it is associated to another entity situated elsewhere in the scene. More specifically, this is the case when the current constraint location is farther than a pre-determined threshold from the constraint location at the previous frame. The newly calculated rotations to be performed by the joints to attain the new constraint's position are then distributed over the appropriate number of frames.

4.5 Results

We used our framework to create some examples of the possibilities of our method. The motion clips for our examples have been sampled at 30 fps. All the animations were generated on an Intel Core 2 Duo 3.0 GHz, 2GB RAM, NVidia GeForce 8800 GT.

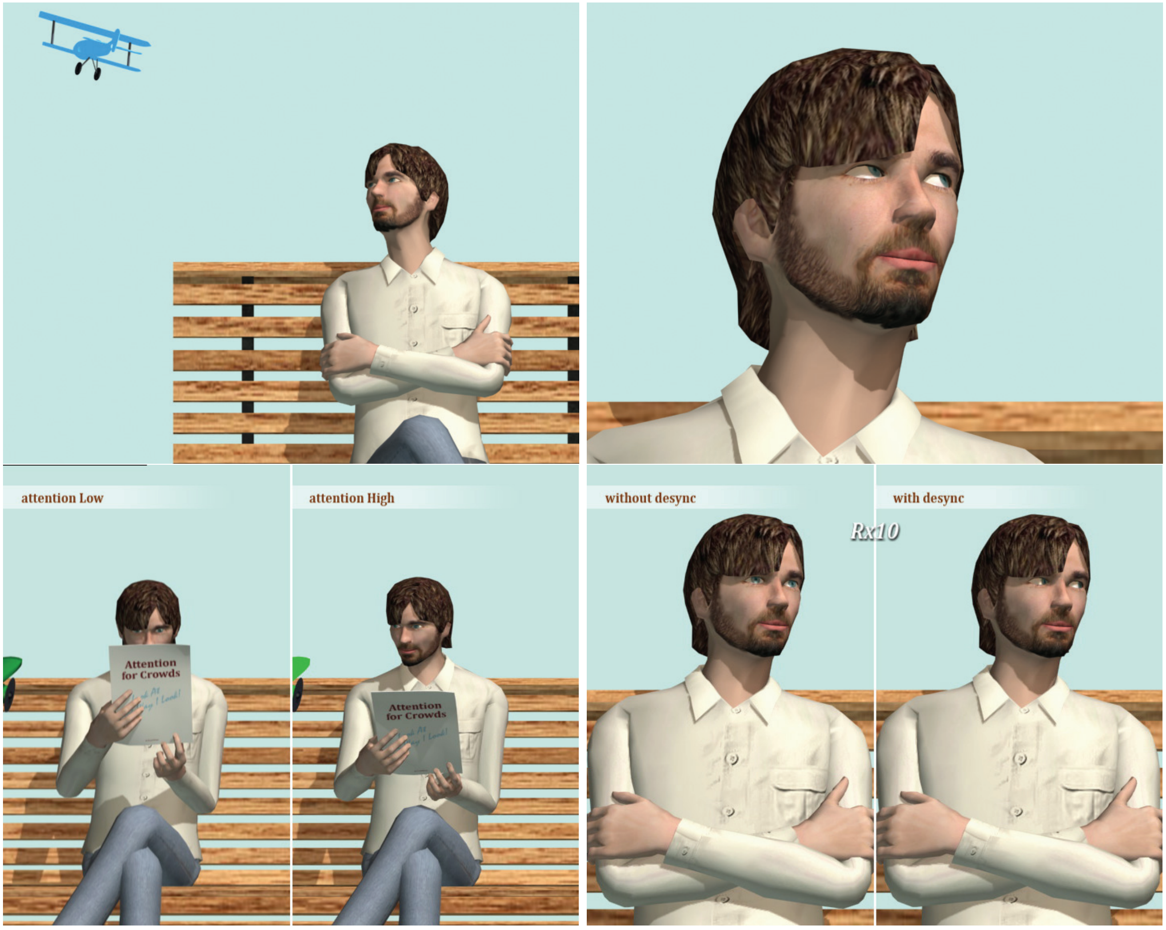

Figure 4.4: Illustration of a character following an interest point with different sets of parameters. *Top Left:* Illustration of the periphery parameter. *Top Right:* Desynchronization between eyes, head, and torso. *Bottom Left:* Illustration of the attention parameter. *Bottom Right:* Illustration of the presence/absence of desynchronization between eyes, head, and torso.

4.5.1 Gaze Behavior for Characters

In this example, we first illustrate the desynchronization between the three groups of joints (eyes, head, and torso). We then illustrate our different parameters by applying them individually to a single character. Figure 4.4 depicts the desynchronization, the attention parameter and the periphery parameter. On the top left, maximal values have been set to the periphery and attention parameters. The character is therefore sensitive to what happens in the periphery and is attentive 100% of the time. On the top right, the eyes of the character converge on the interest point while the head and spine joints have not yet satisfied the constraint. The bottom right image further illustrates the presence and absence of desynchronization. Finally, on the bottom left, we illustrate the difference between a very attentive character and a non-attentive one. The maximum looking duration is not activated in these examples since they aim at demonstrating the motion editing part of our method only.

4.5.2 Gaze Behavior for Crowds

Figure 4.5: Examples of gaze attention behaviors in a crowd animation.

In this example, we illustrate the use of our scoring algorithm together with the motion editing over 130 characters walking up and down a street, standing, or sitting on a bench. This is depicted in Figure 4.5. The maximum distance threshold was set to 10 meters. The attention threshold and the importance of each parameter was randomly generated by our application and is different for each character. For each one, the scoring algorithm is applied to all other eligible characters in the scene. Additionally, it is applied to all eligible scene objects defined as potential interest points (60 in all). We can thus simulate a simple form of top-down attention in the sense that some characters seem to be looking for something or trying to find their way. An interesting aspect emerging from those results is that some characters walking or standing next to each other regularly look at each other. We thus have the impression that they are talking together.

4.5.3 Complexity and Computational Times

The complexity for the automatic interest point detection algorithm is in $\mathcal{O}(n^2)$ with n being the number of characters. Indeed, for each character, we have to evaluate all other entities. However, since we do not compute the interest points for entities out of the character's field

Cs	100	200	300	400	500	600	700	800	900	1000
IPs	0.017	0.033	0.051	0.067	0.088	0.100	0.120	0.136	0.159	0.177

Table 4.1: Interest point detection computational times in milliseconds per character and per frame. Here the number of potential interest points is the same as the number of characters.

of view and farther than a threshold distance from it, this is greatly reduced and depends on population density. For the previous example discussed in Section 4.5.2, the computational time for the interest point detection, per character and per frame, was of 36.00 microseconds. Table 4.1 shows the interest point detection computational times in milliseconds per frame and per character for various numbers of characters. We can see that the automatic interest point detection can be done for hundreds of characters in real-time (almost 500). However, due to the complexity of our method, the computational times for 1000 characters are prohibitive. Nevertheless, the attention behavior animation of 1000 characters is not a necessity since users would only perceive those behaviors in the foreground.

We have also compared our dedicated IK solver with a typical Jacobian-based IK approach [Peinado et al., 2007]. To perform this comparison, the same skeleton has been used in both cases. The Jacobian-based approach takes a mean time of 20 milliseconds per iteration. If we consider that this method needs about 15 iterations before converging, this amounts to approximately 300 milliseconds to solve a constraint. Our method takes a mean time of 300 microseconds to solve a constraint. Moreover, since it is an analytical approach, we do not need more than one iteration to solve it. The complexity of our IK solver is therefore in $\mathcal{O}(n)$ with n being the number of characters.

It is to be noted that all steps undertaken in our method are on a single thread. The multi-threading capabilities of the computer on which the simulations are conducted are left aside for other tasks such as collision detection or rendering.

4.6 Discussion

Score computation. The most time consuming step in our method is the interest point computation. An interesting thing to do would be to use levels of detail (LOD) to reduce the complexity of this phase. Characters farther than a threshold distance from the camera may be assigned random interest points or no interest points at all since it will hardly be noticeable where or what they are looking at.

Motion continuity. An important aspect which should be mentioned is motion continuity, which we do not specifically ensure. However, we compute the displacement map at each frame from the motion captured postures. As long as this sequence is free from discontinuities, there is little chance for our final motion to present some.

Character trajectories. It is not the purpose of our method to edit the trajectories or to perform collision avoidance. These aspects therefore depend on the crowd animation engine which has been used to create the trajectories in the first place.

Motion capture validation. The values we used to distribute the rotation on the joints are empirical. It could therefore be interesting to try to determine them using motion capture and eye-tracking to improve realism. However, we believe that the amount of work needed to precisely determine each joint contribution would be tremendous in comparison to the added value this could convey.

Interest point scalability. A major advantage of our method is that it is extensible. We could easily add extra criteria such as color or contrast without having to modify the existing architecture. Another interesting aspect would be to provide entities with multiple interest points. A character with very flashy shoes would then attract attention to his feet. Similarly, a character waving his hand would attract attention if we consider the body parts' relative velocity. Finally, sound could also be added as it has a very strong attention capture potential.

4.7 Conclusion

In this chapter, we introduced a novel method to enhance crowd animation realism by adding attention behaviors to the characters composing it.

We first proposed an automatic interest point detection algorithm which determines, for each character, *where* and *when* it should look. We additionally presented an extensible and flexible set of criteria to determine interest points in a scene and a method to combine them. Our method also allows the fine-tuning of character attention behaviors by introducing an attention parameter as well as the possibility to modify the relative importance of each criterion if desired.

Secondly, we introduced a robust and very fast dedicated gaze IK solver to edit the character motions. Our solver deals with the spatial and temporal resolution of the gaze constraints defined by our detection algorithm.

Finally, we illustrated our method with visually convincing results obtained with our architecture. We believe that gaze behaviors greatly enhance crowd character believability, and thus, greatly amplify the immersive properties of virtual crowd scenarios in the context of VRET of agoraphobia. To this extent, the next chapter of this thesis tackles the application of such gaze behaviors in an immersive environment and with interaction possibilities.

CHAPTER 5

Interaction with Virtual Crowds for VRET of Agoraphobia

In this chapter, we describe an application which brings together the work we have described in the last two chapters, but also some additional features. This application allows characters to perform gazing motions in a real-time virtual crowd in a CAVE environment. Moreover, it allows for users to interact with those crowd characters.

First, we adapted the model of visual attention described in the previous chapter in order to integrate it in a crowd engine and allow it to function online. We also greatly simplified certain aspects of the automatic interest point detection. Finally, we modified the existing architecture in order to abide with the limitations induced by the online implementation.

Our final application consists of a city scene, projected in a CAVE setup, and in which a crowd of characters walks around. We then use a Phasespace optical motion capture device [Phasespace, 2009] to evaluate where a user is looking and more specifically, which character he/she is looking at. Finally, we further enhance this setup with an RK-726PCI pupil/corneal reflection head-mounted tracking device [Iscan, 2009] in order to evaluate more precisely where a user is looking. Our system then allows the crowd characters to react to user gaze. For example, since we can determine the user's position and orientation in the virtual world, the characters can look at the user.

It is to be noted that the head-tracking alone allows stereographic vision in the CAVE, using red and blue polarized glasses. In the case of head-tracking coupled with eye-tracking, however, stereography is impossible as the glasses prevent correct tracking of the pupil and corneal reflection. More details on the stereographic rendering in the CAVE can be found in [van der Pol, 2009].

Finally, we have used this setup together with a gesture recognition algorithm developed by van der Pol [van der Pol, 2009] in order to obtain increased interaction between user and virtual characters. The gesture recognition is done by tracking the user's hand. We thus use a data glove with Light-Emitting Diodes (LEDs) which is also part of the Phasespace motion capture system.

5.1 System Overview

The overall setup of our system is depicted in Figure 5.1. It illustrates the various elements composing it and how they are interconnected. The various trackers (eye-tracker, head-tracker and data glove) send their position and orientation information to their dedicated software on different PCs. All this information is then gathered by the PC running the simulation and used to correctly render the images on the various CAVE screens and define the character gaze and gesture behaviors.

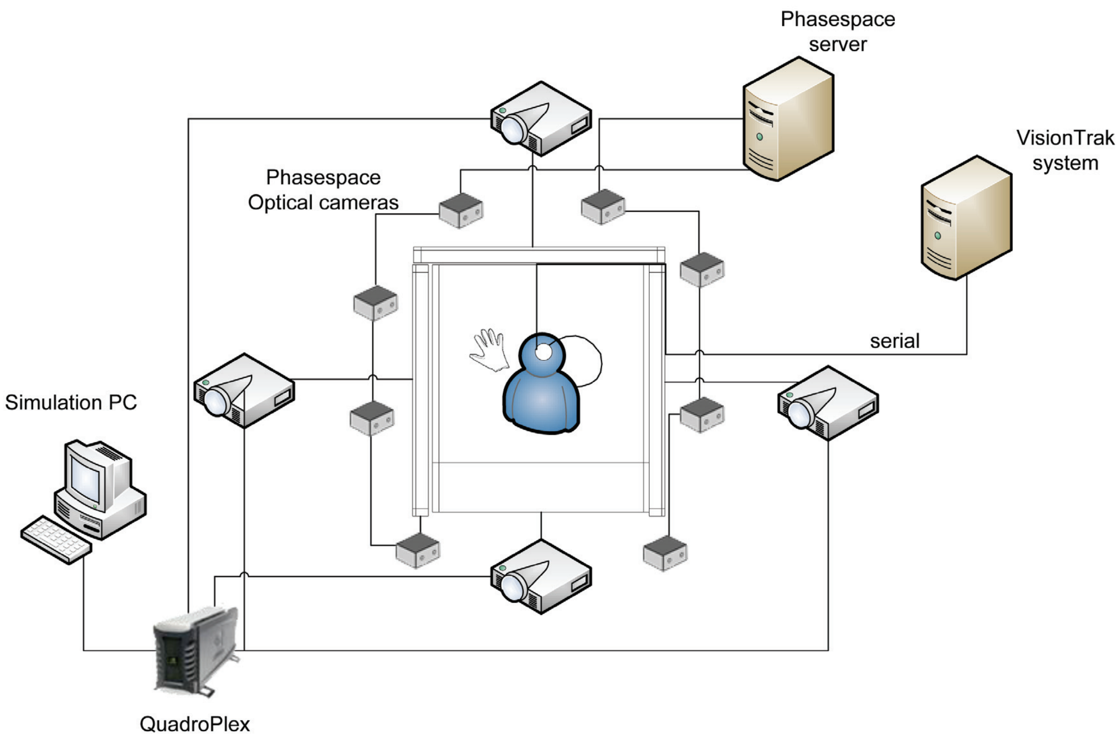

Figure 5.1: Hardware setup in the CAVE environment.

The pipeline for the gaze behaviors is depicted in Figure 5.2. The *Simulation*, *Animation* and *Rendering* phases in the diagram all belong to the crowd simulation engine we have implemented our system in [Maïm et al., 2009]. This diagram shows the real-time simulation loop pipeline. Other steps occur beforehand, such as the loading of the character templates and scene. In the real-time simulation loop, our system integrates after the *Animation* step. The crowd characters are thus animated with walking gestures for example. The motion adaptation that we discuss in the *Gaze* section is therefore applied on top of the current animation. This allows for the characters to maintain their original walking motions as much as possible. In this *Gaze* phase, we first determine the interest point for the current character. We then compute the motion adaptation which would be needed to satisfy this constraint. Finally, we determine the quantity of this adaptation which has to be done at the current phase depending on spatial and temporal factors.

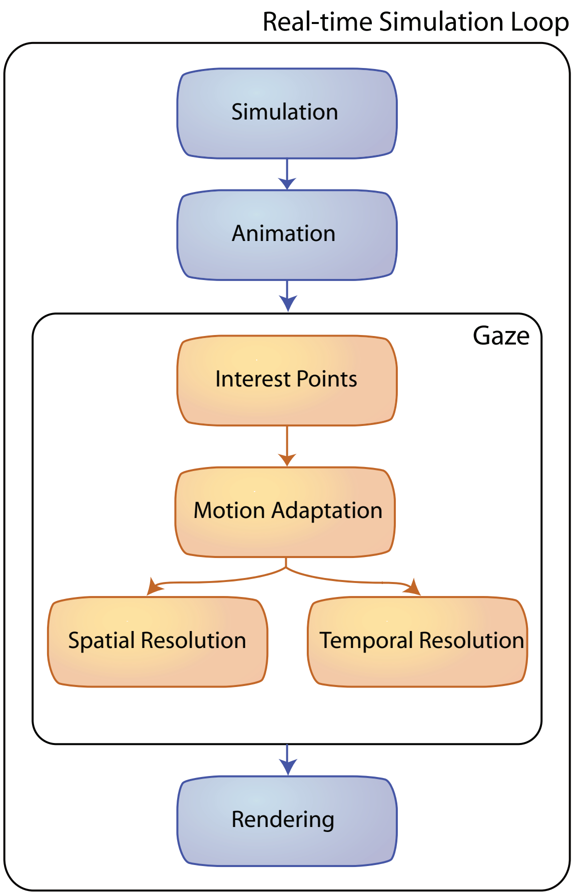

Figure 5.2: Crowd engine real-time simulation loop pipeline.

5.2 Tracking in the CAVE

By definition, the tracking of the user's head in the CAVE is a necessary step. First of all, in order to compute the correct projection matrix for the CAVE screens, the position of the user's eyes needs to be known. This can be approximated by tracking the user's head. In order to do this, we use a regular baseball cap on which we fixed three motion capture markers. Using the positions of these three markers, we approximate the user's eyes position. Second, the tracking of the head also allows us to determine the user's head orientation. We can thus determine the user's gaze direction and by projecting the POR, determine which character he/she is looking at at all times. Finally, the position of the markers allows us to determine where, in the 3D environment, the user is located. We can thus use the user location to define a point of interest and let the virtual characters seem to be aware of the user. We have equally used a setup in which we added an eye-tracker to the head-tracker. The eye-tracker allows us to obtain precise gaze values instead of the approximation obtained with the cap only. In this setup, we use a head-mounted eye-tracking device on which we fixed three optical markers. By projecting the user's gaze on screen and using the available methods in the crowd simulation engine we have used, we determine which character is being looked at by the user at all times. The user's position and what or who he/she is looking at then allow us to define what we consider as the most interesting points in the scene for the virtual characters to look at.

5.3 Interest Point Definition

In creating this application, our aim was to provide an immersive environment in which a user could interact with the virtual characters. We thus hypothesized that the user should be a point of high interest for all virtual characters. Moreover, we hypothesized that what or who the user is looking at should also be considered as a point of high interest for the virtual characters. Indeed, when someone around us looks in a specific direction, we have a tendency to gaze in the same direction to find out what he/she is looking at. Since the tracking of the user's head and gaze allow us to define where he/she is standing in the virtual environment, we just use this position as the interest point to be looked at by the characters. Similarly, when the user is looking at a virtual character, since we know its position in the 3D environment, we can use it (at eye height) as a point of interest. Finally, we also hypothesized that the characters should not always look at the user but should look at other characters walking around in the environment. The third type of interest point is thus a random selection of any other character or object in the scene. We have thus created various modes to determine the points of interest:

- **User-centered:** all characters are interested solely in the participant, and thus look at nothing else but the participant. The only interest points are the user's position.
- **Interest-centered:** all characters, apart from the one being looked at, are interested solely at what the participant looks at. The character being looked at, looks at the user. The main interest point is thus located where the user is looking at. The user's position is the second interest point, only for the gazed at character.
- **User or Interest:** the characters randomly choose to either look at the participant (user-centered) or at what the participant is looking at (interest-centered). The interest point is thus either at the user's position or at the location of the looked at character.
- **Random:** the characters look at any other character in the scene, including the participant. The choice of character or user to look at is completely random. The interest points can thus be at any of the character's or user's location.

We have created these different modes in order to test which seems the most natural. The experiment we have conducted to this effect is described in Chapter 7.

In the user-centered mode, the interest point always remains the same and corresponds to the user's position in the 3D environment. Of course, this position may change over time since the user is free to move about. The user's position is thus sent to the *Gaze* module at each time step. All characters look at the user as long as the user is in the character's field of view. We define a maximum gaze duration in order to avoid characters staring too much. As in the previous chapter, we set this maximum duration to an empirical value of 4 seconds. We based this on the average mutual gaze durations recorded by Lee et al. [Lee et al., 2002c]. This value was of 94 frames at a 30 fps frequency, which we rounded up to the next full second since we take into account the duration of the movement. Thus, if a character has been looking at the user for 4 seconds, the gaze behavior is deactivated and the character returns to its original walking motion. After a random lapse of time, the character may again look at the user, as long as the user is in its field of view. Moreover, we use a minimum gaze duration in order to avoid small saccadic movements. We have arbitrarily set this value to half a second. The way in which we enforce this is discussed in Section 5.4.1.

In the interest-centered mode, the interest point changes. It is positioned wherever the user is looking. In a first step, all characters in the region of interest are selected, whatever their distance from the user. Then, the first one in depth is chosen as point of interest. Indeed, we thought it a reasonable assumption that the character closest to the user would be the one gazed at by the user. Each character will thus look at the user's interest point as long as it is in its field of view. Only the character being looked at will have a different interest point, situated at the user's position. As for the user-centered mode, we use a maximum and minimum gaze duration and a random lapse of time before the character can gaze at the interest point again if the gaze has been deactivated.

In the user or interest mode, the interest point is randomly chosen between the user and the user's interest. In this case, we keep track of the current looked at id, i.e., the id of the character (or user) being looked at. The interest point then remains the same until the gaze is either deactivated or another interest point is chosen. Here as well, an interest point is deactivated either because it has lasted for too long, or because it is no longer in the character's field of view. As in the two previous modes, we enforce a minimum and maximum gaze duration.

Finally, in the random mode, the interest point can be any character or object in the scene. This includes the user. As for the user or interest mode, we keep track of each character's interest point in order for it to remain the same for at least the minimum gaze duration and for it to be changed after the maximum gaze duration. When the character chooses an interest point, if it is not in its field of view, it continues with its original walking motion and waits until the next step of the simulation before selecting another interest point. We thus have characters that are not looking at anything which makes the simulation more natural.

We have also allowed for a certain percentage of the virtual population to be interested. We can thus simulate some form of mood or personality. Some characters will never look at anything while others will be very attentive to characters around them. We can thus simulate a very distracted population in which only a couple of characters will seem interested. On the other hand, we can simulate a very attentive population, in which all the characters will seem very interested in what is going on around them.

5.4 Automatic Motion Adaptation

After having defined the interest points the characters should look at, the original character motion has to be adapted in order for it to perform the gaze motion in a smooth and natural way while keeping a maximum of its original walking motion. The method we use for motion adaptation is very similar to that presented in the previous chapter. However, we modified it for online use. We thus will not go into the details of the method in this chapter, but rather explain the differences with the method presented in the previous chapter. Algorithm 5.1 describes the overall gaze simulation loop. It shows the different steps which are done by our method at each time step and for each character.

Algorithm 5.1: Gaze Simulation Loop

```
Data: character id
Result: character joints' position and orientation
begin
    for character id do
        if not gaze active then
            check mode
            set constraint
            if constraint in field of view then
                compute duration of movement
                compute motion adaptation
                compute percentage of motion adaptation
        else
            if duration >= max gaze duration then
                deactivate gaze
                compute percentage of deactivation
            else
                check last constraint
                update constraint position
                if constraint in field of view then
                    compute motion adaptation
                    compute percentage of motion adaptation
                else
                    if duration < min gaze duration then
                        keep previous posture
                    else
                        deactivate gaze
end
```

The main difference with the method proposed in the previous chapter is that this method works online. We have integrated the gaze motion adaptation directly into the crowd simulation animation loop. We thus do not have any *a priori* knowledge of the duration of an interest point. We therefore do not have a set of pre-determined gaze constraints as in the offline method. This does not have much impact on the spatial resolution but it does on the temporal resolution. The computation of the overall displacement map to be applied to the current character posture remains the same as for the offline method. We thus do not discuss this part of the method in this chapter. However, the computation of the amount of rotation to be applied to each joint at each frame needs to be adapted in order to meet online requirements. These modifications are discussed in the following section.

5.4.1 Temporal Resolution

The main problem of not having any *a priori* knowledge of the constraint durations is that we can not filter out the constraints which last under the minimum gaze duration. We therefore have to deal with this on the fly in order to avoid very small, saccadic movements when a point of interest lasts only a fraction of a second.

Whenever a new point of interest is selected, we define the amount of time the character should take to perform the gaze motion, depending on the angle of rotation which has to be done; the larger the angle, the longer it takes to perform the motion. We then simply interpolate at each time step to determine the amount of total rotation which has to be done by each joint.

However, the problem which can arise is that the interest point can change or exit the field of view before the gaze motion is finished, i.e. before the joints have attained the final gaze posture. The fact that the motion is not finished is not actually a big problem. However, if this is the case, the gaze duration will necessarily be very small. This induces very unrealistic behaviors where characters perform very small saccadic movements. In order to counter this, we have defined a minimum gaze duration of half a second. If the interest point is deactivated before attaining this minimum duration, we artificially maintain the interest at the previous point of interest. Actually, we maintain the previous character posture until the minimum gaze duration threshold is attained.

Another difference in the gaze simulation loop between the offline and online methods is that, in the online method, we adapt each character at the current frame, whereas in the offline method, we adapt each character's complete animation before going on to the next one. We thus have to keep track of each character's previous posture in the online method. Moreover, we also need to keep track of the starting posture when initiating gaze deactivation or a change in interest point. Indeed, at each time step, we start from the character's original walking posture and not from the adapted posture at the previous time step.

Finally, the remainder of the temporal resolution stays the same as in the offline method. More specifically, the computation of the different time values for eyes, head, and spine to satisfy the gaze constraints and the desynchronization between these three sets of joints, remains the same.

5.5 Gesture Recognition

In addition to the crowd character gaze behaviors, we have allowed our virtual characters to respond to waving gestures. The method implemented to do this is described by van der Pol [van der Pol, 2009]. This method allows the recognition of motion captured gestures using point & figure charts. The user wears a glove equipped with 8 LED markers. This glove is also part of the Phasespace system [Phasespace, 2009] as mentioned above. The user then does a waving gesture which is recorded and represented as a point & figure chart. This system consists of representing an increasing value with an X and a decreasing value with an O. The user's gestures during immersion are then also represented as point & figure charts which are compared to the original template waving gesture. If there is a match, the character which is being looked at by the user stops its walking motion and performs a waving gesture in return. A predefined threshold value is used in order to filter out small movements from the chart. The gesture therefore need not be exactly the same as the template gesture in order to be matched. Figure 5.3 shows an example of a character waving back at the user. Here, the user's gesture has been matched to the original waving gesture. The character being looked at therefore stops walking and waves back at the user.

Figure 5.3: A user's gesture is recognized by the point & figure chart as a wave. The looked at character therefore waves back.

5.6 Results

The main results we have obtained are discussed in Chapter 7 in which we describe an experiment we have conducted in order to evaluate our application and its different modes. In this section, we present the visual results we have obtained with our application.

Figure 5.4 depicts the various gaze behaviors depending on the chosen mode. All of them are rendered from the user's point of view (in this particular case, the camera). On the top left of the Figure, all characters look at the user as long as he/she is in the character's field

Figure 5.4: Results obtained with the integration of our gaze behaviors into a real-time crowd animation engine. *Top left:* User-centered gaze behaviors. *Top right:* Interest-centered gaze behaviors. *Bottom left:* User or interest gaze behaviors. *Bottom right:* Random gaze behaviors.

of view. On the top right of the Figure, all characters look at what the user is looking at, as long as it is in the character's field of view. On the bottom left of the Figure, the characters either look at the user or look at what the user is looking at. Finally, on the bottom right of the Figure, the characters look at any other character or object in the environment. The user is also included in the potential points of interest.

What we can see from these images is that it is quite clear that all the characters are looking at the user in the user-centered mode. Also, it is visible that the characters seem to not pay attention to the user at all in the random version of the scene. The user or interest mode (mix between user-centered and interest-centered) is the one that seems to be the most natural. Some characters pay attention to the user and others not. They look at the user but also at other characters. This can be explained by the fact that a user shifts his/her attention quite often. The characters having to look at the user's interest point therefore do not all look at the same thing due to the features of our implementation. Indeed, when an interest point lasts for a too short period of time, we have forced the attention to be maintained on the same point. Some characters thus look at what the user is currently looking at and others may be looking at what the user was looking at beforehand. In the interest-centered mode, however, it is difficult to notice the variations in comparison to the random mode. The interest-centered mode can be perceived in the same way as the random one since the characters don't seem to be paying attention to the user at all. Moreover, and as previously mentioned, all characters will not be looking at what the user is currently looking at but also at what the user was looking at beforehand.

5.7 Discussion

Interest Point Definition. Our current implementation does not compute interest points automatically as in the offline method. Our totally random mode would probably benefit from such an implementation instead of randomly choosing from all characters (or the user). We believe this would allow this mode to seem much more natural than in its current state.

Gaze modes. Other gaze modes could be created by adding an impact factor to some characters or objects as in the offline method. For example, we could allow a sculpture in the middle of a square to have a high impact. Many characters would then look at the sculpture and then switch to looking at other characters or the user. Similarly, we could use the random mode in which the user would have a high impact. Characters would then mostly look at the user but would also look at one another from time to time.

Sound. This application could also greatly benefit from the addition of sound. Indeed, we believe that sound, and more specifically 3D sound, could greatly increase its realism and thus potentially allow the exposure to become much more anxiety provoking.

5.8 Conclusion

In this chapter, we have presented an application that allows interaction with virtual crowd characters in a CAVE environment. By using motion capture, we can determine where a user is positioned and what he/she is looking at. We use this data in order to let the characters perform gaze behaviors and waving gestures. The characters thus seem to be aware of the user and interact with him/her.

We have explained in detail the method we use to determine the points of interest the characters should look at. We have also described how the original character motions should be adapted in order for them to perform the gaze behaviors. Finally, we have explained how this method differs from the one presented in the previous chapter in order for it to respond to online requirements.

CHAPTER 6

Experimental Validations - Social Phobia

In this chapter, we first go over the equipment we used in our various experiments in the context of social phobia. We then describe three different experiments we have conducted in this domain. The first was done in order to validate the scenarios we have discussed in Chapter 3.2. The second experiment was done in order to evaluate the potential of eye-tracking for the diagnosis and assessment of social phobia. Finally, our third experiment evaluates the potential of the application we have described in Chapter 3.3 which allows interaction between a user and virtual characters based on gaze behaviors.

6.1 Experimental Setup

The equipment setup we used to conduct the experiments relative to social phobia is depicted in Figure 6.1. This setup contains various components. The first one is a head-mounted eye-tracking device, depicted on the left of Figure 6.2. This device is an ISCAN RK-726PCI pupil/corneal reflection tracking system [Iscan, 2009]. It is a head mounted system which operates at 60 Hz. It is composed of two cameras, one directed toward the eye and the other, in the opposite direction. It also has a dichroic mirror and an infrared eye illuminator. It tracks the subject's pupil and a reflection from the corneal surface. The distance between the two then allows the determination of the eye position, and the POR. In order for this to function properly, the subject first has to go through a calibration process. This consists of successively looking at five points projected on screen; on the top left, on the top right, in the middle, on the bottom left and on the bottom right. During this process, the user has

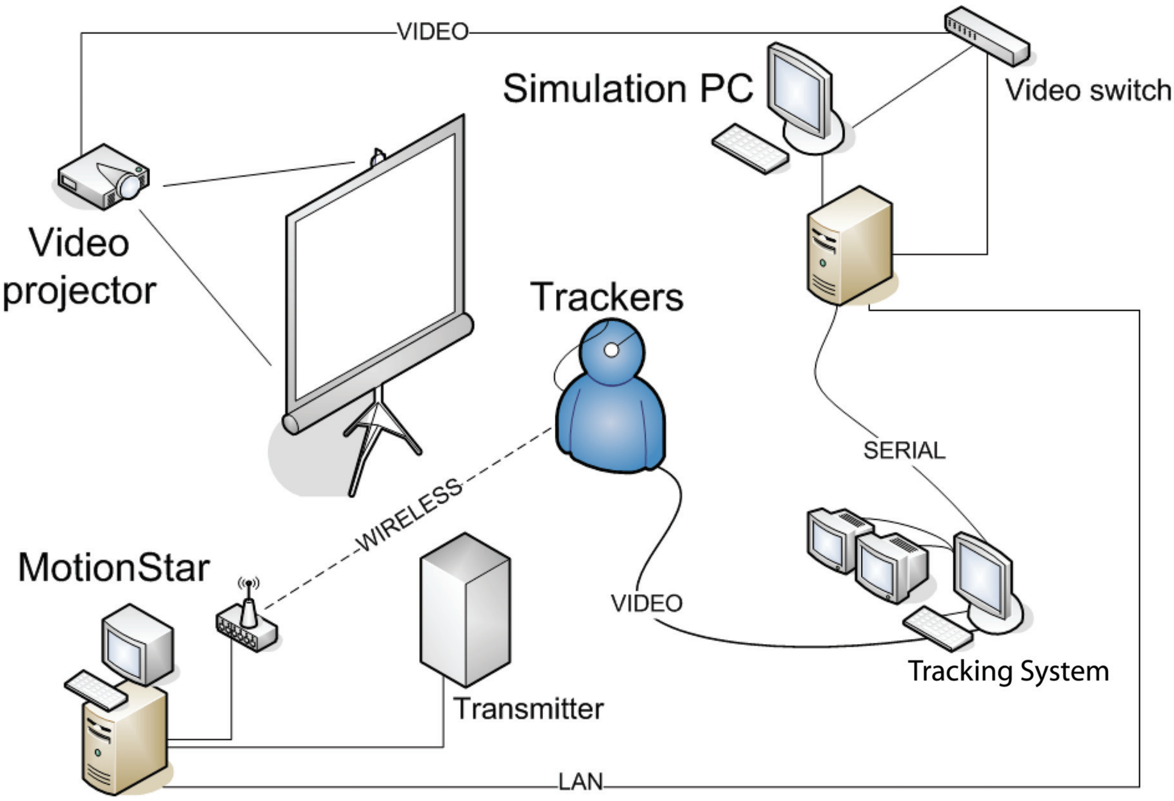

Figure 6.1: System setup

to remain still and not move his/her head. These five measures are then used as reference points. The xy coordinates sent by the system during tracking are then interpolations between these reference points. With this tracking device comes a PC with two specialized PCI cards. These implement the video processing hardware. The system computes the pupil and corneal reflection and sends the result to dedicated software used for the calibration and data acquisition. The eye-tracking device allows the determination of a subject's eye position with an accuracy of 0.3 degrees of visual angle, which corresponds to approximately 2 centimeters on screen. However, we have to account for possible drift from the system due to headband slippage. To this end, Herbelin experimented on this system's accuracy [Herbelin, 2005]. He concluded that objects of approximately 13% of the screen width could be hit by the tracked eye with an 80% probability for a person sitting 2 meters away from the screen. However, as will be shown in the various experimental studies, the virtual characters we use take up a large portion of the screen. This greatly reduces problems which can be encountered due to accuracy issues.

The second main component is a 6-DOF magnetic sensor: a wireless MotionStar from Ascension [Ascension, 2009]. It is added to the head-mounted eye-tracking device as shown on the left of Figure 6.2. This allows the coupling of eye- and head-tracking. By tracking the head position and orientation together with the eye, the subject can move around freely. The sensor accuracy is of approximately 1 centimeter in position and 0.15 degrees in rotation, which corresponds to approximately 1 centimeter on screen. Since this tracker is positioned on the eye-tracking headband, we equally have to account for possible slippage in our ac-

Figure 6.2: Eye-tracking device, *Left:* eye-tracker. *Right:* subject wearing the device, seated in front of the back-projection screen (*Photo: Alain Herzog).*

curacy calculations. Moreover, this type of sensor presents some drift after some time of use. The rotations done by the head around the vertical and horizontal axes induce a rotation around the sagittal axis. We thus have to recalibrate the system at regular intervals in order to correct and cancel this extra rotation.

Finally, the last component of this setup is a large back-projection screen on which our environments are displayed. This can be seen on the right of Figure 6.2. Its size is of 3.2 meters x 2.4 meters on which we display the scenes at a resolution of 1024 x 768. In this setup, the subject is generally seated approximately 2 meters away from the screen, or standing and walking around in its vicinity.

6.2 Clinical Study on VRET for Social Phobia

As preliminary work to that presented in this experiment, Herbelin et al. and Riquier et al. have conducted a study during which they exposed subjects to a VR situation representing a 3-dimensional audience composed of emergent gazes in the dark and surrounding the subject [Herbelin et al., 2002; Riquier et al., 2002]. They experimentally confirmed that the audience was able to provoke more anxiety to social phobics than to non phobics and emitted the hypothesis that eye contact was an important factor of social phobia. Herbelin therefore developed and experimented with an eye-tracking setup integrated in the VR system [Herbelin, 2005]. As mentioned in Chapter 2, he developed a tool that allows to determine not only if virtual characters are being looked at, but which parts of the characters are being looked at by a user wearing an eye-tracker.

We hereby describe the experiment we have conducted in order to validate his work and the various scenarios we have discussed in Chapter 3.2.

The goal of this study was to define a therapeutic program for social anxiety disorders using VR and to assess its efficacy in order to confirm that VR was a promising tool for psychotherapists as part of social phobia treatment. We equally wanted to evaluate the use of eye-tracking as a new tool for the assessment of social phobia and to see if the technology could “provide therapists with an objective evaluation of gaze avoidance and can give tangible feedback to the patients to estimate their progress during gaze behavior exercises.” [Herbelin, 2005](p.62). To do so, we submitted an ethics commission request (see Appendix A) which was approved by the ethics committee. As mentioned, our eye-tracking based architecture allowed us to put into evidence and analyze which zones or character parts were being looked at by a subject or participant.

6.2.1 Technical Information

To evaluate the efficacy and the potential of VRET, we used one of the social situations which is most characteristic of social phobia: the fear of public speaking. We used Hofmann’s model [Hofmann, 1997; Hofmann and Otto, 2008] to define the therapeutic protocol to be followed. This model consists of maximum exposure to social anxiety in the context of public speaking. We have conceived a framework based on that model in which we replaced the group exposure situations by individual exposure sessions to different virtual public speaking scenarios depicted in Figure 3.1. Phobic subjects often rely on avoidance strategies to deal with fearful situations. The aim of these exposures was to confront the subjects to their fear and by habituation, make them cope with anxiety instead of avoiding it.

In order to evaluate the efficacy of our program, we used various scales specific to social anxiety disorders at different phases of the treatment. The first one is the Fear Questionnaire, which consists of rating the main phobia, global phobia, total phobia and anxiety-depression [Marks and Matthews, 1979]. The total phobia is composed of agoraphobia, social phobia and blood-injury phobia. The second scale is the Liebowitz social anxiety questionnaire [Yao et al., 1999]. Its objective is to assess the range of social interaction and performance situations that people suffering from social phobia may fear or avoid. The third scale is the Social Interaction Self-Statement Test (SISST) [Yao et al., 1998]. This consists of 15 positive and 15 negative self-statements relative to difficult social situations. Finally, we have used the Beck Depression Inventory (BDI) [Beck et al., 1961] which measures the severity of depression. Our aim was to obtain an improvement and a normalization of the score values for each subject after treatment as well as to uphold the improvement during the follow-up evaluations.

The setup we used for the eye contact evaluation is described at the beginning of this chapter. The subjects were exposed to two virtual scenes, namely the office and auditorium scenes depicted in Figure 3.1. The equipment used by the therapist consisted in a Kaiser ProView XL50 HMD [Collins, 2009]. This HMD has a resolution of $1024x768$ and a field of view of 50 degrees. It was coupled with an Intersense inertial sensor [Intersense, 2009] in order for the participants to be able to look around in the environment. The sensor allowed the images to change with respect to the participant’s head movements.

6.2.2 Selection and Description of Participants

The study was conducted over 8 subjects recruited via a mailing sent to students in second and third years of college and via ambulatory consultations specialized in anxiety disorders. 5 of these subjects were females and 3 of them males. Our therapist collaborator admitted them to participate in the study after a structured interview regarding socio-demographical

Week	Questionnaires	Session Type	Theme	Equipment
-2	yes	Individual	-	-
-1	yes	Individual	-	-
0	yes	Individual	-	Eye-tracking
1	no	Group	Social phobia	-
2	no	Individual	Hobbies	HMD
3	no	Individual	Profession or education	HMD
4	no	Individual	Memorable event	HMD
5	yes	Individual	Dramatic situation	HMD
6	no	Individual	Conflict situation	HMD
7	no	Individual	Anxiety related to love	HMD
8	no	Individual	Efficient communication	HMD
9	yes	Individual	-	Eye-tracking

Table 6.1: Overview of the experimental protocol.

variables as well as psychiatric and medical antecedents. She led a diagnostic according to the DSM-IV's 5 axes [American Psychiatric Association, 1994] for each subject and presented them with the Mini International Neuropsychiatric Interview (MINI) [Sheehan et al., 1998] in order to verify the prevalence of social phobia and absence of comorbidity. Originally, 10 subjects were selected but two dropped out during the first phase of the treatment.

6.2.3 Experimental Protocol

We evaluated eye contact behavior in a pre-therapeutic phase and at the end of the treatment. During these phases, the subjects were asked to do a 3 minute verbal expression exercise. We evaluated eye contact avoidance from the eye-tracking recordings during exposure to the virtual scenes. We were then able to analyze these recordings and materialize the data as a map showing the zones swept by the gaze as well as the lapse of time contact lasted. These are discussed in more detail in Section 6.2.4.

The main part of the study was a clinical experiment following an A-B protocol, outlined in Table 6.1. During the A phase –the non-intervention phase– we established and analyzed the target symptoms evolution curve through 3 evaluation sessions. During the B phase –the intervention phase– the therapist exposed the subjects to anxiety provoking situations through the HMD on a weekly basis, during 8 weeks. Each session lasted approximately 30 minutes of which 10 in the HMD. These were conducted in the therapist's office.

During the A phase, at weeks -2, -1 and 0, we asked the subjects to fill in the various questionnaires (Fear, Liebowitz, SISST and BDI). We then analyzed them and averaged the results we obtained over the three weeks in order to obtain a before-treatment value for each subject and each scale. Between the A and B phases, the subjects participated in a group session without VR. The therapist instructed them on social phobia and asked them, one after the other, to give a speech on what they had learned about this anxiety disorder.

For the B phase, the subjects were asked to classify 8 social situations from least anxiety provoking to most anxiety provoking. They were then exposed to these various virtual situations throughout the 8 HMD sessions, each more anxiety provoking than the previous one. The proposed virtual scenes were:

- In an office, facing one man or one woman
- In an office, facing 5 people
- In an auditorium, facing one man or one woman
- In an auditorium, facing approximately 20 people or sitting at the back of the room
- In a cafeteria, facing one person but with many people around
- In a bar, facing one person but with many people around

It is a typical trait of social phobia to find it harder to deal with a person of the opposite sex. We therefore considered the office with one person (man or woman) and the auditorium with one person (man or woman) as four different situations. We have equally noticed that vocal interruptions from the therapist during HMD exposure could create breaks in presence. We avoided these by making our virtual characters talk instead. We recorded a number of sentences which could be triggered by the psychotherapist to respectively begin, continue and end the speech sessions. We equally set up the virtual characters with facial animation corresponding to each of the pre-recorded sentences. Finally, we set up our characters with a “look at” function which allowed them to make eye contact at all time and more specifically when talking to the exposed subject. For the first HMD session, each subject was asked to present social phobia once again, as they did for the group session. Then, each week, they were asked to prepare the following week’s session. As homework, they had to prepare the following week’s speech in front of a mirror in order to auto-evaluate their body language. Sessions 2 to 8 dealt with the following themes:

- Session 2: talk about hobbies
- Session 3: talk about professional or educational activity
- Session 4: talk about a memorable event
- Session 5: talk about a dramatic situation
- Session 6: talk about a conflict situation
- Session 7: talk about anxiety related to love
- Session 8: give a lecture on “efficient communication” from given documents

These situations were presented to them in this specific order for each to be more personal than the previous, and therefore, more anxiety provoking. However, since the subjects were not affected by each situation in the same way, the situations were modulated according to each one. As an example, some subjects recited a poem or sang a song for session 7 because talking about their love life was not sufficiently anxiety provoking.

As in Hofmann’s therapeutic program, they were asked, as homework, to prepare and repeat speaking exercises in front of a mirror. They were also asked to try to decrease their avoidance behaviors in real life. Finally, they were asked to fill out a “fearful situation” document (see Appendix B) in which they exposed the anxiety provoking situations to which they had been confronted in the past week as well as the degree of avoidance, the degree of anxiety, and the satisfaction felt throughout these experiences. These documents were then used as a basis for each weekly discussion.

The subjects were asked to fill out the same 4 questionnaires as during the A phase of the treatment at week 5, half way through the treatment and once again at week 9, after the end of the HMD exposure sessions.

	Phobic norm	Before	After
Fear Men	21.4(5.44) - 24.4(8.0)	19.8(7.78)	16.5(4.95)
Fear Women	15.94(8.96) - 23.4(8.4)	22.3(8.46)	16.4(7.83)
Liebowitz	67.2(27.5)	73.3(22.28)	56.7(18.51)
SISST Positive	36.93(7.40)	37(5.71)	42.6(11.01)
SISST Negative	53.46(9.11)	46.9(14.03)	38.6(8.04)
BDI	-	9.7(7.99)	5.3(5.48)

Table 6.2: Mean scores to the various questionnaires. These are compared with known norm values for phobic subjects. Their standard deviation is in parentheses.

We conducted the eye-tracking tests before and after the HMD sessions in order to analyze the progression in eye contact before and after treatment. Before starting with the first session, we exposed the subjects to a 5 minute habituation session. We asked them to write their name on the back projection screen with their eye after the calibration procedure. This was done in order to habituate them to the equipment and to relax before exposure by playing a game. During both sessions, we set the subjects in front of two different scenes, facing one man sitting in his office and facing an auditorium containing an audience of approximately 20 virtual actors. We did these two recordings in order to check whether the eye contact attitude was the same in different situations or not, i.e. when talking to one or several people.

6.2.4 Results

We first noted a general improvement from most subjects through the analysis of the various questionnaires even though some assessments on these scales were quite low, indicating a mild phobia. We have equally seen that the tendencies for each subject were repeated throughout all questionnaires. Visual contact avoidance equally decreased. Our results showed that the subjects presented less avoidance behaviors after treatment than before. We also noted that one person out of the 8 did not follow the improvement pattern, on the contrary. Subject C's progression was the opposite of all others. We hypothesized that this was due to a different cultural background inducing a mental block towards the effects of VRET. If we took this subject out of our study, the mean evolution would have been much higher.

6.2.4.1 Questionnaire Analysis

Table 6.2 contains the mean results we have obtained to the various questionnaires. All norm values for phobic people present in this table have been found in Bouvard and Cottraux [Bouvard and Cottraux, 2002]. The *Before* value is the average score over the three weeks preceding the beginning of the treatment for all subjects. Similarly, the *After* value is the average score after the end of the treatment for all subjects. The mean values to the Fear Questionnaire, the Liebowitz and the SISST negative thoughts have all decreased after treatment as compared to before treatment. Moreover, the positive thoughts score of the SISST has increased after treatment. The results we have obtained to these questionnaires are thus all very promising. For the BDI questionnaire, knowing that a score of $18.7(10.2)$ denotes a

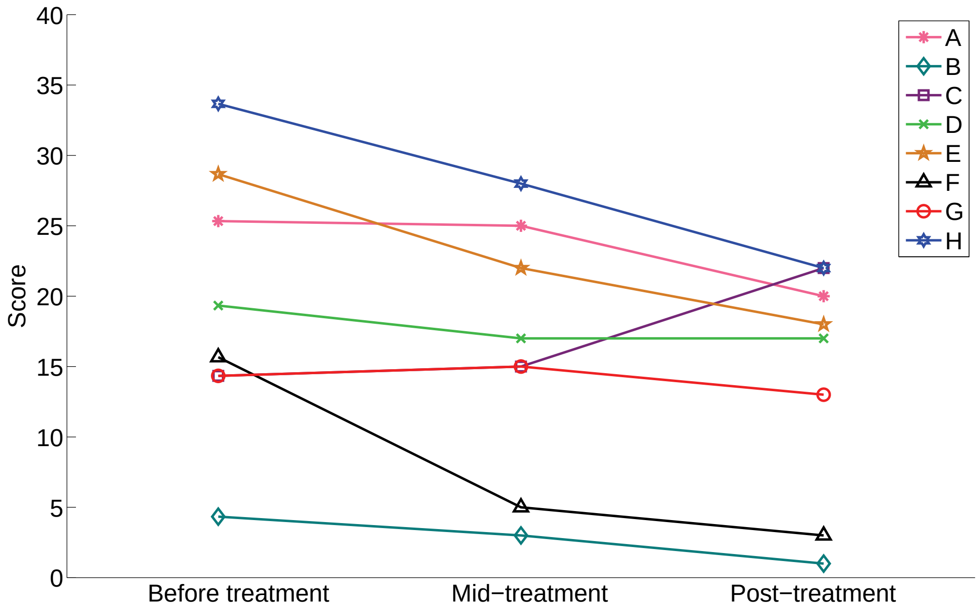

Figure 6.3: Results to the Fear Questionnaire.

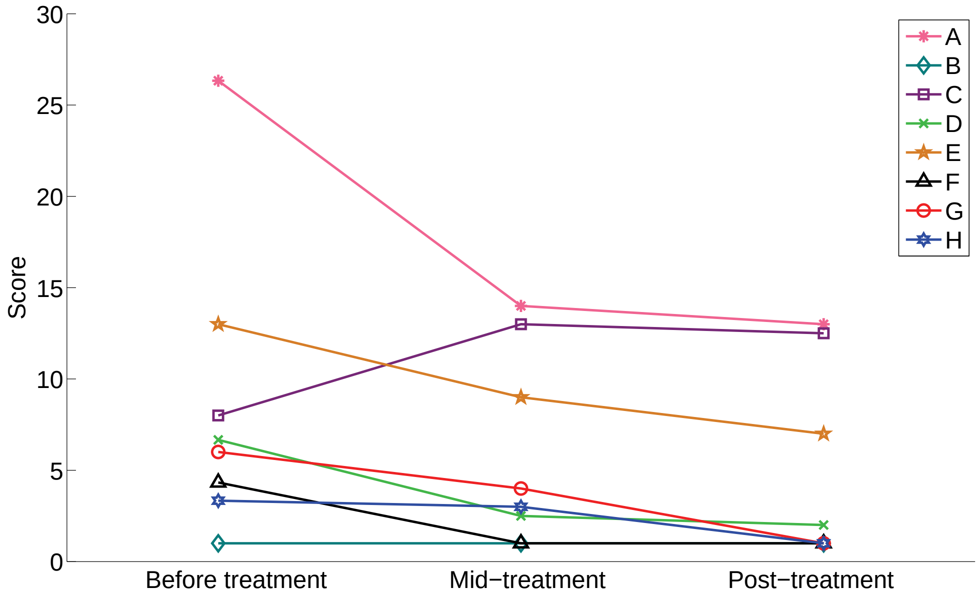

Figure 6.4: Results to the BDI.

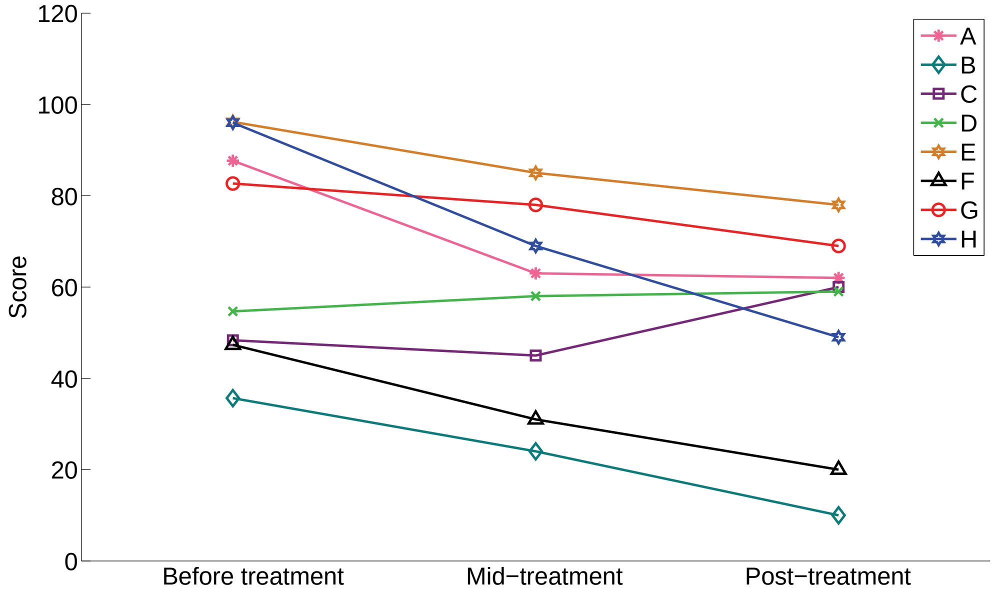

Figure 6.5: Results to the Liebowitz scale.

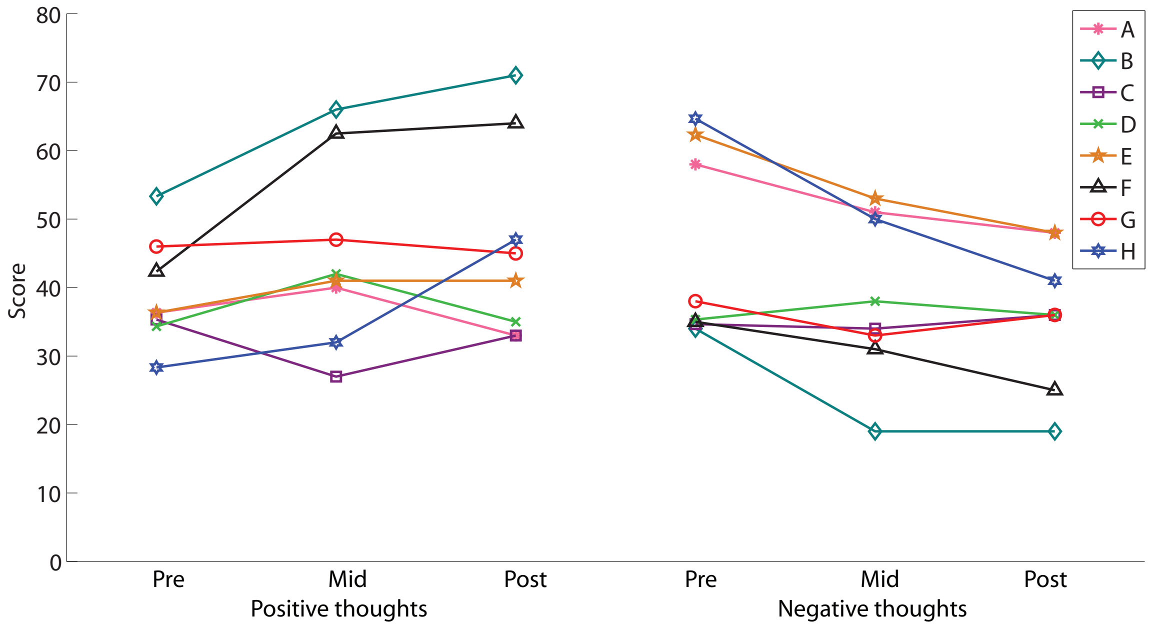

Figure 6.6: Results to the SISST.

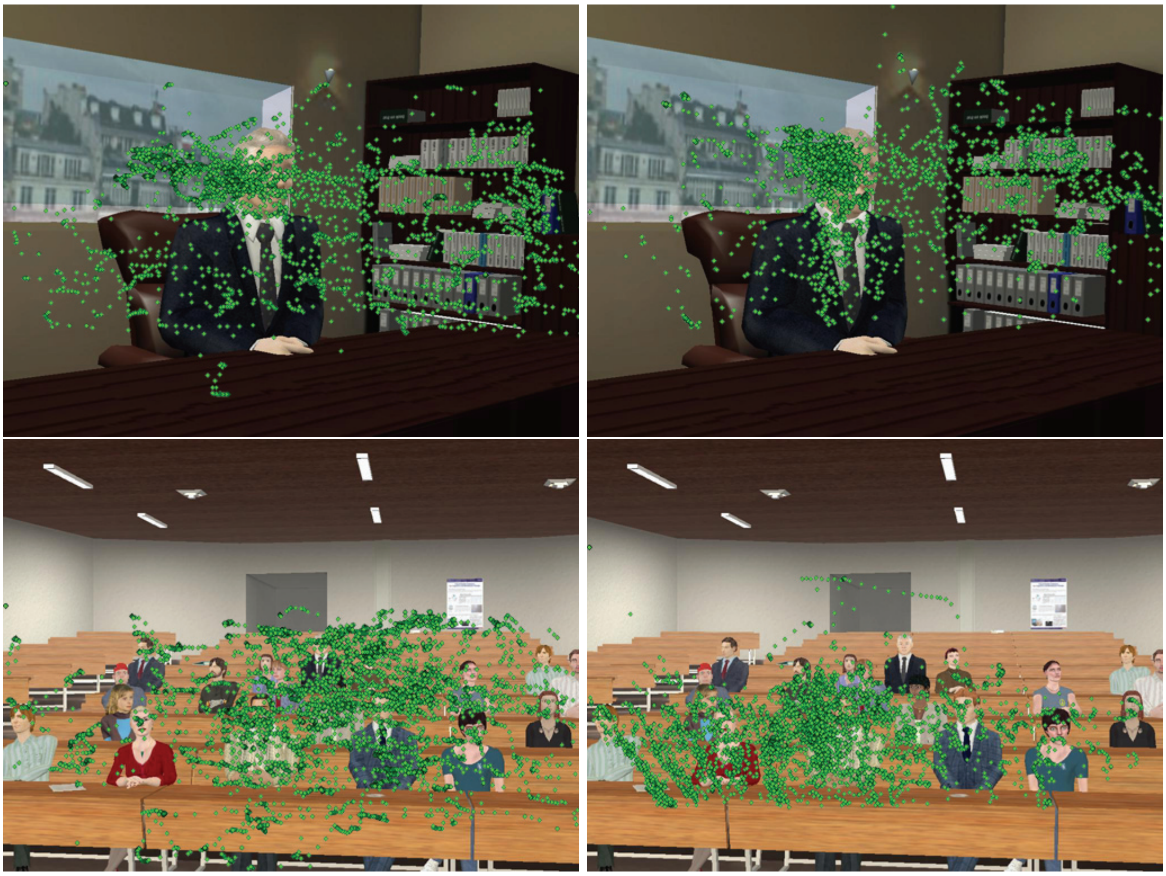

Figure 6.7: Eye contact behavior during exposure. *Top Left:* Before treatment example - office situation. *Top Right:* After treatment example - office situation. *Bottom Left:* Before treatment example - auditorium situation. *Bottom Right:* After treatment example - auditorium situation.

slight depression and that a score of $10.9(8.1)$ denotes no depression whatsoever [Bouvard and Cottraux, 2002], we can conclude that our subjects were not depressive be it at the beginning of the treatment or after its end. This is in line with one of our subject selection criteria, namely, absence of comorbidity.

Figures 6.3 to 6.6 show the results to these questionnaires for each subject. The first value is the averaged three weeks before the beginning of the treatment value. As previously mentioned, one of the subjects, C, demonstrated a score evolution contrary to that of all other subjects. The treatment seems to not have worked at all on this subject.

6.2.4.2 Eye-tracking

Concerning the evolution of eye contact behaviors, it is different for each subject. Indeed, their gaze behaviors were different to start with and thus evolved differently. However, we have noticed a clear tendency towards improvement in visual contact behaviors. From the recordings we have done, we can see that the virtual character's face in the office scenario was looked at much more after the end of the treatment than before treatment. There was a noticeable decrease in salient facial feature avoidance. An example of this is depicted in the top row of Figure 6.7. In the auditorium scene, we can also see that the talking virtual characters, the lady in red and the man with the white sweatshirt in the front row, were equally looked at much more after treatment than before treatment. This is depicted in the bottom row of Figure 6.7. Here as well, there was a noticeable decrease in avoidance of eye contact with the interlocutors. On the whole, the eye-tracking results tend to show that eye contact avoidance has diminished after treatment and that looking at the talking person's face has become less difficult for most subjects.

6.2.5 Discussion

Efficacy of VRET. First, we have noticed a general improvement for most subjects after having analyzed all questionnaires. We have equally noticed that the tendencies for each subject were the same over the questionnaires. However, we have seen, in our results, that one of the subjects apparently was not affected by the treatment at all. This leads us to think that certain people are more reactive to VRET than others. Moreover, our therapist collaborator believed that this particular subject rejected VRET due to a different cultural background. This, however, is only a hypothesis and cannot be corroborated by any data. A different study would need to be conducted in order to verify it.

Validation of VR as therapeutic tool. Due to the limited size of our sample, we cannot conclude to the validation of VR as a treatment for social phobia. Indeed, statistical significance requires a minimum of 30 subjects. Moreover, our study cannot prove the efficacy of VRET since we did not have a control group (following therapy without VR) or a waiting list. Nevertheless, our results support previous studies in the domain and show that it could be a promising therapeutic tool. More extensive experimentation on a larger cohort and with various groups (therapy without VR and waiting list) would enforce the efficacy of such treatments.

Positioning in relation to previous work. As mentioned in Chapter 2.1.1, many studies have been conducted on the evaluation of the potential of VRET to treat social phobia and concluded to the validity of VRET as therapeutic tool. The results we have obtained support those conclusions. Our main contribution though, is to bring in the use of eye-tracking in order to assess and evaluate eye contact behavior progress. Preliminary studies and literature have shown eye contact to be an important factor of social phobia. Our idea in this study was thus to confirm this within a known therapeutic program, and in this sense, we have obtained promising results.

Use of eye-tracking as diagnosis and assessment tool. We have observed different gaze behaviors during VR exposure and have obtained a qualitative appreciation of the patients' gaze avoidance behaviors. After having analyzed the eye-tracking recordings, we have equally noticed an improvement in eye contact behaviors once the therapy was over. This has lead us to think that eye-tracking could be a useful tool in the assessment and diagnosis of eye contact avoidance in people suffering from social phobia. The next experiment presented in this chapter discusses this in detail.

6.2.6 Conclusion

In this section, we have presented a study whose aim was to evaluate the efficacy of a therapeutic program using VR as a tool to treat social phobia. We have presented the VR-based protocol which we have used to conduct a study over 8 social phobic subjects. We have equally presented the results we have obtained through questionnaire analysis and eye-tracking recordings. By using subjective appreciation of social anxiety throughout the whole experiment as well as video recordings of the exposure sessions and the results provided by our eye-tracking system, we have noticed an improvement regarding avoidance behavior and a decrease of anxiety with time and exposure.

6.3 Eye-tracking as Assessment and Diagnosis Tool

As of today, no objective clinical diagnosis test exists when it comes to phobias. It is based on patients' accounts of undergone experiences and therefore remains very subjective. Moreover, the experiment previously described in this chapter raised a number of research directions. More specifically, while recording eye movements, we have noticed that several traits in eye contact behavior were recurrent. For example, we noticed rapid eye blinking rates in some subjects, and very rapid eye movements in others.

For this reason, we propose to use an eye-tracking system, in order to diagnose and assess, in an objective way, various known features often present in people suffering from social phobia, and more specifically, in people suffering from fear of public speaking. The tests we decided to conduct concern the avoidance of salient facial features, the presence of hyperscanning, and the rates and durations of eye blinks. The avoidance of salient facial features is a known parameter often present in people suffering from social phobia [Horley et al., 2003]. Hyperscanning consists of an overly rapid sweeping of the eyes. This factor equally consists of a feature found to be present in social phobics [Horley et al., 2004]. Eye blink rates are known to increase with anxiety [Kanfer, 1960] and can therefore be used as a revealing measure, relative to the anguish caused by the phobia. Finally, we believe eye blink durations to be another good indicator of the presence of visual contact avoidance and therefore investigate this feature as well. People suffering from social phobia do not all present those traits of visual avoidance but many of them do. Our hypothesis was that phobic subjects would not present all traits but that most of them would at least present one of them. We conducted our experiment over five social phobic patients and five non phobic subjects who served as control group. In a first phase, we exposed all ten subjects to two VR public speaking exercises. The first one was in front of a single virtual actor and the second in front of an audience. These scenes are depicted in the middle left and bottom left of Figure 3.1. For these exercises, we used the social phobia setup described at the beginning of this chapter in order to track the patients' and control group participants' eye movements. The phobic subjects then went through ten weekly sessions of therapy based on the cognitive behavioral model [Beck and Emery, 1985]. This was a group therapy following the Boston group protocol [Stern et al., 1999]. Finally, six out of the ten subjects went through a second eye-tracking session, three from each group. This post-treatment exercise was conducted under the same conditions as the first one.

6.3.1 Participants

The five social phobic subjects who participated in this study were patients picked by our psychiatrist collaborator. They completed the MINI [Sheehan et al., 1998] in order for us to verify the prevalence of social phobia and the absence of comorbidity. All subjects were aged 25 to 55, with a mean age of 37 for the phobic group and 38 for the control group. Subjects were not chosen from any specific educational or socio-economical category but were all of Caucasian type. The phobic subject group was composed of three females and two males and the control group of three males and two females.

6.3.2 Experimental Protocol

All subjects first started with a pre-treatment, individual, eye-tracking session. We first explained them how the system worked and that we were going to analyze the way they used their eyes in a public speaking situation. The phobic subjects were neither told which behaviors were typical of social phobia nor that their results would be compared to those of subjects from a control group. Similarly, the control group subjects were not aware that their results would be compared to those of phobic subjects. Once the session was over, we informed them all that they would be coming back for a second session, twelve weeks later. This corresponded to the end of the therapy for the phobic subjects. The five people from the social phobic group then underwent a 10 week group therapy following the Boston group protocol. Three out of the five subjects then came back for the second eye-tracking session. One of the two others started an individual therapy after two sessions with the group and the other stopped the therapy after seven sessions. We therefore decided to exclude them from the second eye-tracking session. Likewise, we only included the second set of results of the three first subjects from the control group into the study in order for their number to correspond to the phobic subjects'.

6.3.2.1 Pre-treatment Eye-tracking

We first equipped the subjects with the eye-tracker and seated them in front of the large back-projection screen. They then went through the calibration process and a phase of accommodation to the equipment. This accommodation phase was a kind of game, during which we asked them to write their name on the screen with their tracked eye. This was done also in order to make them feel more at ease and relaxed. We then asked them to do a public speaking exercise in front of two different VR scenes, during which their eye movements were recorded. The conditions and discussion topics were kept the same for each subject. The first of these VEs, depicted on the middle left of Figure 3.1, was a single person in a bar. The second, shown on the bottom left of Figure 3.1, reproduced an auditorium environment with approximately 20 characters. As for the previous experiment, in order to avoid breaks in presence and to enhance interaction, the discussions were not lead by the therapist. Instead, they were lead by the main character in each of the two scenes. We did this by triggering pre-recorded sentences and thus simulating the virtual character's speech. For each of these scenes, we asked the subjects to talk for a couple of minutes on a given theme. We tracked the subjects' eyes during these exercises in order to analyze their visual behaviors. In this first phase, we analyzed the collected data in order to see if there was a noticeable difference in behavior between the phobic subjects and the subjects from the control group.

6.3.2.2 Group Therapy

The phobic subjects then went through ten sessions of weekly therapy. These sessions were based on the Boston group protocol [Stern et al., 1999]. The subjects were asked to do presentations on given themes in front of the other members of the group. These sessions were filmed and then analyzed by each patient with the psychiatrist. Each presentation comprised a subject more personal and difficult to talk about than the previous. As an example, for the first session, they were asked to talk about their hobbies, and for the last, about a dramatic and difficult situation they had experienced. They were then asked to repeat the exercises in their everyday life and report on the difficulties they encountered and how they dealt with them. All five subjects were asked to evaluate their progress by filling out various scales specific to social anxiety disorders, before and after treatment. These were the Liebowitz social anxiety questionnaire [Liebowitz, 1987], the SISST [Yao et al., 1998], the BDI [Beck et al., 1961], and the Rathus Assertive Behavior Schedule (RABS) [Rathus, 1973]. This latter was used to measure assertiveness and self-esteem.

6.3.2.3 Post-treatment Eye-tracking

Once the phobic subjects finished their ten weeks therapy, we asked all participants to come back for a second eye-tracking session. Three subjects from the phobic group and three from the control group went through this second session. They were exposed to the same two scenes as for the pre-treatment phase and were asked to talk about two different given themes. These differed slightly from the ones in the pre-treatment session in order to avoid repeating a previously done exercise. The goals, in this phase of the study, were the following:

- analyze the differences in gaze behavior between phobic subjects and non phobic ones.
- analyze the improvements in the phobic subjects versus control group ones.
- determine whether there was a habituation to the equipment. This would have been the case if we had noted visible differences in the visual behaviors of the control group subjects as well as the social phobic subjects.

6.3.3 Results

First, and as expected, we have noticed an important difference in eye-contact behavior between phobic patients and non phobic subjects in the pre-treatment phase. Phobic patients demonstrated different types of visual avoidance: shutting the eyes for long lapses of time, hyperscanning, hypervigilance, and avoidance of salient facial features. Non phobic subjects, however, did not demonstrate those types of behaviors. The values used in the results are the raw data given by the eye-tracking system. However, what is interesting to us is to compare these values between phobic and non-phobic subjects as well as between pre-treatment and post-treatment phases for a same subject.

6.3.3.1 Salient Facial Feature Avoidance

In the case of the auditorium scene, non phobic subjects mostly looked at the central character (the one with the white sweatshirt) but also looked at the other characters when speaking, because they were addressing the whole audience. A phobic subject prone to visual avoidance demonstrates one of two behaviors. He/she either eludes the characters or demonstrates hypervigilance [Horley et al., 2003]. In the case of this first scene, there was no obvious avoidance of salient facial features from any of the subjects. However, some phobic subjects

Subject	HSD - auditorium	VSD - auditorium	HSD - bar	VSD - bar
CA	81.42 / 76.03	97.49 / 39.02	62.11 / 50.48	58.00 / 38.07
CB	90.38 / 98.89	54.35 / 80.00	45.53 / 54.82	71.81 / 48.59
CC	42.94 / 43.06	39.75 / 43.78	11.35 / 23.49	20.03 / 33.21
CD	69.93 / -	85.31 / -	51.53 / -	48.21 / -
CE	47.95 / -	38.64 / -	33.86 / -	83.92 / -

Table 6.3: Standard deviation for control subjects. Pre-treatment / Post-treatment.

Subject	HSD - auditorium	VSD - auditorium	HSD - bar	VSD - bar
PA	27.02 / 71.00	24.03 / 40.38	35.38 / 64.51	41.43 / 43.02
PB	25.21 / 39.86	41.65 / 40.67	21.11 / 17.46	44.57 / 35.24
PC	77.77 / 74.60	83.63 / 48.34	75.89 / 64.95	102.07 / 65.05
PD	82.89 / -	58.07 / -	72.38 / -	70.68 / -
PE	55.10 / -	19.43 / -	35.17 / -	39.05 / -

Table 6.4: Standard deviation for phobic subjects. Pre-treatment / Post-treatment.

clearly demonstrated hypervigilance. The subjects from the control group mostly looked at the central character in the first row, but also at the whole audience. On the other hand, only one of the five phobic subjects showed a visual behavior corresponding to the control group subjects'. Four of them focused on one or two characters only. This can be seen in the examples shown in Figure 6.8 where the phobic subject clearly concentrates on a single character whilst the control group one looks at the whole audience.

We have also measured the standard deviation (SD) for each subject. As can be seen in Table 6.3, the values obtained from the control subjects are all above 40 for the horizontal component (HSD) and near 40 and more for the vertical component (VSD). These values are not surprising since looking at the whole audience, i.e., the whole width and height of the screen, translates to high SD values. Table 6.4, containing the data for the phobic group, shows that PC and PD have values similar to those of the control group. PB and PE have one of the two components which is significantly smaller in the pre-therapeutic phase and PA clearly has values very much below those of the control group in the pre-therapeutic phase. However, both PA and PB have values corresponding to those of the control subjects in the post-therapeutic phase. It is important to note that for this scene, the horizontal deviation is much more important than the vertical one due to the image properties; the virtual characters being mostly on the bottom part of the projected image but spread out over its width.

In the second scene, the bar scene, since there is only one main character to which the subject is speaking, non-phobic subjects look at that person in the face, and more specifically, at that character's salient facial features (eyes, nose and mouth). Numerically, this translates in small SD values. On the other hand, a phobic subject prone to visual avoidance behaviors will look at anything but salient facial features. All subjects from the control group mostly looked at that character in the face while talking to him. Four of the subjects from the phobic group equally behaved this way. Subject PC, however, clearly demonstrated an avoidance of the character's salient facial features.

Figure 6.8: Eye contact behavior during exposure. *Left:* Phobic subject. *Right:* Control subject.

Figure 6.9: Eye contact behavior during exposure. *Left:* Phobic subject. *Right:* Control subject.

Figure 6.9 shows the eye-tracking results for PC and CB. Subjects PC and PD showed noticeably higher values than the other subjects in the pre-treatment phase. This means they did not stay focused on the character they were talking to but let their eyes wander. PD, however, did not show clear elusion of salient facial features as PC did. In the case of PC, these values have significantly decreased in the post-therapeutic phase.

6.3.3.2 Hyperscanning

Another type of avoidance behavior present in some social phobic subjects is hyperscanning. Some social phobic subjects seem to be focused on the virtual character(s) present in the scene. However, they move their eyes in such a rapid way that it actually consists of a form of avoidance. In order to measure eye scan velocity, we first calculated the Euclidean distance between each set of two consecutive gaze points. We then averaged this data in order to obtain the mean distance covered at each time step, which can be interpreted as a velocity. Once again, there are no units to these distances since we use the raw data from the eye-tracking device. However, what we are interested in is to compare the results of phobic subjects with those of control subjects in order to verify whether there is a difference in

Subject	auditorium	bar	auditorium	bar
CA	8.98	8.34	8.71	7.16
CB	6.09	6.45	6.61	8.31
CC	1.97	1.64	3.59	4.40
CD	8.14	8.46	-	-
CE	5.77	7.52	-	-

Table 6.5: Eye scan velocity for control subjects. *Left:* Pre-treatment. *Right:* Post-treatment.

Subject	auditorium	bar	auditorium	bar
PA	5.21	6.82	6.50	7.67
PB	8.60	11.16	7.41	7.40
PC	10.37	10.93	11.66	11.14
PD	10.14	11.20	-	-
PE	4.32	4.00	-	-

Table 6.6: Eye scan velocity for phobic subjects. *Left:* Pre-treatment. *Right:* Post-treatment.

behavior between the two. Thus, given that POH_i is the horizontal component of the POR at frame i and that POV_i is its vertical component, we compute the velocity *v* as:

$$v = \frac{\sum_{i=0}^{n} \sqrt{(POH_i - POH_{i+1})^2 + (POV_i + POV_{i+1})^2}}{n} \tag{6.1}$$

As can be seen in Tables 6.5 and 6.6, not all phobic subjects present hyperscanning behavior. PC and PD have scan rates above 10 for the auditorium scene whilst none from the control group attained such values in the pre-therapeutic phase. For the bar scene, PB, PC and PD all have scan values above 10. However, in the post-therapeutic phase, PB has values similar to those from the control group subjects in both scenes. On the other hand, PC demonstrates very high scan rates both before and after therapy. It should be noted that there is a high variability in the results of the control subjects. More specifically, CC demonstrated very low eye scan velocity. We can thus reasonably assume that some of the phobic subjects presented hyperscanning behaviors but can not affirm it.

6.3.3.3 Eye blinks

Normal blinking rates are very different depending on the type of action being undertaken. Doughty et al. conducted tests on a set of people in different contexts: a reading context, a primary gaze context and a conversational context [Doughty, 2001]. Their results show that eye blink rates are much higher in conversational contexts than in other ones. Since our study concentrates on public speaking exercises, we have retained the average values for normal people in a conversational context. This rate is of 23 eye blinks per minute in average but can vary anywhere between 10.5 and 32.5 eye blinks per minute. The eye blink lasts between 100 and 400 milliseconds, which averages to 250 milliseconds [Schiffman, 2001]. Moreover, eye blinks rise with anxiety and tension [Kanfer, 1960].

Subject	Auditorium - Rate	Auditorium - Duration	Bar - Rate	Bar - Duration
CA	50.90 / 68.89	224 / 185	62.24 / 19.12	139 / 202
CB	24.96 / 21.11	250 / 157	33.61 / 17.13	166 / 199
CC	6.26 / 16.71	161 / 110	3.81 / 17.99	80 / 183
CD	45.20 / -	240 / -	41.69 / -	143 / -
CE	32.14 / -	153 / -	34.82 / -	274 / -

Table 6.7: Eye blink rates per minute and durations in milliseconds for control subjects. Pre-treatment / Post-treatment.

Subject	Auditorium - Rate	Auditorium - Duration	Bar Rate	Bar Duration
PA	50.47 / 55.82	168 / 138	56.17 / 55.74	157 / 117
PB	82.62 / 71.27	89 / 96	92.23 / 72.65	73 / 93
PC	38.91 / 38.58	517 / 271	51.12 / 62.69	432 / 345
PD	48.38 / -	195 / -	67.13 / -	259 / -
PE	23.89 / -	168 / -	22.71 / -	147 / -

Table 6.8: Eye blink rates per minute and durations in milliseconds for phobic subjects. Pre-treatment / Post-treatment.

As shown in Tables 6.7 and 6.8, the average number of eye blinks per minute for the phobic subjects and control group subjects are high when compared to the averages as described by Schiffman [Schiffman, 2001]. We believe this is due to the equipment used. Indeed, we have noticed that after 10-15 minutes of eye-tracking use, people started feeling fatigue in their eye. We thus believe that the eye blink rate comes to be increased due to this fatigue. In the pre-therapeutic phase, PA and PB both had a rate above 50 eye blinks per minute in the first exercise and PA, PB, PC and PE all had rates above 50 in the second exercise. In the case of the control group, only CA had blink rates above 50 blinks per minute in both exercises. In the post-therapeutic phase, we have not noted any noticeable difference in this behavior; PA and PB both have values above 50 in the first scene and PA, PB and PC all have values above 50 in the second scene. Concerning eye blink durations in the pre-therapeutic phase, PC had blink durations of 517 milliseconds and 432 milliseconds whilst the highest values in the control group were of 250 milliseconds and 274 milliseconds for the first and second scenes respectively. Figure 6.10 shows a sample of 10 seconds for subjects CB and PC, both for the bar scene. In the post-therapeutic phase, however, PC demonstrated an important decrease in eye blink durations. Moreover, these values are within the range of $100 - 400$ milliseconds. Once again, due to the diversity of our tested population and high variability in results, we can reasonably assume that some phobic subjects presented unusually high blink rates and durations but we can not affirm it.

6.3.3.4 Subjective ratings

Out of the three phobic subjects who went through the complete therapy, PA and PB both had visible improvements in their responses to the various scales they filled out before and after treatment, i.e., the Liebowitz social anxiety questionnaire, the SISST, the BDI, and the

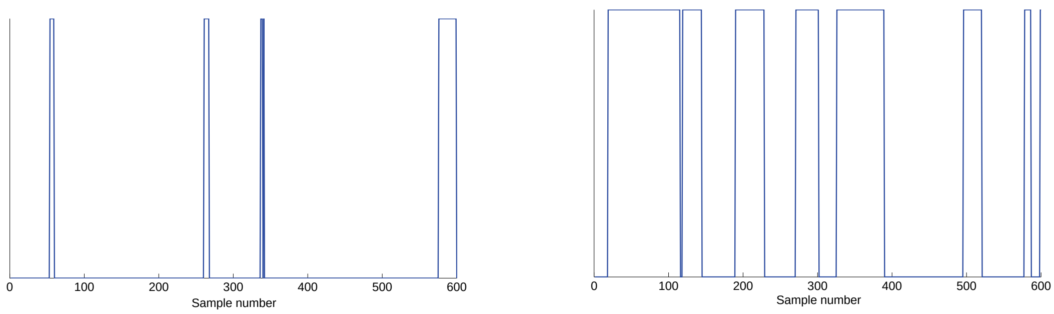

Figure 6.10: Blink rates and durations example. *Left:* Control subject. *Right:* Phobic subject.

RABS. On the other hand, subject PC did not demonstrate the same improvement. These results were confirmed by our psychiatrist collaborator in her personal assessement of the subjects' evolution throughout the therapy.

6.3.4 Discussion

Tables 6.9 and 6.10 show a recapitulation of the present behaviors in each of the subjects before and after treatment. A tick symbolizes the presence of the behavior whilst a cross symbolizes its absence. The wave represents a slight presence of the behavior and the dash, an absence of results. The presence or absence of salient facial feature avoidance and hypervigilance have been subjectively evaluated by visually analyzing the eye-tracking results. For hyperscanning, the threshold was at 10. For eye blink rates, it was at 50 and for eye blink durations, the limit was set at $400ms$.

Facial feature avoidance: Visually and numerically, we can see that some social phobic subjects present facial feature avoidance whilst others do not, should it be in the form of elusion or hypervigilance. However, no subjects from the control group showed this type of behavior. Moreover, when comparing the results of the pre-therapeutic and post-therapeutic phases, the phobic subjects prone to this type of behavior and having gone through both eye-tracking sessions showed improvement. On the other hand, none of the control group subjects demonstrated any noticeable difference for this measure. This analysis therefore seems to be a promising tool for visual contact avoidance diagnosis and assessment.

Eye scan velocity: When comparing the results of eye scan velocity for all subjects before therapy, we can see that the subjects presenting the highest rates are from the social phobic group. When comparing these results to those of the post-therapeutic phase, we can notice that subject PC did not show any improvement for either scene. On the other hand, subject PB demonstrated a decrease in eye scan velocity.

Eye blinks: Regarding the eye blinks, as for previous measures, some phobic subjects show a high eye blink rate and/or long lasting blinks whilst others do not. We have seen that, on the whole, both of these measures were higher in the case of the phobic subjects than in the case of the control group subjects in the pre-therapeutic phase. Moreover, one of the subjects presented very long blink durations which was not the case in any of the control group subjects. On the whole, subjects presenting high blink rates in the pre-therapeutic phase

Subject	Avoidance	Hypervigilance	Hyperscanning	Blink rate	Blink duration
CA	××	××	××	√√	××
CB	××	××	××	××	××
CC	××	××	××	××	××
CD	×−	×−	×−	×−	×−
CE	×−	×−	×−	×−	×−
PA	××	√×	××	√√	××
PB	××	≈≈	××	√√	××
PC	××	××	√√	××	√×
PD	×−	×−	√−	×−	×−
PE	×−	≈ −	×−	×−	×−

Table 6.9: Recapitulation of present symptoms before and after treatment - Auditorium scene. A tick symbolizes the presence of the behavior, a cross symbolizes its absence. The wave represents a slight presence of the behavior and the dash, an absence of results.

Subject	Avoidance	Hypervigilance	Hyperscanning	Blink rate	Blink duration
CA	××	××	××	√×	××
CB	××	××	××	××	××
CC	××	××	××	××	××
CD	×−	×−	×−	×−	×−
CE	×−	×−	×−	×−	×−
PA	××	××	××	√√	××
PB	××	××	√×	√√	××
PC	√×	××	√√	√√	√×
PD	√−	×−	√−	√−	×−
PE	×−	×−	×−	×−	×−

Table 6.10: Recapitulation of present symptoms before and after treatment - Bar scene. A tick symbolizes the presence of the behavior, a cross symbolizes its absence. The wave represents a slight presence of the behavior and the dash, an absence of results.

equally presented high rates in the post-therapeutic phase. On the other hand, subject PC, who had very high blink durations before therapy, showed a very noticeable improvement after therapy. This analysis, as the two others, therefore equally seems to be a promising tool. When analyzing the eye blink results, it is important to note the correlation between eye blink rate and eye blink duration. A person presenting very long eye blinks, such as shown in Figure 6.10, will necessarily have a lower blink rate.

It is clearly visible, when analyzing the overall results presented in Tables 6.9 and 6.10, that the subjects from the phobic group have a tendency to present the symptoms we have described. The control group subjects, however, do not; only one subject presented one of the symptoms before and after therapy. Moreover, we can see that each of the subjects from the phobic group has at least one of the symptoms. We can therefore conclude that the whole series of tests consists of a good indicator of the presence of visual avoidance behaviors.

6.3.5 Conclusion

We have described a study on the use of eye-tracking as a diagnosis and assessment tool for visual avoidance behaviors in social phobic subjects. The results we have obtained are very promising. Even though our study was conducted on a small cohort, our results clearly show the presence of avoidance behaviors in the phobic subjects which are not present in the control group ones. Phobic subjects do not all have visual avoidance behaviors. However, on the whole, our tests show that if there is avoidance behavior, it concerns phobic subjects and not control group ones. First, we have observed different visual behaviors between phobic subjects and non-phobic subjects in the pre-therapeutic phase. Second, this difference between the two groups has decreased in the post-therapeutic phase. Finally, we have not recorded important differences in behavior in the control group subjects when comparing their pre-therapeutic and post-therapeutic results. We can therefore reasonably assume that there was no habituation to the equipment between these two phases. We have seen that not all subjects have visual avoidance behaviors and that the subjects who do, manifest it in different ways. The use of eye-tracking as a diagnosis and assessment tool should therefore consist in a series of tests which should be correlated since some of them are not independent from each other. Even though the therapy undergone by the phobic subjects was very short (two and a half months on a weekly basis), they already showed improvement in their visual behaviors. With a longer therapy, we believe these results would be amplified. We propose the use of eye-tracking as a tool to diagnose and assess the presence of visual avoidance behaviors in social phobics, and in this sense, it gives promising results. Moreover, we believe it could be used as a tool to indirectly diagnose and assess social phobia itself.

6.4 Eye-tracking for Interaction

In this experiment, our aim was to evaluate and validate our application allowing interaction between a user and a virtual character using gaze. As in the previous experiments, we used the combination of eye- and head-tracking described at the beginning of this chapter. The subjects participating in the experiment were seated in front of our back-projection screen on which we projected the bar environment depicted on the middle left of Figure 3.1. As explained in Chapter 3.3, our application then receives the eye-tracking and the head-tracking data and determines whether the user is looking at the virtual character or not. When the user is looking at the character, hence, when the tracking coordinates are within the bounds of the virtual character projected on screen, this character demonstrates positive attitudes, shows interest in what the user is saying and looks back at him/her. On the other hand, when the user is not looking at the character, hence, when the tracking coordinates are outside the virtual human bounds, this character's attitude changes and becomes bored and distracted. The character looks away, at the ceiling or in other directions, sighs, or looks at his/her fingernails or watch.

6.4.1 Experimental Protocol - Healthy Subjects

Our hypotheses before conducting this experiment were the following: first, we believed that the subjects would evaluate a character changing behavior as more realistic than a character being always attentive or always distracted. Secondly, we believed that the subjects would consider a character changing attitude with respect to the user's eye contact behavior as more realistic than a character that randomly changes behavior. In order to evaluate our application, we have conducted an experiment in which 12 healthy people were asked to talk to a virtual character in a bar environment for a couple of minutes. These subjects were not from any specific socio-economical background. However, they were all aged $25-35$. The exposure was four-fold; each of the subjects was exposed during 2 to 3 minutes to four different characters, depicted in Figure 6.11: one which was always attentive and demonstrated a positive attitude, one which always looked away from the user and seemed bored and distracted, one which randomly changed attitude between attentive and distracted throughout the session, and one that changed attitude depending on the subject's eye-contact behavior. More specifically, in this last case, the virtual character looked at the user and seemed attentive and interested when it was being looked at. On the other hand, it lost interest and even seemed bored when it was avoided by eye contact. The four versions of this same scene were set in front of each of the 12 subjects in pseudo-random order. However, the order in which the characters were presented stayed the same.

We included the always attentive and always distracted versions in the study in order to verify that the subjects could identify the differences in behavior of our virtual characters. We then included the random and eye-tracked versions of the scene in order to verify our hypotheses: that a character that changed behavior was more realistic than a character that did not and that it was even more realistic when these changes were not random but depended on human actions. Finally, the subjects were asked to evaluate each of the four characters by answering a set of questions using a five-point Likert scale [Likert, 1932]:

Figure 6.11: *Left:* Attentive version of each character. *Right:* Bored version of each character.

- Did the character on screen seem interested or indifferent? (1: very interested, 2: fairly interested, 3: neither interested nor indifferent, 4: fairly indifferent, 5: very indifferent)

- Did the character on screen seem engaged or distracted during the conversation? (1: very engaged, 2: fairly engaged, 3: neither engaged nor distracted, 4: fairly distracted, 5: very distracted)

- Did the character seem friendly or not? (1: very friendly, 2: fairly friendly, 3: neither friendly nor unfriendly, 4: fairly unfriendly, 5: very unfriendly)

- Did the character seem to behave in a normal way with regards to what you were telling him/her? (1: very normal, 2: fairly normal, 3: neither normal nor abnormal, 4: fairly abnormal, 5: very abnormal)

The first two questions were asked in order to verify the distinction between attentive and bored behaviors. The third question does not indicate much, it was only asked in order to divert attention from the questions specific to our experiment. The fourth question was asked in order to verify our first hypothesis; that a character which changes behavior is more realistic than one who does not. It was equally asked in order to verify our second hypothesis; that a character changing behavior randomly is less realistic than one who changes with regards to what the subject is looking at. A fifth question was asked in order to obtain feedback from the subjects on any point which would not have been addressed by their answers to the previous questions:

- On the whole, how would you describe the differences in the characters' behaviors, if any, in the four scenes?

6.4.2 Experimental Protocol - Case Study

In addition to the validation conducted on these 12 healthy subjects, we have tested our application on a young 14 year-old girl suffering from Asperger syndrome. As explained in the Glossary (see Appendix D), the Asperger syndrome is an Autism Spectrum Disorder (ASD). One of the diagnostic criteria of the Asperger syndrome is a marked impairment in social interaction [American Psychiatric Association, 1994].

As a first point, it is important to note that the experiment did not aim at any therapeutic benefit for the subject. It was aimed to test our architecture with someone having problems with social situations. This young girl accepted to test our setup and give us her feedback. She was accompanied by her therapist and her mother for the experiment. Her mother, however, did not stay in the same room as the experiment went on. We did not tell her what we were looking for. She was thus naive towards our objectives. As a first step, we seated her in front of the large back-projection screen on which we displayed the scene depicted in Figure 6.11. We asked her to wear the coupled eye- and head-tracker and recorded her eye movements throughout the experiment. We then asked her to do a public speaking exercise, and talk to the virtual character for a couple of minutes. The complete session lasted for approximately 10 minutes. We first made sure that the character would be attentive even if not

	Interest	Engagement	Friendliness	Normality
Bored	5.00(0.00)	4.92(0.29)	3.92(0.90)	3.83(0.94)
Attentive	1.67(0.49)	1.92(0.79)	2.33(0.65)	2.50(1.00)
Random	3.50(0.80)	3.58(1.08)	2.92(1.00)	3.33(0.98)
Tracked all	3.50(1.51)	3.67(1.07)	3.42(0.79)	3.17(1.03)
Tracked looked	2.71(1.49)	3.29(1.11)	3.29(0.95)	3.00(1.00)
Tracked looked (absolute)	2.20(1.30)	2.80(0.84)	3.00(0.71)	2.80(1.09)

Table 6.11: Mean values to the four questions. The values in parentheses are the SDs to the mean scores obtained by the 12 subjects. “Tracked all” are the mean values for the 12 subjects. “Tracked looked” are the mean values without considering subjects who did not look during the speaking exercise (7 out of the 12). “Tracked looked (absolute)” are the mean values considering only the subjects who looked at the character throughout the whole talking exercise (5 out of 12).

looked at in order to avoid the subject being even more stressed than what she already was. This part of the exposure session lasted for $5-6$ minutes. The therapist then manually triggered the change in character behavior in order to make it lose its attention and seem bored. It stayed in this mode for 30 seconds to 1 minute. Then, as it triggered eye contact from the subject, the character became attentive once again. Finally, the character remained attentive until the end of the exposure session, which lasted approximately 1 minute more. Once the exposure session was over, we asked the subject to give us her feedback, her impressions on the virtual character and how the session had gone.

6.4.3 Results

In this section, we first present the results we have obtained from the evaluation of the 12 healthy subjects who tested our application. We then present the results to our case study, in which a young girl suffering from Asperger syndrome tested our application.

6.4.3.1 Healthy Subjects

The values we discuss in this section are the overall mean ratings of the 12 subjects for each of the 4 scenes (bored, attentive, random, and tracked). The mean results to the first question are shown in the first row of Table 6.11 and are depicted on the top left of Figure 6.12. They express the interest or indifference of the characters on screen. The mean for the bored version of the scene was of 5 which is the highest possible value and corresponds to a very indifferent evaluation. The attentive version obtained a mean value of 1.67 which corresponds to an evaluation between fairly and very interested. For the random version and the tracked version, both the mean scores were of 3.50. However, when we take out the people who were not looking at the character on screen when talking to him/her (7 subjects out of 12 remaining), this tracked score falls to 2.71. This evaluates to neither interested nor indifferent. Finally, when we only consider the subjects who were looking at the character throughout the whole talking exercise (5 subjects out of 12), this value further falls to 2.2, and evaluates to fairly interested.

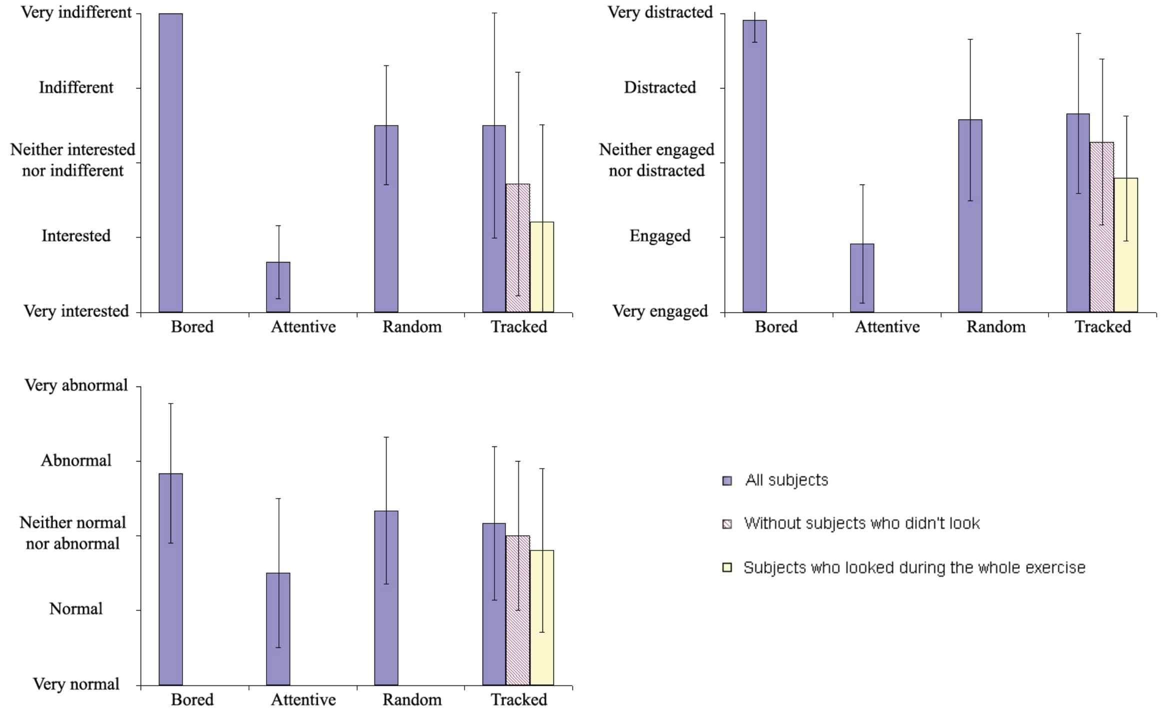

Figure 6.12: Graphical representations of the results for interest, engagement, and normality. The brackets indicate the SDs for each mean.

For the second question, shown in the second row of Table 6.11, and depicted on the top right of Figure 6.12, the results express the engagement or distraction of the characters on screen. The mean value for the bored version was of 4.92, which corresponds to very distracted. The mean value for the attentive version was of 1.92 which corresponds to fairly engaged. The mean value for the random version of the scene was of 3.58, which corresponds to a slightly distracted behavior. Finally, for the tracked version of the scene, this value was of 3.67 when considering all subjects, 3.29 when taking out subjects who were not looking at the character during the talking exercise and 2.80 when only considering the subjects who were looking at the character on screen during the whole talking exercise. This evaluation corresponded to fairly distracted when considering all subjects but to slightly engaged when considering only the subjects who looked at the character throughout the talking exercise.

The results to the third question, concerning the friendliness of the characters on screen, are shown in the third row of Table 6.11. However, this question was asked only in order to distract the subjects from the other questions relative to the experiment. We therefore do not discuss the results to this question any further.

The last row of Table 6.11 and the bottom left chart of Figure 6.12 show the results to the fourth question, determining whether the character on screen was behaving in a normal or abnormal way with regards to what the subject was telling him/her. Here, the score for the bored version was of 3.83, which corresponds to fairly abnormal. The score for the attentive version was of 2.50, which corresponds to quite normal. The random version scored 3.33, which corresponds to slightly abnormal. Finally, the tracked version scored 3.17 when considering all subjects, 3.00 when taking out subjects who did not look at the character, and

2.80 when considering only the subjects who looked at the character throughout the whole exercise. This corresponds to a neither normal nor abnormal evaluation when considering all subjects. However, it corresponds to a fairly normal evaluation when considering only the subjects who looked at the character throughout the talking exercise.

From our results, we can see that the subjects have clearly identified the difference between a bored and an attentive character. Not only are the scores to the bored and attentive characters clearly categorized, but the SDs are also small, which indicates small variations in scoring between the 12 subjects. These values are shown in parentheses in Table 6.11. Moreover, from the responses we obtained for the 5th question, many subjects reacted very negatively to the bored character. We received feedbacks such as “he could not care less about what I was saying”, “I almost made her cry”, or “he was more than distracted, almost condescending”. We also received feedbacks such as “she was making me think of a lover who was drinking my words”, “he was very interested and receptive”, or “he was faking he was interested” for the always attentive character. The results to the random version of the scene were slightly above the mean value of 3 with SDs of 0.80 and 1.08 for the first and second questions respectively. This seems quite logical since the characters for this scene would alternatively be attentive or bored, and that we usually are more receptive to negative attitudes than to positive ones. For some of the subjects, there were more positive attitudes than negative ones and vice versa, which explains the higher SD values. For the tracked version, the results we obtained were very similar to those of the random version. However, when considering only the people who looked at the character throughout the whole talking exercise, these values came closer to those of the attentive version, which equally seems logical. Indeed, we hypothesized that such a closed loop induced by the subject’s gaze behavior characterized an interaction through which the subject is empowered as he/she can perceive he/she has an influence on the outcome of the interaction (note that the subjects were not told about the potential influence of their gaze behavior). However for such a feedback to be built, it is first necessary for the subject to make eye contact with the virtual character. Hence, a lack of motivation from the subject prevents the triggering of such a positive feedback loop. The SDs equally sustain this hypothesis since their values are quite high when considering all subjects and decrease when considering only the subjects who looked at the character during the exercises.

The lack of involvement from some subjects may explain the answers to the fourth question concerning the normality of the characters. The bored version scored slightly under 4, corresponding to fairly abnormal, which is what we expected to obtain. However, for the attentive version, the score corresponds to a slightly normal evaluation. It is not surprising to see a better scoring than for the bored version, but we did not expect it to be the best out of the four. It seems as though the subjects were expecting the virtual character to be always attentive to them. The characters from the random version appeared to be more normal to the subjects than the characters from the bored version, which is what we expected. However, they seemed less normal than characters from the attentive version, possibly for the reason highlighted above (i.e. a bias from the subject towards an expectation of an always attentive virtual character). For the tracked version, the same comments apply. However, it is important to note that the characters seemed more normal to the subjects in the tracked version of the scene than in the random one which is in line with our second hypothesis.

Figure 6.13: 2D gaze map for the case study. *Left:* Attentive character. *Right:* Distracted character.

6.4.3.2 Case Study

Our results show that the subject strongly reacted to the change of behavior in the virtual character. She glanced at the character once in a while at the beginning of the exposure. She then looked down most of the time she was talking. This can be seen on the left of Figure 6.13. It is clear from the eye-tracked points that there is an avoidance of salient facial features. Figure 6.14 provides a clearer view of the tracked coordinates in time. The green points indicate that the character is being looked at, whereas the red ones indicate avoidance. This figure depicts a 20 second sample starting a couple of seconds after the exposure began, as she was still looking at the therapist and not at the screen when the recordings started. We can see that she looked at the character in the face at the very beginning but then looked away most of the time.

When the character behavior change was triggered, and became bored and distracted, we noticed a clear change in eye contact behavior. The subject started looking straight at the character, as if trying to regain her attention. We can see, on the right of Figure 6.13, that some of the points are right on the character's face whilst this was not the case when the character was attentive. Figure 6.15 shows the first 20 seconds of recordings after the change in character behavior was triggered. We can see that at the beginning of character behavior change, the subject is not looking at her. However, after a couple of seconds, which corresponds to the time it takes for the behavior change to be visible, we can see that she is looking at the character right in the face. There are still many points out of the character's face but many more are inside as compared to when the character was attentive. When the character became attentive once again, however, there was no noticeable change in eye contact behavior. Once the session was over, we asked the subject for some feedback on how she perceived the public speaking exercise and the virtual character. She told us that she had felt very anxious. She also told us that after a while, the character became bored and started looking elsewhere, stopped listening to her. She clearly noticed the change in character behavior. She explained that she expected the character to behave in this way from the very beginning so it did not surprise her. She explained that this was the way she expected any real person to react. She also said that she was trying to look at the virtual character at that point on in order to regain her attention. Her subjective feedback therefore corroborated the results we obtained from the eye-tracking measures.

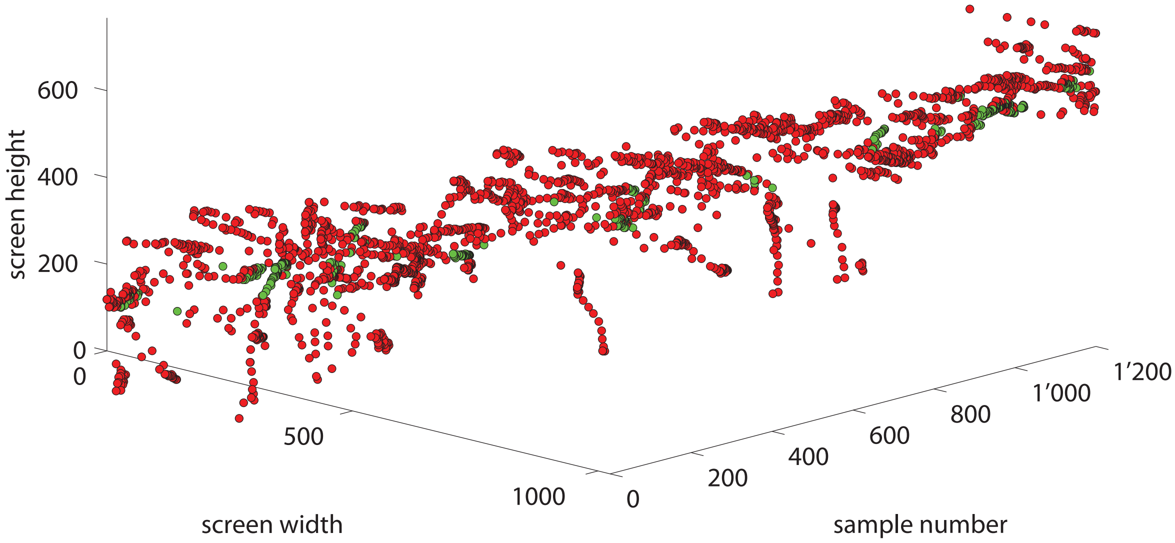

Figure 6.14: 3D gaze points layout sample when the character was attentive - case study.

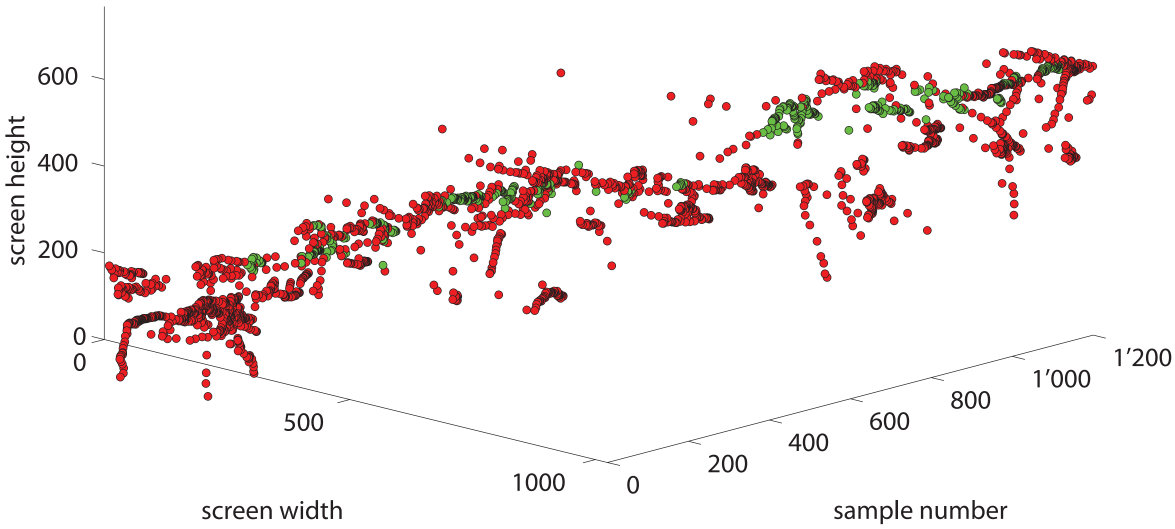

Figure 6.15: 3D gaze points layout sample when the character was distracted - case study.

6.4.4 Discussion

Confirmation of the first hypothesis. Our results on the healthy population have only been able to partially confirm our first hypothesis; that a character would seem more natural when changing behavior than when always maintaining the same behavior. There was a tendency from the subjects to find the tracked version of the scene more realistic than either the bored version or the random version. However, the always attentive version seemed more realistic to the subjects than any other. We believe this is due to a potential bias of the subjects towards the expectation of an “always attentive” virtual character.

Confirmation of the second hypothesis. From the results we obtained with the healthy subjects, we have been able to confirm our second hypothesis; that a character changing behavior with respect to human eye contact behavior seems more normal than a character that randomly changes behavior.

Bias on character behavior. As mentioned, we believe that there was a bias towards an always attentive virtual character in the case of healthy subjects. However, interestingly enough, this bias was reversed in the case of the young girl suffering from Asperger syndrome. Indeed, she thought that it was normal for the character to seem uninterested and distracted. It has to be noted that the healthy subjects were exposed for a shorter amount of time than the case study subject. This bias towards always attentive behaviors may thus have decreased in the case of a longer exposure.

Testing on a phobic population. In this experiment, we have tested our application on healthy subjects mainly. The case study we have conducted on a young patient suffering from Asperger syndrome, however, gave very promising results. Our application should therefore be tested on a phobic population in order to evaluate its efficacy. Moreover, it would allow us to verify whether the same bias towards uninterested and distracted virtual characters would be present in such a phobic population as was seen in our case study.

6.4.5 Conclusion

In this section, we have described the experiment we have conducted over 12 healthy subjects and the results we have obtained from their subjective feedback regarding the realism of the character behaviors in 4 different situations: always bored, always attentive, randomly animated, and animated with respect to the subject eye contact behavior. We have also tested our application on a young girl suffering from Asperger syndrome. In this case study, we used a hybrid approach where the character changed behavior either when triggered by hand or with respect to the subject’s eye contact behavior in order to be able to force the character’s attention at the beginning of the exposure session.

The results we have obtained on the healthy population have partially confirmed our first hypothesis and fully the second. The subjects clearly identified the different behaviors but seemed to be biased towards an always attentive character. Concerning the case study, this bias was reversed. She expected the virtual character to be uninterested. During exposure, the subject reacted very strongly to character behavior change. She clearly identified the change and was affected by it. Moreover, the eye-tracking measures corroborated her impressions during the exposure session. This demonstrates the potential of closing the loop to reveal and trigger more elaborate interactions between the users and the virtual characters.

We strongly believe that this new form of interaction based on gaze behavior will greatly amplify immersion and thus, the potential of VRET for the treatment of social phobia. We believe that such a tool, that can induce such a positive feedback loop, can be of great help to therapists in the treatment of social phobia by CBT. More specifically, we think that it could be used to make patients understand that they can change people’s behaviors with their own.

CHAPTER 7

Experimental Validation - Agoraphobia

In this chapter, we present the experiment we have conducted over 30 healthy subjects in order to validate our application concerning the interaction with crowds of characters in an immersive environment. We first present the system setup in the CAVE environment before explaining the experimental protocol. We then go over the results we have obtained and discuss how our application could further be improved. Finally, we discuss the validity of our application in the domain of VRET of agoraphobia.

7.1 Experimental Setup

The setup we used for the experiment relative to agoraphobia is depicted in Figures 5.1 and 7.1. It consists of several components. The same eye-tracking device is used as for the social phobia setup. However, the magnetic sensor is replaced by a Phasespace optical motion capture system [Phasespace, 2009]. Three LEDs are set on the eye-tracker headband or on a baseball cap. More specifically, for participants wearing glasses or contact lenses, we used the head-tracking only. In this case, the markers were set on the baseball cap. For the participants who did not wear glasses or contact lenses, we used the combination of eye- and head-tracking. In this case, the markers were set on the eye-tracker. These three markers allow the definition of the position and orientation of the head in space. We used these two different setups in order to verify the possible added value of using an eye-tracking device. We actually wanted to check whether or not it was interesting to use the eye-tracking to the detriment of stereoscopic vision. One of our aims with this study was thus to compare the

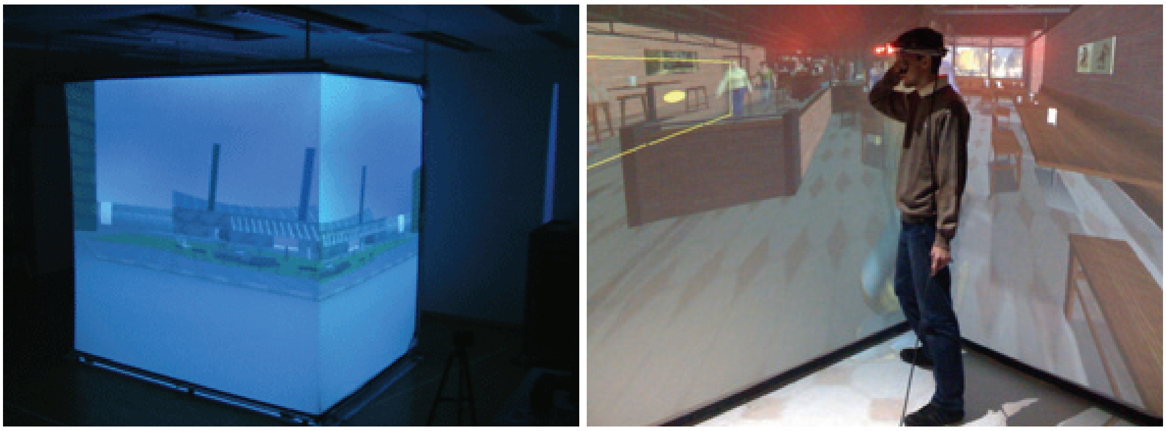

Figure 7.1: CAVE Setup. *Left:* CAVE viewed from outside. *Right:* A user testing the setup for agoraphobia.

appreciation of users testing our system with head-tracking only and with the combination of eye- and head-tracking. We also wanted to measure eye scan velocity and pupil dilation in order to see if there would be any differences in eye contact behavior between the different modes. Since stereographic display proved to be impossible whilst wearing the eye-tracking device, we deactivated stereography for all participants.

In addition to the eye- and head-tracking, we use a glove in order to capture user gestures. It is depicted in Figure 7.2. This glove has 8 LEDs on it which allow the retrieval of hand position and orientation as well as an approximate position of the fingers. 8 optical cameras are positioned around the upper edges of the CAVE in order to receive the signals from the LEDs. The advantage of this system is that it presents much less drift than the magnetic type; it does not suffer from the induced rotation around the sagittal axis.

Finally, we use a CAVE environment. As depicted in Figure 7.1, the CAVE consists of 4 screens: three walls and a floor. Its dimensions are 2.2 meters in length, 2.5 meters in width and 1.8 meters in height. We use 4 projectors, one behind each wall, and one on the ceiling. The scene is thus back-projected on the three walls and forward projected on the floor for a very immersive environment. A more detailed description of the CAVE hardware setup can be found in Peternier et al. [Peternier et al., 2007]. Additionally, a detailed description of the software developed to display the scene seamlessly and with the correct perspectives with respect to the user's position can be found in van der Pol [van der Pol, 2009].

7.2 Participants

To validate our architecture, we asked 30 healthy participants to test it. Most of these participants were from the academic world, as either students or researchers. All of them were aged 20 to 60 years old with an average age of 35. Out of the 30 subjects, 22 were males and 8 females. They were not from any specific socio-economical background. 10 participants tested the system with the combined eye- and head-tracking of which 8 males and 2 females. The remaining 20 tested the setup with head-tracking only.

Figure 7.2: Phasespace impulse glove.

7.3 Experimental Protocol

We started off the experiment by telling the participants that they would be set in a city scene filled with characters walking up and down the street. We told them that they would have to evaluate various modes of this same scene in a CAVE environment by answering a set of questions. Each participant was asked to test the system individually (without any of other participants around) in order to keep them naive towards our expectations. Before starting the experiment, the participants were asked to go through various calibration steps. The first consisted in calibrating the eye-tracker for those who used the combined system. The calibration procedure in this case was the same as discussed in the previous experiments. The second calibration procedure consisted in recording a reference waving gesture in order to be used for the gesture recognition. We then asked them to enter the CAVE environment in which we projected the scene. This scene depicted a city street in which a crowd of characters was walking around, as depicted in Figure 5.4.

With this experiment, we first wanted to make sure that the participants could understand the difference between the various modes. More importantly, we wanted to verify whether there would be the same bias towards always attentive characters as in the previous experiment or not. Finally, we wanted to verify whether immersion and anxiety increased with gaze. In the case of anxiety, however, we did not expect to obtain clear results since we did not test the application on a phobic population but on a healthy one.

We determined each participant's region of interest throughout the experiment by analyzing their POR determined by the coupled eye- and head-tracking, or by simply using head orientation, depending on the setup. Each participant was then presented the four different modes as described in Chapter 5 in pseudo-random order. This was done to remove the order variable. We then asked them to move around in the CAVE environment, telling them to look around, look at the characters, and even wave at them if they felt like it. We basically asked each participant to interact with the characters for $1 - 2$ minutes in each mode and give us their feedback. To do so, we asked them to evaluate the different modes by answering a set of eight questions for each one:

- Do the characters behave in a normal way?
- Are the characters looking at you?
- Do you feel immersed?
- Is it anxiety provoking?
- Do the characters seem friendly?
- Are the characters aware of their environment?
- Are the characters aware of other characters?
- Are the characters aware of you?

They were asked to rate these questions on a Likert type 10 point scale [Likert, 1932]; a rating of 1 meaning "not at all", and 10 meaning "very much". Finally, we asked them to tell us which mode they preferred out of the 4, in other words, which seemed most natural to them, and why. We also asked them if they had any extra comments or anything else they would like to mention, which had not been addressed by the other questions.

The aim of the first question was to evaluate whether characters who were very attentive would seem more natural than characters who did not pay attention to the participant. This question was asked in order to verify whether there would be the same bias towards always attentive characters as was the case in our previous experiment or not. The second and eighth questions are very similar. We wanted to make sure that the participants could make the difference between the various modes and that they would answer in a similar way to both these questions. The third question was asked in order to verify our hypothesis, that immersion would increase with gaze. However, we also wanted to see if too much gaze would decrease the feeling of immersion. The fourth question, concerning anxiety is actually not really applicable in this context since we tested our application on healthy subjects only. Nevertheless, we wanted to see if a situation in which all the characters look at the participant would induce a slight anxiety. The fifth question, concerning character friendliness, was asked in order to see if attentive characters would seem more pleasant than characters who do not pay any attention to the user, knowing that they did not present any difference in behavior between the different modes apart from gaze. The sixth question, relative to character awareness towards their environment was asked to see if the participants would notice differences where there were none. In all four modes, the characters did not differ in their attention towards their environment. Finally, question number seven, concerning character awareness towards other characters, was asked in order to see if the participants would rate this question higher in the random mode and lowest in the user-centered mode. We wanted to see whether the perceived awareness of characters between themselves would increase in the random mode even though they looked at each other in a totally random way. We believed that we should put the emphasis on the user and on what the user looked at, but that the rest would not be of much importance.

	User-centered	Interest-centered	User or Interest	Random
Normality	5.93(1.64)	5.77(1.57)	5.7(1.6)	5.53(1.46)
Gaze	7.57(1.83)	5.37(2.3)	6.8(2.0)	3.53(1.5)
Immersion	6.73(1.96)	5.97(2.11)	6.37(2.25)	5.8(2.17)
Anxiety	3.17(2.2)	2.57(2.05)	2.9(2.22)	3.13(2.46)
Friendliness	5.33(1.35)	5.2(1.75)	5.27(1.86)	4.37(1.54)
Environment	5.6(2.01)	5.7(1.99)	5.83(2.0)	5.37(2.14)
Characters	3.7(1.82)	4.53(2.11)	4.57(2.42)	4.13(2.4)
User	7.13(1.81)	5.8(2.38)	6.53(2.18)	3.93(1.84)

Table 7.1: Mean scores to the various questions. Their standard deviation is in parentheses.

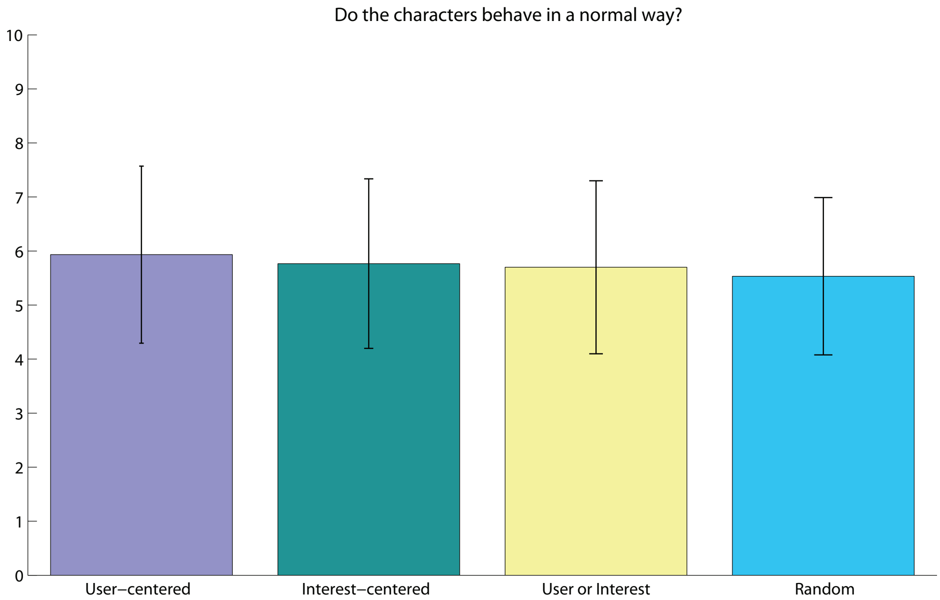

Figure 7.3: Graphical results to the assessment of character normality.

7.4 Results

Table 7.1 gives an overview of the mean results obtained for the various questions.

Concerning the first question we asked, about the normality of the character behaviors, we obtained very similar average results for the four modes. This is depicted in Figure 7.3.

For our second question, which evaluated to what extent the characters were looking at the user, we have a clear preference for the user-centered and the user or interest modes, which is in line with our expectations. This is depicted in Figure 7.4. In the user-centered mode, all characters look at the user when possible. In the user or interest mode, they all either look at the user or at what the user is looking at. However, the character the user is looking at looks back at the user. The participant can thus have the impression of being looked at a lot in this mode as well. In the interest-centered mode, most characters look at

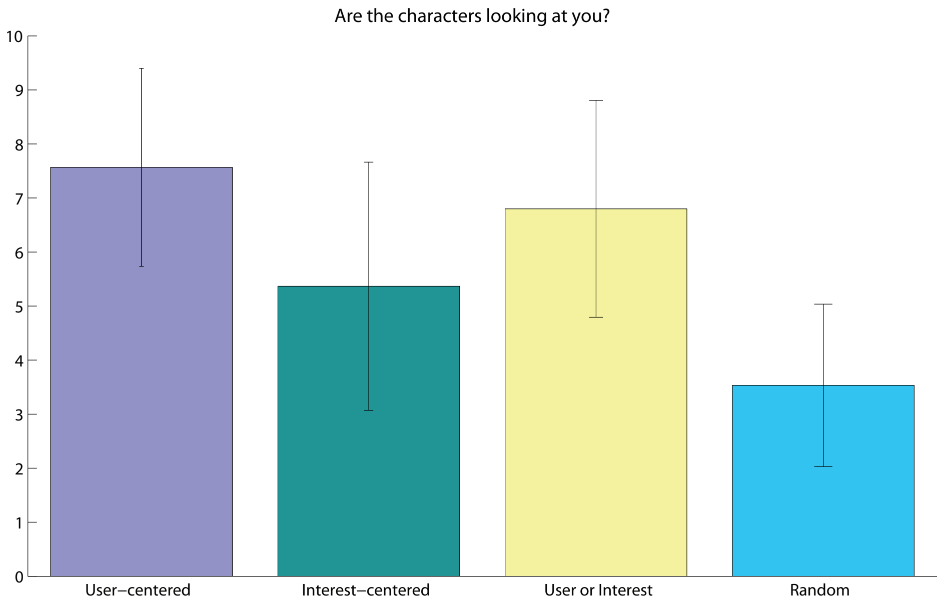

Figure 7.4: Graphical results to the assessment of character gaze.

what the user is looking at but the character being looked at by the user looks back at the user. For this reason, the participant perceives a higher quantity of gaze in this mode than in the random mode.

Regarding immersion, shown in Figure 7.5, we can see that the mean results have the same tendency as for the second question. Immersion therefore seems to increase with character gaze. This is in line with our expectations since we believed that immersion would increase if the participant had the impression that the characters were aware of him/her. This is actually the case when the characters look at the participant.

Regarding anxiety, depicted in Figure 7.6, none of the participants were particularly anxious which is normal since we did not deal with a phobic population. However, we can still notice a slight preference for the user-centered and the random modes. Regarding the user-centered mode, this seems quite logical. Indeed, if we were to walk in town and have everyone stare at us, it would be quite anxiety provoking. However, the random mode equally scored highest in anxiety, as if it was anxiety provoking that no-one looked at us. This may be explained in the case of participants biased towards always attentive characters, it can seem abnormal for characters not to look at them and thus more anxiety provoking.

Regarding friendliness, we can see that when the characters were not looking at the user at all, or only as much as other characters, they seemed less friendly than in the other modes. This can be seen in Figure 7.7. All other modes scored approximately the same for this feature. It is important to note that there was no facial animation in our experiment. Characters therefore did not change expression throughout the exposure. We can thus reasonably assume that they seemed less friendly only because they were not looking at the user.

Regarding the character awareness towards their environment, depicted in Figure 7.8, we can see that the results to all four modes are quite similar. There is however a slight decrease

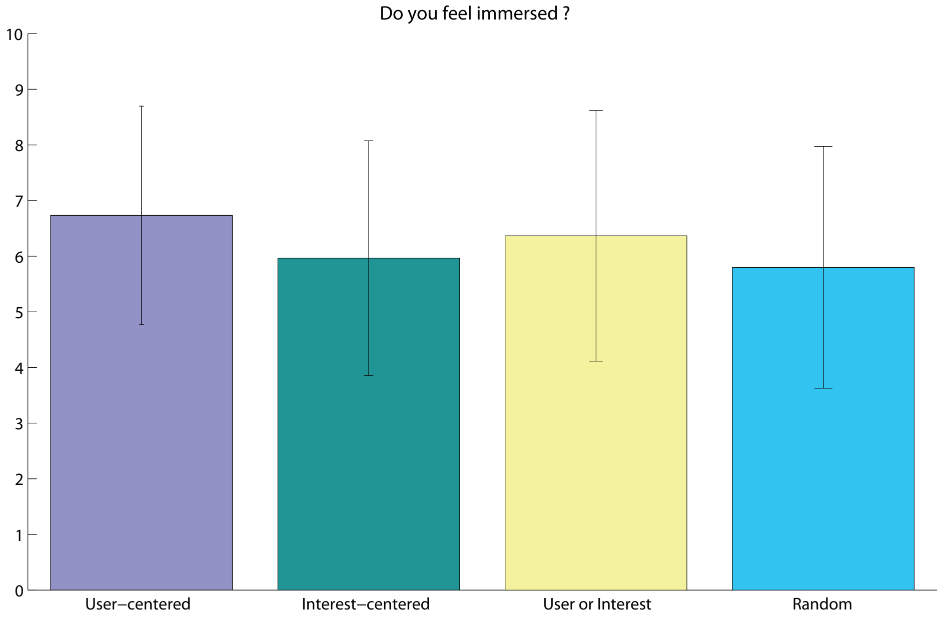

Figure 7.5: Graphical results to the assessment of immersion.

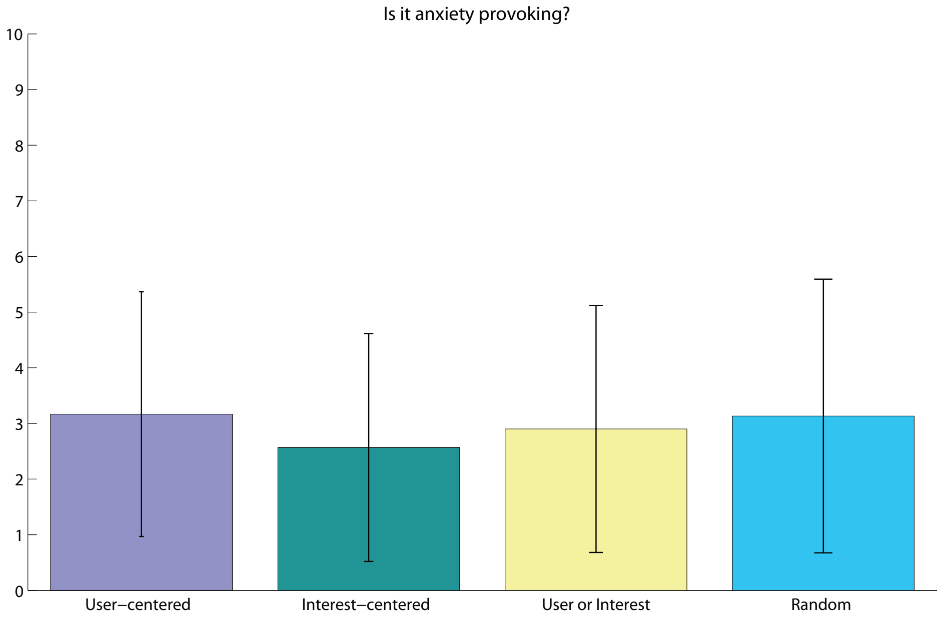

Figure 7.6: Graphical results to the assessment of anxiety.

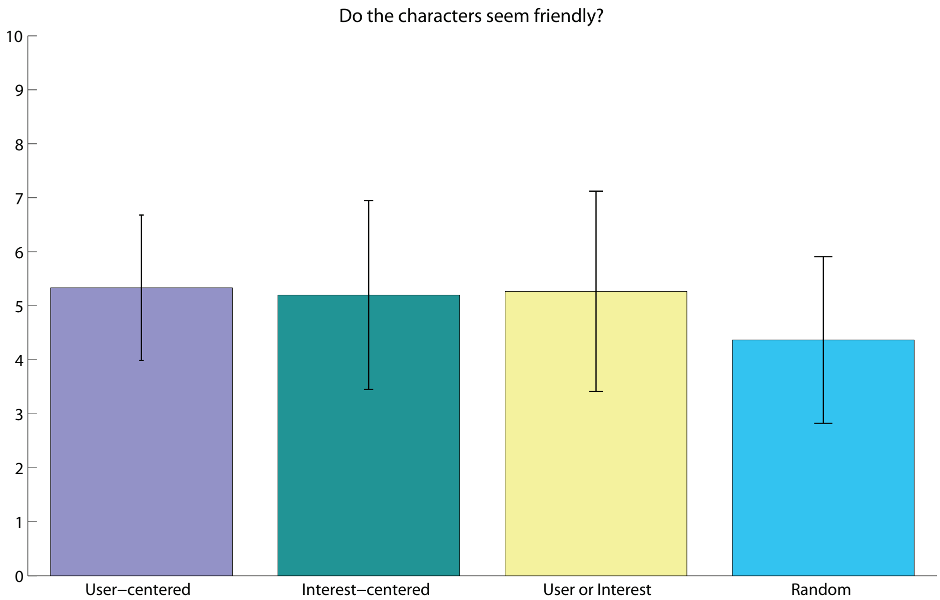

Figure 7.7: Graphical results to the assessment of character friendliness.

in awareness perception in the random mode and a slight increase in the user or interest mode. This parameter however did not change throughout the different modes. It seems as though when characters look at each other and at the user randomly, users perceive that they are less aware of their environment. This may be explained by the fact that the characters seem not to be looking at the participant at all. This could equally be explained by the fact that there is no logic as to where or what the characters look at in this mode; the interest points are arbitrarily selected.

Concerning the evaluation of character awareness towards other characters, the results were not what we expected. We expected the random mode to have the highest results. However, we can see in Figure 7.9 that the two which obtained the highest results were the interest-centered and the user or interest modes. It is normal for the user-centered to obtain the lowest score since the characters would only be looking at the user but not at the other characters. Those which scored highest though, were the modes in which character attention was divided between the user and other characters. Once again, the fact that there is no logic as to what the characters look at in the random mode seems to give the impression that they are not aware of one another.

Finally, concerning character awareness towards the user, the user-centered mode scored slightly above the user or interest one, as depicted in Figure 7.10. We expected the user-centered mode to be clearly scored highest. The tendency however is the same as for the second question, which is in line with what we expected since both questions were alike.

We have equally performed a Student's T-test in order to verify whether there were any statistical significances between the various modes. The results are shown in Table 7.2. If we consider the 95% probability of significant difference between modes, we can see that the assessment of gaze (are the characters looking at you?) falls into this category for all

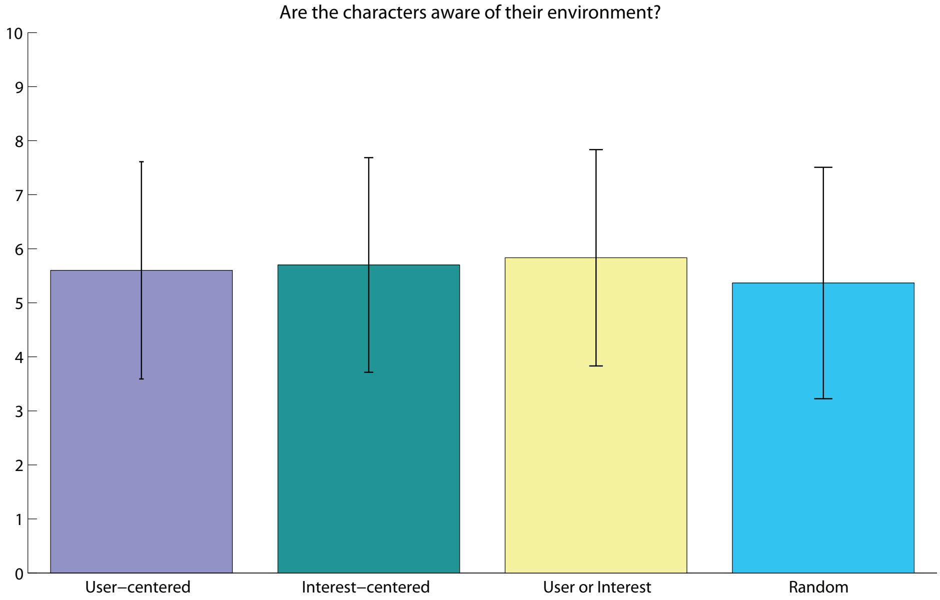

Figure 7.8: Graphical results to the assessment of character awareness towards the environment.

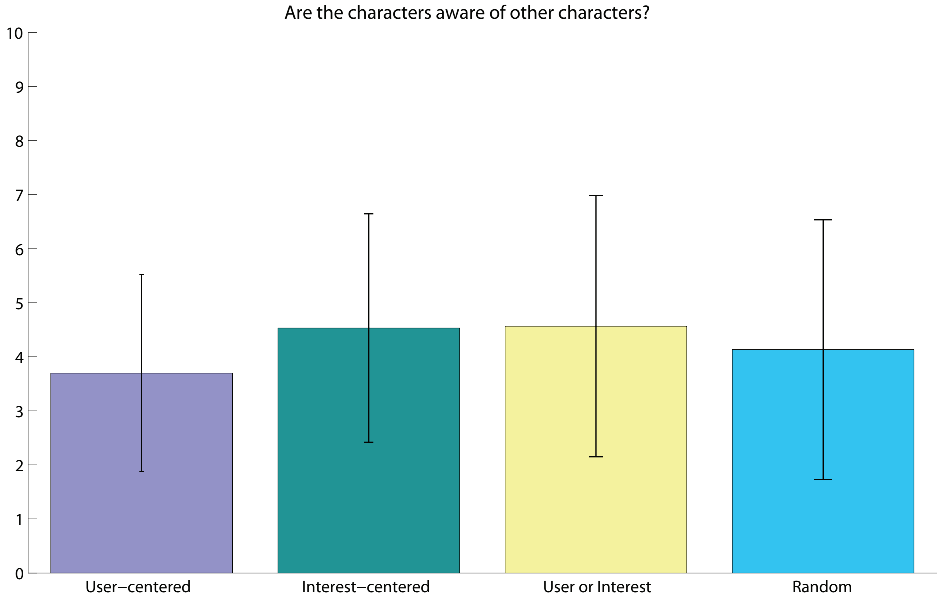

Figure 7.9: Graphical results to the assessment of character awareness towards each other.

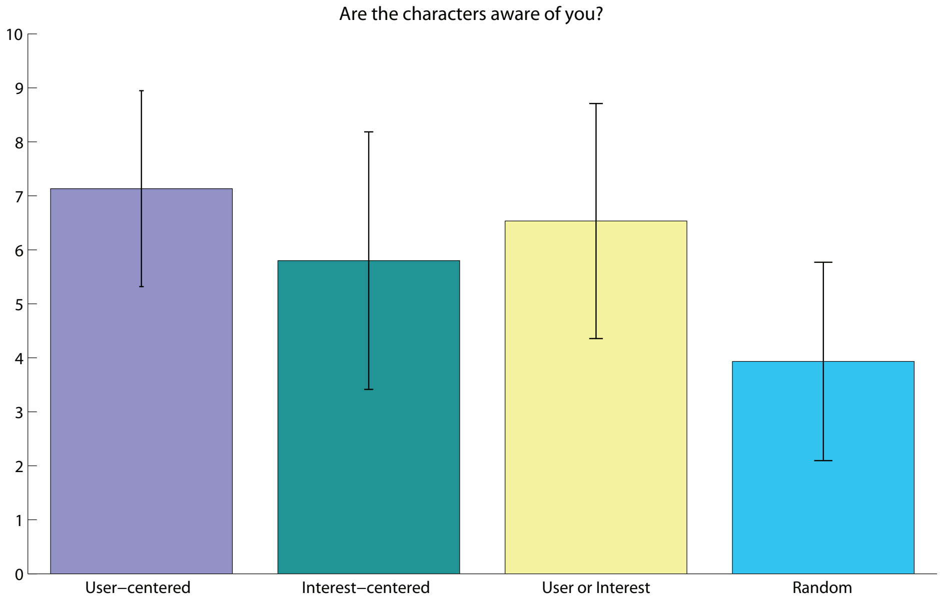

Figure 7.10: Graphical results to the assessment of character awareness towards the user.

compared modes except for the user-centered versus the user or interest-centered modes. The perceived difference in friendliness also falls into this category for the user-centered versus random modes and the user or interest-centered versus random modes. Finally, it is also the case in the assessment of character awareness towards the user for all modes except user-centered versus user or interest-centered modes and interest-centered versus user or interest-centered modes. Even though we did not obtain statistically significant differences in observation for the other questions, we have to take into consideration that these measures remain subjective and that variability may thus be highly increased as compared to objective measurements. Moreover, these results are in line with those we have been able to observe in graphs 7.3 to 7.10.

We have also tested the correlation between various questions. To this end, we have used Pearson's product-moment correlation coefficient r:

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{s_X} \right) \left(\frac{Y_i - \bar{Y}}{s_Y} \right) \tag{7.1}$$

The convention is to use r^2 as measure of association between two variables. For instance, the value of r^2 between the results to the second and eighth questions is of 0.73. 73% of the variance in the perceived gaze of characters can be accounted for by changes in the perceived awareness of the characters towards the user. Thus, these two values are correlated at 73% in average. The correlation between the second and third questions is of 0.53. 53% of the variance in immersion can thus be accounted for by changes in quantity of perceived gaze. We have also tested the correlation between immersion and perceived awareness of the characters towards the users. This value is of 59%. Similarly, the correlation between perceived friendliness and perceived awareness of characters towards the user is of 53%.

	1 - 2	1 - 3	1 - 4	2 - 3	2 - 4	3 - 4
Normality	0.6888	0.579	0.3217	0.8711	0.5527	0.6746
Gaze	0.0001	0.1277	0.0000	0.0126	0.0005	0.0000
Immersion	0.1504	0.5041	0.8614	0.4804	0.7641	0.3252
Anxiety	0.2783	0.6418	0.9561	0.5475	0.3360	0.7010
Friendliness	0.7421	0.8741	0.0122	0.8867	0.0551	0.0456
Environment	0.8470	0.6540	0.6651	0.7965	0.5343	0.3868
Characters	0.1073	0.1222	0.4345	0.9548	0.4963	0.4890
User	0.0179	0.2510	0.0000	0.2185	0.0012	0.0000

Table 7.2: Results to Student's T-test. 1 is the user-centered mode, 2 is the interest-centered mode, 3 is the user or interest centered mode and 4 is the random mode.

However, it has to be noted that all participants did not react in the same way towards the experiment. Some participants focused on other things than expected, such as character size and the fact that there was no collision detection between them and the user. On the other hand, other participants set these factors aside and focused on character gaze attention behaviors. For example, the correlation between perceived immersion and perceived awareness of characters towards the user increases to 66% if we take out those who complained of lack of collision avoidance. We thus believe that this factor can greatly account for low scores in normality, awareness and immersion. This value was above 75% for 10 of the participants and above 90% for 6 of them.

Regarding the most normal, preferred mode, 14 participants out of the 30 opted for the user or interest one. This is almost half of the population we have tested. Out of the 16 remaining, 11 chose the user-centered one. A little more than a third of our tested cohort were therefore possibly biased towards an always attentive population of virtual characters. Finally, 4 participants chose the interest-centered mode and 1 chose the random one. When asked why, almost all the participants having chosen the user or interest mode, answered that it seemed more natural to them because the characters were looking at them but not staring at them. The characters seemed to be interested in them and aware of them but also of other characters in the scene. There are thus really two major trends. On one hand, we found a group of people for whom it seemed more natural of being constantly looked at by the virtual characters. On the other hand, almost half the tested cohort had a preference for the random mode in which they were being looked at but not all the time and not by all characters.

It is interesting to note that the correlations are different if we isolate the preferred mode. For example, the correlation between anxiety and the perceived gaze of characters was 58% for those who preferred the user-centered mode. On the other hand, it was 29% when considering those who opted for the user or interest mode. It is quite interesting to note that for those users who found always attentive characters as most natural, the correlation with anxiety was almost twice as high as for those who preferred characters who looked at them but not always. The same tendency was found between perceived normality and awareness of characters towards the participants, and also between perceived friendliness and perceived awareness towards the user. The correlations in all these cases are practically twice as high in the participants who preferred the user-centered mode.

Finally, we analyzed the eye-tracking data for the 10 participants who wore this device. We have not noticed any difference in the types of assessments to the various questions we have asked between users who wore the combined eye- and head-tracking device and those who wore the head-tracking device only. Eye-tracking therefore does not seem to have a significant added value in the determination of a user's gaze direction. As mentioned at the beginning of this chapter, we also analyzed eye scan velocity and pupil dilation. Indeed, hyperscanning is a symptom found in people suffering from social phobia [Horley et al., 2004] and pupil size is known to increase with anxiety [Marks, 1987]. Our results, however, were not probing. We did not obtain any correlation between eye scan velocity, pupil dilation and the subjective ratings to the questions we asked. However, this is probably due to the fact that we did not deal with a phobic population. We believe that both eye scan velocity and pupil dilation could actually be very good indicators of the presence of increased anxiety depending on the various modes in agoraphobic patients. However, we would need to test our application on such a population to confirm these hypotheses.

7.5 Discussion

Simulation related problems. We have encountered several problems in the display of our simulation which have interfered with the features we wanted to evaluate. First, there was no collision detection between characters and with the user. They thus collided with and intersected the user from time to time. For some participants, this considerably reduced their evaluation of character awareness. Here as well, we can mention the uncanny valley effect as there were clearly inconsistencies between character appearance and behavior. Moreover, the frame rate being quite slow due to the complexity in the four-screen display, the scene lagged from time to time, which we believe provoked a decrease in the feeling of immersion. Finally, also due to those quite slow frame rates, we had to display our characters in a lower level of detail. They thus lacked facial animation, which we also believe considerably reduced the feeling of immersion.

Combination of gaze and gesture. When we evaluated our application, we allowed for the users to interact with the characters using both gaze and gesture. We believe that we should have evaluated these two forms of interaction separately. Indeed, some participants were distracted by the waving and thus did not pay very much attention to the gazing behaviors. Moreover, when the gesture was not recognized, the characters did not respond. Some participants' evaluation of awareness and friendliness therefore suffered from this.

Computation of Interest Points. We have seen from our results that even though the characters were looking at each other in the random mode, the participants did not perceive their awareness towards each other. We believe this is due to a random choice of interest points to look at instead of a logical one. We thus think that this mode of the application could greatly benefit from the implementation of an interest point detection algorithm as the one we presented in Chapter 4.

Validation with a phobic population. Nevertheless, we believe our application to have a high potential in the treatment of agoraphobia with crowds. We believe that it could prove to have very high immersive qualities. Moreover, the application allows the definition of the percentage of the virtual population which should be attentive to its surroundings (should it be other characters, objects or the user). We thus have an application which can be tuned in order to gradually expose patients to increasingly anxiety-provoking crowds.

Interaction precision. We believe that we could further enhance our application by allowing more precise interactions using the eye-tracking device. As in the application we presented allowing interaction in the context of social phobia, we could use gaze to define character behavior. For example, we could make them attentive when being looked at in the face and make them nod or wave in return.

7.6 Conclusion

In this chapter we have presented a study we have conducted over 30 healthy people in order to evaluate the application we have developed in the context of agoraphobia with crowds. We have tested four different situations, one in which all the characters were interested in the user and gazed at him/her, one where all characters were interested in what the user was looking at and paid attention to that object/character, one in which the characters either looked at the user or at what the user was looking at, and finally, a last situation in which the characters randomly gazed at any other character in the scene, user included. The results we have obtained show that there was a clear identification of the user-centered and the random modes. Participants to the experiment noticed that they were being stared at by all characters in the former and that none of them were looking at them in the latter. We have also been able to outline that there were tendencies in correlations between gaze, immersion and perceived friendliness of characters. From our results, we have also seen that our application could benefit from improvements, which we have discussed in the previous section. Nevertheless, we strongly believe that this form of gaze based interaction can greatly amplify the immersive properties of such applications dedicated to the treatment of phobias.

CHAPTER 8

Conclusion

In the past decade, VR has proven to be an efficient tool to be used in the treatment of phobias. Many studies have demonstrated its efficacy for many phobias such as acrophobia, arachnophobia and fear of flights. In this thesis, we have focused on the development of various applications to be used in the treatment of social phobia as well as tools to diagnose and assess them. More specifically, we have worked on tools to be used in the treatment of fear of public speaking and agoraphobia with crowds.

8.1 Contributions

Our contributions can be divided into three major sections: gaze interaction in the context of social phobia treatment, gaze attention behaviors for crowds, and gaze interaction in the context of agoraphobia treatment.

8.1.1 Gaze Interaction for Social Phobia

Our first main contribution resides in the development of an application allowing interaction between a user and a virtual character in the context of social phobia. This application allows the determination of how a virtual character should behave using user eye-tracking data. We allow for characters to respond to the user's eye contact behavior. The character thus seems interested and attentive when being looked at, and distracted when being avoided by eye contact. A phobic person doing a public speaking exercise may thus induce a positive feedback loop when gazing at the virtual character.

8.1.2 Gaze Attention Behaviors for Crowds

Our second main contribution is the development of gaze attention behaviors for crowds of characters. This consists of an extra layer added to an existing crowd animation. Our method automatically detects the points which are interesting to look at for the virtual characters using features such as distance, orientation, and speed. We use the character trajectories only in order to define our interest points. It then consists of adapting the character motions to satisfy the automatically defined gaze constraints in a smooth and natural way. The method we have proposed allows the definition of the *where*, the *when* and the *how* to gaze for hundreds of characters.

8.1.3 Gaze Interaction for Agoraphobia

Our third contribution is the development of gaze attention behaviors for crowds in real time. The main idea of the method is the same as the one we have proposed to create gaze attention behaviors for crowds, but reconsidered in order to meet online requirements. In this system, not only are the characters able to look at each other or at objects in the scene, but also at the user walking around in their environment. This application allows the determination of the amount of characters which should perform gaze behaviors. It also allows the definition of different modes of interest for the characters, i.e. gaze at the user, gaze at the user's interest point, randomly gaze at any character, or a mix of any of these modes. Finally, we allow a very immersive setup in a CAVE environment where a user can interact with characters in a crowd using gaze and gestures.

8.1.4 Experimental Validations

Finally, we have discussed various studies to validate our applications in the context of social phobia and agoraphobia with crowds. Our first validation experiment evaluated the use of VR to treat social phobia and more specifically, fear of public speaking. We have also validated the use of eye-tracking as an assessment tool for therapists, as it allows the visualization of where a patient has been looking throughout an exposure session.

Our second validation experiment focused on the use of eye-tracking as a diagnosis and assessment tool for the treatment of social phobia; and more specifically, fear of public speaking. Since abnormal eye contact behavior is a known feature present in social phobia, this consists in an objective evaluation tool. We validated the use of this tool with an experiment conducted over a phobic population and a control group population.

Our third validation experiment consisted in evaluating the application we have developed for interaction between a user and a virtual character in the context of fear of public speaking. We have tested our application on a group of healthy people and conducted a case study on a young girl suffering from Asperger syndrome.

Finally, our last experiment consisted in validating our application allowing the interaction with crowds of characters in a CAVE environment. We have tested this on a group of healthy people and discussed its application to agoraphobic patients.

8.2 Perspectives and Future Work

There are several different work directions which could be undertaken with respect to the contributions described in this thesis.

8.2.1 Gaze Interaction for Social Phobia

As mentioned in Chapter 3.3, we could enhance our existing algorithm by filtering the eye-tracking data in order to have a more overall view of the gaze behaviors, and thus, a more appropriate response in the virtual character behaviors. Concerning the experimental validation, presented in Chapter 6.4, a direction for future work would consist in conducting this study on a phobic population, as we have only done it on a healthy one. Due to the strong reaction we have obtained from our case study subject, we believe that the effect of these character behavior changes could be much more important on a phobic population than in the case of healthy subjects. Moreover, it would be interesting to see if a phobic population would present the same bias towards always attentive characters, as the healthy population did. However, we do not think this would be the case.

8.2.2 Gaze Attention Behaviors for Crowds

Regarding our application on attention behaviors for crowds, described in Chapter 4, a first research direction would be to try to add other interest criteria such as color or intensity. It could also be interesting to add sound. Characters could thus react to different sounds depending on the direction they come from. We believe this could greatly enhance the realism of character behaviors. Finally, it could be possible to use our method for navigation, in order for characters to change their behaviors with regards to what they see. Also, our architecture could benefit from a set of top-down rules to attend to specific things or seek for objects or other characters.

Another interesting future work direction would be to combine eye-tracking with motion capture in order to animate characters. This would allow the achievement of realistic eye movements and cues on frequency and duration of gaze attention behaviors. This is actually work which is currently in progress. A statistical model of gaze attention behaviors and gestures for crowd characters could then be developed using these captured data. Finally, such a model could be compared to the one proposed in this thesis in order to evaluate the pros and cons of each and to possibly improve simulated gaze attention behaviors.

8.2.3 Gaze Interaction for Agoraphobia

Concerning our application which allows for interaction with virtual crowd characters in a CAVE environment, presented in Chapter 5, several improvements could be done. As previously mentioned, instead of having a random mode, an automatic detection of interest points could be added to the application. Eye-tracking could also be used to create more elaborate interactions between the user and the virtual characters, as was done in our application on interaction for social phobia. Concerning the experimental validation, presented in Chapter 7, the main future work direction which could be extremely interesting, would be to evaluate such an environment with an agoraphobic population.

8.2.4 Eye-tracking for Diagnosis and Assessment

Finally, concerning the two other studies we have conducted and described in Chapters 6.2 and 6.3, regarding the use of eye-tracking as diagnosis and assessment tool, it could be interesting to conduct a study on a large cohort in order to statistically validate our tools.

8.3 Final words

In this thesis, we have focused on the simulation and the use of gaze attention behaviors in the context of social phobia. Working in this area has been a fascinating experience and there is still much to learn in the domain. We believe that a lot remains to be done in the human - virtual human interactions and more specifically in applications dealing with health. As the saying goes, “the eyes are the window to the soul”. We are convinced that the day we will be able to reach real empathy between real humans and virtual characters by using non-verbal behavior, we will have made an important step in the ability to use VR in the treatment of social phobia. We believe that in this thesis, we have been able to make small but nonetheless meaningful steps in that direction.

APPENDIX A

Ethics Commission Request

Protocole de recherche soumis à la Commission d'Ethique de la Psychiatrie du Service des Hospices Cantonaux

1. Titre de l'étude
Utilisation d'un dispositif d'acquisition de mouvements oculaires comme outil de diagnostique et d'évaluation de l'évitement du contact visuel dans le cadre du traitement des troubles de l'anxiété sociale.

Cette étude a été initiée et le protocole a été rédigé par l'investigateur principal (Docteur Françoise Riquier) en collaboration avec le VRLab (Professeur Daniel Thalmann, Helena Grillon, EPFL, Lausanne).

2. Dates
Date de l'envoi du protocole : 15 octobre 2006 Date prévue pour le début de l'étude : 15 novembre 2006

3. Cadre de l'étude
L'investigateur principal est le Docteur Françoise Riquier, psychiatre psychothérapeute spécialisée en thérapie comportementale et cognitive.
Date et signature :

La participation technique est assurée par le laboratoire VRLab à l'EPFL (Ecole Polytechnique Fédérale de Lausanne) sous la direction du Professeur Daniel Thalmann.

4. Mise en perspective de l'étude
a. Etat des connaissances

- **Phobie sociale**
  Le trouble de l'anxiété sociale (ou phobie sociale) est un des troubles psychiatriques les plus répandus. Dans les années 1990, on estimait que 13% de la population occidentale en souffrait. Ce trouble fréquent et chronique débute généralement dès l'adolescence avec une prévalence à vie de 2 à 4%.
  Le trouble de l'anxiété sociale est caractérisé par une crainte intense et persistante des situations de performances sociales dans lesquelles une gêne peut exister (typiquement la crainte de parler en public et/ou des situations où il doit y avoir des interactions avec des personnes). Selon le DSM-IV (Diagnostic and Statistical Manual of Mental Disorders), les sujets souffrant de phobie sociale sont inquiets qu'autrui les jugent anxieux, fragiles, fous ou stupides et perçoit leur malaise. Comme pour les autres troubles anxieux, les sujets cherchent à éviter les situations angoissantes ou menaçantes de différentes manières :

  – Evitement plus ou moins complet des situations anxiogènes (manger, boire en public, serrer la main, utiliser les toilettes publiques).
  – Recherche de réassurance par la présence d'un proche ou le recours à une médication accessible rapidement.
  – Recours à l'alcool ou à des drogues.
  – Au-delà de la souffrance éprouvée dans les situations redoutées, ce trouble s'aggrave avec le temps et peut conduire à une restriction significative de la vie sociale et professionnelle.

  Ce trouble peut exister seul ou en comorbidité avec d'autres troubles psychiatriques ; en particulier le trouble panique, les abus de substances et la dépression majeure. Il représente donc un problème de santé publique.

- **Diagnostic et évaluation de la phobie sociale**
  Il n'existe pas de test diagnostic de laboratoire en ce qui concerne les phobies. Le diagnostic est basé sur le récit du patient à propos de ses expériences vécues. Le DSM-IV fournit des critères de diagnostic pour plusieurs phobies dont la phobie sociale :

  – La personne a peur ou est anxieuse de faire l'expérience d'un embarras public ou d'une humiliation dans une situation sociale ou de performance.
  – De se retrouver dans de telles situations provoque une anxiété intense et possiblement une attaque panique.
  – La personne se rend compte que sa peur est excessive et irrationnelle.
  – Les situations sociales ou de performance sont évitées ou endurées avec une angoisse intense.
  – Leur condition entrave leur habilité à fonctionner au travail ou à l'école et leur fait se retirer d'activités sociales et/ou de relations sociales ou de les endurer avec une angoisse intense.

- Cette condition persiste pendant au moins 6 moins chez les personnes âgées de plus de 18 ans.
- La peur et l'évitement ne sont pas causés par d'autres problèmes mentaux, une condition médicale ou les effets d'une drogue.

La caractéristique centrale de ce trouble est une peur sous-jacente d'être évalué de façon négative ou d'être jugé par d'autres. Les patients atteints de phobie sociale évitent certaines situations soit entièrement, soit partiellement, en minimisant le contact visuel, la communication verbale ou la présence physique.

Les symptômes physiques les plus fréquents en réponse à la situation redoutée sont les palpitations cardiaques, les tremblements, la transpiration, les muscles tendus, l'estomac noué, la bouche sèche, les sueurs chaudes ou froides, les rougissements et les maux de tête.

- **Utilisation d'un dispositif de capture de mouvements oculaires comme outil de diagnostic**
  Certaines études ont déjà été menées concernant l'utilisation de dispositifs de capteurs de mouvements oculaires afin de diagnostiquer certaines maladies et phobies ainsi que pour le traitement de certains dysfonctionnements oculaires ou mentaux.

  En effet, certains types de mouvements oculaires sont symptomatiques de certaines maladies ou phobies. Par exemple, il a été démontré que les tâches exigeant des mouvements en saccade de l'oeil permettent d'identifier le déficit cognitif chez les enfants atteints de schizophrénie. De même, l'évitements du contact visuel est un symptôme de la phobie sociale.

- **Réalité virtuelle dans le traitement de la phobie sociale**
  L'exposition en réalité virtuelle se présente aujourd'hui comme une alternative aux expositions in vivo standard. Plusieurs études confirment l'efficacité de ces expositions dans le traitement de la phobie sociale.

  La réalité virtuelle offre en effet des scénarios anxiogènes difficilement accessibles et disponibles dans la vie quotidienne (avion, sommet d'une montagne, discours devant une assemblée de personnes) et permet de répéter sans limite les expositions.

  Ces études représentent une amorce de tout un travail qui reste à réaliser aussi bien dans l'amélioration des aspects techniques (logiciel de simulation 3D, amélioration et multiplication des scènes virtuelles, confort et maniabilité des équipements comme les casques, les lunettes ou capteurs de mouvements oculaires) que des programmes thérapeutiques employant la réalité virtuelle (développement et définition de protocoles thérapeutiques, évaluation de l'efficacité dans un objectif de reproductibilité).

b. Justification

Le développement d'outils de diagnostic ainsi que de traitement pour le trouble de l'anxiété sociale se révèle nécessaire et important compte tenu de l'impact de cette pathologie en termes de santé publique et chez un individu.

En ce qui concerne le diagnostic, cela se révèle d'autant plus intéressant à prendre en compte en raison du manque d'outils à cette fin. En effet, le seul diagnostic actuellement connu pour la phobie sociale est le récit même de la personne en ressentant les effets.

c. But de l'étude

- **Etudes préliminaires**
  Deux études préliminaires ont été effectuées en collaboration avec l'équipe du Professeur Daniel Thalmann à l'EPFL, spécialisée dans la représentation humaine en réalité virtuelle.

  La première de ces études a impliqué des sujets sains chez lesquels une évaluation de l'inconfort éprouvé dans des situations sociales avait été évaluée. Ces sujets ont été exposés à une situation en réalité virtuelle figurant une audience en 3 dimensions composée de regards émergeants dans le noir et cernant le sujet. Les sujets qui présentaient le plus de malaise dans les situations sociales au quotidien ont éprouvé l'anxiété subjective la plus forte dans ce cadre. Les mesures objectives physiologiques relatives à l'anxiété (rythme cardiaque et conductance cutanée) étaient corrélées aux mesures subjectives. Le rythme cardiaque s'est trouvé être un indicateur particulièrement bon du niveau d'anxiété et a permis de distinguer significativement les sujets sains gênés en situation sociale, des sujets sains qui ne l'étaient pas.

  La deuxième de ces études a découlé directement de la première et a été acceptée par la Commission d'Ethique de la Psychiatrie du Service des Hospices Cantonaux sous l'intitulé de Ütilisation de la réalité virtuelle comme outil thérapeutique pour les expositions comportementales dans le cadre du traitement des troubles de l'anxiété sociale ̇. Elle a consisté en 8 séances d'exposition à des scènes virtuelles variées dans un visiocasque et de deux séances d'exposition à des scènes virtuelles sur un grand écran de projection pendant lesquelles les sujets étaient munis d'un capteur de mouvements oculaires. Ces deux séances ont eu lieu en début et fin de traitement. Les résultats obtenus lors de cette étude ont été très positifs. Nous avons noté une nette différence de comportement en ce qui concerne le contact visuel entre les deux séances pendant lesquelles les sujets étaient munis du capteur de mouvements oculaires.

- **But de l'étude actuelle**
  Le but de cette étude est d'utiliser le cadre de l'exposition virtuelle pour évaluer de manière objective l'évitement du contact visuel, qui est un paramètre de la phobie sociale. Jusqu'à présent, peu d'études sur l'importance du contact visuel dans le trouble de l'anxiété sociale ont été réalisées de manière objective.

  Plus particulièrement, il s'agit d'utiliser un système de capture des mouvements oculaires sur un groupe de personnes souffrant de phobie sociale ainsi que sur un groupe contrôle afin de déterminer si ce système pourrait être utilisé en tant qu'outil de diagnostic.

d. Objectif de l'étude

Un système d'eye tracker (suivi des mouvements oculaires) sera utilisé pour l'évaluation de l'évitement du contact visuel. Ce système sera utilisé couplé avec une exposition à des scènes virtuelles, projetées sur grand écran, pendant laquelle les sujets devront faire un exercice d'expression verbale de quelques minutes. Le système de capture de mouvements oculaires nous permettra alors de mettre en évidence et d'analyser de manière directe, précise et en temps réel, les zones des environnements virtuels regardées par les sujets.

5. Plan général

Ce projet est une étude se déroulant en deux temps. Premièrement, une phase d'habituation au matériel. Les sujets, autant sains que souffrant de phobie sociale, seront exposés à une scène virtuelle et équipés du matériel de capture de mouvements oculaires. Une fois les sujets habitués au matériel, nous passerons à la deuxième phase de l'étude. Celle-ci consiste en une deuxième exposition à 3 scènes virtuelles différentes. Les sujets seront également munis de l'équipement permettant la capture des mouvements oculaires. Pendant cette deuxième phase, les données relatives au positionnement de l'oeil seront enregistrées.

Il s'agira ensuite d'analyser les résultats. Premièrement, il s'agira de voir si le comportement visuel est différent entre sujets sains et sujets souffrant de phobie sociale. Deuxièmement, il s'agira de voir si le comportement visuel est le même pour les différentes situations représentées par les scènes virtuelles pour un même sujet.

6. Sélection des sujets

En ce qui concerne les sujets souffrant de phobie sociale, 6 personnes seront recrutées par voie de consultations ambulatoires spécialisées dans l'anxiété. En ce qui concerne les sujets sains, 6 personnes seront recrutées au sein de la population estudiantine de l'EPFL.

Comme le matériel utilisé émet un certain champ magnétique et que la capture des mouvements oculaires se fait par la capture de la réflexion d'un rayon lumineux sur la cornée, les critères d'exclusion sont :

- Femmes enceintes
- Personnes ayant un pacemaker
- Personnes ayant des problèmes oculaires

7. Méthodes d'investigation
a. Evaluation de l'évitement du contact visuel

Il est effectué à partir de l'enregistrement à distance du mouvement de la pupille et de la réflexion cornélienne par un système de capture de mouvements oculaires (eye-tracking) à l'occasion de l'exposition à plusieurs scènes virtuelles. L'enregistrement est analysé et les données sont matérialisées sous forme d'une carte figurant les zones balayées par le regard et le temps de contact grâce à un logiciel spécifique et conçu à cet effet. Hormis la gêne liée au dispositif durant l'exposition, ce système ne présente aucun danger. Il est toutefois recommandé aux femmes enceintes et aux personnes portant un pacemaker d'éviter cette exposition car le système fonctionne à l'intérieur d'un champ magnétique.

8. Surveillance médicale

Les bilans d'évaluation et de suivi comme les séances d'information seront effectuées par le médecin chargé de l'étude (Docteur Françoise Riquier).

9. Rôle du personnel soignant

Rôle du personnel technique
L'ingénieur qui assistera le médecin dans les expériences d'exposition sera soumis à un contrat de secret professionnel. Il n'aura aucun accès au dossier médical ni au nom des sujets souffrant de phobie sociale.

Rôle du psychiatre
Il sera chargé de l'exposition, de l'accompagnement et des explications concernant le déroulement des séances. Il sera chargé du recueil des données en rapport à l'anxiété (symptômes physiques, évaluation subjective de l'anxiété, relevé des cognitions).

10. Evaluation des risques et enjeux éthiques

11. Formulaires d'information et de consentement

Formulaire d'information et de consentement (cf. annexe)

12. Considération financière

Les sujets n'auront aucun frais à leur charge.
Les sujets ne seront pas rémunérés.

13. Gestion des données

Chaque patient participant à cette étude aura un dossier géré et gardé confidentiellement sous la responsabilité de l'investigateur principal. Pour conserver l'anonymat, chaque dossier recevra un numéro. La clé de cet anonymat sera constituée par une liste des patients avec le numéro de dossier attribué. Cette liste sera établie et gardée confidentiellement par l'investigateur principal. Les données anonymisées seront traitées de façon statistique par ordinateur à la fin de l'étude.

Richard G Heimberg, Social Phobia - Diagnosis, Assessment and Treatment, Guilford Press, 1995.

Randal G. Ross, Shari Heinlein, Gary O. Zerbe1 and Allen Radant, Saccadic eye movement task identifies cognitive deficits in children with schizophrenia, but not in unaffected child relatives, Journal of Child Psychology and Psychiatry, 46;2005:1354.

Kaye Horley, Leanne M. Williams, Craig Gonsalvez, Evian Gordon, Face to face: visual scanpath evidence for abnormal processing of facial expressions in social phobia, Elsevier, Psychiatry Research 127;2004:43-53.

Kaye Horley, Lea M. Williams, Craig Gonsalvez, Evian Gordon, Social phobics do not see eye to eye: A visual scanpath study of emotional expression processing, Pergamon, Anxiety Diorders 17;2003:33-44.

Rameshsharma Ramroll, Cheryl Trepagnier, Marc Sebrechts, Andreas Finkelmeyer, A gaze Contingent Environment for Fostering Social Attention in Autistic Children, Proceedings of the 2004 symposium on Eye tracking research & applications, San Antonio, Texas, p. 19 - 26.

APPENDIX B

Fearful Situation Document

Nom : .... Semaine du .... au ....

Situation	Evitement 0–8	Niveau d’anxiété 0–8	Satisfaction 0–8	Remarque

Situation : Décrire chaque situation anxiogène (quoi, quand, où, avec qui, comment)
Evitement : A coter sur une échelle de 0 à 8
(0 = vous avez fait face, 8 = vous avez évité d’affronter la situation)

Anxiété : A coter sur une échelle de 0 à 8
(0 = pas d’anxiété, 8 = niveau d’anxiété maximum)

Satisfaction : De quelle manière avez-vous eu l’impression de maîtriser la situation
A coter sur une échelle de 0 à 8
(0 = pas satisfait du tout, 8 = entièrement satisfait)

APPENDIX C

List of Abbreviations

APPENDIX D

Glossary

Acrophobia: acrophobia is the fear of heights. It comes from the Greek term akron, which means peak or summit, and phobos, fear. It belongs to the space and motion discomfort category of specific phobias.

Agoraphobia: agoraphobia is the fear of open spaces or of any place which is difficult to escape from. A place crowded with people thus falls in this category. The term comes from the Greek words agora, and phobos, which literally translate to "a fear of the marketplace.

Arachnophobia: arachnophobia is the fear of spiders. It comes from the Greek term arachne, which means spider, and phobos, fear. This fear is among the most common of all phobias.

Asperger Syndrome: Asperger Syndrome is a neurobiological disorder named after the physician Hans Asperger. In 1944, he published a paper which described a pattern of behaviors in several young boys who had normal intelligence and language development, but who also exhibited autistic-like behaviors and marked deficiencies in social and communication skills. It is only in 1994, however, that this syndrome was added to the DSM-IV [American Psychiatric Association, 1994] and recognized by professionals.

Autism Spectrum Disorders: Autism Spectrum Disorders are also known as Pervasive Developmental Disorders and refer to a group of five disorders characterized by varying degrees of impairment in communication skills, social interactions, and restricted, repetitive and stereotyped patterns of behavior.

Claustrophobia: Claustrophobia is the fear of enclosed spaces. It comes from the Latin claustrum which means "a shut place".

Comorbidity: Comorbidity literally means additional morbidity. It indicates the presence of one or more disorders in addition to a primary one.

Hyperscanning: Hyperscanning is a form of hypervigilance. It consists of an overly rapid sweeping of the eyes.

Hypervigilance: Hypervigilance is an increased state of vigilance and sensory sensitivity. It is usually accompanied by an exaggerated intensity of behaviors whose purpose is to detect threat. This consists of a perfectly normal behavior in situations of danger but is used inappropriately in phobic subjects.

Phobia: a phobia is a persistent and irrational fear, occurring in specific circumstances [American Psychiatric Association, 1994]. It comes from the Greek term phobos, which translates to fear. Its main symptom is the excessive and unreasonable call upon avoidance strategies to elude the feared stimuli.

Social phobia: Social phobia refers to an excessive anxiety in social situations causing distress and impaired abilities in various areas of daily life.

LIST OF FIGURES

BIBLIOGRAPHY

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. American Psychiatric Publishing, Washington DC, 4 edition, 1994. 1.2, 3.3, 3.5, 6.2.2, 6.4.2, D

P. Anderson, B.O. Rothbaum, and L.F. Hodges. Virtual reality in the treatment of social anxiety: Two case reports. *Cognitive and Behavioral Practice*, 10:240–247, 2003. 2.1.1

O. Arikan and D. Forsyth. Interactive motion generation from examples. In *Proceedings of ACM SIGGRAPH, Annual Conference Series*, pages 483–490, 2002. 2.3

Ascension. `http://www.ascension-tech.com/`, 2009. 2.1.3.2, 6.1

ASL. `http://www.a-s-l.com/Site/`, 2009. 2.1.3.2

Autodesk. `http://usa.autodesk.com/`, 2009. 2.1, 2.1.1, 3.2

N.I. Badler, J.D. Korein, J.U. Korein, G.M. Radack, and L. Shapiro Brotman. Positioning and animating human figures in a task-oriented environment. *The Visual Computer*, 1(4): 212–220, 1985. 2.3

A.T. Beck and G. Emery. *Anxiety Disorders And Phobias - A Cognitive Perspective*. Basic Books, New York, 1985. 6.3

A.T. Beck, C.H. Ward, M. Mendelson, J. Mock, and J. Erbaugh. An inventory for measuring depression. *Archives of General Psychiatry*, 4:561–571, 1961. 6.2.1, 6.3.2.2

C. Botella, R.M. Baños, V. Guillen, C. Perpina, M. Alcañiz, and A. Pons. Telepsychology: Public speaking fear treatment on the internet. *Cyberpsychology & Behavior*, 3:959–968, 2000. 2.1.1

C. Botella, S.G. Hoffmann, and D.A. Moscovitch. A self-applied, internet-based intervention for fear of public speaking. *Journal of Clinical Psychology*, 60(8):821–830, 2004a. 2.1.1

C. Botella, J. Osma, A. Garcia Palacios, S. Quero, and R.M. Baños. Treatment of flying phobia using virtual reality: Data from a 1-year follow-up using a multiple baseline design. *Clinical Psychology & Psychotherapy*, 11(5):311–323, 2004b. 2.1

C. Botella, H. Villa, A. Garcia Palacios, S. Quero, R.M. Baños, and M. Alcañiz. The use of vr in the treatment of panic disorders and agoraphobia. *Studies in Health Technology and Informatics*, 99:73–90, 2004c. 2.5, 2.1.2, D

C. Botella, A. Garcia Palacios, H. Villa, R.M. Baños, S. Quero, M. Alcañiz, and G. Riva. Virtual reality exposure in the treatment of panic disorder and agoraphobia: A controlled study. *Clinical Psychology and Psychotherapy*, 14:164–175, 2007. 2.1.2

S. Bouchard, S. Côté, J. St.-Jacques, G. Robillard, and P. Renaud. Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3d games. *Technology and Health Care*, 14(1):19–27, 2006. 2.1

R. Boulic, B. Ulicny, and D. Thalmann. Versatile walk engine. *Journal of Game Development*, 1(1):29–43, 2004. 4.4.1

M. Bouvard and J. Cottraux. *Protocoles Et Echelles d'Evaluation En Psychiatrie Et En Psychologie*. Masson, Paris, France, 3 edition, 2002. 6.2.4.1

A.S. Carlin, H.G. Hoffman, and S. Weghorst. Virtual reality and tactile augmentation in the treatment of spider phobia: A case report. *Behaviour Research and Therapy*, 35(2): 153–158, 1997. 2.1

J. Cassell and K.R. Thorisson. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4):519–538, 1999. 1.1, 2.1.3.1

K.-J. Choi and H.-S. Ko. Online motion retargetting. *The Journal of Visualization and Computer Animation*, 11(5):223–235, 2000. 2.3

S. Chopra Khullar and N.I. Badler. Where to look? automating attending behaviors of virtual human characters. *Autonomous Agents and Multi-Agent Systems*, 4(1-2):9–23, 2001. 2.2

Rockwell Collins. `http://www.rockwellcollins.com/`, 2009. 6.2.1

C. Colombo, A. Del Bimbo, S. De Magistris, V. Cantoni, L. Lombardi, M. Mosconi, M. Savini, and A. Setti. Human-computer interaction based on eye movement tracking. In *CAMP '95: Proceedings of the Computer Architectures for Machine Perception*, pages 258–263, Washington, DC, USA, 1995. IEEE Computer Society. 2.1.3.1

N. Courty, E. Marchand, and B. Arnaldi. A new application for saliency maps: Synthetic vision of autonomous actors. In *Proceedings of International Conference on Image Processing*, volume 3, pages 1065–1068, 2003. 2.2

Dassault Systèmes. http://a2.media.3ds.com/products/3dvia/3dvia-virtools/, 2009. 2.1.1, 2.1.2

J. Difede and H.G. Hoffman. Virtual reality exposure therapy for world trade center post-traumatic stress disorder: A case report. *Cyberpsychology & Behavior*, 5(6):529–535, 2002. 2.1

J. Difede, H.G. Hoffman, and N. Jaysinghe. Innovative use of virtual reality technology in the treatment of ptsd in the aftermath of september 11. *Psychiatric Services*, 53(9): 1083–1085, 2002. 2.1

M.J. Doughty. Consideration of three types of spontaneous eyeblink activity in normal humans: During reading, and video display terminal use, in primary gaze, and while in conversation. *Optometry and Vision Science*, 78(10):712–725, 2001. 6.3.3.3

A.T. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, and Computers*, 4(1):455–470, November 2002. 2.1.3

H.E. Egeth and S. Yantis. Visual attention: Control, representation, and time course. *Annual Review of Psychology*, 48:269–297, 1997. 4.3

P.M. Emmelkamp, M. Bruynzeel, L. Drost, and C.A. Van Der Mast. Virtual reality treatment in acrophobia: A comparison with exposure in vivo. *Cyberpsychology & Behavior*, 4(3): 335–339, 2001. 2.1

P.M. Emmelkamp, M. Krijn, A.M. Hulsbosch, S. De Vries, M.J. Schuemie, and C.A. Van Der Mast. Virtual reality treatment versus exposure in vivo: A comparative evaluation in acrophobia. *Behavior Research and Therapy*, 40(5):509–516, 2002. 2.1

D. Freeman, K. Pugh, A. Antley, M. Slater, P. Bebbington, M. Gittins, G. Dunn, E. Kuipers, D. Fowler, and P. Garety. Virtual reality study of paranoid thinking in the general population. *British Journal of Psychiatry*, (192):258–263, 2008. 2.1.2

GarageGames. http://www.garagegames.com/, 2009. 2.1.2

M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *SIGCHI Conference of Human Factors in Computing Systems*, pages 529–536, 2003. 3.3

A. Garcia Palacios, H.G. Hoffman, S. Kwong See, A. Tsai, and C. Botella. Redefining therapeutic success with virtual reality exposure therapy. *Cybersychology & Behavior*, 4 (3):341–348, 2001. 2.1

M. Gillies. *Practical Behavioural Animation Based On Vision And Attention*. PhD thesis, University of Cambridge, 2001. 2.2

G.E. Grossman, R.J. Leigh, L.A. Abel, D.J. Lanska, and S.E. Thurston. Frequency and velocity of rotational head perturbations during locomotion. *Experimental Brain Research*, 70(3):470–476, 1988. 4.3, 4.4.2

E. Gu and N. Badler. Visual attention and eye gaze during multiparty conversations with distractors. *Lecture Notes in Computer Science*, pages 193–204, 2006. 2.8, 2.2, D

S.R. Harris, R.L. Kemmerling, and N.N. North. Brief virtual reality therapy for public speaking anxiety. *CyberPsychology & Behavior*, 5(6):543–550, 2002. 2.1.1

R.G. Heimberg. *Social Phobia - Diagnosis, Assessment, And Treatment*. Guilford Press, 1995. 1.2

B. Herbelin. *Virtual Reality Exposure Therapy For Social Phobia*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2005. 2.1.3.2, 2.6, 2.1.3.2, 3.1, 6.1, 6.2, D

B. Herbelin, F. Riquier, F. Vexo, and D. Thalmann. Virtual reality in cognitive behavioral therapy: A preliminary study on social anxiety disorder. In *VSMM02: 8th International Conference on Virtual Systems and Multimedia*, 2002. 2.1.1, 2.4, 6.2, D

B. Herbelin, H. Grillon, P. De Heras Ciechomski, and D. Thalmann. Coding gaze tracking data with chromatic gradients for vr exposure therapy. In *ICAT '07: Proceedings of the 17th International Conference on Artificial Reality and Telexistence*, pages 7–14, 2007. 3.1, 3.3

R.W. Hill. Perceptual attention in virtual humans: Towards realistic and believable gaze behaviors. In *Proceedings of Simulating Human Agents*, 2000. 2.2

L.F. Hodges, B. Rothbaum, R. Kooper, D. Opdyke, T. Meyer, M.M. North, J. De Graff, and J.S. Williford. Virtual environments for treating the fear of heights. *Computer Graphics and Applications*, 28(7):27–34, 1995. 2.1

S.G. Hofmann. *Self-Focused Exposure Therapy For Social Phobia. Unpublished Treatment Manual*. Boston University, Boston, USA, 1997. 6.2.1

S.G. Hofmann and M.W. Otto. *Cognitive Behavioral Therapy for Social Anxiety Disorder: Evidence-Based and Disorder-Specific Treatment Techniques*. Routledge, 2008. 6.2.1

C.S. Holt, R.G. Heimberg, and D.A. Hope. Situational domains of social phobia. *Journal of Anxiety Disorders*, 6:63–77, 1992. 3.2.1

K. Horley, L.M Williams, C. Gonsalvez, and E. Gordon. Social phobics do not see eye to eye: A visual scanpath study of emotional expression processing. *Journal of Anxiety Disorders*, 17:33–44, 2003. 2.1.3.2, 6.3, 6.3.3.1

K. Horley, L. Williams, C. Gonsalvez, and E. Gordon. Face to face: Visual scanpath evidence for abnormal processing of facial expressions in social phobia. *Psychiatry Research*, 127 (1-2):43–53, 2004. 2.1.3.2, 6.3, 7.4

T.E. Hutchinson, K.P. White, W.N. Martin, K.C. Reichert, and L.A. Frey. Human-computer interaction using eye-gaze input. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1527–1534, November-December 1989. 2.1.3.1

Intersense. `http://www.isense.com/`, 2009. 6.2.1

Iscan. http://www.iscaninc.com/, 2009. 2.1.3.2, 2.1.3.2, 5, 6.1

L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998. 2.2

L. Itti, N. Dhavale, and F. Pighin. Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proceedings of the Symposium on Optical Science and Technology*, volume 5200, pages 64–78, August 2003. 2.2

R.J.K. Jacob. What you look at is what you get: Eye movement-based interaction techniques. In *CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 11–18, New York, NY, USA, 1990. ACM. 2.1.3.1

L.K. James, C.Y. Lin, A. Steed, D. Swapp, and M. Slater. Social anxiety in virtual environments: Results of a pilot study. *CyberPsychology & Behavior*, 6(3):237–243, 2003. 2.4, 2.1.1, D

W. James. *The Principles Of Psychology*. Harvard University Press, 1983. 4

F.H. Kanfer. Verbal rate, eyeblink, and content in structured psychiatric interviews. *Journal of abnormal and social psychology*, 61:341–347, 1960. 6.3, 6.3.3.3

Y. Kim, R.W. Hill Jr, and D.R. Traum. A computational model of dynamic perceptual attention for virtual humans. In *Proceedings of Behavior Representation in Modeling and Simulation*, 2005. 2.2

E. Klinger, S. Roy, P. Légeron, I. Chemin, F. Lauer, and P. Nugues. Pre-demonstrator of the social phobia module and its associated clinical protocol. *European Commission IST-2000-25323*, 2002. 2.1.1

E. Klinger, S. Bouchard, P. Légeron, S. Roy, F. Lauer, I. Chemin, and P. Nugues. Virtual reality therapy versus cognitive behavior therapy for social phobia : A preliminary controlled study. *Cyberpsychology & Behavior*, 8(1):76–88, 2005. 2.3, 2.1.1, D

L. Kovar and M. Gleicher. Flexible automatic motion blending with registration curves. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 214–224, 2003. 2.3

L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. In *Proceedings of ACM SIGGRAPH, Annual Conference Series*, pages 473–482, 2002a. 2.3

L. Kovar, J. Schreiner, and M. Gleicher. Footskate cleanup for motion capture editing. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 97–104, 2002b. 2.3

J.J. Kuffner Jr and J.-C. Latombe. Fast synthetic vision, memory, and learning models for virtual humans. In *Proceedings of Computer Animation*, pages 118–127, 1999. 2.2, 2.7, D

R. Kulpa, F. Multon, and B. Arnaldi. Morphology-independent representation of motions for interactive human-like animation. In *Eurographics 2005*, volume 24, pages 343–352, 2005. 2.3

B. Lance and S. Marsella. Emotionally expressive head and body movement during gaze shifts. *Lecture Notes in Computer Science*, pages 72–85, 2007. 2.2

W.G.T. Lange, K.J. Tierney, A.H. Reinhardt-Rutland, and P. Vivekananda-Schmidt. Viewing behaviour of spider phobics and non-phobics in the presence of threat and safety stimuli. *British Journal of Clinical Psychology*, 43(3):235–243, 2004. 2.1.3.2

B. Le Callennec and R. Boulic. Interactive motion deformation with prioritized constraints. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 163–171, 2004. 2.3

J. Lee and S.Y. Shin. A hierarchical approach to interactive motion editing for human-like figures. In *Proceedings of ACM SIGGRAPH, Annual Conference Series*, pages 39–48, 1999. 2.3, 4.4.1

J. Lee, J. Chai, P.S.A. Reitsma, J.K. Hodgins, and N.S. Pollard. Interactive Control Of Avatars Animated With Human Motion Data. In *Proceedings of ACM SIGGRAPH, Annual Conference Series*, pages 491–500, jul 2002a. 2.3

J.M. Lee, D.P. Jang, D.H. Kim, Y.H. Choi, I.Y. Kim, and S.I. Kim. Virtual reality system for treatment of the fear of public speaking using image-based rendering and moving pictures. *Cyberpsychology & Behavior*, 5(3):191–195, 2002b. 2.1.1

S.-H. Lee and D. Terzopoulos. Heads up!: Biomechanical modeling and neuromuscular control of the neck. In *Proceedings of ACM SIGGRAPH, Annual Conference Series*, pages 1188–1198, 2006. 2.10, 2.3, 4.4.2, D

S.P. Lee, J.B. Badler, and N.I. Badler. Eyes alive. In *Proceedings of ACM SIGGRAPH, Annual Conference Series*, pages 637–644, 2002c. 2.9, 2.3, 5.3, D

M.R. Liebowitz. Social phobia. *Modern Problems of Pharmacopsychiatry*, 22:141–173, 1987. 6.3.2.2

R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 22(140): 1–55, 1932. 6.4.1, 7.3

J. Maïm, B. Yersin, J. Pettré, and D. Thalmann. Yaq: An architecture for real-time navigation and rendering. *IEEE Computer Graphics & Applications Special Issue on Virtual Populace*, 29(4), 2009. 3.5.1, 5.1

I.M. Marks. *Fears, Phobias, And Rituals*. Oxford University Press, 1987. 1.1, 7.4

I.M. Marks and A.M. Matthews. Brief standard self-rating for phobic patients. *Behavior Research and Therapy*, 17:263–267, 1979. 6.2.1

A. Mühlberger, M.J. Wieser, and P. Pauli. Visual attention during virtual social situations depends on social anxiety. *Cyberpsychology & Behavior*, 11(4):425–430, 2008. 2.1.3.2, 2.6, D

K. Moore, B.K. Wiederhold, M.D. Wiederhold, and G. Riva. Panic and agoraphobia in a virtual world. *Cyberpsychology & Behavior*, 5(3):197–202, 2002. 2.1.2

M. Mori. Bukimi no tani (the uncanny valley. *Energy*, 7(4):33–35, 1970. 3.3

U. Neisser. *Cognitive Psychology*. Appleton-Century-Crofts, 1967. 4.3

M. North, S. North, and J.R. Coble. Virtual reality therapy: An effective treatment for the fear of public speaking. *International Journal of Virtual Reality*, 3(2):2–6, 1998. 1.1, 2.1.1, 2.1.2

M.M. North, S.M. North, and J.R. Coble. Effectiveness of virtual environment desensitization in the treatment of agoraphobia. *Presence: Teleoperators and Virtual Environments*, 5(3):346–352, 1996. 2.1.2

S.M. North and J.R. Coble. Virtual environment psychotherapy: A case study of fear of flying disorder. *Presence, Teleoperators and Virtual Environments*, 6(1):87–105, 1997. 2.1

L.M. Optican. Adaptive properties of the saccadic system. *Reviews of Oculomotor Research*, 1:71–79, 1985. 4.3

W. Peñate, C.T. Pitti, J.M. Bethencourt, J. de la Fuente, and R. Gracia. The effects of a treatment based on the use of virtual reality exposure and cognitive-behavioral therapy applied to patients with agoraphobia. *International Journal of Clinical and Health Psychology*, 8 (1):5–22, 2007. 2.1.2

M. Peinado, D. Meziat, D. Maupu, D. Raunhardt, D. Thalmann, and R. Boulic. Accurate on-line avatar control with collision anticipation. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pages 89–97, 2007. 4.5.3

D. Pertaub, M. Slater, and C. Barker. An experiment on fear of public speaking in virtual reality. *Studies in Health Technology & Informatics*, 81:372–378, 2001. 2.1.1

D. Pertaub, M. Slater, and C. Barker. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments*, 11(1):68–78, 2002. 2.1.1

A. Peternier, S. Cardin, F. Vexo, and D. Thalmann. Practical design and implementation of a cave environment. In *2nd International Conference on Computer Graphics Theory*, pages 129–136, 2007. 7.1

C. Peters and C. O'Sullivan. Synthetic vision and memory for autonomous virtual humans. *Computer Graphics Forum*, 21(4):743–752, 2002. 2.2, 2.7, 2.2, D

C. Peters and C. O'Sullivan. Bottom-up visual attention for virtual human animation. In *Proceedings of Computer Animation and Social Agents*, pages 111–117, 2003. 2.2

C. Peters, C. Pelachaud, E. Bevacqua, M. Mancini, and I. Poggi. A model of attention and interest using gaze behavior. *Lecture Notes in Computer Science*, pages 229–240, 2005. 2.2

R.J. Peters and L. Itti. Computational mechanisms for gaze direction in interactive visual environments. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 27–32, 2006. 2.2

Phasespace. `http://www.phasespace.com/`, 2009. 5, 5.5, 7.1

R. Ramloll, C. Trepagnier, M. Sebrechts, and A. Finkelmeyer. A gaze contingent environment for fostering social attention in autistic children. In *ETRA '04: Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, pages 19–26, New York, NY, USA, 2004. ACM Press. ISBN 1-58113-825-3. doi: http://doi.acm.org/10.1145/968363.968367. 2.1.3.2

S.A. Rathus. A 30-item schedule for assessing assertive behavior. *Behavior Therapy*, 4: 398–406, 1973. 6.3.2.2

G.M. Reger, G.A. Gahm, A.A. Rizzo, R. Swanson, and S. Duma. Soldier evaluation of the virtual reality iraq. *Telemedicine and e-Health*, 15(1):101–104, 2009. 2.1

F. Riquier, M. Stankovic, and A.F. Chevalley. Virtual gazes for social exposure: Margot and snow white. In *VRMHR 2002: First International Workshop on Virtual Reality Rehabilitation*, pages 35–49, 2002. 6.2

A.A. Rizzo, J. Pair, P.J. McNerney, E. Eastlund, B. Manson, J. Gratch, R. Hill, M. Roy, and B. Swartout. Design and development of a vr therapy application for iraq war veterans with ptsd. *Studies in Health Technology and Informatics*, 111:407–413, 2005. 2.1

R.G. Ross, S. Heinlein, G.O. Zerbe, and A. Radant. Saccadic eye movement task identifies cognitive deficits in children with schizophrenia, but not in unaffected child relatives. *Journal of Child Psychology and Psychiatry*, 46(12):1354–1362, 2005. 2.1.3.2

B.O. Rothbaum, L.F. Hodges, R. Kooper, D. Opdyke, J.S. Williford, and M.M. North. Virtual reality graded exposure in the treatment of acrophobia: A case report. *Behavior Therapy*, 26(3):547–554, 1995. 2.1

B.O. Rothbaum, L.F. Hodges, B.A. Watson, C.D. Kessler, and D. Opdyke. Virtual reality exposure therapy in the treatment of fear of flying: A case report. *Behaviour Research and Therapy*, 34(5–6):477–481, 1996. 2.1

B.O. Rothbaum, L.F. Hodges, R.D. Alarcon, D. Ready, F. Shahar, K. Graap, J. Pair, P. Hebert, D. Gotz, B. Wills, and D. Baltzell. Virtual reality exposure therapy for ptsd vietnam veterans : A case study. *Journal of Traumatic Stress*, 12(2):263–271, 1999. 2.1

B.O. Rothbaum, L.F. Hodges, S. Smith, J.H. Lee, and L. Price. A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology*, 68(6):1020–1026, 2000. 2.1

B.O. Rothbaum, L.F. Hodges, D. Ready, K. Graap, and R.D. Alarcon. Virtual reality exposure therapy for vietnam veterans with posttraumatic stress disorder. *Journal of Clinical Psychiatry*, 62(8):617–622, 2001. 2.1

B.O. Rothbaum, L.F. Hodges, P.L. Anderson, L. Price, and S. Smith. Twelve-month follow-up of virtual reality and standard exposure therapies for the fear of flying. *Journal of Consulting and Clinical Psychology*, 70(2):428–432, 2002. 2.1

S. Roy, E. Klinger, P. Légeron, F. Lauer, I. Chemin, and P. Nugues. Definition of a vr-based protocol to treat social phobia. *Cyberpsychology & Behavior*, 6(4):411–420, 2003. 2.1.1

H.R. Schiffman. *Sensation And Perception: An Integrated Approach*. John Wiley and Sons, Inc., New York, USA, fifth edition, 2001. 6.3.3.3, 6.3.3.3

D.V. Sheehan, Y. Lecrubier, K.H. Sheehan, P. Amorim, J. Janavs, E. Weiller, T. Hergueta, R. Baker, and G.C. Dunbar. The mini-international neuropsychiatric interview (m.i.n.i.): The development and validation of a structured diagnostic psychiatric interview for dsm-iv and icd-10. *Journal of Clinical Psychiatry*, 59:22–33, 1998. 6.2.2, 6.3.1

H.J. Shin, J. Lee, S.Y. Shin, and M. Gleicher. Computer Puppetry: An Importance-Based Approach. *ACM Transactions on Graphics*, 20:67–94, 2001. 2.3

SICS. `http://www.sics.se/dive/`, 2009. 2.1.1

M. Slater, D.P. Pertaub, and A. Steed. Public speaking in virtual reality: Facing an audience of avatars. *Computer Graphics & Applications*, 19(2):6–9, 1999. 2.1.1, 2.3, 2.1.1, D

M. Slater, D.P. Pertaub, C. Barker, and D. Clark. An experimental study on fear of public speaking using a virtual environment. In *3rd International Workshop on Virtual Rehabilitation*, 2004. 2.1.1

M. Slater, A. Antley, A. Davison, D. Swapp, C. Guger, C. Barker, N. Pistrang, and M.V. Sanchez-Vives. A virtual reprise of the stanley milgram obedience experiments. *Plos One*, 1(1), 2006. 2.1.2

J.D. Smith. Social anxiety and selective attention: A test of the vigilance-avoidance model. Master's thesis, Florida State University, 2005. 2.1.3.2

J. Spira, J.M. Pyne, B. Wiederhold, M. Wiederhold, K. Graap, and A.A. Rizzo. Virtual reality and other experiential therapies for combat-related ptsd. *Primary Psychiatry*, 13 (3):58–64, 2006. 2.1

I. Starker and R.A. Bolt. A gaze-responsive self-disclosing display. In *CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3–10, New York, NY, USA, 1990. ACM. 2.1.3.1

D.N. Stern, N. Bruschwiler-Stern, A.M. Harrison, K. Lyons-Ruth, A.C. Morgan, J.P. Nahum, L. Sander, and E.Z. Tronick. The process of therapeutic change involving implicit knowledge: Some implications of developmental observations for adult psychotherapy. *Infant Mental Health Journal*, 19(3):300–308, 1999. 6.3, 6.3.2.2

M. Sung, M. Gleicher, and S. Chenney. Scalable behaviors for crowd simulation. *Computer Graphics Forum*, 23(3):519–528, 2004. 4.3

superscape. `http://www.superscape.com/`, 2009. 2.1, 2.1.2

V. Tanriverdi and R.J.K. Jacob. Interacting with eye movements in virtual environments. In *CHI '00: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 265–272, New York, NY, USA, 2000. ACM. 2.1.3.1

D. Tolani, A. Goswami, and N.I. Badler. Real-time inverse kinematics techniques for anthropomorphic limbs. *Graphical Models*, 62(5):353–388, 2000. 2.3

A. Treuille, Y. Lee, and Z. Popović. Near-optimal character animation with continuous control. *ACM Transactions on Graphics*, 26(3):7, 2007. 4.2.2

L.R. Valmaggia, D. Freeman, C. Green, P. Garety, D. Swapp, A. Antley, C. Prescott, D. Fowler, E. Kuipers, P. Bebbington, M. Slater, M. Broome, and P.K. McGuire. Virtual reality and paranoid ideations in people with "at-risk mental state" for psychosis. *British Journal of Psychiatry*, (51):63–68, 2007. 2.1.2

R. van der Pol. Using gesture recognition to interact with virtual humans in a cave environment. Master's thesis, Universiteit Utrecht, 2009. 5, 5.5, 7.1

vepsy. `http://www.vepsy.com/`, 2009. 2.1.1, 2.1.2

J.C. Verlinden, D. Kessler, and L. Hodges. The simple virtual environment library. Technical Report GIT-GVU-93-24, Georgia Institute of Technology - Graphics, Visualization and Usability Center, June 1993. 2.1

H. Villa Martin, C. Botella, A. Garcia Palacios, and J. Osma. Virtual reality exposure in the treatment of panic disorder with agoraphobia: A case study. *Cognitive and Behavioral Practice*, 14:58–69, 2007. 2.5, 2.1.2, D

F. Vincelli, L. Anolli, S. Bouchard, B.K. Wiederhold, V. Zurloni, and G. Riva. Experiential cognitive therapy in the treatment of panic disorders with agoraphobia: A controlled study. *Cyberpsychology & Behavior*, 6(3):321–328, 2003. 2.1.2

VirtuallyBetter. `http://www.virtuallybetter.com/`, 2009. 2.1, D

VREAM. `http://www.vream.com/`, 2009. 2.1.1

vrphobia. `http://www.vrphobia.com/`, 2009. 2.2, D

H. Wang, M. Chignell, and M. Ishizuka. Empathic tutoring software agents using real-time eye tracking. In *ETRA '06: Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, pages 73–78, New York, NY, USA, 2006. ACM. 2.1.3.1

WHO. *The ICD-10 Classification Of Mental And Behavioural Disorders: Diagnostic Criteria For Research*. World Health Organization, Geneva, Switzerland, 1993. 2.1.2

B.K. Wiederhold, R. Gervirtz, and M.D. Wiederhold. Fear of flying: A case report using virtual reality therapy with psychological monitoring. *Cyberpsychology and Behavior*, 1: 97–103, 1998. 2.1

B.K. Wiederhold, D.P. Jang, R.G. Gevirtz, S.I. Kim, I.Y. Kim, and M.D. Wiederhold. The treatment of fear of flying: A controlled study of imaginal and virtual reality graded exposure therapy. *IEEE Transactions of Information Technology in Biomedicine*, 6(3):218–223, 2002. 2.1

J.M. Wolfe and T.S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6):495–501, 2004. 4.3

J. Wolpe. *The Practice Of Behavior Therapy*. Pergamon Press, 1969. 1.1

S. Yantis and J. Jonides. Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, 16(1):121–134, 1990. 4.3

S.N. Yao, J. Cottraux, E. Mollard, E. Albuisson, I. Note, I. Jalenques, F. Fanget, P. Robbe-Grillet, M. Dechassey, N. Ektmedjian, M. Bouvard, and A.J. Coudert. The french version of the social interaction self-statement test (sisst): A validation and sensitivity study in social phobics. *Behavioral and Cognitive Psychotherapy*, 26:247–259, 1998. 6.2.1, 6.3.2.2

S.N. Yao, I. Note, F. Fanget, E. Albuisson, M. Bouvard, I. Jalenques, and J. Cottraux. L'anxiété sociale chez les phobiques sociaux: Validation de l'echelle d'anxiété sociale de liebowitz. *L'encéphale*, 25:429–435, 1999. 6.2.1

Q. Yu and D. Terzopoulos. A decision network framework for the behavioral animation of virtual humans. In *SCA '07: Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 119–128, 2007. 2.2

Z. Zhu and Q. Ji. Eye and gaze tracking for interactive graphic display. *Machine Vision and Applications*, 15(3):139–148, July 2004. 2.1.3.1

Helena Grillon
Chemin de la Roche 1
1020 Renens
Tel : +41/76.398.30.33
helena.grillon@gmail.com
http://vrlab.epfl.ch/~helena

Born on July 18th 1979
Swiss national
Single

Education

PhD in Computer Sciences
VRLab, EPFL, Lausanne
Title : Simulating Interactions with Virtual Characters for the Treatment of Social Phobia

July 2005 – June 2009
Creation of virtual characters and environments, development of a gaze attention simulation model, experimental validations on social phobic patients, use of motion capture, inertial sensors and eye-tracking

Masters in Visualization and Infographic Communication
EPFL, Lausanne

October 2003 – May 2005
Creation of a 3D airport – modeling with Discreet® 3ds Max® 6, texturing with Adobe® Photoshop® CS

Bachelors in Information Systems And Communication
University of Geneva

October 1999 – June 2003
Creation of a computer graphics clip of 4 minutes – Discreet® 3ds Max® 5, Adobe® Premiere® 6.0, Adobe® Photoshop® 6.0

Professional Activities

VRLab, EPFL
Research assistant

July 2005 – July 2009
Assistant for a first year C++ course, supervision of student semester and master projects

Mercer Human Resource Consulting, Geneva

December 2000 – October 2003
Employee in technical pension fund administration

Institut Victor Ebner, Geneva
English educator

March 2000 – June 2001
Private and group courses, preparation to the FCE

Bridge Information Systems SA, Geneva, Secretary - receptionist

April 1999 – November 1999

Young & Rubicam, Moscow
Intern

July 1997
Help in developing a slogan for a Philip Morris campaign

Atelier d'architecture Alain Meylan, Geneva, Intern

July 1995
Drawing of plans, archiving

Languages

Oral expression, written expression, European language standard evaluation

French, mother tongue
English, excellent, excellent, C2
German, good notions, good notions, B1
Spanish, intermediate notions, intermediate notions, A2

Language training courses

6 weeks in Germany
6 weeks in Austria

Computer Graphics Qualifications

3D Programming Gaze attention model	C++ development with Microsoft® Visual Studio® 2005
Computer graphics & animation Virtual humans and environments	Modeling with Autodesk® 3ds Max® 8, rigging with Kinetix Character Studio™ 4 and Digimation Bones Pro 3, animation with Autodesk® MotionBuilder™ 7.5
Design Textures and posters	Adope® Photoshop® CS3, Adobe® Illustrator® CS2
Other	Phasespace motion capture, ISCAN® eye-tracking, InterSense inertial sensors

Hobbies

Music	11 years of piano at the *Associated Boards of the Royal Schools of Music* - Grade 7
Sports	Ski, Tennis
Reading	Detective books, S. King, S. Larsson, E. George

Awards

Best student paper award for *Eye Contact as Trigger for Modification of Virtual Character Behavior*	**September 2008** Award sponsored by GestureTek® **October 2008** Award granted by EPFL

Reference

Prof. Daniel Thalmann
Directeur de thèse
EPFL IC ISIM VRLAB Station 14
Ch-1015 Lausanne
+41/21.693.52.14
daniel.thalmann@epfl.ch

Publications

Journals

[1] H. Grillon and D. Thalmann. *Simulating Gaze Attention Behaviors for Crowds.* Computer Animation and Virtual Worlds, 20, pp. 111-119, 2009.

[2] H. Grillon, B. Yersin, J. Maïm and D. Thalmann. *Gaze Behaviors for Virtual Crowd Characters.* Lecture Notes in Computer Science, 2009.

[3] H. Grillon, F. Riquier, B. Herbelin and D. Thalmann. *Virtual Reality as Therapeutic Tool in the Confines of Social Anxiety Disorder Treatment.* International Journal on Disability and Human Development, 5(3), pp. 243-250, 2006.

[4] B. Herbelin, P. Benzaki, F. Riquier, O. Renault, H. Grillon and D. Thalmann. *Using Physiological Measures for Emotional Assessment: A Computer-Aided Tool for Cognitive and Behavioral Therapy.* International Journal on Disability and Human Development, 4(4), pp. 269-277, 2005.

Conferences

[5] H. Grillon, F. Riquier and D. Thalmann. *Eye Contact for Interaction in Virtual Reality Exposure Therapy for Social Phobia: A Case Study.* In Proceedings of Virtual Reality International Conference, pp. 175-181, 2009.

[6] H. Grillon and D. Thalmann. *Eye Contact as Trigger for Modification of Virtual Character Behavior.* In Proceedings of IEEE Virtual Rehabilitation, pp. 205-211, 2008. **Best student paper award**.

[7] H. Grillon, F. Riquier and D. Thalmann. *Eye-tracking as Diagnosis and Assessment Tool for Social Phobia.* In Proceedings of IEEE Virtual Rehabilitation, p. 138-145, 2007.

[8] J. Pettré, H. Grillon and D. Thalmann. *Crowds of Moving Objects: Navigation Planning and Simulation.* In Proceedings of IEEE International Conference on Robotics and Automation, pp. 3062-3067, 2007.

[9] B. Herbelin, H. Grillon, P. De Heras Ciechomski and D. Thalmann. *Coding Gaze Tracking Data with Chromatic Gradients for VR Exposure Therapy.* In Proceedings of IEEE International Conference on Artificial Reality and Telexistence, pp. 7-14, 2007.

[10] H. Grillon, F. Riquier, B. Herbelin and D. Thalmann. *Use of Virtual Reality as Therapeutic Tool for Behvioural Exposure in the Ambit of Social Anxiety Disorder Treatment.* In Proceedings of International Conference on Disability, Virtual Reality and Associated Technologies, pp. 105-112, 2006.

Workshops

[11] H. Grillon. *Serious Games for Virtual Reality Exposure Therapy.* Workshop on Challenges in Serious Games, Villars-sur-Ollon, March 2-3, 2009.

[12] H. Grillon. *Eye-tracking for Feedback in Virtual Character Behavior.* Workshop on Data Acquisition for 3D Simulation, Zermatt, December 6-8, 2007.