# VIRTUAL REALITY EXPOSURE THERAPY FOR SOCIAL PHOBIA

THÈSE N$^{O}$ 3351 (2005)

PRÉSENTÉE À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS

Institut des systèmes informatiques et multimédias

SECTION D'INFORMATIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Bruno HERBELIN

DEA en informatique, Université Louis Pasteur, Strasbourg, France
et de nationalité française

acceptée sur proposition du jury:

Prof. D. Thalmann, directeur de thèse
Prof. O. Blanke, rapporteur
Dr D. Mestre, rapporteur
Dr F. Riquier, rapporteur

Lausanne, EPFL
2005

*Dedicated to*

Gerard Herbelin
Pierre Moquet

# Abstract

This thesis presents researches and experiments performed in collaboration with a psychiatrist in order to validate and improve the use of virtual reality in social phobia psychotherapy.

Cognitive and behavioral therapies are strongly based on the exposure to anxiety provoking stimuli. Virtual reality seems to be appropriate for such exposures as it allows for on-demand reproduction of reality. The idea has been validated for the treatment of various phobias but is more delicate in the case of social phobia; whereas the sense of presence provoked by the immersion in a virtual environment supports the emergence of fears linked to a location, we had to verify that we can reproduce social phobia related anxiety-provoking stimuli by simulating virtual humans.

Therefore, and in order to provide therapists with an efficient virtual reality system dedicated to the exposure to social situations, we have developed software solutions supporting different immersion setups and enabling realistic simulations of inhabited virtual environments. We have experimented with public speaking scenarios within a preliminary study, three clinical case studies and a validation study on 200 subjects. We have been able to confirm that our virtual reality platform fulfilled therapeutic exposure requirements for social phobia.

Moreover, we have been able to show that virtual reality exposure has additional advantages such as the possibility to improve clinical assessment with embedded monitoring tools. Our experiments with physiological measurements and eye tracking technology during immersion leaded to the validation of systems for objective and reliable assessment of patients' safety behaviors. The observation of such phobic reactions has confirmed the simulation impact and may provide therapists with enhanced pathological progression monitoring.

During our experiments, we have also been able to observe that subjects' reactions during immersion were so much influenced by their sensitivity to fearful stimuli that their cognitive reactions were 'overloaded' by the arousal of anxiety and emotions. This has allowed us to consider that the sense of presence was more importantly related to the subjective impact of the content than to the technological process.

# Version abrégée

Cette thèse présente les recherches et les développements menés en collaboration avec un psychiatre pour valider et améliorer l'utilisation de la réalite virtuelle dans la psychothérapie de la phobie sociale.

Les thérapies comportementales et cognitives se basent fortement sur l'exposition à des stimuli générateurs d'anxiété. La réalite virtuelle semble appropriée pour de telles expositions car elle permet de reproduire la réalite à la demande. L'idée fut validée pour traiter différentes phobies mais le cas de l'anxiété sociale est plus délicat; alors que le sentiment de présence provoqué par l'immersion en environnement virtuel permet l'emergence des peurs liées à un lieu, nous avons du vérifier qu'il est possible de reproduire les stimuli anxiogènes de la phobie sociale en simulant des humains virtuels.

Dans ce but, et pour fournir aux thérapeutes un système de réalité virtuelle dédié à l'exposition en situations sociales, nous avons développé des solutions logicielles sur différents équipements d'immersion et permettant la simulation réaliste d'environnements virtuels habités. Nous avons ensuite expérimenté des scenarios virtuels de prise de parole en public lors d'une étude preliminaire, de trois cas d'étude clinique, et d'une étude sur 200 personnes. Nous avons pu confirmer que notre plateforme de réalité virtuelle remplissait les conditions d'utilisation thérapeutique.

De plus, nous avons pu montrer que l'exposition par réalité virtuelle avait d'autres avantages tels que la possibilité d'améliorer l'évaluation clinique en intégrant des outils de mesure. Nos expériences sur l'enregistrement de données physiologiques et l'intégration de technologies de suivi du regard pendant l'immersion nous ont permis de valider des systèmes permettant une observation objective et sûre des comportements de réassurance chez les patients. L'observation de telles reactions phobiques a permis de confirmer l'impact de la simulation et pourrait offrir aux therapeutes un meilleur suivi de l'évolution de la pathologie.

Lors de ces expériences, nous avons aussi pu observer que les réactions des sujets pendant l'immersion etaient si fortement influencées par leur sensibilité aux stimuli que leurs reactions cognitives etaient submergés par la montée des emotions et de l'anxiété. Nous avons donc pu considérer que le sentiment de présence etait alors plus lié à l'impact subjectif du contenu qu'au procédé technologique.

# Acknowledgements

This work was carried out at the Virtual Reality Laboratory of the Ecole Polytechnique Federale de Lausanne under the direction and thanks to the motivation of my supervisor Prof. Dr. Daniel Thalmann. I am deeply indebted for his invaluable support and his confidence in my research work.

I wish to express my gratitude to Dr. Francoise Riquier who made this research possible and who generously shared her knowledge and ideas. Special thanks goes to her colleages Anne Francoise Chevaley and Miroslava Stankovic for their collaboration on clinical studies.

I want to thank the senior researchers Ronan Boulic, Frederic Vexo and Julien Pettre for their guidance in all the time of research and writing of this thesis.

I would like to express my gratitude to all my colleages for many discussions and friendship; Mireille Clavien and Stephanie Noveraz for sharing their creativity and motivation; Jan Ciger, Sebastien Schertenleib, Pablo De Heras Ciechomsky, Tolga Abaci, Benoit Le Callennec and Pascal Glardon for sharing common research and personnal interrests; Olivier Renaud, Etienne De Sevin, Anderson Maciel, Patrick Lemoine and Sofiane Sarni for sharing ideas and good mood. I am also grateful for the support of my former colleages Michal Ponder, Tom Molet, Branislav Ulicny, and Luc Emering.

Josiane Bottarelli and Zerrin Celebi deserve my warmest thanks for their patience and administrative support.

I wish good luck to Helena Grillon and to new assistants for their research.

My brain was opened to virtual reality for the first time in 1997 by Ivan Chabanaud; as usual, he left the door open!

I am so grateful for the love and support of Myriam, Margot and my familly.

# List of abbreviations

| | |
|---|---|
| **BiP** | Break in Presence |
| **CBT** | Cognitive and behavioral therapy |
| **GUI** | Graphical user interface |
| **IVE** | Immersive virtual environment |
| **LSAS** | Liebowitz social anxiety scale |
| **SAD** | Social Anxiety Disorder |
| **SP** | Social Phobia |
| **SoP** | Sense of presence |
| **VE** | Virtual environment |
| **VH** | Virtual human |
| **VR** | Virtual reality |
| **VRE** | Virtual reality exposure |
| **VRET** | Virtual reality exposure therapy |

# Contents

# Chapter 1

# Introduction

In order to simulate an artificial reality, VR pioneers have created specific hardware capable of stimulating human senses in a similar way than in reality. One could have argue that only an accurate reproduction of reality could actually cause the illusion of an artificial world. However, what discovered these pioneers was much more encouraging. The experiments from Ivan Sutherland or Jaron Lanier for instance have shown that, despite the low quality of the synthetic graphics and the cumbersomeness of the devices in the sixties, our brain was compensating for the lack of precision and quite flexible regarding the potential emergence of a subjective feeling of being in another place. Thanks to these new technologies, a virtually mediated *presence* became not only a reality but also freely reproducible. Therefore, in parallel to developments related to the stimulation of the senses (computer graphics, acoustics, haptics), researchers in virtual reality tried to explain how someone can actually feel the existence of an object although its reproduction is of low quality, or how our brain can adapt to unrealistic paradigms of interaction or deal with contradictory perceptions. In an effort to propose definitions, virtual presence has for example been has been compared to illusions (optical, acoustic, and proprioceptive) or associated to the ability to escape from reality when reading books or watching movies.

However, there is still no consensus today on the meaning and the evaluation of presence. Since the complexity of the problem stands in the understanding of the human mind, this research involved more and more collaborations with experts in psychology and human sciences. Recently, this tendency was even formally expressed at a European level during the preparation meeting for IST 6th Framework Programme for research projects on *Presence and Interaction in Mixed-Reality Environments* (FET information event[1], Brussels, Jan. 13th 2005): “Research on presence is at the intersection of technology, neuroscience and psychology” said Mel Slater, invited as public speaker for his expertise on VR and presence.

---

[1] http://www.cordis.lu/ist/fet/pr.htm

Another example is given by the ENACTIVE[2] European network of excellence which currently involves laboratories in robotics, digital media, virtual reality, biomechanical physics but also psychology, social sciences, music, and more, in order to conduct joint researches on a new generation of human-computer interfaces based on the proactive involvement of human body in mediated tasks.

Following a similar approach, the research presented in this thesis proposes an interdisciplinary work between VR and psychiatry. Within a substantial computer science background involving graphics, behavioral animation and human-computer interactions, we have been able to investigate the psycho-therapeutic effect of immersion in virtual environments. More specifically, as our collaboration with mental health professionals required a strong medical framework, we focused on the analysis, design and improvement of virtual reality exposure techniques for the cognitive and behavioral therapy of social anxiety disorders.

## 1.1 Motivations and objectives

Also known as cybertherapy, the emerging field of computer-supported rehabilitation regroups a large variety of techniques exploring the new possibilities offered by Internet, multimedia systems and virtual reality to provide better and/or more flexible therapies. Its application domain in health care covers medicine, psychology and neuroscience. Concerning mental health, the possibilities offered by simulations have been investigated about ten years ago to determine if the immersion into virtual environments could support the cognitive and behavioral therapy (CBT) of anxiety disorders. The principle of virtual reality exposure therapy (VRET) is to replace the usual in-vivo exposure to the anxiety provoking situation with a mediated experience. Exposures are behavioral exercises directed by therapists and involving a confrontation of the patient to the fearful stimuli (e.g. a spider for arachnophobia).

Following the firsts experiments' success with people suffering from disorders such as fear of flying, acrophobia or post traumatic stress disorders, this area of research quickly expanded and involved more and more collaborations between VR and psychiatry. However, the simulation of such situations and the immersion of patients involving complex simulation engines and specific VR devices, these systems remain experimental and the applications limited. There is a real need for adapted and versatile VR systems that could be applicable to a larger scope of pathologies, but the current systems have already shown their limits. Typically, as the latest experiments on therapeutic exposure for social phobia are promising, there are new expectations from the medical side to integrate social interactions into VRET. However, the simulation of virtual humans being a complex and heterogeneous problem, the current systems can not cover all these aspects together yet: optimized but believable

---

[2] http://www.enactivenetwork.org

rendering, dynamic character animation, programming of natural behaviors, and interaction with the patient.

Our research work therefore first consisted in the development of and the experimentation with systems for the simulation of inhabited virtual environments in order to apply exposure therapy to the treatment of social phobia. Our primary objective was to confirm that virtual reality exposure was efficient in this case and to propose an integrated software solution to this end.

The second objective was to determine the benefits of the VR approach in comparison to the more classical in-vivo, in imagination or mediated solutions. Our goal was not only to compare its usability or efficiency, but to determine where software engineering could potentially influence or even change the therapeutic process. To do so, one direction of research we have choosen was the development of computerized tools for clinical assessment which could improve therapy.

## 1.2 General approach

Firstly, as our research had to be validated and confronted to clinical usage, the collaboration with a therapist was fundamental. This provided us with a medical framework for our VR experiments and allowed for the development of clinical applications.

Secondly, since the elaboration of a VR platform and the simulation of virtual humans involve multiple domains which are also at a research stage, we opted for a collaborative approach based on the development of a component-based framework. This optimized the development phase and extended the opportunities of further improvements.

Finally, we deeply relied on experimentation with interaction and assessment devices during our developments. This allowed us to better explore the potential usage of VR hardware and computerized tools in virtual reality exposure therapy and to propose novel approaches to VRET.

## 1.3 Document structure

This document is organized as follows. In chapter 2, we present a general overview of virtual reality in psychotherapy. We first clarify the basic concepts involved in virtual reality exposure therapy before recounting the evolution of this research activity in the past ten years. This then progressively leads to social anxiety disorders therapy and to more precise objectives. Connections are then made with the problematic of presence and social presence.

Chapter 3 presents the VR platform we designed and experimented with by successively consider the hardware and the software aspects. The former includes the design of a device abstraction solution and the proposition of two immersion solutions. The latter covers the programming of a flexible component-based real time engine dedicated to the simulation of inhabited virtual environments.

Chapter 4 describes the developments and experimentation we did with physiological and eye-tracking devices to provide a computerized and VR-integrated clinical assessment. Our research on the identification of affective states with physiological measurements was intended to monitor the internal arousal of a patient when exposed whereas the experimental tracking of behavioral gaze avoidance was designed specifically for immersion to provide therapists and patients with an objective observation of unconscious eye behaviors.

Chapter 5 relates the three main experiments we conducted. The first one was a preliminary study intended to establish the basis for our collaboration with the therapist partner. The second one consisted in three clinical case studies regrouped to synthesize the expertise we gained during exposure sessions with social phobic patients. The last one was a validation study on a large cohort to confirm our expectations and the reliability of our system.

Before concluding, we present in chapter 6 a global synthesis in order to answer to our objectives —the ability of VR to replace exposure for social phobia and the improvements in clinical assessment. We will also show more advanced simulations and control modes for exposure which may be useful in the management of a therapy.

# Chapter 2

# Virtual reality in psychotherapy

This chapter is an introduction to the use of virtual reality in psychotherapy. After a short presentation of the principles of immersion in virtual environments, we will emphasize on the question of Presence in a virtual environment because this concept is central in the development of virtual reality exposure therapy. An overview of the emergence of this field will then lead us to an analysis of the benefits of this particular technique before entering deeper in the treatment of social anxiety disorders.

## 2.1 Virtual reality

The word 'virtual reality' was originally made-up to describe the experiments in the simulation of artificially generated elements that could fool the human perception. This story started with Sutherland's first virtual world in 1965 and his invention of the head mounted displays (HMD) soon after in 1968. Since then, VR systems have been developed and used for various applications. Although we cannot retrace the whole history of VR here, there are some particularities in the technology for performing truly immersive simulations that should be described in order to better understand the added value of VR for therapy.

### 2.1.1 Immersion in virtual environments

#### Immersion

The principle of VR is based on the human perception-action loop (fig.2.1). The mediation of human perceptions (outputs) and reactions (inputs) through computerized models is a necessary (but not sufficient) condition to make the user feel that a synthetic object/place actually exists. Outputs should cover multiple senses –visual, auditive and haptic being the most important. The role of inputs is to provide the computer with necessary information to

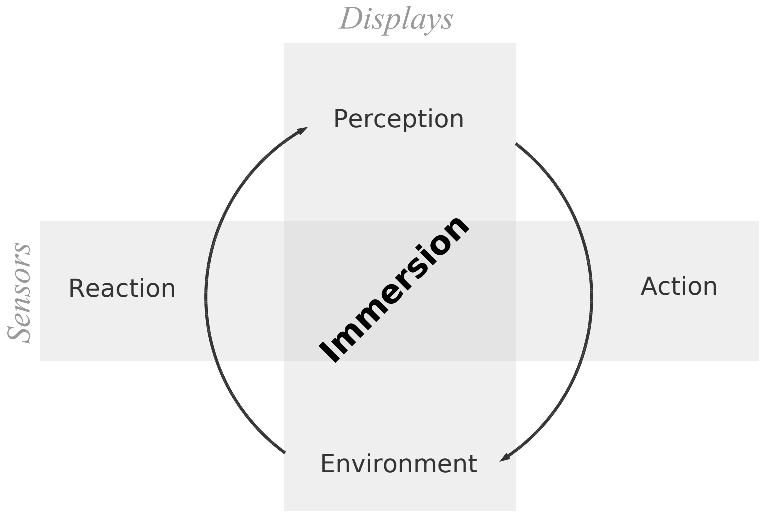


Figure 2.1: Perception/action loop for immersion

update the outputs according to the user's movements and actions. This loop has to be fast enough so that the user should not mention it.

The specificity of immersion, as opposed to augmented reality (AR) for example, is to consider a perceptual displacement of a person rather than bringing virtuality towards him. This tele-presence can take place in a non-real place called virtual environment (VE).

### Immersive virtual environment

Immersive virtual environment (IVE) was defined by Slater & Usoh (1994) as a VE designed for the immersion of people together with specific devices and tools. The latter are all involved into the sensory stimulation/substitution process, each being dedicated to one of those features: (i) producing stimuli (visual, auditory, haptic, force, etc..), (ii) transmitting the stimuli (the physical output devices), (iii) getting user feedback (the physical input devices) and (iv) reacting to the user(s), adapting/regulating the simulation. The whole process is a simple loop over these steps. Whereas VEs are today common and widespread thanks to 3D games and internet, IVEs target deeper bodily immersion and wider surrounding of the participant thanks to specific equipment.

### Hardware

As previously introduced, the head mounted displays ensure the user will only see the computer-generated images. A large variety of these devices is available commercially today and the price range varies according to the screens quality: since the 210x140 pixels screens originally in the 'Eye-Phones' developed by Jaron Lanier in 1983, the resolution increased up to 1600x1200px for LCDs on the ProView series from Kayser ElectroOptics Inc.

The only specificity of other headsets are: light weight (e.g. Sony Glasstron glasses), larger field of view (up to 127 degrees horizontal on the HMD for Tornado jet simulator by CAE Electronics), and pass-through capabilities for augmented reality.

The other way to immerse users in 3D is to literally surround them with screens. The CAVE system was developed around 1992 at the University of Illinois in the Electronic Visualization Laboratory (EVL) at Urbana Champaign, USA. Originally with 3 projection walls and one floor, CAVE systems can have more or less screens (e.g. a corner) and can immerse more than one user at a time. As the cost and the cumbersomeness of these systems prevented them from a large dissemination, intermediate solutions also pointed out, generally as simple extensions of a desktop setup by enlarging the size of the displays (e.g. wide back-projection screen, workbench, or the ImmersaDesk from EVL in 1995) and by making them stereoscopic.

The capability to simulate depth and to make objects rise out of the screen strongly influences the perception and enhances the sense of being present in the virtual environment. More details on the influence of stereoscopy in IVEs can be found in (Gaggioli & Breining, 2001). For HMD, it is enough to provide a different image for each eye ('passive' stereo). For 'active' stereo projection (e.g. in CAVE), two main commercial solutions exist: the high frequency projectors with shutter glasses (shutter glasses by Stereographics Corp. started around 1985) or the dual projection with passive polarization glasses (often custom projection design). Stereo for computer graphics used to be dedicated to high-end workstations or rendering machines (typically Silicon Graphics) but can nowadays be available as a simple option on commercial professional graphic boards for PC (NVidia, 3DLab).

### 2.1.2 Sense of presence

The question of 'presence' is central for any VR experience as the problem is to know if it was actually capable of producing a sufficient implication of the user, compared to a similar real situation. The sense of presence (SoP) is the ability of someone to be aware of the presence mechanism that occurs during immersion.

#### Theories on Presence

Considering SoP as the participant's subjective “sense of being there” (Heeter, 1992) was often used by VR specialists as it focuses on the opposition between real and virtual place, thus giving a reference point for the evaluation of a simulator. For similar reasons, Lombard & Ditton (1997) proposed to rather consider SoP as the “perceptual illusion of non-mediation”; by focusing on the opposition with/without equipment (mediation), presence could be better understood and measured with technological variables. However, we believe that presence is much more than "being there" as it involves different levels of bodily mobilization. It should be clearly distinguished from immersion and better considered as “a ’response’ to a system of a certain level of immersion” (Slater, 2003). The “suspension of dis-belief” definition of SoP from Slater & Usoh (1993) is indeed rather well accepted in the community —it is based on people ability to forget for a time that they are in a world other than where their real bodies are located, in short to forget the artificiality. As it applies for any mediated experience (Books, TV, movies) it is tangible and easily understandable, even for non VR specialists.

However, he debate on presence is far from being closed. Typically, a recent debates on the on-line journal Presence-Connect between Slater (2003) and Waterworth & Waterworth (2003) still raised questions about the terminology, the influence of “form” and “content” on presence, or the underlying mental or biological mechanisms at the origins of this feeling. But the more important point we would like to point out of these discussions is the current tendency to rely deeply on psychology and neuroscience.

According to Lee (2004), studies on presence used to focus more on the effects (arousal, physiological, mood, memory, persuasion, skills) rather than understanding its causes. He considers that presence can occur when the IVE is compatible with the human “innate or rapidly developed knowledge of causal relations in the world —folk physics” (p.499). As such, breaks in presence (BiP) can be explained as artifacts of folk physics that lead to a come-back into a consistent reality. Then, bellow this physical action-reaction expectation, presence was also confronted to the psychological theory of play by Klimmt & Vorderer (2003); the felt ’responsibility’ in what happens, the involvement for a goal, the capacity to neglect certain aspects of reality, and the exaltation of being transported into another world, may all together constitute the bridge between children imaginary games and SoP in VE. Finally, even deeper in the human mind, Riva & Waterworth (2003) proposed a unified model of different aspects of presence by associating its mechanism with archetypal functions of the self standing deeply in our brain.

### Evaluation of SoP in IVE

The problem when studying the sense of presence is to find valuable means of evaluation. In fact, trying to give a single value to SoP may even be nonsense. What is concretely done is the analysis of the three quantitative and qualitative parameters of the subject’s reactions during an immersion: cognitive response, overt behavior and emotional states.

Firstly, SoP is generally evaluated with questionnaires (Slater *et al.*, 1994; Lombard *et al.*, 2000). However, the evaluation of “sense of being there” through questionnaire relies strongly on the assumption of a persistence of the feeling and on a subjective cognition of this state. This hypothesis is highly controversial. First, Usoh *et al.* (2000) have shown that there is no significant difference between the answers to the classical presence questionnaires from Witmer & Singer (1998) for subjects having a real experience compared to those having a virtual one. Second, it has been shown by Slater (2004) that any arbitrary concept (i.e. the colorfulness) can be associate to a lived experience and evaluated with questionnaires without any valuable conclusion out of it. One can thus wonder what has been evaluated as being 'presence' in studies using questionnaires...

Secondly, observations of subjects' overt behaviors, although very useful for therapists, are hard to conduct and to quantify. In practice, monitoring performance in the achievement of a task can provide numerical estimations (e.g. navigation and orientation in space done by Prothero *et al.* (1995)), but will not indicate how much the subject was in presence in the VE.

And finally, despite a potential objectivity and universality, physiological measurements are rarely used due to the extremely complex and empirical interpretation of data. Even though, Dillon *et al.* (2000, 2001) proposed to use the physiological measurement of arousal to indicate the presence during immersion. Wiederhold & Wiederhold (2003) also concluded that "percentage change in heart rate and skin resistance had a high level of correlation with Presence, degree of realism, and immersiveness". However, the main problem here is the potential confusion between presence and the observable physiologic parameters of emotions and anxiety, SoP being only a potential factor of the latter.

## 2.2 Virtual reality exposure therapy

For the purpose of our study, we will basically consider that presence is the difference between *looking at an image* and *being inside the scene*. Presence is indeed a matter of personal feeling, a whole body response as if the surrounding events during immersion were real. The original motivation of using VR for mental health was based on this principle. If someone could experience a situation as lived through (i.e. with a sufficient sense of presence), the psychological impact should be the same than in reality. This section will present this area of psychotherapy relying on such behavioral therapy with VR, and make an overview of the work done in virtual reality exposure therapy.

### 2.2.1 Virtual reality for cognitive and behavioral therapy

The treatment of anxiety disorders is very often supported by pharmacotherapy. However, their positive effects end soon after the end of the treatment. On the contrary, psychotherapies have slower but more persistent effects because they aim at covering multiple aspects of the pathology.

Cognitive behavioral therapy is a mixed approach of psychotherapy which rose in the seventies upon the idea that cognitive and behavioral experiences are intertwined and must

Table 2.1: Advantages/Disadvantages of different exposure solutions

| Exposure | In-vivo | Imagination | Single media | Multimedia | Virtual Reality |
|---|---|---|---|---|---|
| Covers all sensitivities? | yes | yes | no | more/less | more/less |
| Focus on stimulus? | no | no | yes | yes | yes |
| Active participation? | yes | more/less | no | no | yes |
| Applicable to any phobia? | yes | yes | no | more/less | yes |

be considered by therapists as an interacting pair. Treatments were adapted to this principle by combining the cognitive approach (changing the pessimistic ideas, unrealistic expectations, and overly critical self-evaluations), with behavioral exercises (exposure to the feared situation, stress and anxiety management).

There are two ways to conduct behavioral exposure. First, during repeated exposures to the fearful situation, the patient is asked not to avoid the stressful factors and to develop better coping skills. Along the sessions, the anxiety becomes less intense. This is the habituation. The second approach is to preliminarily relax the patient and then to progressively present specific stimuli provoking anxiety. The objective is to teach the patient to associate stimuli with a relaxation state. This is the de-sensitization.

Originally conducted in-vivo, the exposure sessions to stressful stimuli can also be based on various media, such as pictures and videos, and more recently with virtual reality. From a therapeutic point of view, we can compare virtual reality exposure (VRE) with other exposure techniques (table 2.1). First, in-vivo exposure is not always the best solution. A side effect of covering perfectly every cause of anxiety is the difficulty for therapists to isolate stimuli in order to reduce the number of variables influencing the patient. To reach this goal, they often use single media exposure with pictures, sounds or touch. But in these cases, the sensitivity of the patient to one medium or one stimuli limits the exposure impact. With multimedia technologies (e.g. video), this issue is improved. However none of mediated exposure require an active participation from the patient. Exposure in imagination has the opposite approach but does not ensure that the patient actually forces himself to face the feared situation. According to these criteria, virtual reality seems to propose a good compromise: it can focus on specific stimuli, it covers many media and sensitivities, and it ensures that the patient is active.

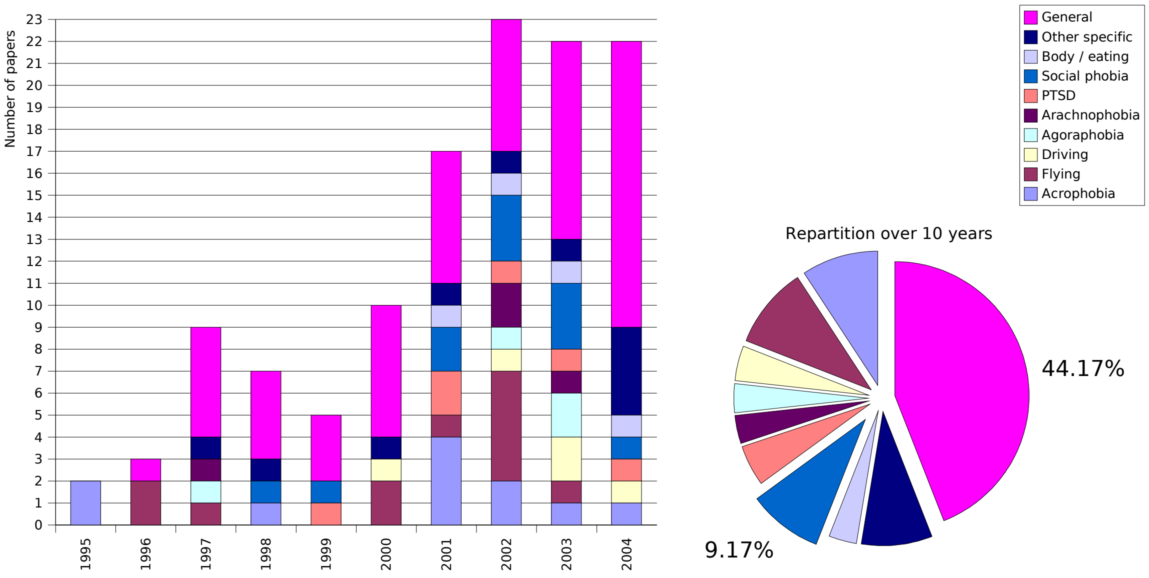


Figure 2.2: Past 10 years of publications in Virtual Reality Exposure Therapy

## 2.2.2 Ten years of experiments

### Overview

Before entering into a detailed description of the research done in VRET, it is interesting to have a global overview of this field's history. We could retrace the evolution of the research activity in this field by collecting every publications related to VR in psycho-therapeutic context that are referenced in the major databases (Pascal[1], BIOSIS[2], MEDLINE[3], Web of Science[4] and CSA[5]) or in international conferences. Our database covered the past ten years and shows the evolution and the repartition of scientific work in this area (fig.2.2). This analysis was made possible by the limited amount of publications in this emerging and specific field.

Over the 120 references collected, we could observe a linear increase of publications since 1995 which reflects the growing interest and the success for VR in psychotherapy. In addition, although there has always been more articles discussing the general aspects of VRET (~44% in average), the variety of disorders studied proves the flexibility of the VR technique. This repartition is shown in figure 2.2. Not surprisingly, phobias whose exposure are related to well known simulation areas were the first and the most studied (flight simulator for fear of flying and geometrical environments for acrophobia).

---

[1] http://www.inist.fr

[2] http://www.biosis.org

[3] http://www.nlm.nih.gov

[4] http://www.isinet.com

[5] http://www.csa.com


Virtual airplane
Virtual heights
Virtual Audience
Virtual Vietnam

*www.virtuallybetter.com*

Figure 2.3: Therapeutic VEs at Virtually Better Inc


Fear of Flying
Social phobia
Acrophobia

*http://www.vrphobia.com*

Figure 2.4: Therapeutic VEs at Virtual Reality Medical Centers

However, publications only represent a single facet of the development of VRET. In parallel, virtual reality techniques became a speciality for psychological clinics or hospital psychiatry services. Among them, Virtually Better, Inc.[6] is an example of private center created in 1996 to apply the results of successful researches conduced by Barbara O. Rothbaum and Larry F. Hodges, particularly known as pioneers in VRET with the treatment of acrophobia (Hodges, 1995) or fear of flying (Hodges *et al.*, 1996; Rothbaum *et al.*, 1996) (see fig.2.3). Similarly, the Virtual Reality Medical Centers[7] today propose a large set of simulators for the treatment of various phobias (fig.2.4). Their technology and experience in the use of VR in therapeutic programs is the result of researches conduced by Dr. Brenda K. Wiederhold of the Interactive Media Institute (IMI). With similar goals, the European Union funded the VEPSY[8] research project which started in 2001 to join the efforts of multiple countries to validate and disseminate tele-medicine and portable virtual environments for clinical psychology. More specifically, medical institutes in Italy (G. Riva, Istituto Auxologico Italiano, Milan), Spain (C. Botella, Universitat Jaume I, Castellón), and France (P. Légeron, Hôpital Sainte-Anne, Paris) who focus on clinical psychology promote VRET by providing free software solutions[9]; figure 2.5 shows VEs used for treating panic attacks, and figure 2.6 shows the free tools available for the design and immersion into architectural environments usable for various phobias. Their experience in the application of virtual reality in medicine, neurology and psychology is explained in details by Riva *et al.* (2003b, 2004).

### Major contributions

We would like to emphasize here the work done by the most well-known researchers in this field. Typically, the work of Rothbaum *et al.* (1999) on the treatment of Vietnam veterans suffering from post traumatic stress disorders (PTSD) was so successful that today Rizzo *et al.* (2004c) apply similar therapy to soldiers coming back from the Iraqi war. In a similar way, the clinical experiences of Wiederhold on various disorders such as acrophobia (Jang *et al.*, 2002), fear of flying (Wiederhold *et al.*, 2002b), or agoraphobia (Vincelli *et al.*, 2003), became a reference for the elaboration of clinical studies and use of biofeedback assessment tools. North *et al.* (1998, 2002) were also pioneers in the development of VRE systems with one of the first studies on fear of public speaking, whereas Riva *et al.* (2001, 2003a) explored the use of VRET for body image disturbances and eating disorders .

---

[6] http://www.virtuallybetter.com

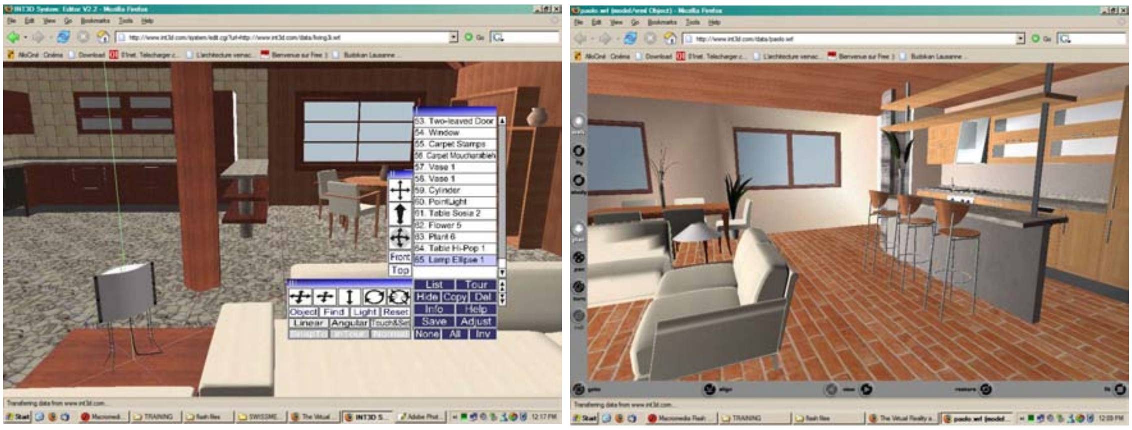
[7] http://www.vrphobia.com

[8] http://www.cybertherapy.info

[9] http://www.vrtherapy.net

http://www.vrtherapy.net

Figure 2.5: Therapeutic VEs for panic disorder (free tools for the VEPSY project)



http://www.int3d.com

Figure 2.6: VRML scenes available for immersion and edition (free models for the VEPSY project)

**Synthesis**

In addition to the fear of flying, acrophobia, PTSD, agoraphobia, social phobia and eating disorders mentioned above, VRE has been experimented more recently on driving phobia (Wald, 2004), arachnophobia (Garcia-Palacios *et al.*, 2002; Hoffman *et al.*, 2003), and specific animals phobia (Parrott *et al.*, 2004).

In almost all applications, researchers use head mounted displays and head tracker. Joysticks are often used for the navigation in large IVEs. Additional scenic equipments were rarely used for some specific applications to improve the immersion (e.g. the steel guard rail for the acrophobia environment in (Jang *et al.*, 2002) or the wooden ledge in (Meehan *et al.*, 2002)).

The expansion of VRET in the medical field supports the optimistic therapeutic results presented in the studies. However, we can observe that publications generally present case studies or non clinical experiments. This lack of long term followup studies is of course partially due to the youth of the field but mainly to the difficulties in the organization, the validation by ethic committee, and the management of interdisciplinary projects. Nevertheless, the efficiency of VRET has been strongly validated for some anxiety disorders. Rothbaum *et al.* (2002) and Wiederhold & Wiederhold (2003) with their respective twelve months and three years follow-up studies have certainly proved VRE was efficient in the treatment of fear of flying.

## 2.3 Virtual reality exposure therapy of social phobia

In the study from Lincoln (2003) on the CBT treatment of social phobia (SP), the analysis of prevalence indicates “social phobia to be one of the most frequent chronic psychological disorders” in western countries (e.g. 13.3% in U.S.A in1996, 16.1% in Switzerland in 1992). However, we could observe that VRET have not been proportionally applied to socially related phobias in the past (i.e. ~9% of the publications in figure 2.2). The technological limitations may have been the major cause for this lack, but the recent increase of experiments in this area indicates that we can now really target efficient VRET for social phobia.

### 2.3.1 Treating social anxiety disorders

Social anxiety disorders (SAD), or social phobias, are characterized by intense and persistent fear of social performance situations in which embarrassment may occur, typically fear of public speaking and/or situations where interactions with others will occur (DSM-IV, 1994). These chronic disorders usually appear during adolescence and have a life-time prevalence of 2% to 4%. SAD can be divided in specific and general types: in the former case, the patient complains of a single specific performance fear with avoidance behavior, and in the second case he may complain of several fears and avoidance behaviors —the term generalized social anxiety disorders (GSAD) is used in this case. This disease is often associated with other subtypes of anxiety disorders with an average of 50% of SAD patients reporting phobic or panic disorders, generalized anxiety, as well as frequent depressive disorders and substance abuses (Merinkangas & Angst, 1995). This complexity leads to the need for efficient diagnostic tools. Here are the mostly used self-evaluation questionnaires:

**Social Avoidance and Distress Scale (SADS):** developed by Watson & Friend (1969) to measure the tendency to experience distress in social situations and to avoid them.

**Fear Questionnaire –Social Phobia subscale (FQ-Social):** short questionnaire by Marks & Mathews (1979) including anxiety-producing situations rated for degree of avoidance.

**Liebowitz Social Anxiety Scale (LSAS):** commonly used questionnaire from Liebowitz (1987), experimentally and clinically validated. See Appendix A on page 125 for more details.

Concerning the typical behaviors of social phobic people, several elements are interesting regarding the clinical evaluation of anxiety and the design of exposure sessions. To begin, it is well-known by therapists that safety behaviors are symptomatic of the phobic will to decrease risks of negative evaluation from others. Typical examples of safety behaviors for social phobia are; avoid drawing attention, avoid eye to eye contact, speak very fast or fill the pauses, etc. The observation of visual scan-path done by Horley *et al.* (2003) allowed to objectively establish this eye-to-eye avoidance behavior. However, and contradictorily, Foa *et al.* (2000) suggested that this avoidance behavior does not lead to a decreased ability to obtain significant social information (relatively to non-phobic people), but would rather be associated with a higher memory for all facial expressions; social phobic are very sensible to facial expressions. A possible explanation may be given by Hofmann *et al.* (1997); they explain that the frequency of eye contact seems significantly lower for phobic than for non-phobic in passive situations (talking with an experimenter or sitting in front of an audience), whereas it is almost identical during speech (exposure session). They have also been able to observe behavioral parameters in the silent pauses and *ah-ratio* (a measure for filled pauses); "social phobic had greater ah-ratio, showed longer pauses, paused more frequently, and spent more time pausing" (p.582).

The behavioral aspect of the CBT of SAD stands in social-performance exposure sessions. They generally consist in a public speaking situation, often presented as role-playing tasks, or even an impromptu speech task on various topics like Beidel *et al.* (1989) proposed.

Objective (e.g. heart rate) and subjective (self evaluation) observation of stress are good indicators of the reactions of phobics during exposure (Levin *et al.*, 1993; Sheffer *et al.*, 2001).

In parallel of the behavioral exposure sessions, Kim (2005) recommends rational cognitive feedback on the bad influence of safety behaviors to improve the therapeutic effect of CBT. The goal of exposure is to force the patient to experience real fear and discomfort when in situation, without avoidance, in order to perform efficient habituation and de-sensitization. The patients wouldn't be treated properly by exposure if they were not guided and didn't have a verbal feedback from the therapist.

### 2.3.2 Major contributions

We have seen that North *et al.* (1998) has been one of the first to apply VRE against public speaking anxiety. According to their recent results presented by Harris *et al.* (2002), they could significantly reduce the fear of public speaking of four university students after only four VR sessions. They made progressive exposures; first to get used to the HMD and the apparatus, second to talk in front of a growing assembly with positive attitude, and two more sessions in front of a growing and negative audience. Measurements of heart rate during sessions appeared to be a good indicator of stress during sessions and its decrease across sessions was interpreted as a good sign of a better management of stress. Another VR platform for public speaking was proposed by Lee *et al.* (2002). The system presented offers various technological improvements to make the simulation more realistic and more interactive. First, the VE is displayed using image-based rendering of a real place. As many character as requested can be added to the scene by adding video impostors to the assembly, thus providing live and expressive figures of people having various attitudes (refer to (Jo *et al.*, 2001) for more details). Second, a sort of video-conference with a live video integrated to the scene let the therapist interact freely with his patient.

Public speaking sessions may be effective for SP therapy, but the simulation of other social situations would allow to cover different aspects of GSAD. In order to confirm the potential impact of such VRE, James *et al.* (2003) compared subjective measurements of social anxiety after immersion in a socially neutral situation (travelling in the London underground) and after a socially intense experience (discussing with people in a wine bar). They could conclude that, despite the lack of realism, "social anxiety can be generated within a virtual social setting" (p242). In parallel, a clinical evaluation of the interest of covering different aspects of GSAD is presented by Roy *et al.* (2003). The goal of this study was to compare a VRET protocol with classical cognitive and behavioral group therapy (CBGT) and with no therapy (the waiting list). They designed a full protocol based on 12 sessions organized in three phases conduced by a therapist; *assessment*, *spontaneous* and *instructed.* The VRE consists in on-screen navigation into different VEs targeting *performance, intimacy, scrutiny*, and *assertiveness* social aspects. The VEs are 3D models of rooms populated with impostors (pictures of people). The preliminary results on ten subjects indicated this VRET as effective as CBGT.

The visible success of these experiments shows that VRE sessions have cognitive and behavioral impacts on the patients and may fulfill the requirements for the CBT of SAD. However, the use of VR for social phobia still require larger validations and clinical experiments. Moreover, the 3D simulation engines used were not taking advantage of the recent improvements in computer graphics for the rendering of virtual humans, neither providing them with believable behaviors; for the specific case of social interactions, the link between presence in the IVE and impact of the exposure session cannot be made as simply as for situation-based simulations (e.g. flying, acrophobia). Concepts of co-presence and social presence need to be investigated in order to better understand the underlying mechanisms of VRE for social phobia.

### 2.3.3 Social presence

We can all experience the 'tele-presence' of someone when confronted to mediated communication with phone or on-line chat. We also make a clear distinction between the presence of someone by phone and the presence of people in the TV set. Typically here, despite the fact that the stimulation is more "immersive" in the latter case, the social presence is stronger in the first case. Consequently, the concept of presence introduced for the immersion cannot be applied directly to social presence. Together with Biocca (1997), we consider that there is certainly more than the "sense of being with another", or "the sense of being together". He proposed this definition of social presence; "The minimum level of social presence occurs when users feel that a form, behavior, or sensory experience indicates the presence of another intelligence." (i.e. we quickly determine if its a phone answering machine or someone). However, we do not consider that social presence is only related to a kind of Turing test we would unconsciously perform. Simulating social relationships would rather be, as Heeter (2003) says, an adequation between the perceptual stimuli and the expected social cognition. Lee (2004) calls this implicit knowledge about how the social world works "folk psychology".

In order to experiment with such social cognition, Garau *et al.* (2003) conducted specific tests with non-phobic subjects to determine the level of responsiveness of virtual agents toward real humans that is required to actually provoke a sufficient co-presence in IVE. Their results suggest that non-verbal responses (e.g. look at you) already provoke a "significantly higher sense of personal contact with the agents" (Garau *et al.*, 2005, p.114).

Therefore, before using VR for social phobia, Slater & Pertaub wanted to answer a critical question: if someone is anxious with real people, will he also be anxious when facing simulated people despite knowing that the avatars are computer generated? The experimentation of Pertaub *et al.* (2002) with three public speaking IVEs aimed at establishing the correlations between the changes in the assembly attitude (neutral, positive and negative attitudes) and the response of participants to personal report of confidence as speaker (PRCS) questionnaires (Paul, 1966). Results have shown that not only social anxiety was induced by the virtual audience, but also that the degree of anxiety experienced was directly related to the type of virtual audience feedback the speaker received (Pertaub *et al.*, 2001). In a recent study described by Slater *et al.* (2004), they were able to directly observe significant differences of anxiety levels (heart-rate) between confident speakers and phobic speakers. According to them, the social phobic group have expressed higher anxiety when talking to an assembly than speaking in an empty room, whereas non-phobic have not shown much difference between the situations.

## 2.4 Discussions

Once the reliability of VR for conducting exposure sessions for various anxiety disorders and for social phobia in particular is established, the problem is to determine if it is really worth the effort. We already exposed in table 2.1 the potential interest of using VR instead of in-vivo or other mediated exposure techniques. However, as suggested by Roy (2003), “this ’technicization’ of the psychotherapy, as attractive as it is, does not modify the theoretical and methodological bases on which VRT rests” (p.182). Doubtlessly, building more complex tools could rather be an argument against VRET, as therapists would naturally prefer the simplest approach.

VRE was indeed introduced to make therapists’ lives easier; the organization of exposures being often hardly manageable and sometimes impossible in-vivo, the availability of ’on-demand realities’ should facilitate their management. As noticed by Rizzo *et al.* (1998), simulation is even sometimes the only solution for exposure to specific stimuli. This was for example the case of Difede & Hoffman (2002) who had to treat PTSD of survivors of Nov. 11th catastrophe in New York. Simulating ’altered realities’ for the needs of a particular phobia may also open new areas for therapy where cognitive and behavioral skills can be freely, but carefully, explored. As such, Rizzo *et al.* (2004a) suggest that VR may potentially influence the therapeutic methodology by providing more ecologically valid exposure scenarios, that is, a greater relevance relative to the ’real’ world than with pictures, video or even role-playing tasks. Therefore, we have decided to first improve the simulation of social situation for the therapy of SAD in order to reach the exposure impact of simulators for fear of flying, acrophobia or PTSD.

However, other technological added value have to be considered. For instance, Roy (2003) pointed out the augmented capability to control the course of exposure sessions; intensity of the stimuli, progressive exposures, immediate interruption or pause of the simulation, etc. If not the main, this argument in favor of VRE is also combined with the possibility to provide therapists with enhanced control tools, like for example the management of the patient's point of view (Schuemie, 2002). Finally, Rizzo *et al.* (2004a) also mentioned the new capability to deliver immediate feedback in course of the exposure as well as to capture and record performance for review and analysis. This is why our second direction of research consisted in the experimentation with assessment tools to be integrated into VRET.

# Chapter 3

# Virtual reality platform for social phobia therapy

The observation of the technological developments done for VRET could already give us a global overview on the needs and on the improvements to bring to existing solutions. The main difficulties encountered stand in the complexity of VR technologies, both from hardware and software points of view. As a consequence, the gap between experimental setups —targeting high-end solutions such as CAVE and realistic rendering— and clinical systems —focusing on affordable and simple technologies— is nowadays still so important that it is not proved that the good results obtained in the former case will be as good when used under everyday clinical conditions. Our effort was therefore to build a platform with strong VR capabilities and still enough flexibility to be practically used by a therapist.

This chapter presents our approach for the development of an efficient exposure system for social phobia. While focusing on the therapeutic requirements, we also wanted to investigate into VR solutions which could enhance the quality and the efficiency of VRE. We will first present the hardware setups we designed and then the software platform we developed to be dedicated to the simulation of social situations.

## 3.1 Immersion setups

As introduced in section 2.1.1, the quality of the immersion is a key issue for the successfulness of a VR experience. However, we could observe in the literature that the equipment used in VRET is very heterogeneous and its potential influence on presence rarely considered. Typically, some results were not comparable since the immersion setups were too different (e.g. CAVE used by James *et al.* (2003) v.s. desktop and joystick used by Roy *et al.* (2003)).

Therefore, we designed different setups with high immersion capabilities but which could still be installed in a medical context. From the hardware side, we did not build new device and relied on the assembly of existing commercial solutions. However, the lack of standard in the software support of VR devices required the development of evolving solutions for the integration of tracking technologies, immersion paradigms and output support.

### 3.1.1 Tracking devices support

The requirements for tracking human movements during immersion are basic. For the visual feedback loop, you need at least to track the head orientation with a three degrees of freedom (DOF) sensor. Six DOF tracking is of course preferable to enable translations. Additionally, tracking hands or other body parts can be used for more involving interaction.

There is a large panel of position and orientation tracking solutions with all their proprietary drivers and software approach. To unify the problem and enable the selection of technology with minimal software dependency, we designed a generic software interface layer. The shared input device (SID) system is based on the sharing of a memory block containing sensors' data. A server opens the access to the shared memory. Device drivers are clients for writing and applications are clients for reading. A basic mutual exclusion mechanism ensures the consistency of data. The implementation done with the Adaptive Communication Environment (ACE)[1] library is also multi-platform (MS Windows and Linux).

SID enables a basic device abstraction and simplifies the management of devices drivers. Typically, a simulation application can access data from any kind of sensor without need of change in its code. For our needs, we implemented drivers for the following commercial devices:

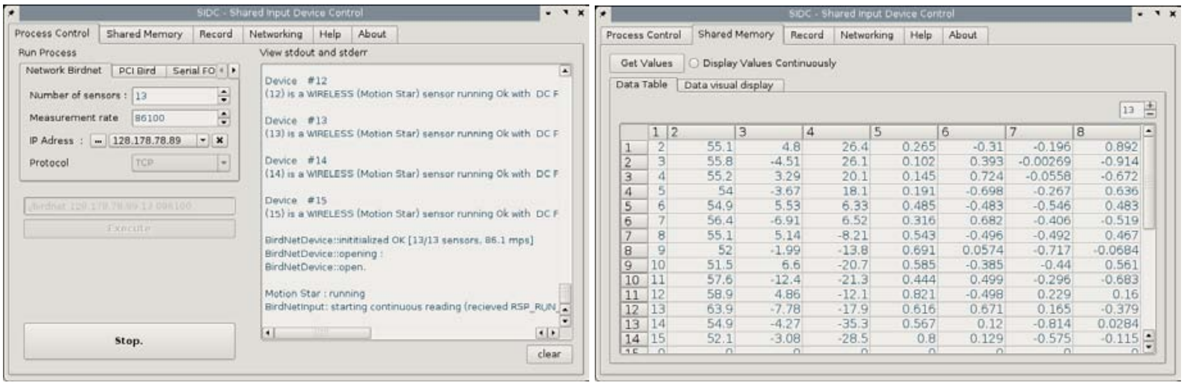
- *Flock Of Birds (FOB):* 1st generation of magnetic 6-DOF sensors by Ascension Technology[2]. Many FOB units can be serially plugged, each unit handling one sensor accessible through serial port.
- *Motion Star*: professional magnetic motion capture system from Ascension Technology with up to twenty 6-DOF sensors, possibly wireless (backpack), and accessible through TCP or UDP using proprietary Birdnet protocol.
- *PCI Bird*; same as a FOB unit but integrated directly in the PC (PCI slot).
- *InertiaCube3:* Inertial tracker from Intersense Inc.[3] providing yaw pitch and roll (3-DOF), accessible through serial port.

---

[1] http://www.cs.wustl.edu/~schmidt/ACE.html

[2] http://www.ascension-tech.com

[3] http://www.isense.com



Number of sensors : 13
Measurement rate 86100
IP Adress : 128.178.78.89
Protocol TCP

```
Device   #12
(12) is a WIRELESS (Motion Star) sensor running Ok with  DC F

Device   #13
(13) is a WIRELESS (Motion Star) sensor running Ok with  DC F

Device   #14
(14) is a WIRELESS (Motion Star) sensor running Ok with  DC F

Device   #15
(15) is a WIRELESS (Motion Star) sensor running Ok with  DC F

BirdNetDevice::inititialized OK [13/13 sensors, 86.1 mps]
BirdNetDevice::opening :
BirdNetDevice::open.

Motion Star : running
BirdNetInput: starting continuous reading (recieved RSP_RUN_
```

○ Display Values Continuously

13

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 55.1 | 4.8 | 26.4 | 0.265 | -0.31 | -0.196 | 0.892 |
| 2 | 3 | 55.8 | -4.51 | 26.1 | 0.102 | 0.393 | -0.00269 | -0.914 |
| 3 | 4 | 55.2 | 3.29 | 20.1 | 0.145 | 0.724 | -0.0558 | -0.672 |
| 4 | 5 | 54 | -3.67 | 18.1 | 0.191 | -0.698 | -0.267 | 0.636 |
| 5 | 6 | 54.9 | 5.53 | 6.33 | 0.485 | -0.483 | -0.546 | 0.483 |
| 6 | 7 | 56.4 | -6.91 | 6.52 | 0.316 | 0.682 | -0.406 | -0.519 |
| 7 | 8 | 55.1 | 5.14 | -8.21 | 0.543 | -0.496 | -0.492 | 0.467 |
| 8 | 9 | 52 | -1.99 | -13.8 | 0.691 | 0.0574 | -0.717 | -0.0684 |
| 9 | 10 | 51.5 | 6.6 | -20.7 | 0.585 | -0.385 | -0.44 | 0.561 |
| 10 | 11 | 57.6 | -12.4 | -21.3 | 0.444 | 0.499 | -0.296 | -0.683 |
| 11 | 12 | 58.9 | 4.86 | -12.1 | 0.821 | -0.498 | 0.229 | 0.16 |
| 12 | 13 | 63.9 | -7.78 | -17.9 | 0.616 | 0.671 | 0.165 | -0.379 |
| 13 | 14 | 54.9 | -4.27 | -35.3 | 0.567 | 0.12 | -0.814 | 0.0284 |
| 14 | 15 | 52.1 | -3.08 | -28.5 | 0.8 | 0.129 | -0.575 | -0.115 |

(a) Device control

(b) Data monitoring

Figure 3.1: SID graphical user interface.

A graphical user interface for SID allows for easy management of various device drivers (fig.3.1(a)) and data input control (fig.3.1(b)). Moreover, we implemented the recording and replay of raw data (to simulate a device when not available) and the TCP sharing of memory blocks (to expose local devices on network).

### 3.1.2 Immersion solutions

Among the various solutions for conducting VRE, some are best for sensory isolation, while others improve user involvement. Head mounted displays (HMD) are often used for VR because they provide a good visual immersion and are easy to setup. Solutions based on video-projection used to be quite rare because too expensive and also quite cumbersome. The recent improvements in their technologies and the reduction of their cost may soon propose interesting solutions and we wanted to experiment with a back-projection system in order to imagine an alternative to HMD for VRE. Here are the technical specifications of the two setups we could experiment with.

#### Head mounted displays

We experimented with two models of HMD (fig.3.2):

- The *ProView* XL50 from Kaiser Electro-Optics Inc[4] is a professional high resolution HMD ($1024 \times 768$) with $50°$ diagonal field of view (FOV) and separate inputs for passive stereo.

- The *i-glasses* PC3D Pro from IO Display Systems[5] have medium resolution displays

---

[4] http://www.rockwellcollins.com/keo

[5] http://www.i-glassesstore.com

(a) Kaiser XL50 with MotionStar wireless sensor (b) I-glasses with InertiaCube3 sensor

Figure 3.2: Two models of HMD equipped with head trackers.

$(800 \times 600)$ and only $29°$ diagonal FOV. It should supports up to 100Hz active stereo with synchronization through Display Data Channel (DDC) signal.

According to these specifications, the former model provides the best immersion. However, when considering their respective cost (~30 000 € for Pro-View vs. ~1 200 € for i-glasses), the latter seems to be a good compromise. In order to keep this distinction between high-end expensive solutions and acceptable low-cost systems, we placed different models of head trackers on these HMDs; a wireless MotionStar 6-DOF sensor on the ProView and an InertiaCube3 3-DOF tracker on the i-glasses. In addition, we added plastic shields on both to obstruct the peripheral vision.

Our objectives regarding the use of HMD will be to confirm their usability for VRE and to determine if some compromise can be done on the hardware quality.

### Large stereoscopic back-projection

Placing the user close to a large screen is an other easy and good solution to cover a large part of the field of vision (fig.3.3). When facing the $2.8 \times 2m$ screen at $1.5m$, the image covers $86°$ horizontally and $56°$ vertically, which is already larger than with HMD. To provide active stereoscopic projection, we used a BARCO[6] Cine7 projector supporting high frequency computer sources ($800 \times 600@100Hz$, $1024 \times 768@85Hz$) and we equipped the user with a pair of CrystalEyes shutter glasses from StereoGraphics Corp.[7]. In addition, the opaque material of the back-projection screen together with the line-quadrupler video mode of the Cine7 projector avoid the pixelization of the computer image when looked from a short distance.

---

[6] http://www.barco.com

[7] http://www.stereographics.com

Figure 3.3: Large back-projection screen setup; the user wears CrystalEyes glasses and a MotionStar head tracker.

As we wanted the virtual humans' scale to look natural from the user's point of view, the 3D rendering configuration had to be tuned to keep compatible proportions between real and virtual places. As for an anamorphosis, the projection on the screen should give the illusion to extend the real place through a "window" to another place.

The large back-projection setup can be used with or without stereoscopy (less immersive in the latter case), but will anyway remain far from the full immersion in a CAVE. Moreover, it requires a dedicated navigation paradigm that the user need to get used to before being able to explore freely the VE.

### Navigation paradigms

The main advantage of the HMD solution is its naturalness of usage; people simply have to look around and move the head to explore the IVE. The one-to-one mapping between sensor orientation (and position) and the 3D camera matrix is done in software at each update frame. The main disadvantages are the cumbersomeness and inertia of the helmet and the limited freedom of movement —i.e. size of the VE is bound by the size of the physical place and length of the cables.

To the contrary of HMD navigation, users can't simply move the head to explore the IVE because he needs to keep looking at the screen. Originally designed for a VR training application developed with Ponder *et al.* (2003a), the "head-window" navigation paradigm proposes to extend the screen-and-joystick gaming principle by letting the user 'become' the joystick. Basically, the head movements control a "window" into the VE, thus requiring full body motions and enhancing the involvement of the user. In practice, a single magnetic tracker is attached to the user's head. In order to "walk around" the virtual environment, he needs to step into navigation rings which in effect triggers camera motion in the desired direction (physical navigation rings are visible in figure 3.3). The user can still move inside the central area of the ring without causing any camera motion. In order to "look around", a solution needed to be introduced in order to maximize the functional visual field with a natural paradigm. In a similar approach as Peruch & Mestre (1999), we used the head-mounted sensor orientation to rotate the camera. When the user looks in the margins of the projection screen, this analogously triggers horizontal and vertical camera rotation and. This let the user look all around with head movements while keeping gaze on the screen.

Figure 3.4 shows the computation of the translation vector from the user's displacements in the navigation rings and the smooth camera rotations based on the position of the gaze target on the screen. The camera translation direction is given by the vector starting at the center of the navigation rings and ending at the 2D projection on the ground of the user's head sensor. Secondly, the translation speed is computed by applying a linear transformation (sigmoid function) to the user's distance to the center of the rings. He can thus control precisely his displacements by moving his body in the area in-between the two navigation rings. The computation of horizontal and vertical camera rotations requires to first determine the gaze target on the screen. By computing the intersection of the head sensor direction on the 2D plan of the screen, we obtain the 2D coordinate of the user's head target in the screen frame. The angular speed of the horizontal camera rotation is given by a linear transformation of the horizontal coordinate of this point. Concretely, the more the user orients his head on the side of the screen, the faster the camera rotates in the direction of his gaze. For example, when looking to an object on the right side of the screen, the camera will rotate right, the object will move to the center of the screen, and if the user keeps focus on the object, his head will slowly come back to the center and slowdown the camera rotation up to the moment he can observe the object in front of him. The same mechanism is applied vertically, but with harder constraints as we observed the users were quickly disoriented when the camera was too sensible to up and down movements.

### Sound

Although we did not develop complex sound systems, the diffusion of sounds for ambiance or events is mandatory for immersion. Depending on the immersion setup, we provided the user with normal earphones for binaural stereo (HMD) or equipped the place with 5.1 surround home cinema speakers (back-projection). Dynamic acoustic sound effects are supported in hardware by Sound Blaster[8] Audigy sound boards and could be controlled dynam-

---

[8] http://www.soundblaster.com

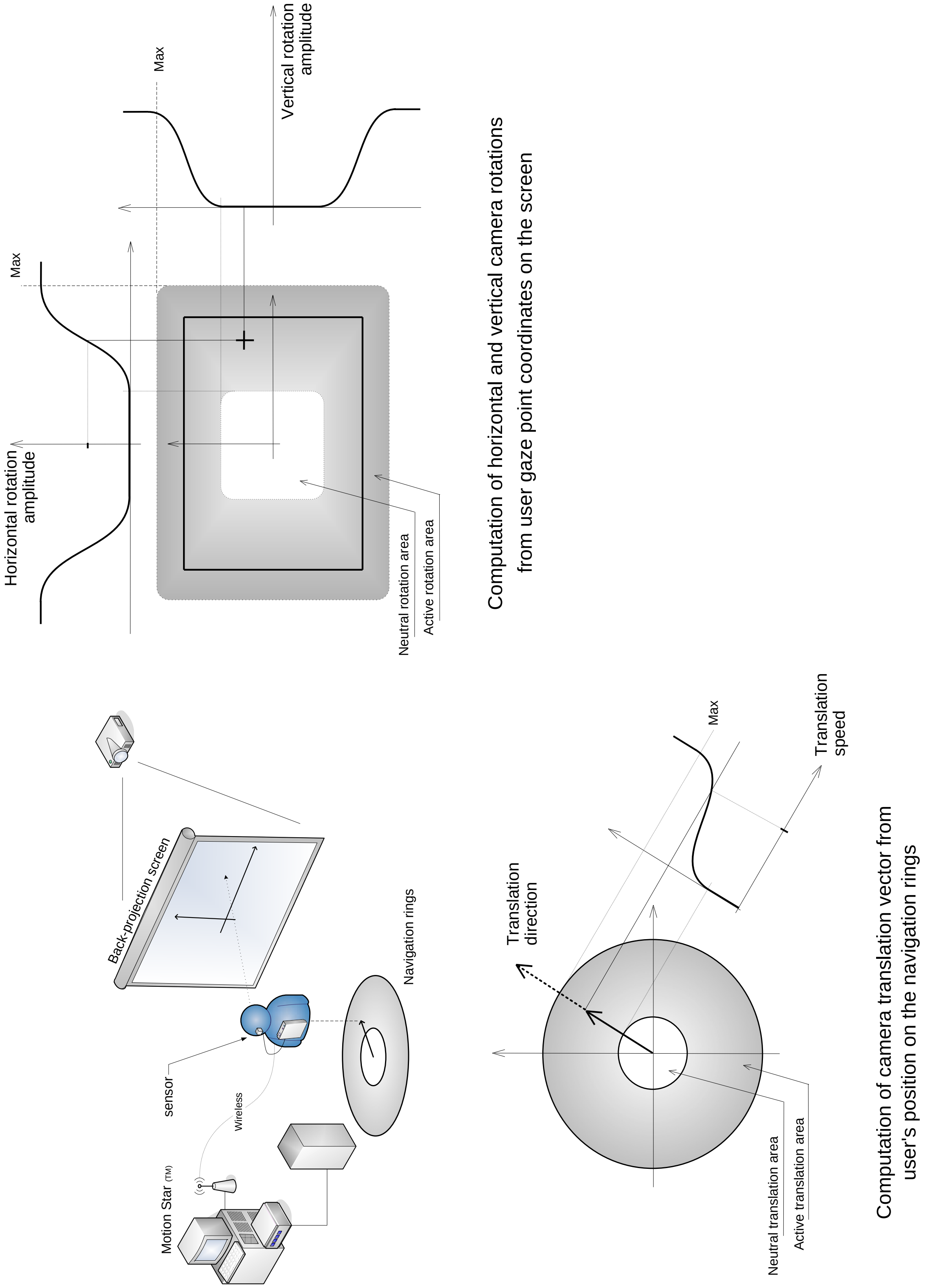


Figure 3.4: Head-Window navigation: computation of camera movements from user's displacements and gaze direction

ically in the simulation software.

## 3.2 Generic software platform for social simulations

Since technologies required for simulating inhabited IVEs are complex and heterogeneous (real time engine, computer graphics, character animation, artificial intelligence, as well as support for immersion and interaction), compromises have to be made for software development.

When targeting specific immersion platform, it is usual to rely on dedicated software. For example, the experiments made by Pertaub *et al.* (2002) in a CAVE have involved custom 3D engines with complex (multiple walls in stereoscopy) but not very realistic rendering of virtual characters. On the opposite, the solution often used for VRE is the customization of existing game engines dedicated to people simulation (military, spying, etc.). For example, Rizzo *et al.* (2004b) used Unreal Engine 2.0 (Epic Games[9]) for their experiments on the treatment of attention disorders. In a similar way, the Cyberpsychology Lab of the University of Quebec in Outaouais[10] uses the engines of Half Life (Valve software[11]) and Max Payne (Remedy Entertainment[12]) as 3D simulators for spider or height phobias. The graphics quality is generally good with these solutions, but the software architecture is bound by the production pipeline of a games with high dependency on hand-made design and special effects. Moreover, the support of immersion hardware is not handled by default and has to be emulated or lowered —software tricks to replace joystick by head tracker, no support for stereoscopy, no direct control of the navigation, etc.

To combine the two solutions, the approach we had with Ponder *et al.* (2003b) was to integrate in a unique framework the different components dedicated to simulation and VR immersion support. We will shortly describe here this generic VE software platform named VHD++, considering the general architectural aspects as well as all the specific functionality requested for the simulation of social situations.

### 3.2.1 VHD++ software architecture

The VHD++ component-based development framework is an open architecture enabling efficient integration of various VR/AR basic and advanced components. As opposed to game engines or VR toolkits, a software framework allows for the integration of application pieces and the abstraction of modules for the creation of multiple applications with minimal pro-

---

[9] http://www.epicgames.com

[10] http://www.uqo.ca/cyberpsy

[11] http://www.valvesoftware.com

[12] http://www.remedygames.com

gramming and optimal reuse of existing components. Components can be considered as pieces of software embedding specific features and collaborating with each others through a central kernel. The role of this run-time engine is not only to allow for communications but also to orchestrate the creation and suppression of modules and guarantee real time performances. Practically, the kernel runs first and reads in an XML configuration file the list of modules to be loaded and their initialisation parameters. Still, any module can be stopped, suppressed or created at any time. The synchronization and the mutual exclusion of the multiple threads running in the program is done through the central management of the shared properties. Every property embeds a data container and follows a formatted API for the uniformity and genericty of data management.

Every run-time component in VHD++ has to be built on the same template ensuring its integration into the framework. Here are the major features to comply with:

- Handling of creation and suppression; the *module loader* enables the loading and the instantiation of the module from a dynamically linked library.
- Initialisation and termination methods; allocation, initialisation, and deletion of module's proprietary data structures.
- Run and update methods; algorithms operating on the first run and at each update of the module.
- Handling of properties creation or removal; communication with the kernel to request the access to properties to operate on.

In addition to the kernel functionality, we developed the support for graphical user interfaces (GUI) required for the control and interaction with application users. A GUI manager built with Trolltech[13] QT library was embedded into VHD++ to provide thread safe exchange of data and real time interactive performances. Figure 3.6 shows a typical screen-shot of an application developed with VHD++.

Within this framework, we had to get maximum advantage of the re-usability of the components, but we also had to develop the software we needed. More specifically, we had to integrate into the generic simulation engine the rendering of virtual humans, their animation and the scripting of their behaviors.

### 3.2.2 Rendering module

In addition to the technological challenges of 3D rendering, the question of visual quality of graphics needs to be considered. On one hand, we agree with Garau *et al.* (2003) who could

---

[13] http://www.trolltech.com

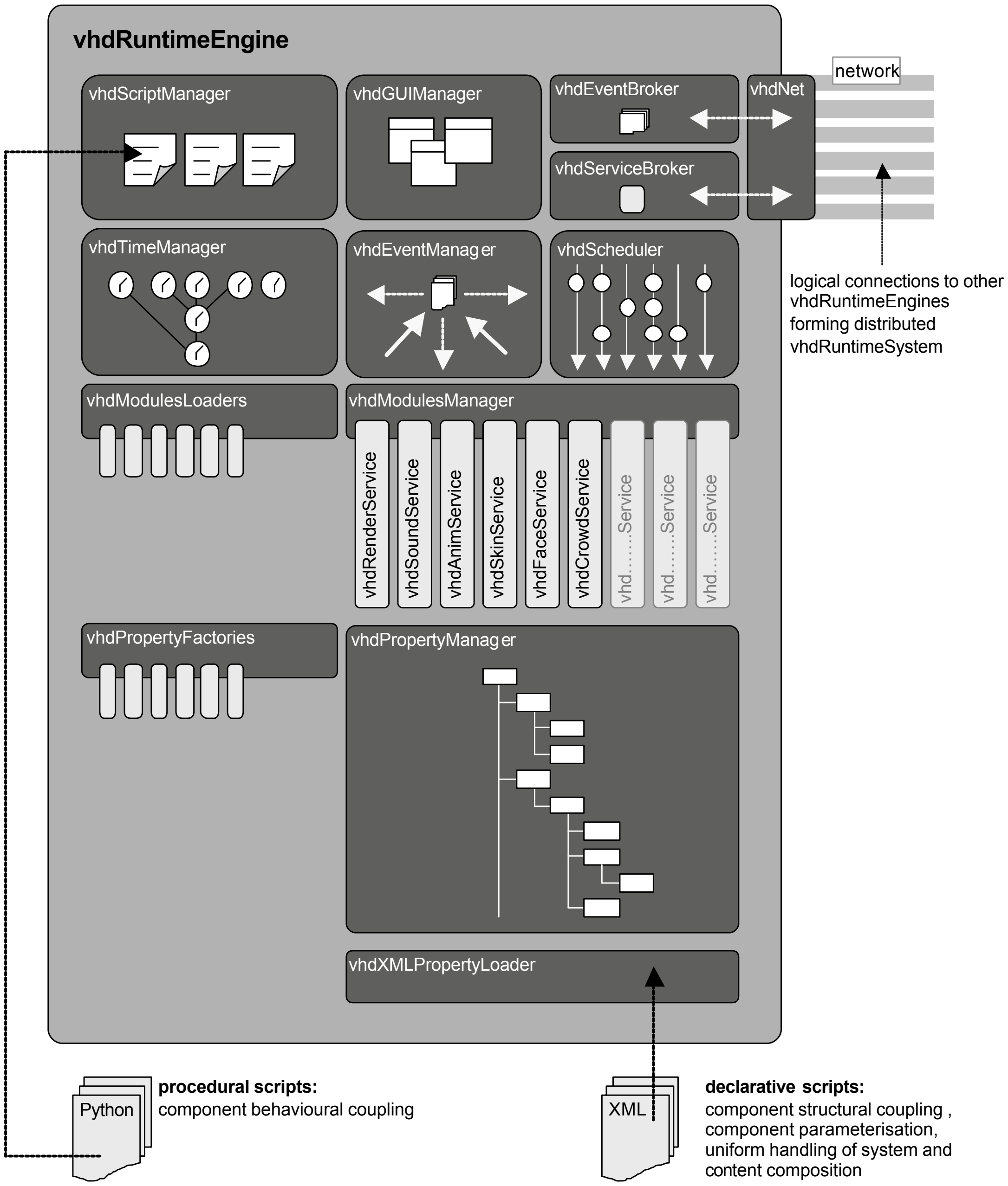
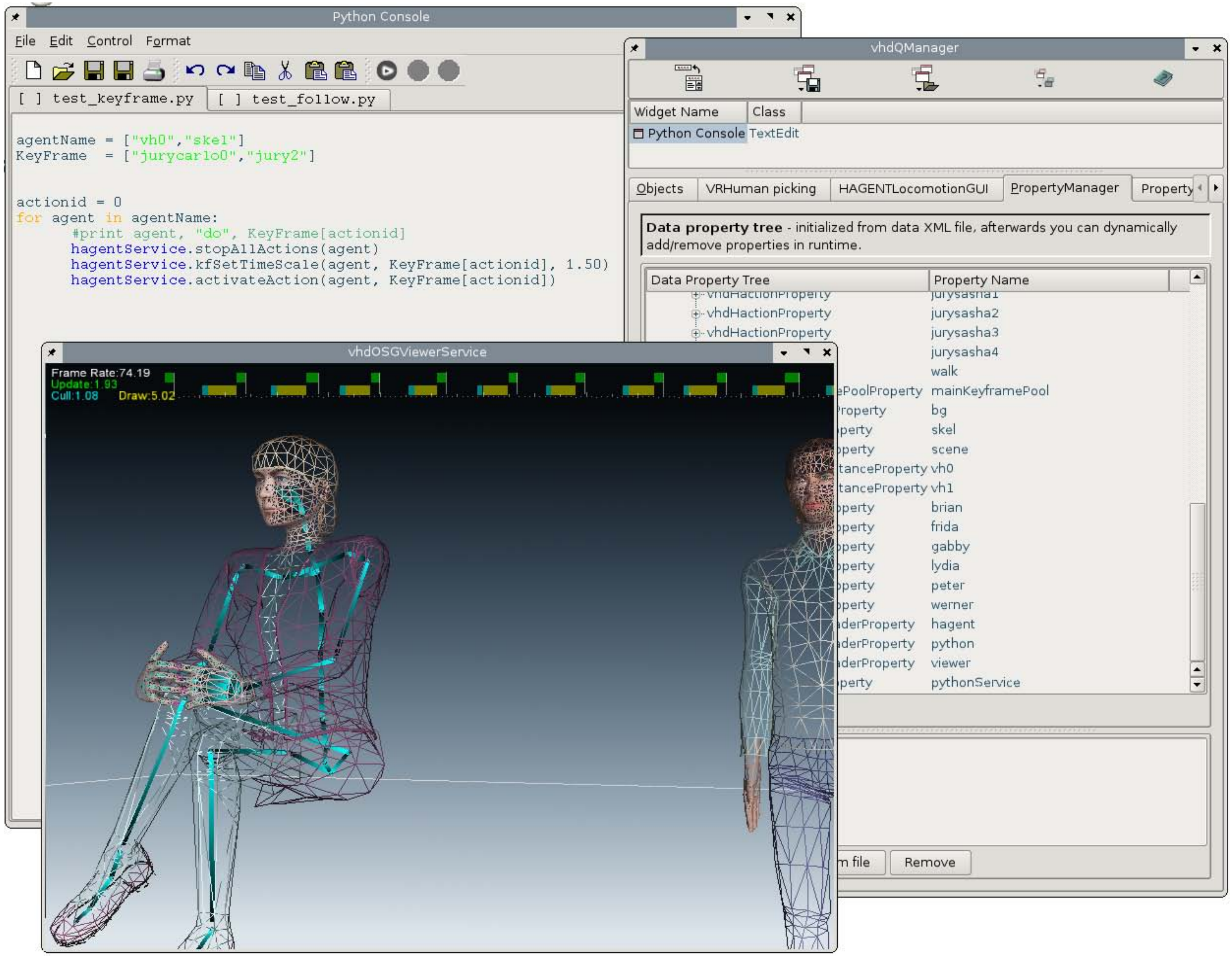


Figure 3.5: VHD++ software architecture (from Ponder (2004)).



```
agentName = ["vh0","skel"]
KeyFrame  = ["jurycarlo0","jury2"]


actionid = 0
for agent in agentName:
        #print agent, "do", KeyFrame[actionid]
        hagentService.stopAllActions(agent)
        hagentService.kfSetTimeScale(agent, KeyFrame[actionid], 1.50)
        hagentService.activateAction(agent, KeyFrame[actionid])
```

**Data property tree** - initialized from data XML file, afterwards you can dynamically add/remove properties in runtime.

| Data Property Tree | Property Name |
|---|---|
| vhdHactionProperty | jurysasha1 |
| vhdHactionProperty | jurysasha2 |
| vhdHactionProperty | jurysasha3 |
| | jurysasha4 |
| | walk |
| ePoolProperty | mainKeyframePool |
| roperty | bg |
| perty | skel |
| perty | scene |
| tanceProperty | vh0 |
| tanceProperty | vh1 |
| perty | brian |
| perty | frida |
| perty | gabby |
| perty | lydia |
| perty | peter |
| perty | werner |
| aderProperty | hagent |
| aderProperty | python |
| aderProperty | viewer |
| perty | pythonService |

Figure 3.6: Scripts development environment in VHD++.

observe that realism is not a key factor for presence and communication in IVE. But on the other hand, the population of SP patients we target consists mostly of post-adolescents used to video games and animation movies, and thus very critic and demanding toward computer generated images. For example, using static photographs or polygonal models to represent humans is possibly sufficient to simulate a social communication, but will provide much less engagement from the patient than visually appealing characters. We tried to be pragmatic but dedicated to social phobia therapy; using standard 3D features, the rendering should provide pleasant ambiances while focusing on engaging virtual humans (VH).

### Optimized 3D rendering

Compromises between real time constraints and visual quality must be found in order to guarantee sufficient believability of the content for the user. the rendering software should get advantage of hardware optimizations and perform efficient culling in a scene graph as well as local optimizations (e.g. levels of details (LOD), impostors, etc..). We programmed the viewer component of VHD++ with the Open Scene Graph (OSG)[14] library as it provides all these features and can handle very complex scenes.

Today, there is no important hardware specificity to support 3D rendering as powerful machine with high graphic capabilities are affordable with a decent budget. However, stereoscopic rendering requires some special functionality from the graphics card. First, the computational power has to be almost double than for single viewpoint rendering. Second, a quad-buffer is necessary to enable the simultaneous OpenGL dual-buffers of the two points of view (e.g. NVidia Quadro models support active stereo for shutter glasses or passive for HMD). Finally, the rendering module has to enable and control this functionality through OpenGL since tuning mathematical parameters of 3D projections is required to adapt to the display configuration (Cahen, 1990). For the needs of VR applications, our OSG viewer module was also responsible for the management of cameras projection settings (including stereoscopy) and the rendering of basic OpenGL visual effects (lights, fog). All our developments were done on a PC with dual Pentium 4 at 1.7GHz, with 1GB RAM and NVidia[15] Quadro FX 1000 graphic card.

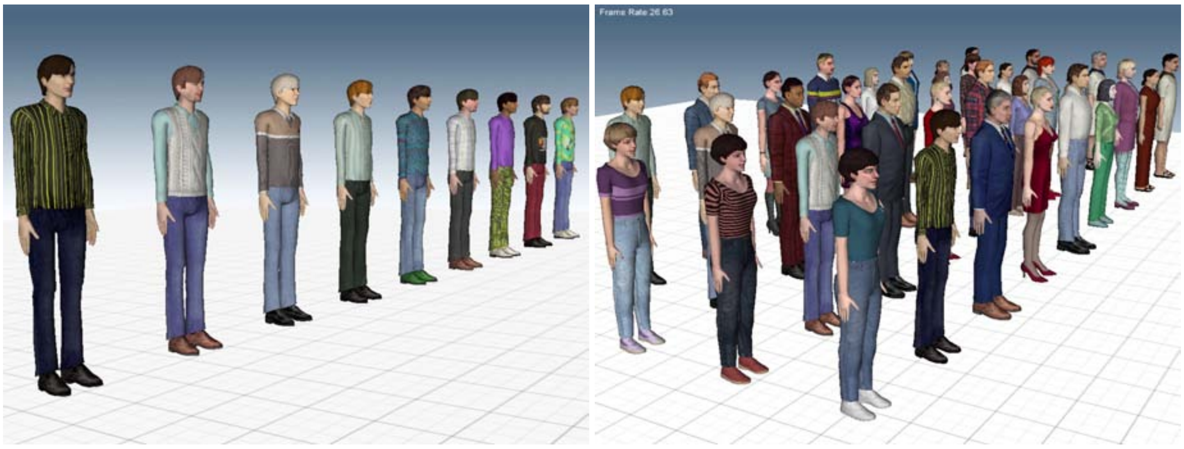
### Rendering virtual humans

The display of animated characters is one of the most resource consuming 3D rendering process. To display properly a virtual human shape, the rendering engine has to apply at each frame the transformations from the animated skeleton to the body mesh while respecting visually correct skin deformations. As this is another research area, we only integrated in

---

[14] http://www.openscenegraph.org

[15] http://www.nvidia.com



(a) Multiple textures of a model (b) Multiple models with various textures

Figure 3.7: Variety of virtual humans.

our viewer the optimized OpenGL rendering of virtual humans done by de Heras Ciechomski *et al.* (2004). It integrates weighted dynamic mesh deformation, dynamic levels of details, variety in the texture mapping, and interface for the skeleton animation.

From the design point of view, the elaboration of humanoid 3D models should respect limited mesh complexity, small texture sizes, and careful definition of mesh deformation data —more precisely around the articulations. The elaboration of each model is long and requires a good expertise, but thanks to the management of texture variety in the rendering module the designers could give various appearances (clothes and skin) to each mesh (fig.3.7(a)). Finally, with more than nine different meshes and four to height appearances each, we disposed of a quite large set of virtual humans which can be involved in a unique environment (figure 3.7(b) gives an idea of this variety).

Concerning the performances, the engine could display on screen up to 30 animated high LOD VHs with interactive frame rate (fps > 25 in fig.3.7(b)). However, the rendering was in general much faster as in real simulations, humans are not all together close to the camera —either hidden by the scene or far and in a lower LOD. The bottleneck is generally not rendering, but rather the skeleton animation which was in general more CPU consuming.

### 3.2.3 Character animation module

The simulation of virtual characters requires animation engines in order to have VHs perform life-like attitudes and gestures. Since we consider that quality of animation has a great impact in a VRE for social phobia, we needed a rich and configurable animation engine. Systems used in game engines (e.g. HalfLife or Quake) providing pre-designed key-frame animations of mesh are not generic enough. Newer game engines now use skeletal animation (e.g.

HalfLife2, OGRE[16]) which allows multiple characters models to play the same animations. However, even the most flexible animation libraries like Cal3d[17] have strong limitations in the management of actions —typically, they rely on the full-body weighted blending of key frames requiring precise design and preparation of the animations.

To overcome these limitations, we use a customized implementation of the integration framework for heterogeneous motion generators originally described by Emering *et al.* (2000). We also adapted it to the rendering layer by integrating the properties between the two modules in VHD++.

Our library provides a flexible management of multiple simultaneous motion-generation algorithms for a single agent. It handles the hierarchy of 72 joints of the HANIM[18] standard skeleton, including eyes and mouth. Simultaneous actions are handled in a priority stack and share their influence on the skeleton into scopes of action (e.g. an agent walking can play a higher priority key-frame on the upper body scope without any need for changes on the full-body walking action). Automatic and configurable motion blending for transitions make the visual result smoother. The other main advantage of using a proprietary engine is the extensibility of motion generators. We dispose of a dozen of specialized actions (SA) coming from state-of-the-art research in computer animation. The most widely used SAs allow for key-frames execution and high level control of walk thanks to the versatile engine of Boulic *et al.* (2004). The most advanced and specific SAs integrate algorithms from Baerlocher & Boulic (2004) and Le Callennec & Boulic (2004) for full body inverse kinematic, from Glardon *et al.* (2004) for generic locomotion by extrapolation of motion-captured data, and from Peinado *et al.* (2004) for real time motion capture.

For the needs of VRE for social phobia, we developed advanced gaze control and facial expression actions (fig.3.8). The gaze control SA animates neck and eyes to follow a given target or direction of gaze. It is quite realistic as it respects the cervical vertebrae constraints, the repartition of movement between eyeballs, eyelids and neck. It also performs automatic eye blinks and randomized eye saccades. The basic facial expression SA performs simple facial animations by moving the eyebrows and opening the mouth. It is mainly intended to animate faces of agents when speaking, but can also give some expressions to the face of the VH.

The management of numerous agents was handled at a higher level in a VHD++ animation module. It was provided with various graphical interfaces to help the developer or the designer in the control of VHs animations. Naturally, there was a performance limit in the amount of VHs to be integrated in a scene. For our simulations, we considered the situations with small to medium groups of humans (< 50).

---

[16] http://www.ogre3d.org

[17] http://cal3d.sourceforge.net

[18] http://h-anim.org

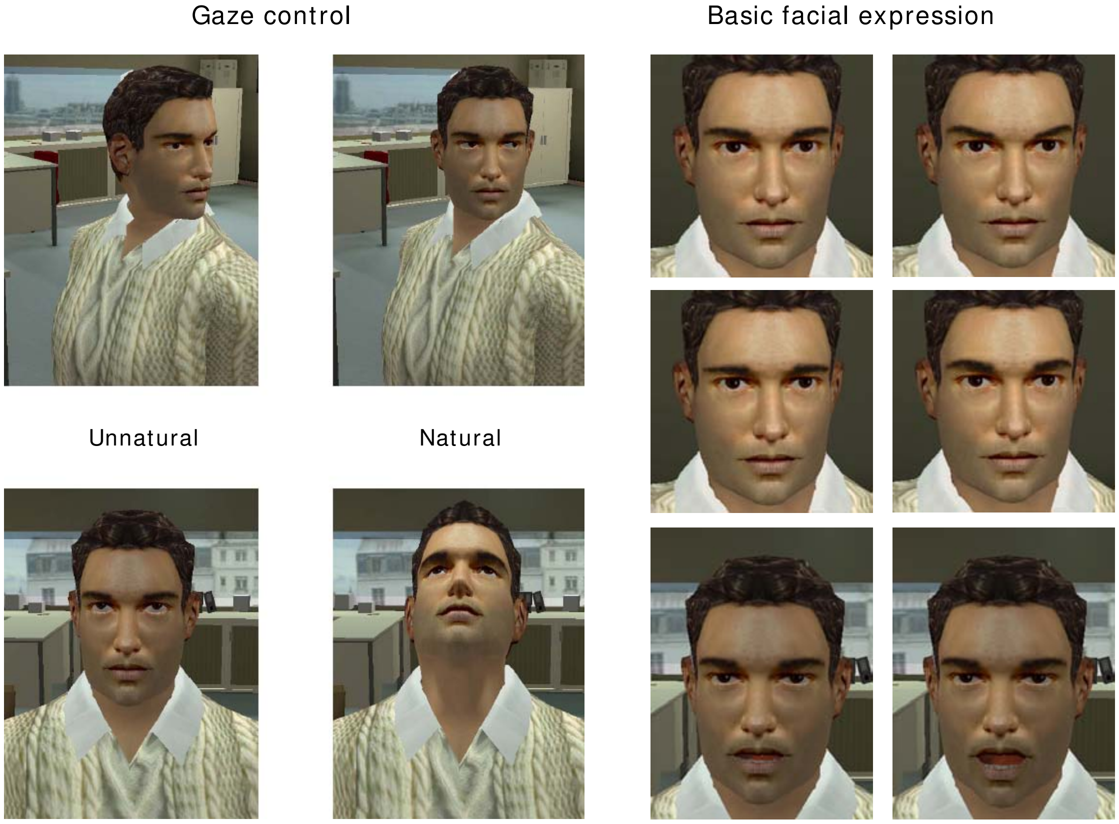


Figure 3.8: Gaze control and facial expression actions.

### 3.2.4 Scripting modules

To satisfy real time performances, 3D rendering and characters animation had to be develop at a low level with compiled languages (C++). But for many higher level tasks, the main constraint is not performance but rather flexibility and adaptability. Typically, characters behavior can be orchestrated in higher level processes which just operate on the animation engine.

We developed two different solutions for the middle layer interface of VHD++ functions. The first one consists in running scripts into an integrated Python[19] interpreter which can directly collaborate with other modules. The development of scripts was fully integrated in the platform and was simplified by a graphical user interface (fig.3.6). However, Python interpreters do not manage very well with multi-threading and the execution of scripts inside the engine slowed down the global performance (i.e. frame rate slows down). The other solution could overcome this problem by remotely calling modules through CORBA[20] from any program that could connect to the simulation engine (over network or on the same machine). This technique is complementary to the former as it provides a better sharing of resources

---

[19] http://www.python.org

[20] http://www.omg.org, http://omniorb.sourceforge.net

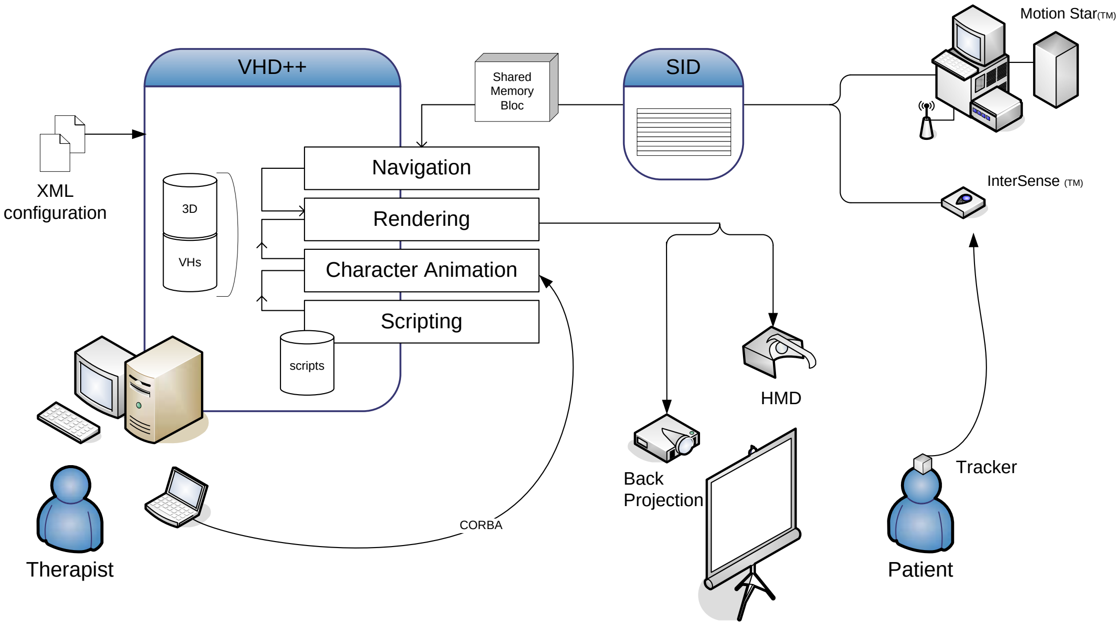


Figure 3.9: VRE platform software architecture.

into multiple applications and will not slow down the rendering. The bottleneck in this case was the limited communication bandwidth with CORBA and the necessity to engineer the software into well balanced processes.

Together, these two scripting possibilities offered all the flexibility for fast and efficient programming. The complexity of the scripts depended on the needs of scenarios but was not limited by the architecture. A typical usage of the Python scripts was to perform the initialisation of the scene within the embedded interpreter module and to run interactive scripts into dedicated processes using CORBA —potentially for managing sounds and high level control of behaviors.

## 3.3 Synthesis

The integration of all these modules into the VHD++ framework gave shape to a fully functional simulation engine. However, in order to become a VR platform, it had to integrate the support for the immersion setups. This was done by developing navigation modules connecting the hardware support (SID) and the rendering pipeline.

An overview of the software architecture of the VRE platform is shown in figure 3.9. All together, we obtain a modular system where multiple IVEs can be arranged by only changing run-time parameters. A unique executable can be adapted to the immersion setup by specifying the modules to be loaded with parameters in a single XML configuration file. A second XML file describes 3D models, VHs, animations and scripts to be loaded.

In this context, we called 'scenario' an imaginary situation proposed by a therapist as an exercise for a patient and supported in an IVE. The virtual environment gives shape to this scenario by providing the context and the actors and the VR platform brings to life the VE and immerses the patient into an involving experience. With our software platform, we covered the whole chain of development and design of VRE scenarios. We had the possibility to imagine and experiment with various exposure scenarios. We will see in chapter 5 how the different aspects of this system were used for the specific needs of VRET.

The technological complexity of VR being still a limiting factor for VRET, we proposed innovative solutions to improve the genericity of hardware support and the flexibility of software simulation. The achievement of this VRE platform required a progressive integration of joint developments which were also validated within other European projects, as detailed by Manganas *et al.* (2005) or Papagiannakis *et al.* (2005) for instance. Our objective was to provide VRET with high quality immersive systems with the same requirements than for any professional VR simulation field. Compared to software platforms recently developed for the VRET of social phobia (Lee *et al.*, 2002; James *et al.*, 2003; Roy *et al.*, 2003), we could enhance the simulation of VHs with more advanced visualisation and animation capabilities allowing for a larger variety of scenarios.

We consider this contribution to the field of VRE as an opportunity for therapists to go further into the experimentation with VR. The same stands for clinical use of assessment tools which may be well-known in medical and in engineering fields but which are rarely used by therapists because too constraining for everyday work. For example, in the experiments performed by Hofmann *et al.* (1997) and Horley *et al.* (2003), the observation and the analysis of gaze behaviors were either very laborious or required particular conditions. This is the kind of behavioral observations, done for the purpose of experimental psychology, that cannot be performed as precisely in the course of a therapy because of technical limitations. However, we saw in this chapter that such technological framework is already involved in the context of VRE. We will see in the next chapters how assessment tools could potentially be integrated into VR setups and why VRET could propose not only a replacement for in-vivo exposure but an integrated exposure and clinical assessment tool.

# Chapter 4

# Computerized clinical assessment tools

When conducting behavioral exposures, therapists have to interpret every signs in patients' reactions and may rely on different means to observe them. First, cognitive responses indicate the patient interpretation of what he experiences. We can obtain information on how he perceives himself in situation through spontaneous reports, interviews or questionnaires. Second, overt behaviors regroup the external expressions that can be directly observed in the patient's attitudes; we may focus on safety behaviors and on conditioning behaviors. Third, affective states concern the patient's internal reactions; they are best observed with physiological measurements.

The cognitive assessment is done by therapists and cannot be done with computers. However, the integration of emotional and behavioral assessment systems into VRE could be welcome, as far as such tools provide meaningful information for diagnostic or clinical evaluation.

## 4.1 Identification of affective states with physiological measurements

The objective of our first experiment was to provide an automatic evaluation of affective states that could be used during immersive VRE sessions. To this end, we investigated in the field of affective computing (AC) the possibilities to develop a system providing a simplification of the physiological data.

### 4.1.1 Affective computing; limitations and expectations

Picard (1997) has defined affective computing as computing that "relates to, arises from, or deliberately influences emotions"[1]. This covers the examination of media content –the

---

[1] http://affect.media.mit.edu

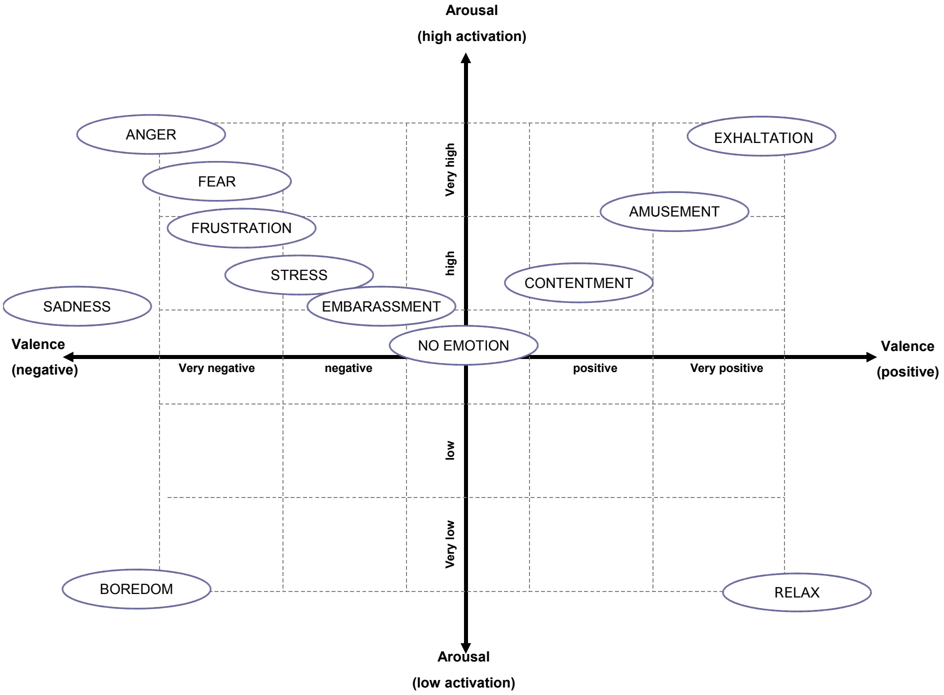


Figure 4.1: Arousal/Valence model of affective states.

stimuli– as well as the analysis of affective states –the reaction. The fully computational extraction of affective aspects in videos done by Hanjalic & Xu (2001) perfectly illustrates the first case. For the second case, the use of physiological measurements is often chosen to represent internal affective states. For instance, Wang *et al.* (2004) correlate the galvanic skin response to the intensity of the emotion. Affective computing experts propose several means and techniques to perform a transformation of human factors into computable data. Among them, the Arousal/Valence model retained our attention: it covers a large spectrum of emotions, it is widely used in psychology, and it has already been related to the Sense of Presence (Dillon *et al.*, 2000).

The simplicity if the Arousal/Valence model is indeed very attractive. As shown in figure 4.1, a large scope of emotions can be labelled with only valence (unpleasant or "negative" to pleasant or "positive") and arousal (drowsy or peaceful to exited or alert). However, this apparent simplicity is extremely subjective and the computation of those parameters hides a great complexity that Hanjalic & Xu (2001) simply overcame by selecting 'arbitrary' digital features. To avoid this, other researchers (Healey & Picard, 1998; Rani *et al.*, 2003) designed protocols to learn and optimise the correlations between physiological data and fixed classes of affective states. Whatever the algorithm used (respectively statistical or fuzzy), the principle of experimental affective computing remains the same: record various physiological signals, compute several parameters, and operate a classification/recognition.

According to these observations, it seemed possible to establish a computerized model of arousal and valence from physiological measurements, or even more precisely an identification into classes of emotions. However, in order to build such a tool, several constraints had to be dealt with: variability among individuals, identification of emotions, selection of physiological signals, selection of computational models, evaluation of reliability, etc. As a first approach, we designed an emotion induction protocol involving one actor over several sessions during which we measured multiple physiological signals. Statistical analyses have then be performed to correlate these data with emotional classes or at least the arousal and valence scale.

### 4.1.2 Experimental physiological identification of affective states

#### Principle

If isolated from the CBT-VRE context, the experiments we conducted simply consisted in recording the physiological signals on a person trying to self-induce five classes of emotions situated at the extremes of the theoretical arousal-valence model. In addition, the subject estimated his subjective arousal and valence each time to provide a reference point. We then tried to find the best correlations between several features derived from the physiological measurements and the theoretical or estimated arousal and valence.

#### Protocol

There is no ideal way to induce specific emotions and to ensure the person has actually felt the expected emotion. However, the Velten (1968) Mood Induction Procedure is still widely used to experimentally induce different moods; the subject is simply instructed to try to feel the mood expressed on a card and the experiment relies on his capability to reproduce it with memories. In order to take advantage from the ability of professional actors to deeply feel emotions when playing a role, Healey & Picard (1998) chose an actress for their experiments with emotions. We therefore equally asked a professional actor to perform our experiments. Moreover, by referring to a single person, we "maximize the chances of getting a consistent interpretation for each emotion" (Picard *et al.*, 2001, p1179).

During several self-induction sessions, the actor was asked to concentrate sequentially on five distinctive affective states, labelled according to basic emotions: “neutral / no emotion” (low arousal, neutral valence), “fear / panic” (high negative), “boredom” (low negative), “joy / meditation” (low positive), “exaltation” (high positive). Simultaneously, a Physio-Recorder from the Vienna Test System Corp.[2] was used to measure six physiological signals:

---

[2] http://www.schuhfried.at

(a) Physio Recorder (TM) and its sensors (b) Actor in a quiet corner for self induction sessions

Figure 4.2: Experimental setup for physiological observations of affective states.

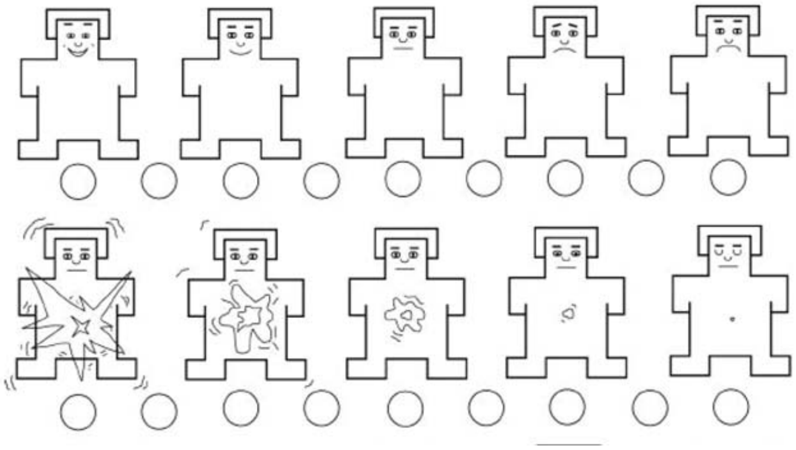


Figure 4.3: Iconographic SAM rating for Valence (top) and Arousal (bottom) (Morris, 1995).

- **SCL**: skin conductance level (electro-dermal activity, in micro-Siemens, $\mu S$),
- **HR & PVA**: heart rate and pulse volume amplitude (measured with a photoplethysmo sensor strapped to a finger, respectively expressed in beats-per-minute and in percents of volume change),
- **EMG**: frontal electromyography (*venter frontalis* EMG, in micro-Volts, $\mu V$),
- **BF**: breathing frequency (abdominal and thoracic respiration together, in resp-per-minute),
- **ST**: skin surface temperature (on non-dominant hand finger, in Celsius degrees, $^{\circ}C$).

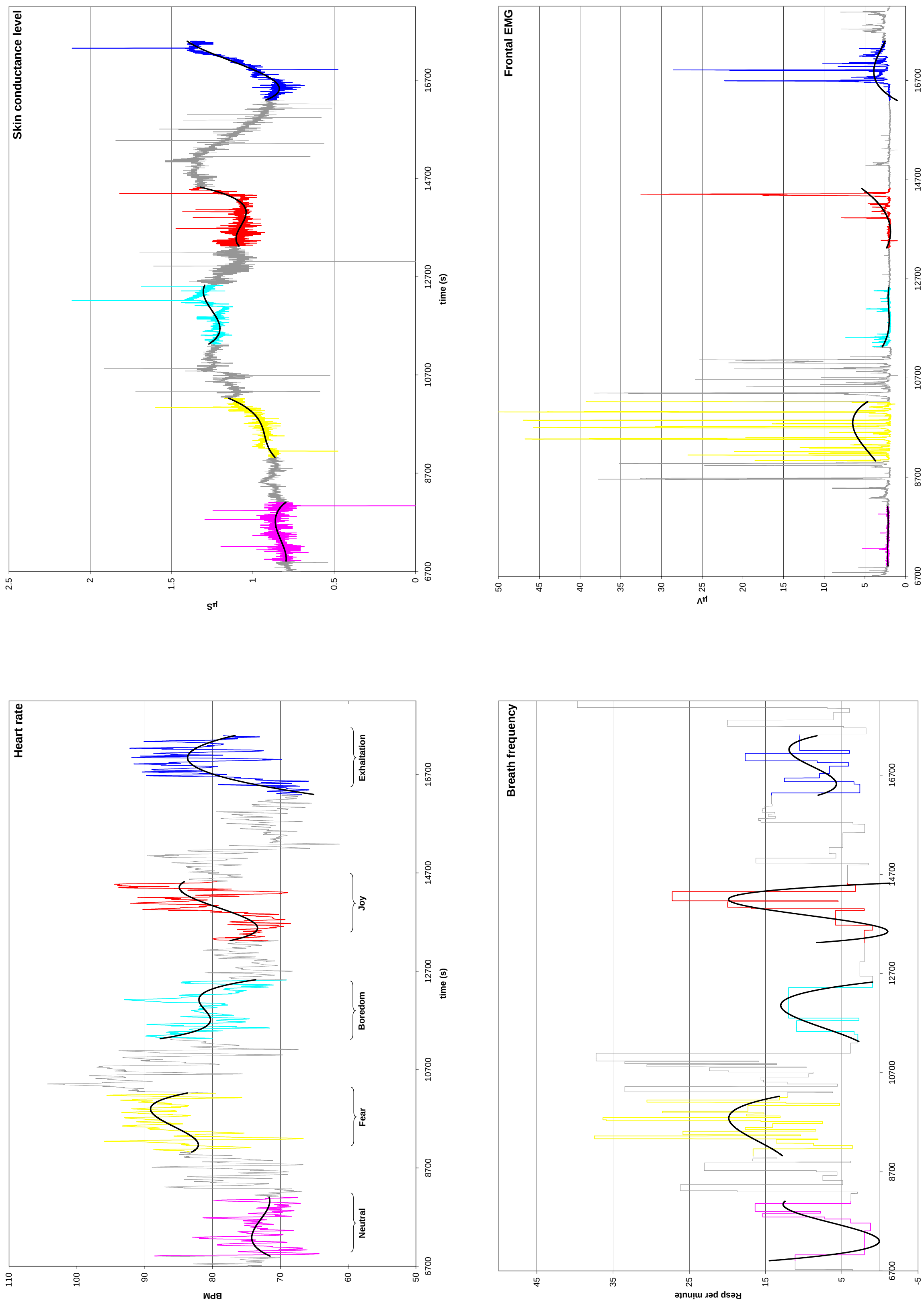


Figure 4.4: HR, SCL, BF and EMG signals during a session of self-induction of 5 basic emotions.

The setup for the recording sessions is shown in figure 4.2. Each session started with a relaxation phase requested by the actor and for sensors to reach stable assessment values (e.g. 10 minutes for surface temperature). Then, lying down, the actor started the induction procedure by acting on a guiding software: this application displayed the current emotion to be induced and timed two minutes before beeping. Then, a subjective arousal and valence evaluation screen proposed to select values on a 1-9 scale associated with the Self Assessment Manikin from Bradley & Lang (1994); Morris (1995) (fig.4.3). For each emotion, the timing and the subjective arousal and valence were stored. Examples of signals recorded during a session are shown in figure 4.4.

### Features extraction from physiological data

During the four months of experiments, we regularly made recording sessions of approximately thirty minutes (including the relaxation phase). Due to sensor failures, the first recordings were considered as a training period for the actor, and we finally obtained ten full sessions.

The physiological signals contained in the database were splitted according to the time sheet and regrouped by emotions (ten times five records of two minutes each). For each record $X$ we compute the six parameters proposed by Picard (2001) on the $N = 1200$ values (120 seconds at 10Hz): the mean of the raw signals (Eq.4.1), the standard deviation of the raw signals (Eq.4.2), the mean of the absolute values of the first differences of the raw signals (Eq.4.3), the mean of the absolute values of the first differences of the normalized signals (Eq.4.4), the mean of the absolute values of the second differences of the raw signals (Eq.4.5) and the mean of the absolute values of the second differences of the normalized signals (Eq.4.6).

$$m_X = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{4.1}$$

$$s_X = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\left(X_n - m_X\right)^2} \tag{4.2}$$

$$d_X = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|X_{n+1} - X_n\right| \tag{4.3}$$

$$\tilde{d}_X = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|\tilde{X}_{n+1} - \tilde{X}_n\right| = \frac{d_X}{s_X} \tag{4.4}$$

Table 4.1: Features selected for their best Pearson correlation with arousal and valence.

| | Arousal | | Valence | |
|---|---|---|---|---|
| *Signal* | *Feature* | | *Feature* | |
| Heart Rate | $s_{HR}$ | $r = .424,\ P < .005$ | $d_{HR}$ | $r = .255,\ P < .1$ |
| Pulse Volume Amplitude | $d_{PVA}$ | $r = -.371,\ P < .01$ | $\tilde{e}_{PVA}$ | $r = -.263,\ P < .05$ |
| Skin Conductance Level | $\tilde{d}_{SCL}$ | $r = -.491,\ P < .001$ | | - |
| Electromyography | $\tilde{d}_{EMG}$ | $r = .421,\ P < .005$ | | - |
| Skin Temperature | $m_{ST}$ | $r = .389,\ P < .01$ | | - |
| Breathing frequency | | - | $e_{BF}$ | $r = .242,\ P < .1$ |

$$e_X = \frac{1}{N-2}\sum_{n=1}^{N-2}|X_{n+2} - X_n| \tag{4.5}$$

$$\tilde{e}_X = \frac{1}{N-2}\sum_{n=1}^{N-2}\left|\tilde{X}_{n+2} - \tilde{X}_n\right| = \frac{e_X}{s_X} \tag{4.6}$$

### 4.1.3 Results

Among the 36 features collected (6 for each signal) over 50 records (10 sessions of 5 emotions), we needed to extract the most representative and a way to exploit them. We conducted two approaches in parallel.

#### Best features for arousal and valence

The first approach consisted in determining if numerical features could be correlated to arousal and valence as suggested in the literature. A simple Pearson correlation was computed between every feature and self-estimated arousal/valence in order to identify the most representative. Table 4.1 shows the ordered list of signals for which the computed features were correlated at best.

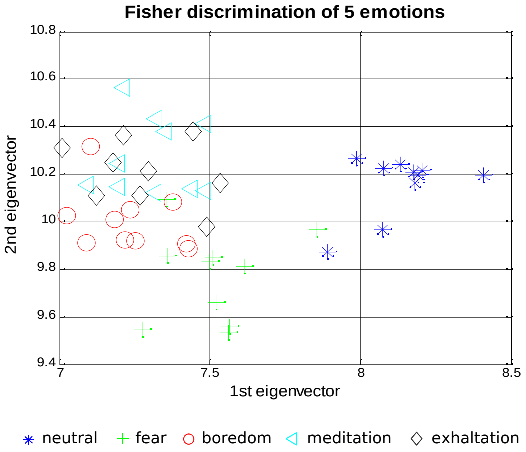
Our results are compliant with numerous studies that suggest that pulse (HR and PVA) and SCL signals are the most involved in the recognition of affective states. To be more precise, we first found that the standard deviation of HR is significant for the evaluation of arousal. In other words, the heart rate is regular when little aroused and is irregular when the affect is intense. As a side effect, the arousal of the affective state is also correlated with $m_{ST}$ and inversely correlated with $d_{PVA}$ –which represents an average variation of blood volume between two successive measures. This seems logical as when the heart rate is regular, the blood circulates better and the temperature is lower (and reversely). But, the best correlation for arousal was found with the SCL signal: the feature $\tilde{d}_{SCL}$ is high when the skin conductance varies in large steps ($d_{SCL}$) and regularly ($1/s_{SCL}$). A strong negative correlation with arousal suggests that intense emotions can be observed by irregular small variations of the skin conductance. Similar conclusions can be made from the positive correlation with $\tilde{d}_{EMG}$; high arousal can be observed by regular high variations of the EMG.

On the other hand, the correlations between physiological parameters and the valence of affective states are very low. This confirms the lack of publications proposing such evaluation. Once again, pulse is a good indicator but with different features than for arousal. Positive emotions (joy and exaltation) seem to provoke lower variation amplitudes in heart rate (low $d_{HR}$) than the negative ones (boredom and fear). Although quite low, this correlation is coherent with what Simons *et al.* (1999) suggested: "The relationship between stimulus valence and heart rate was linear [...] with the greatest deceleration associated with negative images and the least with the positive images" (p.622). The $\tilde{e}_{PVA}$ feature is harder to interpret; it is high when pulse volume varies in very large steps ($e_{PVA}$) and regularly ($1/s_{PVA}$). The negative correlation with valence suggests that negative states could be identified by irregular small variations of blood volume. We also found that breathing frequency was more correlated to valence than to arousal for every feature, with highest score for $e_{BF}$; the respiration frequency varies probably more with negative emotions than with positive ones.

To sum up, physiological signals indicate quite the level of arousal well but the valence of affective states rather badly. However, this lack of direct correlation could be overcome by a more complex combination of features.

### Classification into affective states

The second approach consisted in computing the discriminant function analysis of physiological signals similarly to Nasoz *et al.* (2003) and Vyzas (1999). Since we wanted to isolate the two functions which best fit to arousal and valence, we chose the two-group analysis called Fisher linear discriminant analysis. It allows dimensionality reduction by finding a linear projection of the entry data to a space of fewer dimensions where the classes are hopefully well separated. According to the dimensions of our learning feature matrix (fewer training points per class than the number of total features; the matrix is rank deficient), we had to apply a variation of the traditional Fisher projection algorithm by first projecting the data matrix into an ortho-normal basis $[N \times N]$ (where $N$ is the number of training points) and produce a matrix of full rank. We then tested this projection considering three kinds of class labels: 5 classes of emotions, 3 classes for arousal (low, middle and high values) and 3 classes for valence (negative, neutral and positive). Then, we focused on a resulting discrimination of the physiological features mapped on arousal and valence. Evaluation of



| original class | | KNN classification | | | | | |
|---|---|---|---|---|---|---|---|
| | | Neutral | Fear | Boredom | Meditation | Exhaltation | **Total** |
| | Neutral | **6** | 2 | 0 | 1 | 1 | **10** |
| | Fear | 1 | **3** | 1 | 1 | 4 | **10** |
| | Boredom | 1 | 2 | **2** | 1 | 4 | **10** |
| | Meditation | 2 | 1 | 5 | **1** | 1 | **10** |
| | Exhaltation | 2 | 3 | 1 | 4 | **0** | **10** |
| | **Total** | **12** | **11** | **9** | **8** | **10** | **50** |

| original class | | KNN classification (%) | | | | | |
|---|---|---|---|---|---|---|---|
| | | Neutral | Fear | Boredom | Meditation | Exhaltation | **Total** |
| | Neutral | **60%** | 20% | 0% | 10% | 10% | **100%** |
| | Fear | 10% | **30%** | 10% | 10% | 40% | **100%** |
| | Boredom | 10% | 20% | **20%** | 10% | 40% | **100%** |
| | Meditation | 20% | 10% | 50% | **10%** | 10% | **100%** |
| | Exhaltation | 20% | 30% | 10% | 40% | **0%** | **100%** |

Figure 4.5: Fisher discrimination of 5 emotions and its confusion matrix.

the discrimination has been carried out with the K-nearest-neighbors classification algorithm and the leave-one-out method has been chosen for cross validation. Here is the simplified procedure applied to each data point:

- the data point to be classified is excluded from the original data set and the remaining data considered as the training set,
- the subsequent fisher projection matrix is computed from the training set and both training data and testing point are projected down to the two best eigenvectors of the Fisher projection matrix,
- the data point is then classified according to the KNN principle based on the Euclidean common measure distance,
- and finally confusion matrices are calculated for the various classifications considered.

Results of the various classification protocols have very low classification rates. For example, when considering the five emotional classes (fig.4.5), only 24% of all the decisions of the algorithm lead to their original label. Neutral and Fear are best identified by our algorithm, although with weak success rates respectively 60% and 30%. Other classes are largely less than a random guess.

When considering valence as three modal classes, we obtained better results than for the five emotion classes. Figure 4.6 shows the results of Fisher discrimination of arousal and valence and allows to compare between theoretical (top) and self-reported classes (bottom) of arousal (left) and valence (right). With an acceptable classification rate of 45%, the three classes of negative/neutral and positive affective states appear as coherent, well distributed areas on the 2D Fisher plot. However, the classification with self reported arousal is too

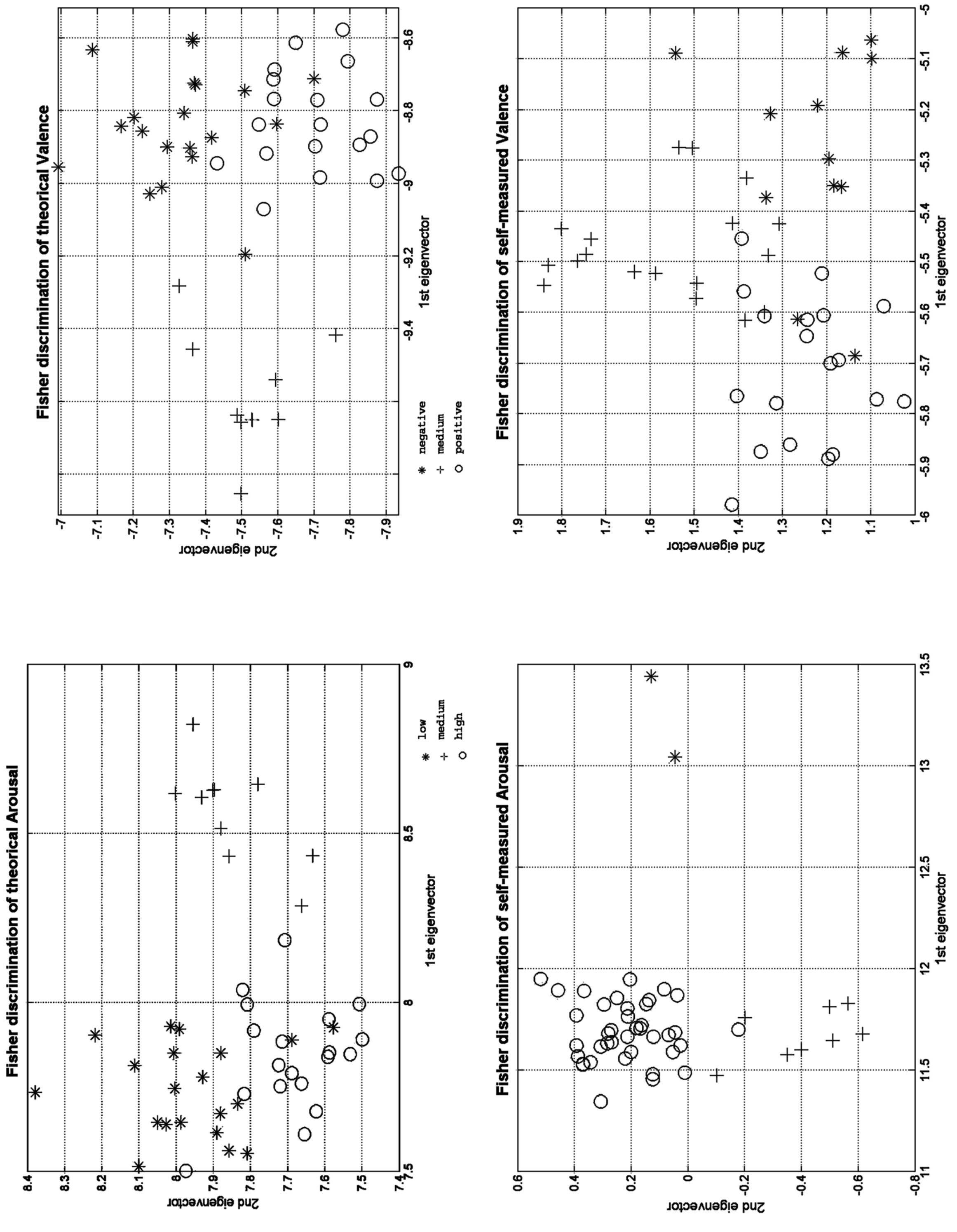


Figure 4.6: Fisher discrimination of arousal and valence.

unbalanced for a meaningful interpretation. In fact, we obtained almost one unique class in which 85% of the data is classified because the self reported arousal was always highly scored by the actor. This is in contradiction with the expected theoretical arousal values that should be, according to the repartition of the 5 emotions, equally splitted into low, middle and high arousal. This bias is verified by the very low correlation between the theoretical and evaluated arousal ($r = 0.19$). In the mean time, we could also notice a very high correlation between theoretical and evaluated valence ($r = 0.96$) which validates our previous conclusions regarding the validity of the classification of the valence component and which confirms its correct cognitive assessment.

### 4.1.4 Discussion

We have been experimenting with bio-feedback devices to develop computerized interpretation of physiological signals into arousal, valence and classes of affective states. Following a long term study on one person, we could establish numerical features providing a good estimation of the arousal. However, our procedure did not lead to statistically generalizable computation of valence or classification into affective classes. If our objectives had been to investigate the medical aspect, we could have pushed our investigations which already offered promising perspectives. For instance, the Fisher linear discriminant analysis on the 36 features could certainly be optimized to guarantee higher classification rates for the three classes of valence. Additional parameters and signals, such as heart rate variability (McCraty, 2002) and electroencephalogram, could also be considered to improve these results. But our goal was to design an on-line physiological assessment tool that could easily be integrated in an exposure session. As such, the features selected as best correlated to arousal give an objective numerical evaluation of the intensity of affective states. We did not manage to establish meaningful linear correlations of valence, but only a correct classification into three ranges. This would not help therapists much, who can obtain such non quantitative information more easily with cognitive assessment.

However, this experimentation allowed us to better establish the limitations of affective computing. We have been able to acquire an overview of what a review of the literature couldn't have told us; how to manage a procedure with biofeedback, how complex the interpretation of signals is, and which precisions we can realistically expect in 'normal' conditions —i.e. not in the specific medical experimental framework of the publications. First, in order to ensure a good quality of recognition, the data set has to be trustworthy. If with 10 sessions the learning set was too small to establish statistical validity, applying this procedure in therapy would be in contradiction with the goal to limit the amount of sessions. Second, results would not be conclusive if the method was used to identify emotions in an unknown context, but the system we built was sufficient to correlate reactions of the patient with a known stimulus. It could be used to assess the patient's response to visual or auditory events during VRE, and be eventually completed with behavioral observations.

## 4.2 Tracking of behavioral gaze avoidance

Behavioral therapists are used to request an active involvement of patients during exposure sessions which is also valid for VRE. The main advantage in the latter case is the possibility to track, record and analyze the patient's behavior during the simulation with computerized tools which give more information than the usual recordings on video tapes.

Based on this idea and referring to the observations made by Horley *et al.* (2003), the gaze behavior of social phobics should be observed during immersion as a characteristic of 'eye to eye' avoidance. For these reasons, we experimented with eye tracking technology to open the possibility to establish a statistical analysis of the patient's gaze focus. The objective of this study was to propose a better integration of gaze tracking into VRE conditions in order to detect when the patient is looking at a VH and more specifically looking at it in the eyes.

### 4.2.1 Eye-tracking in immersion conditions

Eye tracking consists in following eye movements and computing gaze direction into a computer system. According to the state of the art made by Yang *et al.* (2002), this technology became really usable in the late nineties. It is today available as commercial products. Meanwhile, various applications were experimenting with gaze-control simulations (Krapichler *et al.*, 1998; Rizzo *et al.*, 2002), interactive multimodal systems (Tsui *et al.*, 1998; Kaur *et al.*, 2003; Sibert & Jacob, 2000), or rehabilitation (Lin *et al.*, 2004).

To perform gaze analysis during a VR experiment, the solution originally developed for aviation was to integrate tracking cameras directly in the HMD. The accuracy and the reliability of these systems compensate for their high cost. Experiments conducted by Renaud *et al.* (2002b) with such equipments leaded to very detailed analysis of the behavioral dynamic of users' visual exploration in IVEs. Moreover, as the authors are also working in the field of VR therapy, their experiments confirmed the interest for performing gaze tracking during immersion. The analysis of gaze avoidance of spiders in the behavior of patients was a good example of the possibilities offered by this technique (Renaud *et al.*, 2002a).

#### Setup

The VisionTrak eye tracking system we used consists in several components: first a head-mounted assembly (ISCAN ETL-500) with two cameras, a dichroic mirror and an infrared eye illuminator, and second a PC with two specialized PCI cards implementing the video

(a) VisionTrack device with mounted 6 DOF sensor (b) User wearing VisionTrack and facing the back projection screen

*Photo Alain Herzog*

Figure 4.7: Eye tracking device for back-projection immersion.

processing hardware. The tracking system computes pupil and corneal reflection positions from the video data and computes the gaze direction on one eye to provide results to the ISCAN software used for calibration and data acquisition.

The large screen immersive setup was well suited for eye tracking as important eye movements are necessary to cover the entire screen. The user wearing the VisionTrack has to stand in front of the back-projection screen where the 3D scene is rendered. In order to let the user move during immersion, we need to track not only the eye orientation but also the head position and orientation in the workspace. We therefore added a 6-DOF magnetic sensor (wireless MotionStar from Ascension) to the head-mounted assembly (fig.4.7(a)). The whole setup is described in figure 4.8.

### System calibration and gaze target computation

Since there are differences in the way the VisionTrak device is worn by different users, the eye tracking needs its own calibration to compute $(x_{eye}, y_{eye})$ as coordinates of the target point in a $512 \times 512$ pixels square covering the field of view. This procedure requires that a supervisor enters five (or nine) calibration points on the VisionTrak PC while the user is asked to look at some calibration targets —one in the center and others to cover the field of view. The correspondence between calibration targets and points is visually estimated and adjusted by the supervisor. Moreover, in order to compute the 2D coordinates of the gaze target point on the display, the absolute physical reference frame of the magnetic system (the transmitter) is used as global frame in which screen, user head and gaze direction can be expressed. The screen configuration (corners coordinates) are statically set in the system by measurements. The position $\overrightarrow{p_{world}}$ and the orientation $q_{world}$ of the sensor on the user head

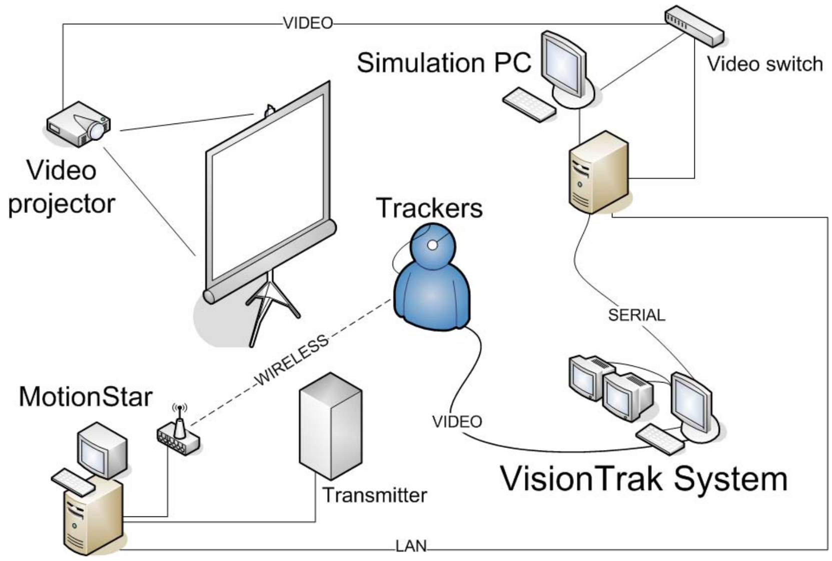


Figure 4.8: Experimental setup for eye tracking and back-projection.

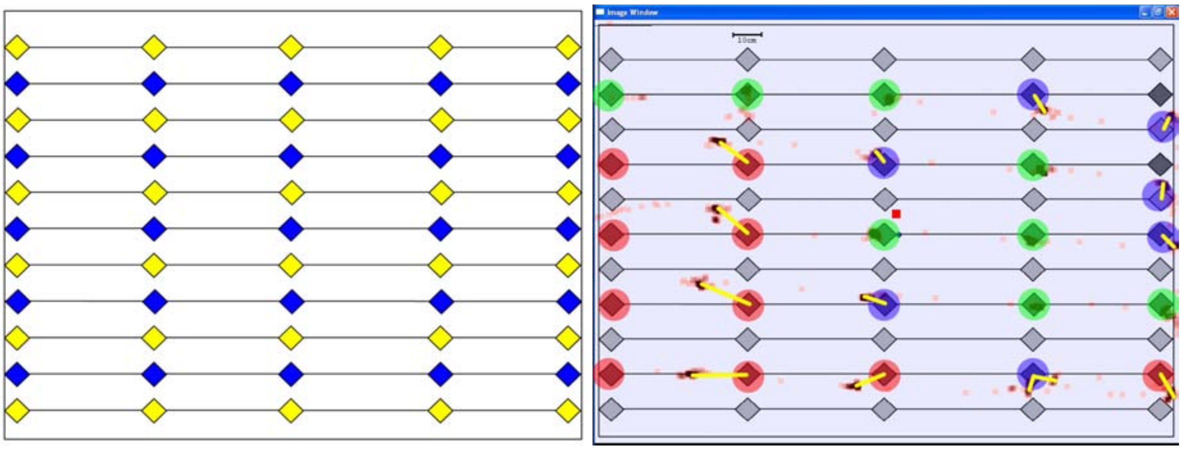
are directly expressed in this world frame. An additional calibration is required to specify the way the sensor is worn by the user. This is simply done by keeping the sensor orientation so that unit quaternion (no rotation) represents the user looking directly at the screen. We combined the two calibration procedures into one session where the user sitting in front of the screen is asked to look at five calibration points without moving the head.

The gaze target is then calculated. First, the gaze vector $\overrightarrow{g_{eye}}$ is defined as the direction of the segment starting in the eye and crossing the point $(x_{eye}, y_{eye})$ relative to the eye frame. Second, the head transformation matrix given by $\overrightarrow{p_{world}}$ and $q_{world}$ is applied to obtain $\overrightarrow{g_{world}}$ in the world frame. Finally, the gaze target $(x_{screen}, y_{screen})$ is computed as the intersection of $\overrightarrow{g_{world}}$ with the screen (equation of a plane).

### Experimental assessment of tracker accuracy

The accuracy of the combined tracker depends on many factors. Both devices introduce their own inaccuracies and there are also subjective factors affecting accuracy –placement of the device on the head, position changes (slips) while in use, concentration and overall fatigue of the user.

We first used a test pattern to assess the accuracy of the combined tracker. The pattern consists of alternating yellow and blue squares and the test subjects were asked to focus on every blue square for five seconds (fig.4.9(a)). The resulting data is represented as clusters of dots and displayed on top of the test pattern to measure the error distance between square targets and measured points. Figure 4.9(b) shows the result of an accuracy test; errors $e > 6.5\%$ are displayed in red, $3\% < e \leq 6.5\%$ in blue, and $e \leq 3\%$ in green.



(a) Test pattern (b) Tracking data in overlay

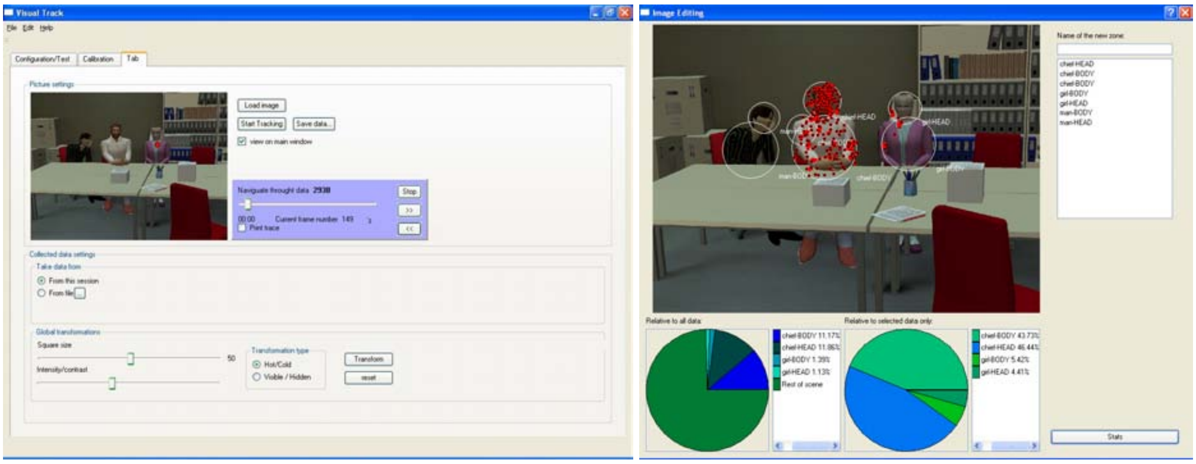
Figure 4.9: Gaze tracker accuracy test.

Table 4.2: Results of the reliability test over 20 subjects (diameter is in % of the screen width)

| Diameter | 13% | 6.5% | 3% | 1.5% |
|---|---|---|---|---|
| with cursor | 79% | 62% | 33% | 4% |
| without cursor | 74% | 59% | 27% | 2% |
| Standard deviation | 4% | 5% | 7% | 1% |

For another reliability test, the subject had to gaze successively at the centre of four differently sized circles which are shown in different zones on the projection screen. We measured the percentage of time during which the calculated gaze point stayed in the center of the circle. The procedure was performed twice, first with a cursor indicating the calculated gaze position and second without the cursor. This allowed us to assess the effect of visual feedback and the ability of subject to consciously “drive” the calculated gaze point to the indicated position.

The tests were performed with 20 subjects. The accuracy tests have shown that the error was not uniform and averaged 5%. The results of the reliability test summarized in the table 4.2 shows two facts. First, the presence of the visual feedback with a cursor had a very limited effect; therefore it may still be useful tool to improve the accuracy, but we can avoid it without loosing accuracy. Second, the standard deviation was low and the results were reliable with good repeatability and random errors were rare. In short, objects around 13% of the width of the screen could be hit with a 80% probability on average.



(a) Gaze record/replay control GUI (b) Statistics by zones of interest

Figure 4.10: Gaze attention analysis tools.

### 4.2.2 Basic gaze behavior analysis

In order to experiment with eye tracking during public speaking simulation, a graphical tool was originally designed with three main features: record/replay the gaze point (fig.4.10(a)), display the gaze trace in overlay as clusters of colored points, and compute statistics on the time spent in user-designed zones (fig.4.10(b)). The therapist started to use this tool to keep records of patients' performances and to observe their improvements. We have observed that the device offered sufficient precision to distinguish not only who the user was speaking to but also which part of the body he was looking at –although distinguishing the part of the face (e.g.the eyes) is only possible if the user gets closer to the virtual character.

### 4.2.3 Gaze map picking on virtual humans

The eye tracking technology has been validated in immersion hardware conditions, but the gaze analysis was still done on a 2D screen area (drawing static interest regions) and did not support camera movements during the immersion. Following is the 3D picking technique solution we developed allowing free exploration in IVE and gaze tracking on moving articulated humanoid targets.

#### Texture picking for humanoids

Picking in 3D is generally used for vertex selection and mesh deformation in modeling software. To have the ability of picking on any part of the 3D model, the mesh has to be extremely dense. Indeed, using OpenGL picking method `glRenderMode` with `GL_SELECT` or `GL_FEEDBACK` only gives per primitive precision such as entire object or triangles on

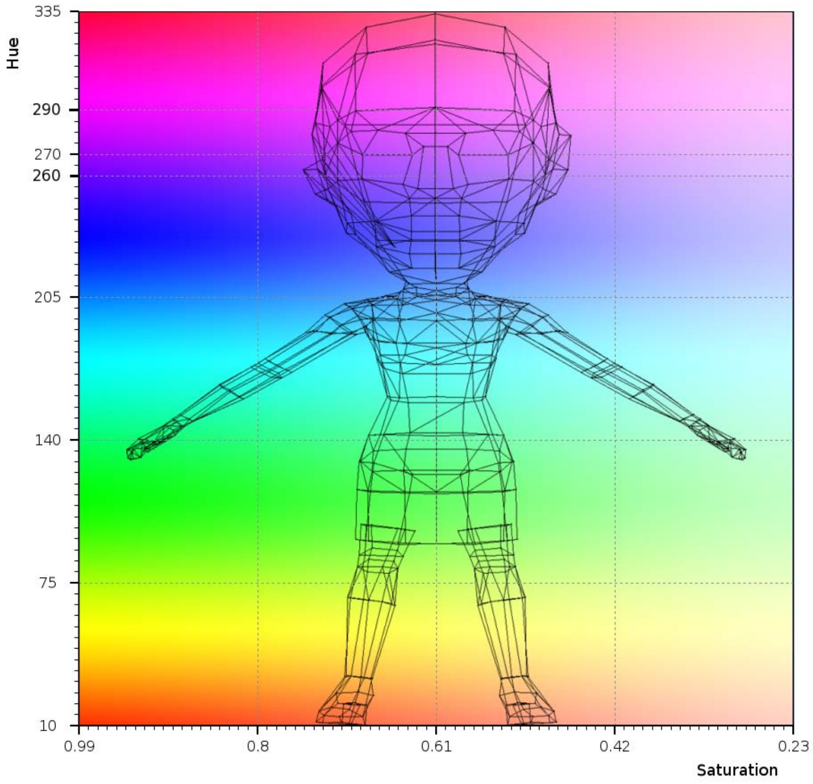


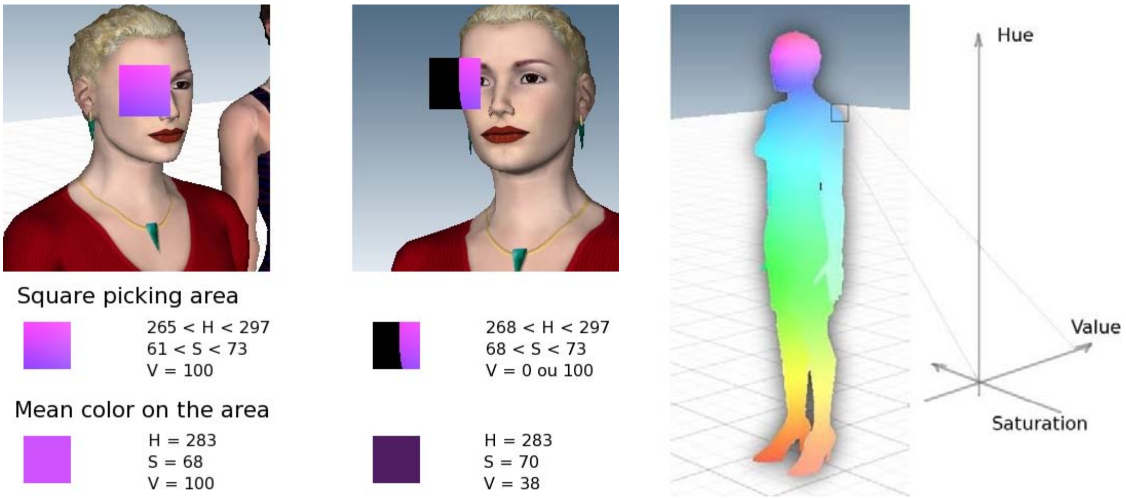
Figure 4.11: Gaze Map on a humanoid shape; the Hue-Saturation gradients texture has a unique color for each body point.

the mesh. To improve this technique, Geiger *et al.* (2003) color the vertices, render using OpenGL color smoothing between them, and convert the color of the pixel picked into coordinates on the 2D local surface.

Texture based picking offers similar advantages as in this case the picking precision is as detailed as the resolution of the picking texture itself, no matter the resolution of the mesh. Its principle is simple; a bitmap is designed to identify some zones of interest using specific colors and the picking is done by reading the color of the pixels at the desired coordinates on the rendered view. The mapping between color and mesh is then straight forward. Since we focus only on humanoids, this texture is applied carefully on the virtual humans in order to fit the interest zones onto the interesting parts of the body.

### Color hue-saturation body map

The main idea of the gaze map is to have a unique color for each part of the body so that the HSV components on the texture identify a unique body part. This is done with a color gradient texture with vertical variations of Hue and horizontal Saturation changes. Figure 4.11 shows how the texture is applied on the 3D model to cover the entire body; low hue values



(a) Value changes when the area does not cover the model but Hue and Saturation are preserved.

(b) Value; the third dimension of picking data.

Figure 4.12: Mean HSV color on a large picking square.

Figure 4.13: Mean HSV color on a large picking square; singularity when 2 humanoids are picked.

in the feet and high for the head, low saturation values for the left hand and high on the right. It is also quite easy for the designer in terms of mapping since a very simple texture is applied with a front-view planar mapping (symmetric for the back). We added some distortions around the face to have a more detailed view of facial regions such as eyes and mouth. Note that the rendering is done in RGB, but the conversion to HSV is simple (c.f. algorithm in Foley *et al.* (1990)). Low saturation was carefully avoided in the mapping since for $S = 0$, color is white for any value of $H$ and the conversion from RGB could introduce artifacts.

### Area picking with pseudo-distance

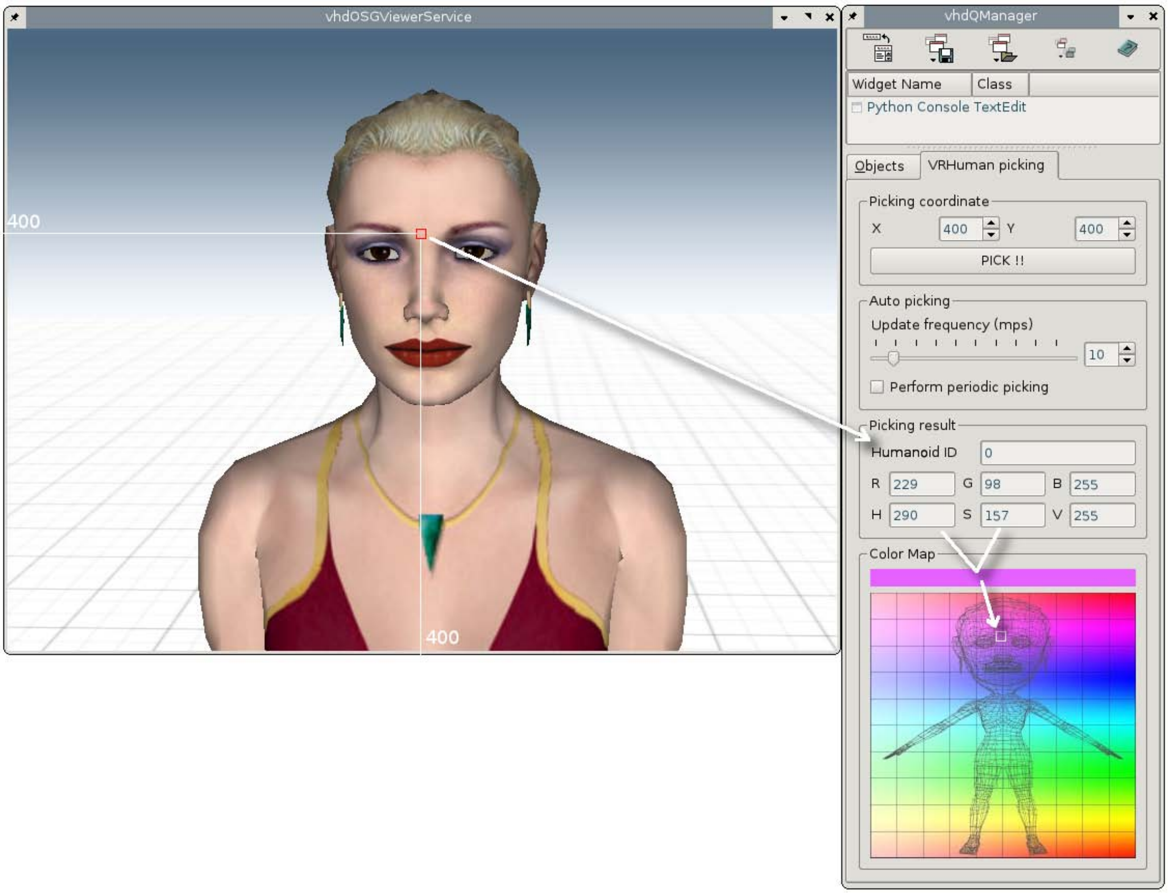
In the methodology explained above, single pixel picking was considered for simplification. However, for gaze tracking applications, we need more flexibility to compensate for the

Figure 4.14: Low LOD picking meshes animated with the humanoid skeleton.

low precision and the instability of the gaze target. For this reason, we render a square region of size $D$ pixels around the estimated picking point and compute the mean color of the picking area. When computing this, the (*H*)ue and (*S*)aturation values still indicate the average location on the body map. Moreover, as the background is cleared to black, only the Value component of the HSV color is affected when the area does not cover the model totally (fig.4.12). The value of $V$ is a percentage of pixels covered by the rendered 3D model in the picking area ($V = 0\%$ in the background and $V = 100\%$ on the map). As a consequence, $V$ can be considered as a pseudo-distance to the model, with $V = 100\%$ means the picking is done at less than $\sqrt{2}.D/2$ pixels inside the model, and $V = 0\%$ means the picking is done at more than $\sqrt{2}.D/2$ pixels. However, this feature requires that picking to be done on individual picked areas per humanoid. Indeed, the mean color of a unique picking area covering two characters would not indicate a body location anymore, and the color value would also be altered (fig.4.13). Therefore, we perform the picking individually for each humanoid. For a given picking location, multiple human picking data can be non null. Only the depth buffer can then resolve which humanoid is actually seen.

### Implementation

Color picking is performed on demand inside the rendering loop during an additional but invisible rendering phase into the back buffer. To pick on a humanoid, a fast rendering of the mesh with its picking texture is done inside a dedicated OpenGL viewport fitting to the picking area. The averages of the Red, Green and Blue components are computed over every pixels, and the values are stored in a table. Since the rendering to the picking buffer is made independent from the final visual representation, we use a simplified version of the mesh for the display of the picked human. The same skeleton is used in both cases so the animation looks the same and the coverage of the mesh is almost the same thus giving a boost in performance (fig.4.14). To have accurate colors we turn off lighting and any other color modulating steps performed on the frame buffer, and the texture file used is in an



☐ Python Console TextEdit

Picking coordinate
X 400 Y 400

Auto picking
Update frequency (mps) 10
☐ Perform periodic picking

Picking result
Humanoid ID 0
R 229 G 98 B 255
H 290 S 157 V 255

Figure 4.15: Gaze Map picking application with control panel GUI.

uncompressed TARGA format. Otherwise we would not be able to map the pixel colors one-to-one with the original picking texture.

### 4.2.4 Results

The implementation of the gaze map picking runs in real time and provides the unique identifier of the virtual human together with the color coordinated of the picked body part. Figure 4.15 illustrates how it was integrated into the VHD++ environment. However, we had to conduct more experiments in order to validate its reliability when used with the eye tracking system.

#### Validation of gaze map with eye tracking

For the needs of the validation of the gaze map picking, we used the virtual environment figuring a public speaking exposure in front of a jury. The screen resolution was $800 \times 600$ and the picking area size $D = 2^5 = 32$ pixels, thus ensuring the max picking precision is

Table 4.3: Repartition of gaze per agents: mean color Value for each viewpoint over the whole exposure duration.

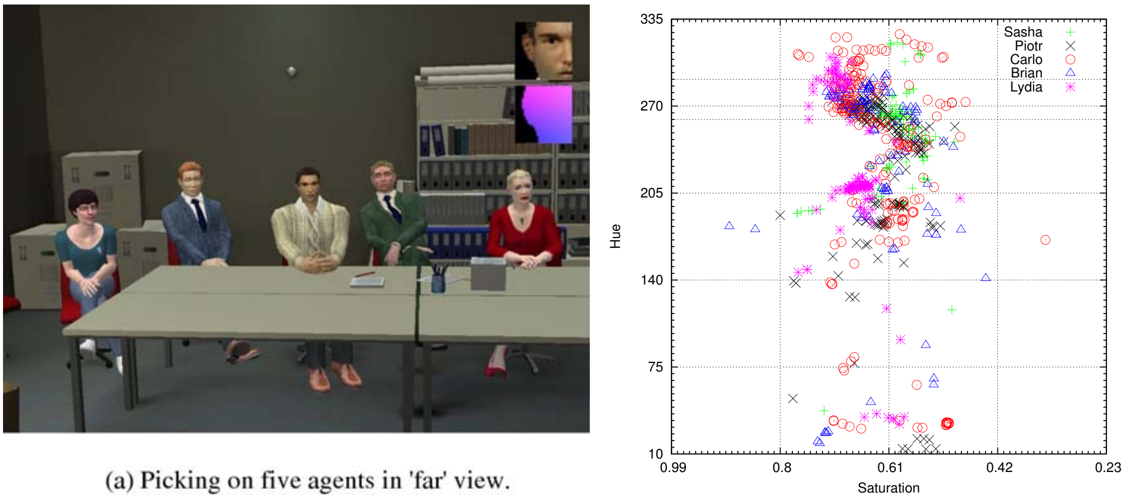
| | Sasha | Piotr | Carlo | Brian | Lydia | Total |
|---|---|---|---|---|---|---|
| Far | .10 | .18 | .22 | .10 | .11 | .78 |
| Medium | .0 | .0 | .39 | .26 | .25 | .89 |
| Close | .0 | .0 | .95 | .0 | .0 | .95 |

Table 4.4: Repartition of gaze per body parts: $\%$ of picking for each viewpoint over the whole exposure duration.

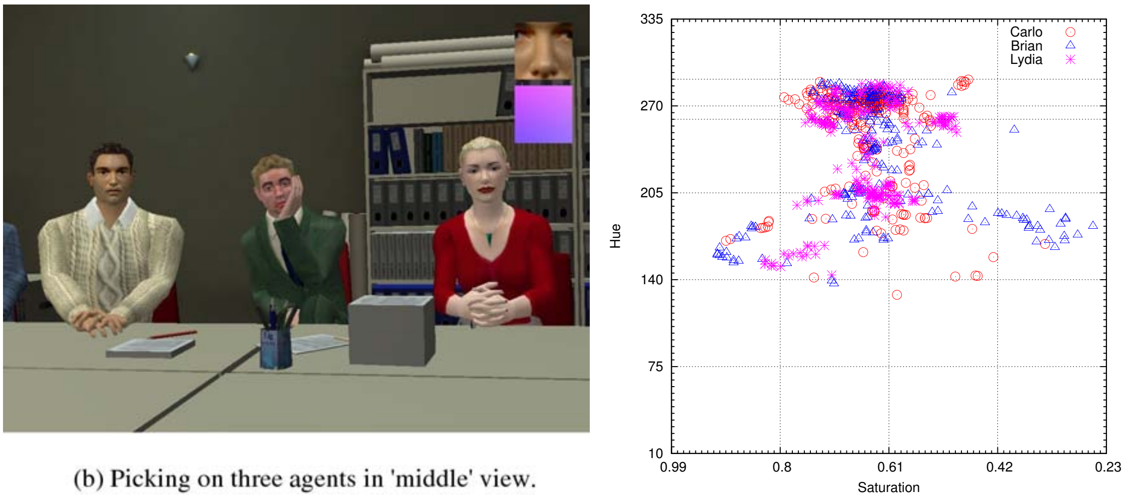
| | Hairs | Eyes | Mouth | Neck | Shoulders | Torso | Hip | Thigh | Knees | Feet |
|---|---|---|---|---|---|---|---|---|---|---|
| Far | 7.6 | 28.2 | 33.6 | 9.7 | 15.9 | 2.6 | 1.0 | 0.3 | 0.3 | 0.8 |
| Medium | .0 | 43.6 | 28.3 | 6.6 | 14.8 | 6.0 | 0.6 | .0 | .0 | .0 |
| Close | 2.2 | 57.1 | 21.0 | 2.2 | 11.8 | 4.5 | 1.2 | .0 | .0 | .0 |

greater than the tracking accuracy ($\sqrt{2}.D/2 = 5.6\%$ of the screen width). Three subjective positions toward the jury were selected to experiment with various targets of size 100 pixels (approx. $12.5\%$ of the screen width). In the far view (fig.4.16(a)), the user can see the five VHs (named, from left to right, Sasha, Piotr, Carlo, Brian and Lydia) and each one is around 100 pixels large. In the medium view (fig.4.16(b)), the user is close and sees only three humanoids and their head is less than 100 pixels large. In the close view (fig.4.16(c)), 100 pixels cover approximately a quarter of the face of a humanoid.

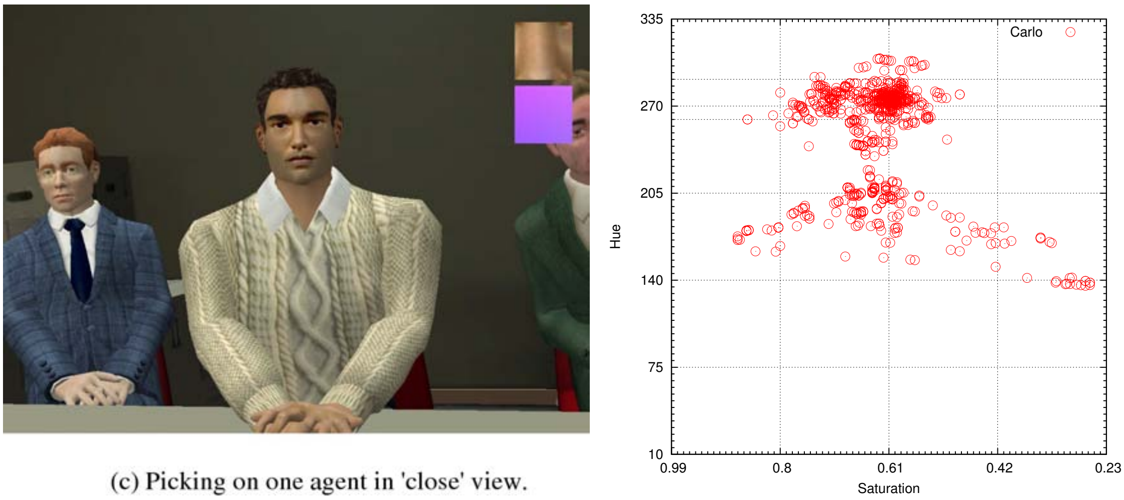
Data were recorded during three immersive sessions lasting approximately 2 minutes for each subjective point of view. Subjects were not social phobics and were asked to address every character while trying to keep focus on their eyes. The values given here are taken from the sessions done with a 25 year old male subject. Table 4.3 gives the means of the $V$ component per humanoid to indicate the relative intensity of gaze over the agents. As we managed to avoid overlap in the picking area between humanoids, the total over every humanoid (last column) represents the overall normalized distinction between humanoid and non-humanoid gaze targets (complement to 1 of the total). Table 4.4 gives the time spent to look at each body part normalized over the whole exposure duration. Each body part of this decomposition corresponds to a $30^o$ chunk of $H$. Finally, the $(S, H)$ points for the whole session for every visible humanoid are plotted on gaze maps in figure 4.16.



(a) Picking on five agents in 'far' view.



(b) Picking on three agents in 'middle' view.



(c) Picking on one agent in 'close' view.

Figure 4.16: Gaze map of picking data with eye tracking for three points of view.

### 4.2.5 Discussion

Concerning the need for therapists to automatically establish the repartition of gaze targets over humanoids, table 4.3 shows that the Value component can numerically represent the repartition of attention over the characters. Moreover, this is much more flexible and reliable than a logical picking as $V$ indicates a percentage of pixels picked in the area, and thus its mean over the exposure session really represents the intensity of the gaze on a character. The percentage of time avoiding them can also be provided by computing the complement to 1 of $V$ over the total. The second requirement was to automate the analysis of gaze centers of interest over the body. By taking advantage of the linearity of the texture gradient, a simple decomposition in regular intervals of Hue already provides an easily readable repartition of attention over body parts (table 4.4). The $(H, S)$ plots can then be used for more subtile interpretations, such as resolving the ambiguity between hip and hands, or comparing data between humanoids or over sessions.

We could also observe that the values obtained with these measurements are consistent with the subjects gaze behavior during the session. First because we could continuously check the picking area to make a parallel observation of the gaze behavior –displayed in the top right corner of the screen in figure 4.16. Second because data matches the subjective appreciation of what the subjects did during the sessions. For example, the subject related that he had successively looked at each person during the same time, and then came back longer to the central character. This is confirmed by the uniform repartition of $V$ over characters with a higher value for Carlo (first line in table 4.3).

The eye tracker and the gaze map picking technique provide sufficient precision for an analysis of gaze avoidance behavior. Taking the example of the close-view session, the last line of table 4.4 very clearly shows that the non-phobic subject could focus on the eyes region ($260 \leq H \leq 290$) and figure 4.16(c) provides even more visual details on other gaze targets (e.g. arms and hands of the agent). We could also observe the precision limits of the eye tracking system when looking at small targets. In close view we can distinguish gaze movements over the face, whereas when getting far from the characters the repartition of gaze targets becomes more fuzzy (fig.4.16(a)). One could argue, but none could verify, that the subject actually had this fuzzy gaze behavior; it seems more reasonable according to the accuracy tests we did to consider that this was due to the eye tracker's inaccuracy. As such, we recommend the middle point of view as a good compromise between precision and reliability of the values (fig.4.16(b)).

In order to improve the system, it could be useful to add a filtering of the gaze behavior before interpreting in order to remove eye saccades or to recalibrate according to the potential targets. The gaze-map algorithms should also be improved. A first enhancement of the graphical algorithm would be to use the original mesh in whatever LOD being deformed and re-use the unique mesh deformation for the picking using knowledge of shared vertices. Moreover, as normals are currently still deformed for the picking humans, a second enhancement would be to discard them completely for the picking. Combining picking results with depth test should also resolve conflicts in crowds of characters.

## 4.3 Synthesis

We have been investigating two possibilities to integrate assessment technologies in the process of virtual reality exposure sessions. We have first shown that we can help therapists in the difficult interpretation of physiological signals by providing an objective, quantitative and meaningful computerized model of the arousal. We are optimistic regarding the future improvements in the field of affective computing which may soon offer simplified and richer interpretations of affective state's valence.

We have then evaluated a gaze tracking technique to perform behavioral analysis during immersion. Considering the appropriate conditions we defined experimentally, eye tracking can be used during VRE to 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. The gaze-map picking technique we proposed has been validated within this context and for the analysis of eye-to-eye contact avoidance with social phobic people. Moreover, since applying this color mapping technique to generic models would be easy, we believe that such a tool could improve the efficiency of therapeutic protocols for various phobias.

There are definitely many interests for such assessment during therapy. As we will see in the next chapter, the first beneficiary are the therapists who can obtain enhanced feedback on their patients disorder and perform more efficient observations and diagnostics. The patients could also take advantage of this feedback in the form of encouraging evaluation of their performance and progress. As pointed out by Rizzo & Jounghyun Kim (2005), such a performance feedback is already successfully applied for VR physical rehabilitation and we believe that similar principles could be applied for CBT.

# Chapter 5

# Therapeutic experiments and results

Various studies were designed and performed to validate the efficiency of our VR tools in exposure sessions and results were used as feedback to the research process. The VRE platform as well as the computerized assessment tools were developed accordingly, in parallel to these experiments. We tried to provide answers to questions regarding the usability of a VR system and its potential impact for psychotherapy. Moreover, we equally experimented with our tools in order to validate their pertinence and their efficiency in clinical usage. This chapter will present the three types of studies we carried out; a first experimentation to validate the basic principles of VRE, followed by clinical case studies involving our developments progressively, and finally a validation study on a larger cohort. The global analysis of all the results will be presented in the next chapter.

## 5.1 Preliminary study

This study was the starting point of our collaboration with Dr. F. Riquier from the University Department of Adult Psychiatry[1] in Lausanne. At this time, both medical and technical experts had numerous questions regarding the possibilities offered by VRET but both agreed on the need for a study on the use of VR for the therapy of SAD.

### 5.1.1 Objectives

The main goal of this preliminary study was to verify the possibility of provoked anxiety during immersive VR exposure session. Moreover, as we already targeted the treatment of social phobia, we wanted to experiment with the specific stress of a public speaking situation. When comparing with existing systems targeting other phobias such as fear of spiders (Carlin *et al.*, 1997; Garcia-Palacios *et al.*, 2002; Hoffman *et al.*, 2003), we noticed that realism

---

[1] http://www.chuv.ch/psy/dupa.htm

was not necessarily the key to success, but rather the reproduction of the anxiety provoking stimulus (spiders in this cases). According to Riquier *et al.* (2002), the key stimulus for social phobia is gaze, and more specifically the subjective feeling of being observed. Our approach was then to design an IVE which could place the subject at the center of attention while having to give a talk.

In the mean time, we wanted to validate the use of biofeedback device during immersion for assessing anxiety arousal.

### 5.1.2 Setup

#### Hardware Configuration

The immersion setup consisted in a HMD (Kaiser ProView) together to with a head tracker was mounted (Ascension Technology Motion Star)[2]. The computer graphics rendering was done in real time stereoscopy on a PC with an accelerated graphics board (NVidia GForce2). In addition, the bio-feedback device (Physio-Recorder from Vienna Test System) recorded pulse and skin conductance level on the non-dominant hand[3].

#### Virtual environment

The "Phobia" IVE is a generic and symbolic virtual environment which places the subject at the center of a virtual audience by converging gazes towards him (fig.5.1). To ensure that the user couldn't avoid the gazes, symbolic spectators are placed all around him/her in concentric circles. They are represented by pairs of eyes hanging in the dark. The scene was configurable though a graphical user interface: number and geometry of circles, number and repartition of virtual humans among circles, and total size of the room. This allows exposure intensity variation, starting with a discussion in small group and ending with a presentation in an amphitheater. As ambient sound is also necessary for better immersion, we use a sound track recorded in an amphitheater full of people.

As explained above, we chose to use only gazes to figure the presence of spectators. We placed photographs of eyes taken from the facial-expression book from Ekman & Friesen (1978). As the same database of facial expression was used in Foa *et al.* (2000), we could get advantage of their conclusions on the higher sensitivity of social phobics towards negative expressions and selected the pictures that had the "darker" gazes (shown in fig.5.1). We managed to maintain the illusion of gazes fixing the user while the latter was in motion by using the bill-boarding technique.

---

[2] See section 3.1.2 on page 23 for more details on the setup.

[3] See section 5 on the preceding page for more details on physiological measurements.

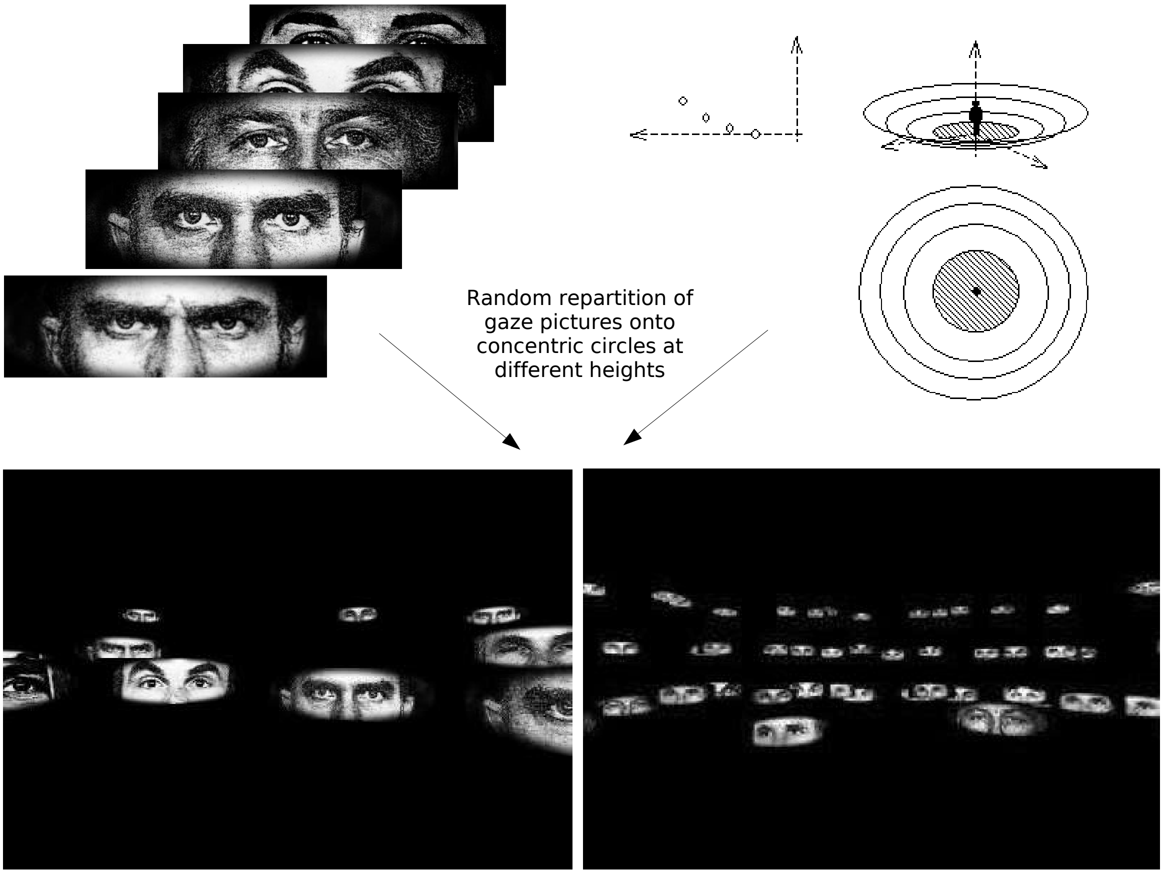


Figure 5.1: Composition of the symbolic IVE used for the preliminary study.

### 5.1.3 Measurements and procedure

Different assessment tools were used to estimate the social discomfort of the subjects and the severity of their anxiety during exposure:

- *Liebowitz social anxiety scale (LSAS)*[4]: self-assessment questionnaire of the degree of anxiety of the subject in many usual social situations.
- *Subjective stress evaluation*: subjects are asked to estimate their subjective degree of anxiety on an analog scale from zero to ten (zero for relaxed, and ten for highly anxious).
- *Physiological measurement*s: heart rate (HR) and skin conductance level (SCL) are monitored during the sessions to allow for arousal evaluation.

The Phobia IVE was configured for the experiment in order to represent a quite large assembly with numerous gazes. After a short introduction to the objectives of the study and a presentation of the devices, each subject wore the equipment (HMD, sensor, earphones and bio-feedback) and sat on a chair for the rest of the session. VRE sessions were organizes in four phases:

- **Phase 0** Preparation and equipment, $\dashrightarrow T_0$
- **Phase 1** Relaxation with quiet music and restful picture, $\dashrightarrow T_1$
- **Phase 2** Immersion in the virtual assembly without speaking, $\dashrightarrow T_2$
- **Phase 3** Speech in front of the virtual assembly. $\dashrightarrow T_3$

The subjective stress was evaluated at the end of each phase at times $T_0$,$T_1$,$T_2$ and $T_3$. The goal of the relaxation phase was to recover from the additional stress generated by the first phase in order to establish a reference at $T_1$ for data interpretation. For phase 1, a restful picture was selected by the subject among a set of landscapes displayed in the HMD.

### 5.1.4 Results

Subjects consisted in ten voluntary EPFL students in computer science, aged between 20 and 30, constituted of four females and six males. In this section, we will present the global synthesis of the analysis we performed in collaboration with our therapist partners. More details on statistical significance are given in (Riquier *et al.*, 2002).

---

[4] See section 2.3.1 on page 15 for more details.

Table 5.1: Scores for Liebowitz social anxiety scale with respective diagnostic.

| LSAS score | Diagnostic |
|---|---|
| Greater than 95 | Very severe social phobia |
| 80 – 95 | Severe social phobia |
| 65 – 80 | Marked social phobia |
| 55 – 65 | Moderate social phobia |
| 40 – 55 | Mild social discomfort |
| Below 40 | No discomfort |

### Data interpretation

In clinical situation, the Liebowitz questionnaire represents the severity of patients' social anxiety disorder. As our subjects didn't suffer from SAD, we had to interpret LSAS scores below the pathological level. Table 5.1 shows the two diagnostic sub-classes we added to the original Liebowitz scale in order to represent mild social discomfort and no discomfort. This distinction leaded to a clear separation of the sample into two groups;

- Group A — score above 40 — 5 subjects with mild social discomfort.
- Group B — score below 40 — 5 subjects without social discomfort.

The first observation which could be made when observing the subjective anxiety in table 5.2 was the coherence with the distinction into two groups by LSAS. Firstly, the global and baseline cognitive assessment of anxiety had higher values for people presenting social discomfort. At $T_2$, exposure to the VE have produced a significant increase of the subjective anxiety level compared to the baseline, only for group A (while marginal for group B). Finally, both groups reported an increase of subjective anxiety at $T_3$ (after having to give a talk) relatively to the previous state $T_2$, with a slightly more important for group B. However, the difference between the final state $T_3$ and the baseline condition $T_0$ was significant only for group A.

Concerning the interpretation of physiological measurements (HR and SCL in table 5.2), the distinction into two groups was globally respected. First, despite baseline signals at $T_0$ which are not significantly different between groups, group A clearly had higher HR values overall. At $T_1$, the relaxation phase significantly lowered HR in group A ($T_0 - T_1$), as opposed to group B where values did not change —i.e. they were not stressed before. We could therefore consider that at $T_1$, physiological signals had stabilized in both groups. Between $T_1$ and $T_2$, the immersion in the VE produced a significant increase of HR and SCL for both groups. However, when compared with the respective baseline measures at $T_0$, the increase is more significant for group B. This is correlated to higher global anxiety in

Table 5.2: Overview of subjective anxiety, HR and SCL during exposure sessions.

| | Group A<br>Subjects with mild social discomfort | | | Group B<br>Subjects without social discomfort | | |
|---|---|---|---|---|---|---|
| Time | Range | Mean | Median | Range | Mean | Median |
| Subjective anxiety ( $[0 : 10]$ ) | | | | | | |
| $T_0$ | 2.5 – 7.0 | 4.4 | 4 | 1.0 – 5.0 | 2.0 | 1.0 |
| $T_1$ | 2.5 – 6.0 | 3.7 | 3.2 | 0.5 – 3.0 | 1.1 | 0.5 |
| $T_2$ | 3.5 – 7.0 | 5.9 | 6.0 | 0.5 – 3.0 | 1.6 | 1.5 |
| $T_3$ | 6.5 – 7.0 | 6.8 | 6.8 | 1.5 – 5.5 | 3.5 | 3.0 |
| Heart rate ($bpm$) | | | | | | |
| $T_0$ | 58 – 111 | 87.0 | 94 | 71 – 81 | 74.8 | 74 |
| $T_1$ | 52 – 98 | 77.2 | 77 | 61 – 78 | 71.6 | 74 |
| $T_2$ | 74 – 121 | 97.6 | 98 | 73 – 87 | 81.8 | 82 |
| $T_3$ | 78 – 132 | 105.2 | 114 | 75 – 91 | 85.4 | 86 |
| Skin conductance level ($\mu S$) | | | | | | |
| $T_0$ | 1.39 – 6.41 | 3.53 | 2.37 | 1.09 – 6.27 | 3.62 | 3.96 |
| $T_1$ | 1.17 – 5.96 | 3.23 | 2.0 | 1.52 – 6.17 | 4.12 | 4.40 |
| $T_2$ | 2.56 – 7.08 | 4.17 | 2.56 | 1.39 – 7.50 | 4.82 | 4.79 |
| $T_3$ | 2.26 – 7.66 | 4.85 | 3.79 | 2.05 – 9.30 | 5.82 | 5.89 |

group A, even before the exposure (subjects were already stressed at $T_0$). Finally, at $T_3$, the public speaking situation produced an increase of both signals which is significant relatively to the baseline in both groups.

Moreover, as seen in section 4.1, physiological signals can also be interpreted in terms of specific features representing arousal and valence. Thus, we also computed the two most meaningful parameters for arousal in table 5.3. As the sample is too small for deeper statistical interpretation, we will just point out the obviously higher arousal for subjects presenting social discomfort, which is expressed mainly with $s_{HR}$ and less importantly with $\widetilde{d}_{SCL}$.

Table 5.3: Arousal estimation for the two groups.

| | Group A<br>Subjects with mild social discomfort | | | | | | Group B<br>Subjects without social discomfort | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | 1 | 2 | 3 | 4 | 5 | **Mean** | 6 | 7 | 8 | 9 | 10 | **Mean** |
| $s_{HR}$ | 15.0 | 15.2 | 9.75 | 3.30 | 20.8 | **12.8** | 5.85 | 9.18 | 6.22 | 7.68 | 4.65 | **6.71** |
| $\widetilde{d}_{SCL}$ | 0.87 | 0.76 | 1.23 | 0.93 | 0.96 | **0.95** | 0.78 | 0.74 | 1.01 | 0.74 | 0.73 | **0.80** |

#### Subjective appreciations

After each exposure, the subjects were interviewed about the experiment. Here are the most common observations they reported:

- "The situation by itself was stressful; wearing HMD and being plugged to biofeedback sensors is very unusual"; although people considered that the relaxing phase helped in getting used to the equipment, they suggested that they would feel much more confident with it if they could do it a second time.
- "The sound ambiance was essential to the immersive sensation"; we shall not have spoken to them directly while they were inside the HMD when asking for subjective anxiety, but should rather have used a microphone and spoke through earphones to avoid BiPs.
- "The VE was too static in general"; some subjects suggested that pictures should be animated or replaced by videos, but none of them criticized the symbolic reduction of the assembly into gazes.

### 5.1.5 Discussions

This experiment allowed us to confirm the capability of VR immersion to generate an emotional arousal specific to the targeted anxiety. We have been able to validate this conclusion for social phobia thanks to the clear correlation between the social phobia tendency estimated with clinical questionnaire and the intensity of subjective and objective anxiety responses during VRE. The estimation of subjective anxiety was globally higher for people presenting social discomfort than for non phobic subjects. The presence of gazes in VR was sufficient to generate a significant increase of anxiety for the formers only. The estimation of objective anxiety with physiological signals have also shown different levels of reactions between the subjects. All subjects experienced a normal heart rate increase when having to give a talk, but the arousal overall was much higher for people with social discomfort. They had already a higher heart rate when starting the experiment compared to the other group, but managed to stabilize the physiological parameters after relaxation. This allowed us to consider that, despite their global nervous state, subjects presenting social discomfort had a significant increase of heart rate when exposed in VR, which denotes a higher sensitivity to the stimuli than non phobic subjects. The heart rate and self-reported anxiety variations are comparable to the in-vivo exposure conditions presented by Sheffer *et al.* (2001).

The judicious selection of a specific social stimulus (gaze) was able to provoke stress when used as the evocation of human presence. Within these conditions, the circular repartition used as a symbolic auditorium was enough to figure a public speaking situation. Despite the limited size of the sample, we could verify that patients' psychological vulnerability influenced their perception of the situation. Our results presented in (Herbelin *et al.*, 2002) were confirmed by Slater *et al.* (2004) when they compared the reaction of phobic and non-phobic subjects during public speaking VRE.

From the medical point of view, the use of a symbolic environment targeting specific stimuli may be used in a suggestive manner (comparable to hypnosis) to support an active form of exposure in imagination. However, both therapist and subjects suffered from the system's lack of flexibility and the impossibility to support tangible scenarios of social interaction constraints too much the potential impact of VRE. We may therefore expect better clinical results with exposure scenarios based on life-like simulations.

## 5.2 Clinical case studies

During the three years of collaboration with Dr Françoise Riquier, we provides VR public speaking exposure sessions for the therapy of well selected patients. Three people with important social anxiety disorder tested the VRE system and were successfully treated afterwards. Here is the description of those case studies.

### 5.2.1 Objectives

As the literature was very optimistic regarding the success of VRET, we expected that the use of VR would be favorable to a quicker and/or better rehabilitation. However, we still needed to practice clinical usage of VR with our public speaking exposure system and to objectively confirm the therapeutic success. We also needed to establish a protocol which would easily integrate the CBT procedure already in use by the therapist. Finally, we wanted to experiment with various immersive setups in order to determine if one would more appropriate than the others.

### 5.2.2 Setup

#### Hardware configuration

We selected the Pro-View HMD setup to be compared with the back-projection solution (fig.5.2)[5]. When using HMD, we noticed no important difference between the user standing up or sitting on a chair — as anyway he did not need to explore the large virtual office. In this case, the starting point of view was calibrated to be close to the assembly. As immersion

---

[5] See section 3.1.2 on page 23 for more details on the setups.

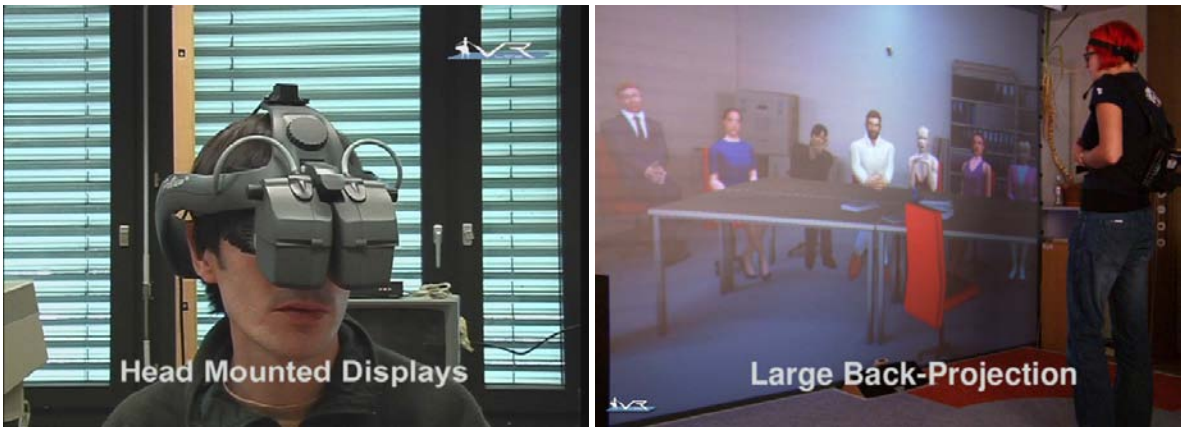


Figure 5.2: Immersive public speaking setups for case studies.

Figure 5.3: Office environment for public speaking exposure (first version).

with back-projection screen is more appropriate for displacements in the environment, we opted for free exploration of the environment in this case.

### Virtual environment

An early version of our VR platform was used to support various public speaking situations according to the patient's needs (examination for a diploma, job interview).

The office environment consisted in a work place where a meeting with several people is organized. Up to seven VHs were placed around a table in order to face the subject. They were all animated with dedicated animations according to their gender and style; women fixed their hair or checked their nails, a well-dressed man sat quietly as a younger one took more relaxed postures by crossing his legs and arms. A sound ambiance recorded in similar real conditions supported the immersion feeling — people clearing their throat, chairs creaking, far street noises.

In this version of the office VE, humanoid rendering and animation engine were in the first stages of development. As we could not animate the gaze yet, we compensated by placing them according to the targeted position of the subject (i.e. their gaze converged towards a central point). Only seven models of VHs were available. Seven animation key-frames were designed per character. The assembly's behavior was performed by regular random selection in these pools of key-frames. The office environment and the audience are shown in figure 5.3.

### 5.2.3 Measurements and procedure

The main clinical assessment done during these sessions was the subjective anxiety evaluation. In practice, the patient gives a number between 1 and 10 expressing his/her internal feeling of anxiety at a particular moment. Several times during behavioral exposure sessions (usually in-vivo), this subjective anxiety was expressed and its progression was analyzed with the patient during cognitive sessions.

Here is the procedure for an exposure session. Before starting, the patient is asked his/her subjective anxiety (*before-time*). After being equipped with immersion devices, he/she can already explore the environment to get used to the immersion. At the same time, he/she may mentally prepare a talk before starting to speak. In fact, according to the therapist's recommendations, he/she may focus on the assembly and wait for his/her anxiety to be high enough before talking. He/she expresses his/her subjective anxiety just before starting the talk (*start-time*). At the end of the presentation, the patient takes off the equipment. The subjective anxiety is given once again *(end-time*).

The immersion with HMD and its limited field of view was used to force conscious gaze movements (head rotation) when looking successively at each member of the assembly.

With the back-projection setup, the patient started at a position far from the table and had to 'walk' towards people before addressing them. He/she could train using the navigation paradigm and was only indicated an appropriate position for giving the talk (where he/she could see everybody on screen).

The experiments were done with three patients –one female and two male. Subject 1 was a 24 year-old male student with generalized social phobia, highly penalized in his studies and social life. Subject 2 was a 24 year-old woman ending her diploma but having clear SP and needing support for coping with social performance anxiety. Subject 3 was a 15 year-old boy with SP and school phobia.

Table 5.4: Subjective anxiety for subject 1 during VRE sessions.

| Session | Scenario | Equipment | before time | start time | end time |
|---|---|---|---|---|---|
| 1 | Exam | HMD | 4-5 | 8 | 4-5 |
| 2 | Exam | BACKPROJ | 4-5 | 6 | 3-4 |
| 3 | Exam | HMD | - | 8 | 4 |
| 4 | Job interview | BACKPROJ | - | 6-7 | 3-4 |

### 5.2.4 Results

We have been able to use VR in a complete way with subject 1. The patient had generalized SAD with LSAS score above 90 (severe SAD according to table 5.1 on page 67). During the first session, he discovered and tried the VR system and agreed on trying such a public speaking exposure. The therapist proposed him to train his final exam presentation as he had a growing anxiety regarding the approaching deadline. Table 5.4 shows the subjective anxiety observed during the next four sessions. The clear increase of anxiety at *start-time* shows that the simulation had the potential to generate exposure stress. In addition, the patient clearly expressed that he felt the agents were less aggressive after session 3 than in the first session. In order to stimulate him a bit more for the last session, the therapist asked him to present himself for an impromptu job interview. This change provoked a little more stress than in the previous session in similar immersion conditions. We could also observe a higher anxiety in HMD compared to back-projection conditions. We tried to obtain an explanation by asking the subject which setup he would prefer to continue the therapy with. He said he would prefer HMD. This could be related to the different capabilities of HMD and projection setups to affect user's SoP, performance, side effects and exertion (Rand *et al.*, 2003). However, we would rather suggest that the selection of the setup should be done according to the patient's preference and after a preliminary observation of the subjective anxiety in both situations. The end choice is the therapist's responsibility as it leads to different observations. For example, in subject 1's case, his global tendency to always look on the right (a kind of safety behavior) could better be observed in the back-projection setting than in HMD.

Subject 2 had a social phobia targeted on social performance situations. She had problems ending her diploma and a growing stress when thinking about the job interviews that would follow. The therapist proposed VRE for self presentation in job interview scenarios. In practice, the therapist relied on the VR simulation as she was used to with classical cognitive and behavioral group therapy. The patient expressed a preference for back-projection immersion because she was reluctant to wear the HMD. For the second session, she was asked to fill the Witmer & Singer (1998) presence questionnaire (PQ) and immersive tendency ques-

Table 5.5: Answers to customized Witmer & Singer (1998) PQ and ITQ by subject 2 after back-projection VRE session.

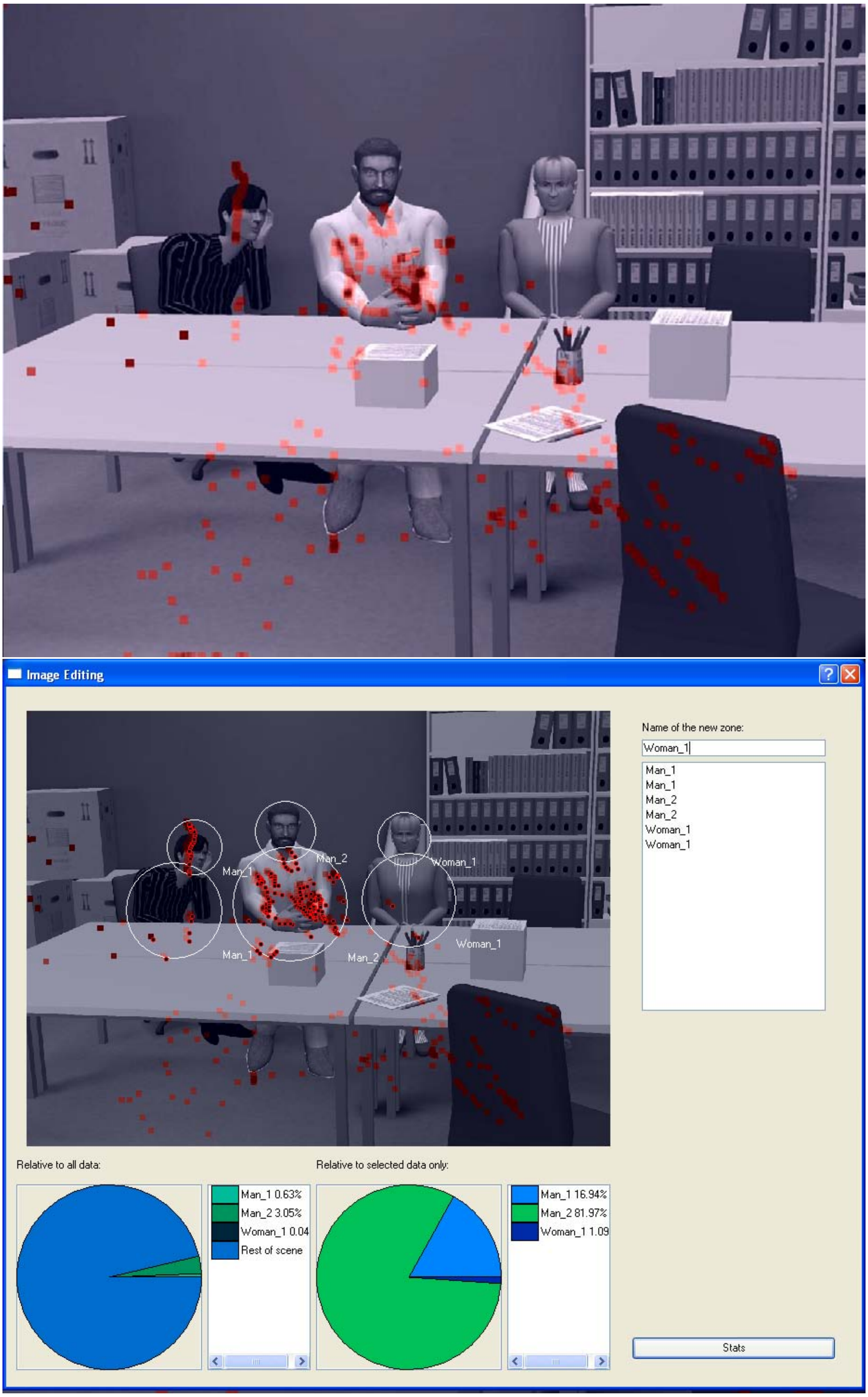
| | Nbr questions | Max score | Score | Percentage |
|---|---|---|---|---|
| **PQ** Total | 20 over 32 | 98 | 84 | 85.7% |
| - Involvement *and control* | 5 over *11* | 35 | 28 | 80.0% |
| - Naturalness | 3 over 3 | 28 | 13 | 46.4% |
| - Interface quality | 2 over 3 | 14 | 4 | 28.6% |
| - Others (auditory *and haptics*) | 3 over *5* | 21 | 10 | 46.6% |
| **ITQ** Total | 28 over 28 | 203 | 141 | 69.5% |
| - Tendency to become involved | 7 over 7 | 49 | 45 | 91.5% |
| - Tendency to maintain focus | 7 over 7 | 49 | 38 | 77.6% |
| - Tendency to play video games | 2 over 2 | 14 | 9 | 64.3% |

tionnaire (ITQ) after a back-projection VRE session. On the 32 items of the original PQ, 12 were removed (1,2,10,17-22 & 27) because they were either irrelevant in our case (about objects manipulation or haptics) or already dropped by the authors as not significant. The ITQ was kept as is. Table 5.5 shows the scores obtained with the original decomposition in sub-scales. We have first been able to observe that her high tendency to become involved was verified with a good feeling of involvement in VR (80%). The gap between the low scores of her judgment of naturalness (<50%) and quality of interface (<30%) could be interpreted in terms of a high presence of the virtual humans provoked by a subjective over sensitivity to this stimuli (social phobia) rather than by VR system itself. However, the therapist considered ITQ as more interesting because more generic; it could allow her to evaluate the potential acceptance of the VRET by a subject before choosing a VR therapeutic approach.

In the case of subject 3, the VRE was used three times to simulate a meeting with his teachers, asking him to expose his problems regarding school attendance. The benefits of these sessions were mainly in the immediate reactions of the boy. As he could better explain his feelings on the spot than when relating the events of the day before, VRE sessions were particularly rich in behavioral and cognitive feedback for the therapist. The therapist was encountering difficulties with him to correct his gaze avoidance behaviors. In order to confront him to an objective observation of his gaze avoidance behavior, the last session was conducted with the eye tracker[6] in front of the large back-projection screen. The patient was asked to look at people in the eyes when addressing them, and to look at them all. After a brief exposure (1'20"), the boy reported that he had been looking at them in the eyes, and mainly the man in the center. The therapist could then present him the visual trace of his gaze (fig.5.4) and compare; the gaze statistics confirmed his preference for the central char-

---

[6] See section 4.2.2 on page 54 for more details on eye tracking.



Name of the new zone: Woman_1

Man_1
Man_1
Man_2
Man_2
Woman_1
Woman_1

Figure 5.4: *Gaze avoidance behavior for subject 3.*

acter, but the extremely high avoidance score (95%) confronted him with the fact that he was not actually looking at them. The therapist was then able to work with him on his safety behaviors. Finally, the patient expressed a great interest in these simulations and considered them as useful for training to cope with stress.

### 5.2.5 Discussions

Our experiments with these patients were beneficial for several reasons. First, we have seen positive results in their rehabilitation after only few sessions; subject 1 finally succeeded in his diploma oral presentation, and the two others ended their therapy soon after. However, there is no measurable way to determine how much VR took part in their rehabilitation.

What is certain, is that we have been able to make some correlations between perception of the virtual assembly and fears in reality. Similarly to Pertaub *et al.* (2001), the therapist noticed that the subjects had been very sensitive to the virtual audience by observing the patients' behaviors and during cognitive post-exposure sessions. All of them considered the presence of the virtual assembly as 'real' enough to generate stress; they felt similar feelings as in reality (difficulty to speak, shaking hands) and had typical safety behaviors (in hand movements, body displacements and gaze avoidance). They also expressed similar cognition as with real people; when reporting their feeling after exposure, patients were using verbal formulations relative to real people when referring to the jury; "This girl was looking at me in a strange way", "I did not like the attitude of this person, I felt he was judging me", or even "They saw I was embarrassed". In comparison, non-phobic visitors usually criticize the absence of reaction and the lack of realism. The fact that social phobic people were more sensible to irrespective behaviors in the assembly was an important anxiety provoking factor which contributed to the efficiency of the exposure sessions.

From the therapeutic point of view, these VR sessions were also very rich and dense in terms of lived experience. As such, the habituation to social phobia stimuli was possible as efficiently as with in-vivo exposure. VRE sessions easily integrated the therapeutic procedure. Our medical partner, who was discovering the tool, quickly found the way to take advantage of the possibilities offered by the system. In addition, patients have shown interest in the possibilities offered, and the system has even been used in a pedagogic way to better explain them their behavioral bias. Only the complexity of the system, still under development at the time and requiring technical support, did not allow for a more intensive usage.

Figure 5.5: Office environment for public speaking exposure (second version).

## 5.3 Large scale validation study

We had the opportunity to participate in a national Swiss festival called "Science et Cité" and to present the immersive exposure system to a large public. Thanks to this opportunity, we have been able to conduct an experiment with a large cohort of volunteers and validate our system.

### 5.3.1 Objectives

The first objective was to validate the use of a simple and low cost VRE setup which could be used in therapy. The system first had to prove its usability during an intensive use by non VR specialists (two students were employed to welcome visitors and manipulate the system). The VR exposure also had to demonstrate its capacity to provoke cognitive or emotional reactions relative to the social phobia tendency of the subjects (in a similar way as for the preliminary study in section 5.1).

The second objective was to determine if it was still possible to perform a behavioral analysis without eye tracker during HMD immersion. To do so, we made the hypothesis that the gaze direction could be approximated by the head direction — as the field of view is very limited in HMD, users should indeed have negligible eye movements. The goal was to accept or reject this hypothesis by establishing a database of such exploratory behavior during exposure in different configurations. Behind this, there was the possibility to establish a simple behavioral criteria (gaze v.s. head movement) which could be used for clinical assessment.

### 5.3.2 Setup

A second version of the office VE was used for public speaking exposure. The system integrated various features: gaze animation, characters selection in the pool of VHs, and possibility to simulate speaking actions. We selected five VH models to build an heteroge-

(a) Positive attitude (b) Negative Attitude

Figure 5.6: Different attitudes of the audience in the office environment.

(a) Close distance (b) Average Distance (c) Far distance

Figure 5.7: Three calibration viewpoints for the HMD.

neous assembly. Figure 5.5 shows a reconstructed panoramic in the point of view of a user immersed with HMD. In addition to the usual body movements (same as in the first version), the characters could look at the subject straight in the eyes or exchange brief looks between them. We optimized the behavior script in order to vary between a very positive audience with people looking at and listening to the subject (fig.5.6(a)) and a negative audience with people always moving and not looking at the speaker (fig.5.6(b)). The office environment could be used for various scenarios; depending on the subject, the assembly can be seen as a jury for an academic examination, experts for a job interview, colleagues in a professional meeting, etc.

The central character, the president of the jury, was capable of pronouncing a few pre-recorded sentences. This was controlled by the experimenters in order to address the participant and guide the interview by asking to start and to end the talk. The supervisor could also trigger some encouraging sentences from time to time (e.g. “Good, please continue”, or “Yes, can you tell us more about that?”).

(a) Facing a character (b) Between two characters

Figure 5.8: Large picking area for approximated gaze tracking in HMD.

In order to experiment with clinical conditions, we used the low cost HMD setup[7] consisting in a pair of Virtual I/O i-glasses equipped with a 3DOF InterSense tracker (fig.5.9(a)). The calibration of the navigation system was done by placing the camera at a pre-defined viewpoint once the subject was equipped with the HMD and by resetting the tracker's relative rotation. Three calibration positions were defined to experiment with various distances between the subject and the assembly in order to observe the influence of the field of vision on subjects' performance. The average-distance calibration position (fig.5.7(b)) is a compromise between the two extrema (figs. 5.7(a) and (c)).

To observe the subjects' gaze behavior, the gaze-map picking system[8] continuously tracked the targets in the center of the field of view during the exposure sessions. We estimated that the usual center of interest when looking in a HMD was horizontally centered but vertically above the center of the display. The picking area had to be quite large to compensate for the absence of eye tracking; it was designed to be a $128 \times 128$ pixels square (16% of the screen width) corresponding in the average distance viewpoint to the size of a VH head (fig.5.8(a)) and to the space between them (fig.5.8(b)). In other calibration positions, additional imprecision shall be taken into account.

### 5.3.3 Measurements and procedure

The basic idea of the procedure was to ask people coming to visit the exhibition for an impromptu speech task in VRE conditions. A first form was filled before immersion in the VE, and a second one after the session (fig.5.9(b)). The first form was intended to inform us

---

[7] See section 3.1.2 on page 23 for more details on the hardware configuration.

[8] See section 4.2.3 on page 54 for a description of the gaze picking technique.

(a) Immersion with HMD (b) Subject filling questionnaires

Figure 5.9: Subject participating to the validation study.

Table 5.6: Post-exposure short cognitive questionnaire (Q.III)

| | | no | slightly | moderately | strongly | very strongly |
|---|---|---|---|---|---|---|
| Q.III.1 | Did you feel you were the center of attention? | | | | | |
| Q.III.2 | Were you disturbed by having to talk ? | | | | | |
| Q.III.3 | Did you look at everybody ? | | | | | |
| Q.III.4 | Did you look the people in the eyes ? | | | | | |

on the subjects' profile (Q.I) and on their degree of social discomfort (Q.II). In questionnaire Q.I, people indicated their gender (Q.I.g) and age (Q.I.a). Questionnaire Q.II was a French translation of the Liebowitz social anxiety scale. We computed the LSAS global score, but also the four sub-scores to distinguish anxiety (LSAS.F), avoidance (LSAS.A), social performance (LSAS.P) and social situations (LSAS.S) related factors. See Appendix A on page 125 for more details.

The second form evaluated the subjects' on-the-spot reactions after immersion. The short cognitive questionnaire Q.III (table 5.6) consisted in four questions specific to the experimentation. The answers rated from "no" to "very strongly" the subjective feelings relative to the act of giving a talk (Q.III.1-2) and the subjective appreciation of the performance (Q.III.3-4). Finally, the corporal sensations questionnaire Q.IV was taken from Cottraux *et al.* (2004) to provide indications on various physiological parameters. Its 22 items score from 0 to 4 the absence or the very strong intensity of a very large scope of corporal feelings like racing heart, chest pressure, shivers, cold sweat, vertigo, lump in the throat, etc. The computation

Table 5.7: Age and gender repartition in the population of the validation study.

| | Q.I.a *(Age)* | | | | |
|---|---|---|---|---|---|
| Q.I.g *(gender)* | 0<br>*< 25* | 1<br>*< 35* | 2<br>*< 45* | 3<br>*< 65* | 4<br>*> 65* |
| 0 *Female* | 20 | 29 | 21 | 14 | 7 |
| 1 *Male* | 42 | 26 | 16 | 17 | 8 |
| Total | 62 | 55 | 37 | 31 | 15 |

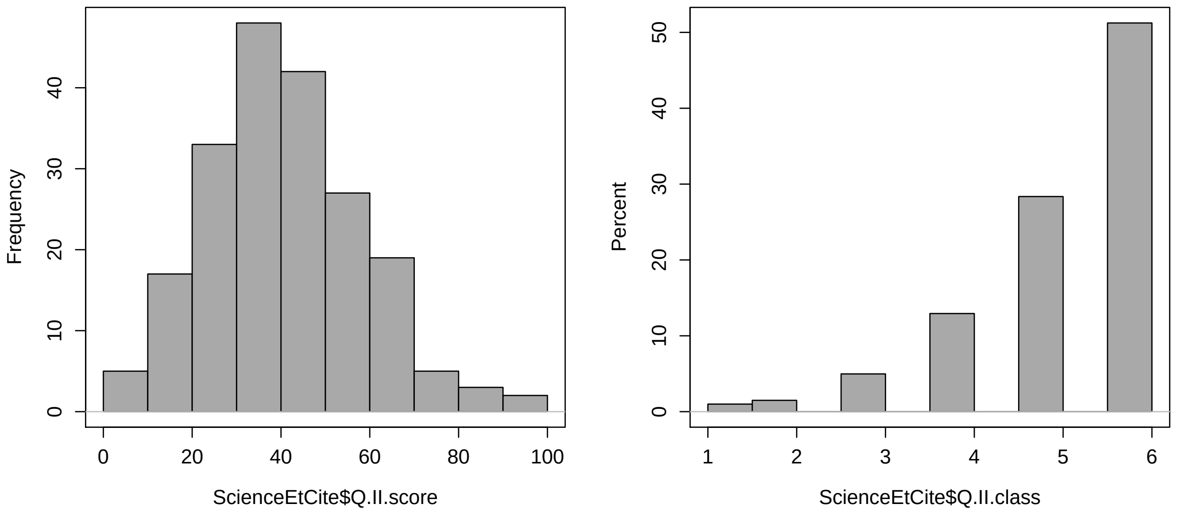
of Q.IV global score ($\leq 88$) can be considered as an indirect monitoring of the symptoms of anxiety. The original french version of these forms are in Appendix B on page 127.

The experimental procedure was as follows. We explained the basic principles of VRET to every visitors. People who were interested in participating to the study were asked to sign a participation agreement and to fill the first form. Then, they were suggested some ideas of scenario (job interview, professional meeting, academic examination, etc.) and given few minutes to prepare the talk to be given in front of the virtual assembly. Before starting, the supervisor also reminded the subjects to gaze at every member of the jury in order to better convince them. After a short adaptation period with HMD (free exploration of the environment, ensure they saw the five characters), the central character of the jury was triggered to ask to start the talk. Later, when the supervisor estimated that the subject was done, he triggered the central character to thank him and terminated the session. During exposure, the gaze picking targets were recorded and saved to disk. The second form was then printed out with a reference to the session for further correlation between questionnaires and recorded data. After filling it out, people were finally thanked for their support. No clinical interpretation of the questionnaire was given to the subjects.

To sum up, each session is described by four questionnaires and a record of visual exploration during immersion. We made as many sessions as possible during the nine days of exhibition, with only one constraint concerning the age of the participants who had to be more than fifteen.

### 5.3.4 Questionnaires results

Two hundred people actually participated in the experiment. The population of experimenters was well separated between genders (110 males, 90 females) and covered all ages (table 5.7). The average duration of the session was 1' 45" (standard deviation 40 seconds). People were glad to participate in our experiment as it gave them the opportunity to try a VR device for the first time. The setup was placed in an isolated corner, but the exhibition room was full of people and quite noisy; we had to amplify the sound in the earphones in order to



(a) Scores to the Liebowitz social anxiety scale

(b) Classes of social anxiety disorder (1:high, 6:low)

Figure 5.10: Social discomfort in the population of the validation study.

better isolate the subjects.

To experiment with the three viewpoints, we changed the setup during the exhibition. The first 131 subjects were placed at the average distance from the jury (fig.5.7(b)), then 41 were close (fig.5.7(a)), and 28 were far (fig.5.7(c)). This formed the three groups with different experimental conditions.

### Cohort representativeness

The answers to Q.II indicated a normal repartition of LSAS scores around the no social discomfort criteria (score = 40 in figure 5.10(a)). A repartition into the six classes of LSAS scores as defined in table 5.1 on page 67 is shown in figure 5.10(b). It appears that 50% of the cohort had no discomfort, 30% a mild social discomfort, and the remaining 20% a moderate to severe social phobia. According to the LSAS scores, we separated the data into three groups representing the different levels of social anxiety disorder: no social discomfort (100 subjects), mild social discomfort (58) and marked social phobia (42).

We have performed chi square tests for statistical significance on subjects distribution in the LSAS groups and in the experimental condition groups according to gender and age of subjects. We could not select the population, but as numerous people actually participated, we were lucky and the results of the chi square test indicate a correct distribution in almost every groups. Only the close distance group, where chi square significance is low ($p < 0.05$), has a non equal distribution of females between mild social discomfort and marked social

Table 5.8: Average values of answers to questionnaires.

| | Q.II (*LSAS*) | | Q.III.1 $[0..4]$ | | Q.III.2 $[0..4]$ | |
|---|---|---|---|---|---|---|
| | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| Whole cohort | 40.9 | 17.6 | 2.24 | 1.17 | 1.29 | 1.16 |
| No Social Discomfort group | 26.9 | 9.0 | 2.16 | 1.19 | **0.95** | 0.98 |
| Mild Social Discomfort group | 46.1 | 4.18 | 2.39 | 1.23 | **1.56** | 1.24 |
| Marked Social Phobia group | 66.4 | 10.6 | 2.23 | 1.04 | **1.74** | 1.2 |
| Close Distance group | 40.9 | 19.5 | **2.44** | 1.25 | 1.39 | 1.24 |
| Average Distance group | 40.6 | 17.6 | **2.23** | 1.14 | 1.28 | 1.18 |
| Far Distance group | 42.1 | 15.1 | **2.00** | 1.20 | 1.21 | 0.94 |

| | Q.III.3 $[0..4]$ | | Q.III.4 $[0..4]$ | | Q.IV $[0, 88]$ | |
|---|---|---|---|---|---|---|
| | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| Whole cohort | 2.53 | 1.18 | 2.24 | 1.13 | 7.5 | 7.7 |
| No Social Discomfort group | 2.59 | 1.21 | **2.46** | 1.1 | **4.6** | 4.5 |
| Mild Social Discomfort group | 2.56 | 1.02 | **2.08** | 1.16 | **9.0** | 8.2 |
| Marked Social Phobia group | 2.33 | 1.32 | **1.93** | 1.1 | **12.1** | 10.3 |
| Close Distance group | **2.95** | 1.02 | **2.78** | 0.94 | 7.1 | 7.6 |
| Average Distance group | **2.44** | 1.27 | **2.15** | 1.17 | 7.9 | 8.2 |
| Far Distance group | **2.31** | 0.81 | **1.86** | 0.99 | 6.2 | 5.3 |

phobia groups (more women in the former). Moreover, there were too few people older than 60 to perform significant comparison with this criteria.

### Global analysis

Table 5.8 presents an overview of the average values to the questionnaires. An apparent good repartition of answers to Q.III is indicated by the quite large standard deviation for all the questions ($\simeq 30\%$ on average). On the opposite, the answers to Q.IV are very low and have a low variation. This is due to the null score to a large majority of items in Q.IV as people did not have every feelings simultaneously. The most representative were beating heart, chest pressure, dry throat and warm/cold feelings. However, $80\%$ of scores being below 8, the answers to Q.IV show a globally low expression of corporal feelings during immersion.

Here are the additional observations which can be made regarding the changes between experimental groups:

- The subjective feeling of being the center of attention (Q.III.1) decreased between groups of growing distance to the assembly (factor of $\simeq 10\%$ on average between each group).

Table 5.9: Questionnaires Pearson *r* correlation matrix.

| | Q.I.a | Q.I.g | Q.II | Q.III.1 | Q.III.2 | Q.III.3 | Q.III.4 | Q.IV |
|---|---|---|---|---|---|---|---|---|
| Q.I.a | 1.00 | - .05 | - .12 | - .29 ** | - .24 * | - .14 | -.09 | - .30** |
| Q.I.g | | 1.00 | .13 | .03 | - .01 | .14 | - .06 | - .02 |
| Q.II | | | 1.00 | .05 | **.27** ** | - .00 | -.23 * | **.43** *** |
| Q.III.1 | | | | 1.00 | .22 * | .26 * | .13 | .23 * |
| Q.III.2 | | | | | 1.00 | - .05 | - .18 | .46 *** |
| Q.III.3 | | | | | | 1.00 | .36 *** | .08 |
| Q.III.4 | | | | | | | 1.00 | - .12 |
| Q.IV | | | | | | | | 1.00 |

Q.I.a=age, Q.I.g=gender, Q.II=LSAS, Q.III=cognitive feedback, Q.IV=corporal sensations
*Significance P < .01 two tailed, ** P < .002, *** P < .0001

- The disturbance when giving the talk (Q.III.2) increased between groups of growing LSAS scores (factor of $\simeq 30\%$ on average between each group).

- The subjective expression of having addressed everybody (Q.III.3) was higher in the group closer to the assembly (around “strongly”), and lower for the average and far distance groups (around “moderately”).

- The subjective expression of eye contacts (Q.III.4) decreased between groups of growing distance to the assembly (factor of $\simeq 20\%$ on average between each group) and decreased between groups of growing LSAS scores (factor of $\simeq 10\%$ on average between each group).

- The score of corporal feelings (Q.IV) strongly increased between groups of growing LSAS scores (factor of $\simeq 50\%$ on average between each group).

It appears that the LSAS criteria is related to different reactions during exposure, with apparently higher disturbance and corporal feelings for people with higher social discomfort. On the opposite, the distance criteria influenced more importantly the ability to look at people (in the eyes).

### Exposure impact on anxiety

In order to verify the apparent influence of social phobia tendency on the subjective experience during exposure, the computation of Pearson’s *r* correlation between the answers to the questionnaires should determine if the VRE actually provoked the expected anxiety according to LSAS scores (Q.II). The correlation matrix (table 5.9) shows the relations between questionnaires in the average distance group. Here are the correlations between answers before and after the exposure related to the subjective anxiety:

- The answers to Q.IV indicate that the increase of physiological symptoms of anxiety is significantly positively correlated to the LSAS score ($r = 0.43, P < .0001$). We computed this correlation with LSAS sub-scores; it appears to be more significant with LSAS.F ($r = 0.51, P < .0001$) than with LSAS.A ($r = 0.20, P < .001$)
- The answers to Q.III.2 are correlated with the severity of SAD too. It is also more significant with LSAS.F ($r = 0.36, P < .0001$) than with LSAS.A ($r = 0.12, P < .001$).

This confirms our hypothesis on the higher disturbance during VRE according to the fear in social situation. The fact that Q.III.2 and Q.IV are correlated confirms the link between subjective sensation of being disturbed and corporal feelings of stress.

Additional tests within the close and far distance groups have shown similar correlations, with apparently no impact of the distance factor. To validate our results, we also compared the answers to Q.III.2 and Q.IV between the 'no social discomfort' group and the 'mild social discomfort' - 'marked social phobia' groups (without distance factor). It appears that the correlation with LSAS.F retains the null hypothesis in the former case —non-phobic did not react— but confirms the correlation in the latter case (Q.III.2 : $r = 0.34, P < .0001$, Q.IV : $r = 0.42, P < .0001$). In other words, the social discomfort expressed during immersion is significantly correlated with the anxiety in usual social situations for subjects with mild to high social discomfort.

### Additional reaction factors

To go further in the interpretation, we can notice in table 5.9 a slight negative correlation between the LSAS score and the answers to Q.III.4 which would confirm our hypothesis that social phobics avoided gazes in VRE as in reality. However, one may observe an apparent contradiction of having a correlation between Q.II and Q.III.4, another between Q.III.4 and Q.III.3, but none between the former and the latter. The same observation may be done for Q.II, Q.III.1 and Q.III.2. This is due to the fact that Pearson correlation does not indicate any causality between the parameters, but only denote an apparent linearity between them — which could be due to a third factor. An additional analysis of correlations between the items of Q.III for each group of LSAS scores better explains this. The most obvious difference between the three groups is the variation of correlations between Q.III.1 and Q.III.3-4. In the no social discomfort group, people who had felt that they were the center of attention also considered they had look at everybody ($r = 0.36, P < .0002$ between Q.III.1 and Q.III.3).

But this correlation looses its significance in the mild social discomfort group ($r = 0.20, P > .1$) and is absolutely null in the marked social phobia group. The same phenomenon occurs with Q.III.4. This shows that people with mild or important social discomfort associated the fact of being the center of attention to the fact of looking at everybody, and a fortiori, looking at people in the eyes. This confirms that, as opposed to non phobics, groups with higher social anxiety did not address everybody and avoided gazes although subjects had the feeling of being at the center of attention.

It is also interesting to note that gender of subjects did not influence at all the impact of exposure. The age seems to have a bit more influence the answers to Q.III.1, Q.III.2 and Q.IV, thus suggesting a stronger feeling of being in front of an assembly for older people.

### 5.3.5 Picking data results

In order to obtain a meaningful interpretation of subjects' exploratory behavior during sessions, we have computed several features on top of the raw data provided by the gaze map picking algorithm. As explained in section 4.2.3, the data recorded for each VH contains for each frame the *hue*, *saturation* and *value* color components on the picking map (fig. 4.11 on page 55). Here are the principal parameters we have been able to extract from these raw data:

- $\Delta_H = H - 260$ : vertical difference from the canonical posture (head of the VH centered in the picking area as on figure 5.8). $\Delta_H = 0$ when the subject looks straight at a VH, $\Delta_H > 0$ when looking above the eyes and $\Delta_H < 0$ when looking bellow.
- $\Delta_S = S - 0.61$ : horizontal difference to the center of a VH. $\Delta_S = 0$ when the subject looks straight at a VH, $\Delta_S < 0$ when looking on the left side, and $\Delta_S > 0$ when looking on the right side.
- $V$ : density covering the VH on the picking area. As the picking area was larger than a VH head, $V$ not only determines the distance to the VH but can equally be used to determine which VH was picked.
- $A \in [0..5]$: identifier of the picked agent, 0 for none.

The four parameters above are computed at each frame. Then, statistical analysis was performed to provide a unique numerical feature for each session. For all of them, the mean (eq. 5.1), the sample standard deviation (eq. 5.2), the skewness (eq. 5.3) and the Kurtosis excess (eq. 5.4) were computed. These last two statistical tools indicate the difference with a normal distribution. Data is skewed left if $w_X < 0$ and skewed right if $w_X > 0$. The distribution is flat if $k_X > 0$ and is peaked if $k_X < 0$.

$$m_X = \frac{1}{N} \sum_{n=1}^{N} X_n \tag{5.1}$$

$$s_X = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (X_n - m_X)^2} \tag{5.2}$$

$$w_X = \frac{\sum_{n=1}^{N} (X_n - m_X)^3}{(N-1)\, s_X^3} \tag{5.3}$$

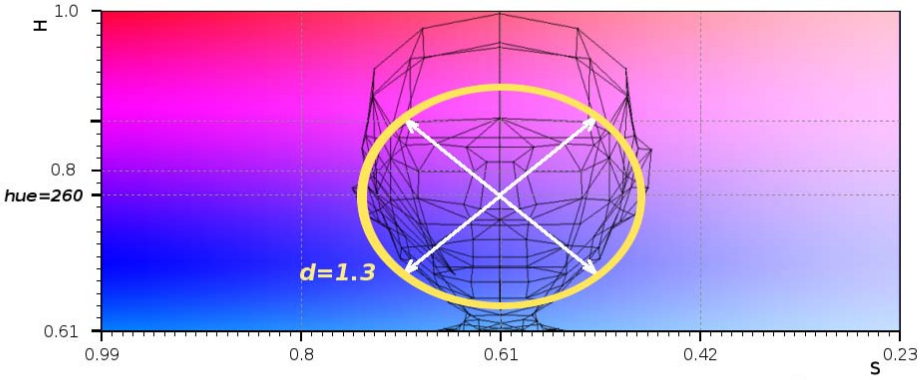
$$k_X = \frac{\sum_{n=1}^{N} (X_n - m_X)^4}{(N-1)\, s_X^4} - 3 \tag{5.4}$$

In an effort to provide more dynamic parameters and to focus on the specific centers of interest for the analysis of gaze avoidance behaviors, we additionally computed the four parameters described below:

- $d$ : Euclidean distance on gaze map with normalized $H$ and $S$ (eq. 5.5). As $\Delta_H$ and $\Delta_S$ are null when the face of the VH is well centered, $d$ is a distance to the center of the face of the agent;

$$d = \sqrt{\Delta_H^2 + \Delta_S^2} \tag{5.5}$$

- $f$ : duration of a fixation period on the face of an agent. We consider that the face of the agent is picked when $d < 1.3$. This threshold value corresponds to the diameter of the face on the gaze map, as shown in the figure below;



- $v$ : mean of the absolute values of the first differences of $V$ (eq. 5.6). The variations of $V$ indicate the amount of change between two frames and may be interpreted as a velocity;

$$v = \frac{1}{N-1} \sum_{n=1}^{N-1} |V_{n+1} - V_n| \tag{5.6}$$

Table 5.10: Principal parameters' features average values in the distance groups.

| Group | Feature | $\Delta_H$ [ -250, 75 ] | $\Delta_S$ [ -0.38, 0.38 ] | $V$ [ 0, 1 ] | $A$ {0,1,2,3,4,5} |
|---|---|---|---|---|---|
| Close distance | $m_X$ | 7.5 | -0.01 | 0.34 | 1.77 |
| | $s_X$ | 28.5 | 0.11 | 0.22 | 1.34 |
| | $s_X$ in % | 8.77 | 14.4 | 22.7 | 26.8 |
| Average distance | $m_X$ | -2.4 | -0.02 | 0.31 | 2.11 |
| | $s_X$ | 30.3 | 0.11 | 0.18 | 1.14 |
| | $s_X$ in % | 9.32 | 14.4 | 18.6 | 22.8 |
| Far distance | $m_X$ | -32.6 | -0.03 | 0.21 | 2.6 |
| | $s_X$ | 33.0 | 0.14 | 0.15 | 1.12 |
| | $s_X$ in % | 10.1 | 18.4 | 15.7 | 22.4 |

- $\tilde{v}$ : the mean of the absolute values of the first differences of the normalized $V$ (Eq.5.7). The feature $\tilde{v}$ is high when $V$ varies in large steps ($v$) and regularly ($1/s_V$);

$$\tilde{v} = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|\tilde{V}_{n+1} - \tilde{V}_n\right| = \frac{v}{s_V} \qquad (5.7)$$

These four additional features were computed for each frame and we considered their mean (eq.5.1) and standard deviation (eq.5.2) on the whole session in order to provide a unique value for each.

Finally, we also added two counters to determine if it could be useful to evaluate the frequency of occurrence of visual events:

- $C_A$: number of times $A$ changed in a session; switches in picking between VHs ($A > 0$), without counting the times no agents were picked ($A = 0$).
- $C_f$: number of fixation phases $f$ in a session; switches in attention to the face(s) of the VH(s), including the times picking was done out of the face and came back on the same character.

All together, we obtained 24 features per session that describe the gaze behavior of the subject as much as possible. The next phase should then confirm their reliability and their ability to characterize the safety behaviors. We should also take into account the impact of changing the experimental conditions.

### Reliability of picking data

Tables 5.10 and 5.11 give an overview of the parameters obtained over the 200 sessions in the various distance groups. Globally, the picking was centered on the agents' face ($m_{\Delta H}$ and

Table 5.11: Additional parameters' features average values in the distance groups.

| Group | Feature | $d$ | $f$ | $v$ | $\tilde{v}$ | Feature | $C_A$ | $C_f$ |
|---|---|---|---|---|---|---|---|---|
| Close distance | $m_X$ | 0.22 | 39.7 | 0.03 | 0.12 | Count | 62 | 33 |
| | $s_X$ | 0.23 | - | 0.05 | - | | | |
| Average distance | $m_X$ | 0.25 | 34.7 | 0.02 | 0.11 | Count | 56 | 32 |
| | $s_X$ | 0.26 | - | 0.03 | - | | | |
| Far distance | $m_X$ | 0.39 | 16.3 | 0.01 | 0.09 | Count | 59 | 44 |
| | $s_X$ | 0.35 | - | 0.02 | | | | |

$m_{\Delta S}$ very low, $m_d < 1.3$) and performed on one to three agents on average ($m_A \simeq 2 \pm 1$). The main limitation we may already consider concerns the low variations on the picking targets; subjects were almost static vertically ($s_{\Delta H} < 10\%$) and were facing the agents in the general case ($s_V < 20\%$). This is confirmed by the very low values of $m_v$ and $m_{\tilde{v}}$. However, $C_A$ and $C_f$ are quite large over all sessions, thus suggesting that people took into account our recommendation to look at every agents.

The impact of the distance toward the assembly on the gaze picking data may be summarizes in two observations. First, the difference in configuration itself influenced the vertical picking position ($\Delta_H$) and the distance to the face ($d$). However, some behavioral tendencies could also be observed in the number of agents addressed ($A$) and the count of fixation phases ($C_f$). In order to avoid such influences, we will consider the average distance group for further analysis.

### Gaze behavior avoidance characterization

The goal here was to determine if we could observe the gaze avoidance during public speaking exposure. As we did not rely on eye tracking, we were not able to observe quick changes of focus but only head movements when addressing the audience. Two types of avoidance could then potentially be observed; the ability to address every member of the jury and the potential shifts in the head orientation when talking.

The first case corresponds to behavioral exercises requested from patients during HMD VRE in order to force them to look at each person when addressing an audience; we expected to observed this in the count of picked agents and in horizontal rotation parameters.

The second case corresponds to subjects' tendency to lower the head or shift it on one side when addressing someone — a phobic subject may look at a person sidelong and with the head not facing. This should be observable with the distance to the face or the duration of fixation periods.

In order to verify these hypothesis, we had to trust the answers given to the question-

Table 5.12: Correlations between Q.III.3 and the 24 features in the average distance group.

| | LSAS < 40 | LSAS $\geq$ 40 | | LSAS < 40 | LSAS $\geq$ 40 | | LSAS < 40 | LSAS $\geq$ 40 |
|---|---|---|---|---|---|---|---|---|
| $C_A$ | **0.335**** | **0.436***** | $m_f$ | 0.054 | 0.125 | $s_{\Delta S}$ | **0.418***** | 0.219 |
| $k_A$ | - 0.186 | - 0.183 | $k_{\Delta H}$ | 0.023 | 0.089 | $k_V$ | - 0.040 | - 0.015 |
| $m_A$ | - 0.113 | 0.059 | $m_{\Delta H}$ | 0.054 | - 0.204 | $m_V$ | - 0.112 | 0.004 |
| $w_A$ | 0.249* | 0.196 | $w_{\Delta H}$ | – 0.060 | – 0.014 | $w_V$ | 0.148 | 0.081 |
| $s_A$ | 0.216 | 0.251* | $s_{\Delta H}$ | 0.016 | 0.227 | $s_V$ | 0.073 | 0.222 |
| $m_d$ | - 0.018 | 0.300* | $k_{\Delta S}$ | - 0.169 | - 0.184 | $m_v$ | 0.272* | 0.365** |
| $s_d$ | 0.171 | 0.250* | $m_{\Delta S}$ | - 0.060 | - 0.111 | $m_{\tilde{v}}$ | 0.210 | 0.201 |
| $C_f$ | **0.326*** | **0.467***** | $w_{\Delta S}$ | - 0.048 | - 0.015 | $s_v$ | 0.268* | 0.260* |

*Significance P< .05 two tailed, **P<.005, ***P<.0005

naires; Q.III.3 should indicate how much the subject was looking at everybody and should be a good comparison point regarding the 'facing avoidance'. As answers may be have been very different according to the social phobia tendency, we tested the correlations with non phobic subjects first (66 subjects with LSAS < 40), and with mild to high phobic subjects afterwards (65 subjects with LSAS $\geq$ 40). Results are shown in table 5.12. It is interesting to confirm that many correlations were finally verified in all groups (Q.III.3 was indeed not correlated with LSAS), and that the list of significantly correlated features is quite limited and corresponds to our expectations:

- The agents count $C_A$, and, as a consequence, also $C_f$ are significantly correlated with the answers to Q.III.3. The correlation is definitely lower for non phobic subjects.
- The saturation $S$ being the only horizontal picking parameter, $s_{\Delta S}$ is logically higher when Q.III.3 is higher since more people are picked. However, the correlation is much less significant for subjects with social discomfort.
- Finally, the variations of $V$ expressed in $m_v$ and speed variation $s_v$ are also slightly correlated with the viewpoint displacement when more agents are picked.

We can conclude that $C_A$, $C_f$, $s_{\Delta S}$, $m_v$and $s_v$ all together represent the ability to look at every member of the jury and can be considered as facing-avoidance features. The apparent differences in subjects behaviors below and above the mild social discomfort criteria (LSAS = 40) indicate that it would be possible to assess the facing avoidance behavior for social phobic patients.

Concerning the second case of avoidance behavior, the answers to Q.III.4 should indicate how much the subject was looking at people in the eyes and may be used as a reference as in the previous case, but we did not obtain significant correlations when computing Pearson

$r$ in the same way. We though this may be due to the important variations of Q.III.4 in table **??** between the three experimental conditions groups. Basically, as gaze behavior should be very different when seeing only one agent in front (close distance) than when having two or more agents in the field of vision (average to far distance group), the distance toward the assembly may be considered as an important factor in the ability to look at VHs in the eyes. However, there was no significant or consistent correlation between Q.III.4 and the features in any of the distance group neither. It is most likely that, independently from their social phobia tendency, subjects actually faced the agents in the same way or in a way so similar that the differences were too small to be observed with our approximation in HMD picking.

### 5.3.6 Discussions

To validate our system for VRE, we have conducted experiments with a basic VR equipment under difficult conditions. The stability and the reliability of the VR platform have been fully proved during a week of intensive usage by non technical users. However, the experimental conditions were not optimal for a precise clinical study; as subjects were visitors of an exhibition we did not have the possibility to establish a clinical between-groups procedure. As the population was globally not phobic, the differences in the reactions were not flagrant, and as the experiment took place in an public place and with a low cost VR setup, the feeling of immersion may not have been as intense as expected.

The sufficient size of the experimental cohort (200 people) however compensated a bit for these limitations and we have been able to make some important conclusions regarding the impact of VRE. We have observed a clear correlation between the social phobia tendency and the subjective expression of anxiety during VR exposure; the subjective disturbance of having to give a talk (Q.III.2) together with the expression of physiological anxiety (Q.IV) indicate that the VRE sessions was capable of provoking discomfort and stress according to people fear in social situations (LSAS.F).

In the same way people with higher LSAS scores have shown a higher level of anxiety during our preliminary study, subjects in the mild social discomfort and marked social phobia groups have expressed higher discomfort and higher indirect anxiety symptoms than the non phobic subjects. The “mild social discomfort” criteria we introduced (LSAS = 40) was significant in both experiments. The act of giving a talk in front of a virtual assembly had a significant effect only for people scoring above this criteria, thus confirming that the simulation content provoked this anxiety.

However, we did not observed important gaze avoidance behavior within our VR immersion setup. We have to reject the hypothesis that the head direction would be able to approximate the gaze direction in HMD, because of the too large approximations made on the picking area and/or because of the absence of avoidance in people head rotation behavior.

Nevertheless, the setup and the measurements were sufficient to monitor the repartition of attention in the audience and to indicate the “facing avoidance” behavior. The HMD and gaze picking platform could be used during VRE sessions to automatically and objectively assess the ability of a subject to address everybody when giving a talk.

To sum up, although this study did not follow strict clinical procedures, it allowed us to validate the reliability of our system and to confirm the possibility to perform clinical VRE with a simulated public speaking environment.

# Chapter 6

# Synthesis and work in progress

The three kind of experiments we have conducted during our researches have leaded us to a better knowledge of the possibilities offered by the use of VR for social phobia therapy. All of them tested the use of our VRE platform and evaluated the reliability of our computerized assessment tools. We will now make a global synthesis of our research work according to our primary objective —validate the use of VRE for SP — and give additional observations regarding the benefits of VR on clinical assessment and therapeutic process management.

## 6.1 Validity of virtual reality exposure

### 6.1.1 Fulfilling the conditions of therapeutic exposure

One major condition to perform an efficient exposure therapy is to provoke similar feelings as those which would be felt in real situations. As for in-vivo exposure, VRE places the patient under reproduced conditions to remind or evoke the usual circumstances provoking anxiety. The success of exposure should be indicated by the degree of anxiety provoked and by its correlation with the diagnosed phobia. We have been able to observe both of these conditions in our experiments with VRE.

Both preliminary and validation studies have shown that immersion into a public speaking situation provoked anxiety for people above a certain degree of social discomfort (LSAS $> 40$). Moreover, as this anxiety was not observed for non phobic subjects, we could safely conclude that the anxiety was not related to the VR immersion by itself but due to the impact of the virtual experience. The statistically significant positive Pearson correlations between the social phobia tendency (expressed by the LSAS score) and the measured or expressed anxiety during exposure (biofeedback or subjective scale) allowed us to conclude that exposure stress was clearly related to the targeted anxiety disorder. The special attention paid to the gazes' reproduction was certainly counted as a major factor of success in our experiments.

Moreover, the three case studies performed within the context of clinical therapy have shown that patients with diagnosed social phobia expressed similar anxiety and safety behaviors during VRE sessions as in reality. The therapist has been able to conduct complete behavioral therapies with VR obtaining the same efficiency as with group or in-vivo therapies.

### 6.1.2 Providing various immersion setups

The overview we made in chapter 2 on the use of VR for therapy has shown that the most widely used immersion technique was HMD. The use of HMD was well adapted for public speaking exposure as such scenarios does not require navigation in a large environment and is mainly based on visual exploration of an assembly. The high quality setup (Kaiser Proview + 6DOF sensor) was appreciated for its high quality of graphics and provided a good immersion feeling in our preliminary and case studies. The low end setup (I-Glasses + 3DOF) used during the validation study supported the feeling of being at the center of attention but the globally low intensity of people reactions suggests that the immersion feeling was not optimal.

Regarding the differences between projection and HMD immersion, we agree with Rand *et al.* (2003) who concluded that “the sense of presence was influenced more by differences between the participants [..] than by the attributes of the two platforms” (p.111). The choice between the two shall be made according to patients’ preference and to assessment constraints. In our case, the large back-projection display has proved its efficiency during case studies and was more appropriate for the observation of gaze behaviors than HMD.

Throughout our experiments with more than 200 people, we did not encounter any problem related to potential cybersickness or perceptual disturbance during immersion with these setups. We have been able to observe that both platforms were valid for the immersion in VRE sessions as they provided successfully the mediation requested for the exposure to the simulated situations.

### 6.1.3 Providing ecologically valid exposure

According to Rizzo & Jounghyun Kim (2005), one of the main strengths of VR rehabilitation is the enhanced ecological validity of the therapy with “VEs that are ’replicas’ of relevant archetypal functional environments” (p.121). In the context of social phobia therapy, we have experimented with various scenarios taking place in the office environment. The reproduction of environments in which patients had problems dealing with (job interview, academic examination) was a powerful mean to reproduce the most fearful situations for them. In addition, this allows for the manipulation of parameters which cannot usually be controlled by

(a) The jury of an acedemic examination... (b) ... with a larger assembly.

Figure 6.1: Amphitheater environment.

the therapist in an in-vivo or group therapy:

- VHs are unknown individuals, such as examiners for an interview usually are,
- VHs can have typical attitudes of people feeling at ease in the position of jury, unlike other patients in a group therapy,
- the assembly can be selected and personalized for each individual case, in size and variety.

Thanks to the flexibility of our VRE software platform, the elaboration of various IVEs does not require programming, but mainly an important design job and a little scripting. To demonstrate this flexibility, we have implemented two more environments for social phobia, each one targeting a particular aspect of the disorder.

The *amphitheater* environment is similar to the office one except that it takes place in an amphitheater (fig.6.1). It is dedicated to scholar or academic presentations. In addition, it can be populated by a large assembly.

The *cafeteria* environment targets a different aspect of social phobia. Instead of having to speak to an assembly, the patient have to experience the anxiety of having to enter a place where others are already seated (item 14 in Liebowitz questionnaire, appendixA). In addition, the therapist can trigger VHs to look at the patient in order to exacerbate the feeling of being the center of attention (fig.6.2).

(a) When entering the room... (b) ..people look at you.

Figure 6.2: Cafeteria environment.

### 6.1.4 Limitations

Regarding the technological limitations in our simulation of inhabited environments, we mainly emphasise on the lack of natural motions and expression of VHs. One solution would be to involve more design time into the elaboration of key-frames in order to provide a large database into which the behavioral scripts could retrieve animations on demand. A more procedural approach would be to synthesize the motions with inverse kinematics or with configurable blending between key animations as we started to investigate with Egges *et al.* (2004).

Another limitation we have been conscious of since the beginning was the limited amount of VHs which could be simulated simultaneously. Although we have been able to design appropriate scenarios with limited groups size, crowd simulation could be useful for the therapy of GSAD or agoraphobia. It would, for example, be possible to combine the stressful effect of being surrounded by gazes (as in preliminary study) with the impact of simulated 3D characters in order to elaborate unusual but frightening scenarios.

In chapter 2, we have pointed out thanks, to an overview of 10 years of VRET, that there was a lack of long term follow up studies to validate the use of VR for therapy. We have experience the difficulty to setup studies in the field of mental health and our results also have the limited validity of individual cases or non-clinical procedures. However, our work allowed our therapist partner to setup and propose a between-groups clinical study involving our system and integrated with therapy. The study was recently accepted by the local ethic comity[1] and is planned to start at the beginning of the 2005-2006 academic year.

---

[1] Commission d’éthique de psychiatrie du Service des Hospices Cantonaux du canton de Vaud, Suisse.

## 6.2 Improvements in clinical assessment

The observation of patients' reactions during exposure is mandatory to determine and ensure the therapy efficiency. We have been investigating the possibilities to more efficiently integrate this factor in the therapeutic process by developing computerized assessment tools for affective and behavioral reactions. Our results have shown that such evaluation tools could be useful for the behavioral aspect of therapy, but also that VR experiences were a good substratum for cognitive feedback.

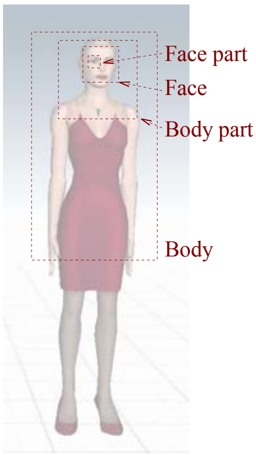
### 6.2.1 Computerized monitoring of reactions

Our experiments with biofeedback devices (sections 4.1 and 5.1) have proved that real time monitoring of arousal is definitely possible with computerized analysis of physiological measurements and can be used to objectively indicate anxiety. The most appropriate features to compute an estimation of arousal appeared to be the ones monitoring irregular small variations of skin conductance and irregularities in heart rate. According to further investigations on emotions recognition, monitoring other affective states components (e.g. valence) requires more advance software analysis and cannot be done with only pulse and skin conductance level sensors. However, our evaluation of arousal could potentially be correlated with events in the simulation content and be applied in a real-time bio-feedback loop between the patient and the VE. As a future research direction, we suggest that VEs dynamically react to the level of arousal to calm down or rather amplify the patient's anxiety during exposure. Nevertheless, this process would have to be studied with careful medical supervision and deeper knowledge of the psychological factors.

Regarding the behavioral analysis, we have been experimenting with eye tracking technology since psychological experts (Horley *et al.*, 2003) suggested that eye-to-eye avoidance was one of the main safety behaviors observed with social phobic people. Within our preliminary study, we have first confirmed with Riquier *et al.* (2002) that the presence of gazes, and more specifically the feeling of being observed, was a key stimulus for social phobia. Then, with on-screen analysis of gaze trace during public speaking sessions performed as case studies, we have also shown the benefits of gaze analysis for therapy; objective assessment of eye-to eye avoidance, correction of biased cognition, and feedback for readjustment exercises. However, in order to be used during immersion, we have had to develop a smarter way to perform eye tracking in VEs. The gaze-map picking principle allowed for accurate tracking of gaze targets with dynamic point of view and moving articulated humanoid targets. The reliability of the technique has been proved with experimental data and its flexibility has been confirmed within our validation study. However, its accuracy highly depends on experimental conditions and dynamic factors. Table 6.1 synthesizes the expectable accuracy depending

Table 6.1: Gaze picking precisions

| Setup | Eye tracker on projection screen | | | HMD |
|---|---|---|---|---|
| *Picking area* | *< 5%* | *> 5%* | *> 10%* | *> 15%* |
|  | Face | Body part | Body | Body |
|  | Face part | Face | Body part | Body |
|  | Face part | Face part | Face | Body part |



on the size of the picking area (in percent of screen width) and the relative distance to the humanoid target (subjective viewpoints). In short, an accurate eye-to-eye avoidance detection can be made with eye tracking in back-projection immersion conditions. In this case, the behavioral bias can be determined from almost any distance to the VH with a picking area of approximately 5%. We have been experimenting with very large picking areas (>15%) in the HMD immersion conditions and concluded that a very poor gaze avoidance analysis can be done in this case; with a simple HMD without integrated eye tracker, the tracking of targets centered in the field of vision does not allow for eye-to-eye avoidance analysis, and can only be used to evaluate the ability of patients to face people when addressing them.

Both assessment techniques were easily integrated into the VRE procedure. By taking advantage of the immersion conditions, the monitoring of patients' reactions was done with accurate devices and in ecologically valid situations. A similar assessment in life-like in-vivo exposure would have required wearable devices or laborious correlations with video recordings.

### 6.2.2 Cognitive validity

As introduced in chapter 4, overt behaviors and physiological states are two means to observe patients' reactions. The third one, cognitive responses, cannot be computerized and was left as therapist's responsibility. However, this does not mean that the exposure mediation does not influence the way therapists can perform their cognitive assessment.

According to discussions we had with the therapist during case studies, it seems that the simulations provided more material for the post-exposure cognitive sessions than expected. It was a unique opportunity for therapist and patient to experience a situation together instead of recounting souvenirs and discussing them on the basis of biased points of view.

Therapists have a very particular role during VRE sessions. Firstly, they are direct observers of how patients manage during exposure. Therapists may perform some observations during sessions and relate them to patients in order to make conscious some involuntary reactions. Some sessions during clinical case studies were very rich in terms of such discussions and the improvement in the ability to cope with anxiety was significant from one session to another. Secondly, as therapists are present next to their patient during immersion, they also indirectly accompany them into the VE. The fact that patients are conscious of 'bringing' a witness gives them the feeling of being supported and better understood. The subjective description of feelings is indeed much more precise on-the-spot than after a long time.

In short, the fact that both patient and therapist can be present at the same time and see the same experience during a VR session is an important cognitive added value of VRET. The large screen projection setup had the advantage to *pseudo*-immerse the therapist during the session —he/she can see and ear within the immersion setup although he/she does not control the viewpoint. In order for therapists to have a feedback when using HMD immersion, additional visual and auditory devices are needed (e.g. desktop display).

### 6.2.3 Limitations

We have been evaluating the technological limitations of our developments and confronting them to the requirements for affective and behavioral assessment. For both, the reliability was sufficient to objectively assess subjects' involuntary reactions during exposure, but our tools were not able to provide an absolute estimation of the reaction to a stimuli that would be comparable amongst individuals. In other words, the assessment techniques we proposed have only been validated to indicate anxiety symptoms and evaluate the variations between sessions. Only larger studies with statistical analysis over a large population would be able to provide an estimator which could be used to compare reactions intensity between patients.

Moreover, concerning practical aspects, if VR facilitated the assessment into exposure sessions, we still had to rely on additional equipments for our experiments. Existing hardware which proposes a better integration of technologies, such as HMD with integrated eye tracker, could be advantageously used to really provide a 'all-in-one' exposure platform.

## 6.3 Towards clinical control and management of exposure

Giving therapists more powerful tools also lead to more prerogatives for them. Typically, relying on simulated environments required to control the scenarios during simulations in order to ensure the coherence between patients' performance and virtual actors' reactions. This control was not only a matter of automatic response, but rather a way for therapists to drive the exposure session and maintain a dramatic tension to raise anxiety.

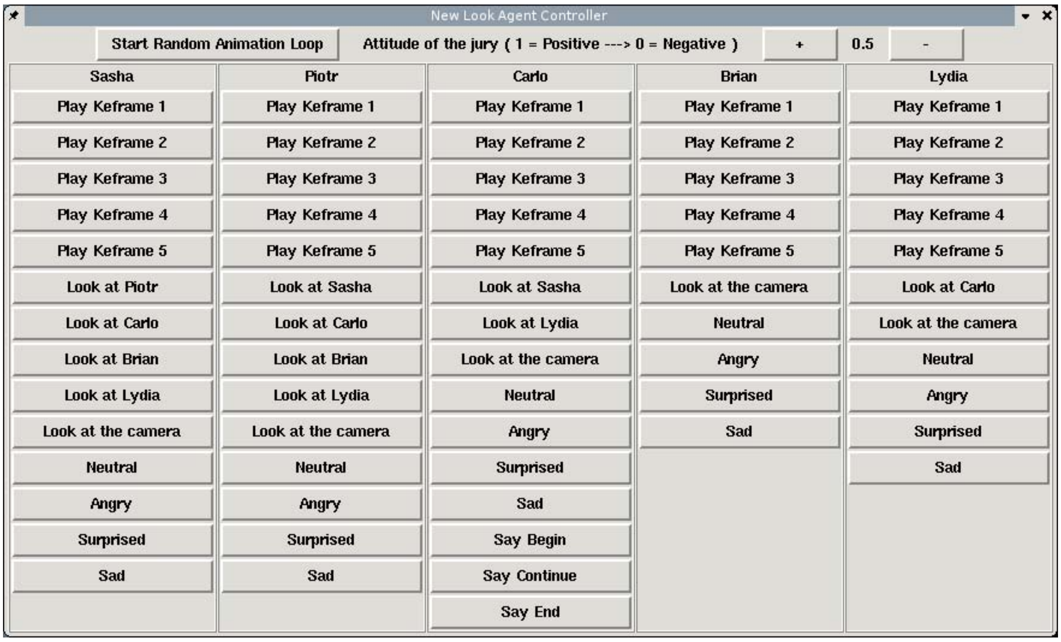


Figure 6.3: Direct control of the office assembly with GUI.

However, we have not been able to validate the use of real time scenarios control within clinical studies. The main reason for this lack was the impossibility for our medical partner to learn how to use our software prototypes (non intuitive interfaces, command lines), and, on the other hand, the long engineering effort which would have had to be done to achieve complete and robust interfaces. Nevertheless, we have investigated the design of control tools for VRE in various directions. Three proposals are given below with a synthetic description of our working implementation. A future study will be to use them under clinical or experimental conditions in order to evaluate and improve them.

### 6.3.1 Direct control

Controlling each individual action in the VE is the simplest approach. It was easily implemented with Python in a process communicating with the simulation engine through CORBA. Figure 6.3 shows how a control panel used for the office environment to trigger individual actions have been used for testing, and also why it was too low level for clinical use.

We have equally experimented with a simplified and hardware version of this control panel by implementing the connection with a button box. This second solution was retain for the validation study as we needed quick and identical exposure sessions for everybody with an elementary scenario. In this case, the buttons on a SpaceBall[2] device were triggering

---

[2] SpaceBalls were used for CAD/3D application under SGI workstation. They are still manufactured by

simple sentences pronounced by a central character in the assembly in order to welcome, motivate or thank the subject.

However, such low level control with too thin granularity would be too constraining during therapeutic exposure. Although it would be possible to compose more complex actions into scripts and provide a larger set of procedures, a better solution would be to allow therapists to edit and orchestrate scenarios in an intuitive way.

### 6.3.2 Storytelling

Interactive narratives could be used to structure the execution of a scenario by letting the user perform choices all along the execution of a pre-designed narration. This way, therapists could author various scenarios of different length and complexity. The execution of a scenario would then not be linear and have branches in a decision tree selected according to patients' answers. Of course, waiting for what happens next by displaying a menu on screen would provoke a BiP. As suggested by Heeter (2003), "a less disruptive interaction would be for an on-screen character to ask the viewer, 'should I go out with this person or not?' The viewer simply says 'yes' or 'no', and the action continues"(p.342). This is what we presented in Herbelin *et al.* (2005) as the *interactive storytelling through social channel.*

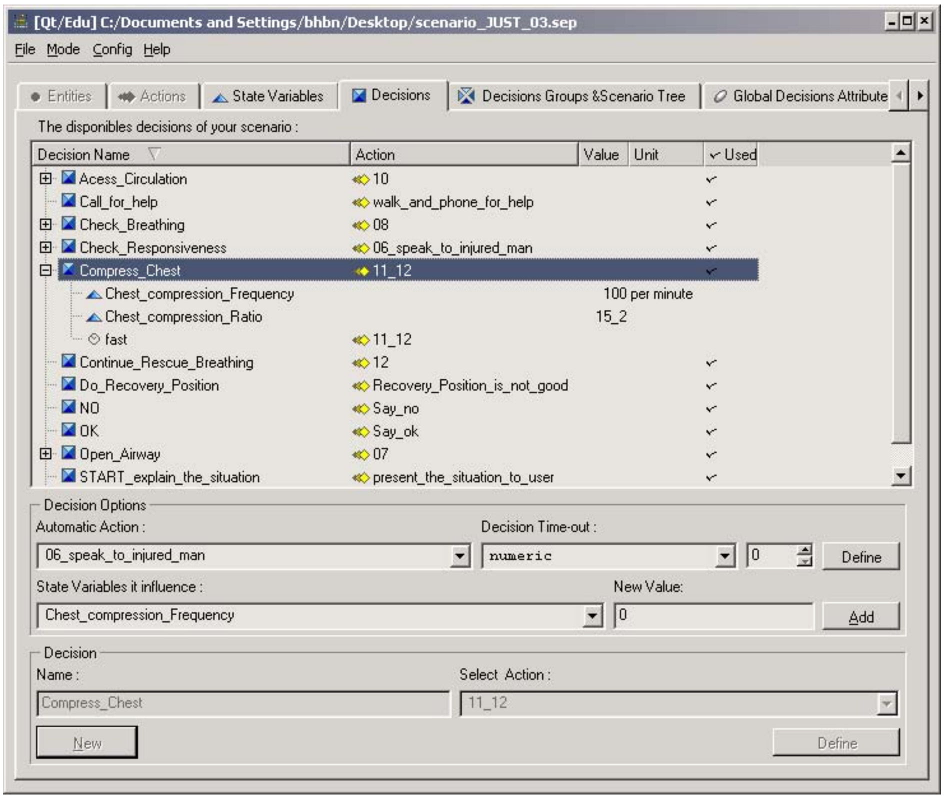
The principle of interactive narration space is to formalize the descriptive elements of a scenario in order to clarify the interconnections between story-line, virtual environment and users. Expressed by a formal semantic, the interactive decision tree makes the connections between decisions made by the patient and actions performed in the 3D simulation environment (scripts called remotely). Tree nodes are represented by a set of possible decisions made by users according to possible interactions with the VHs populating the VE at this moment. The VHs may dialogue with the patient to encourage and suggests decisions, or even reject orders or interpret them in their own way. The flexibility of social interaction tolerates this kind of relationship (people are less strict with humans than with machines) and resolves the contradiction between an involving interaction and a strong supervision.

Concretely, we have validated this kind of narrative within the context of a medical training project described by Manganas *et al.* (2005). Figures 6.4 and 6.5 show the authoring and execution GUIs used for the description of the safety procedure to be applied in an emergency situation (heart attack). At run-time, the supervisor clicks on the execution GUI to progress in the scenario. For this application, the trainee was very strongly guided by the expected procedure, but still had the feeling of taking decisions.

For the exposure of social phobic patients, these mechanisms could be used to control a virtual job interview or an oral examination with advanced dialogue possibilities. Therapists

---

http://www.3dconnexion.com.



The disponibles decisions of your scenario :

| Decision Name | Action | Value | Unit | ✓ Used |
|---|---|---|---|---|
| ⊞ Acess_Circulation | 10 | | | ✓ |
| Call_for_help | walk_and_phone_for_help | | | ✓ |
| ⊞ Check_Breathing | 08 | | | ✓ |
| ⊞ Check_Responsiveness | 06_speak_to_injured_man | | | ✓ |
| ⊟ Compress_Chest | 11_12 | | | ✓ |
| Chest_compression_Frequency | | 100 per minute | | |
| Chest_compression_Ratio | | 15_2 | | |
| fast | 11_12 | | | |
| Continue_Rescue_Breathing | 12 | | | ✓ |
| Do_Recovery_Position | Recovery_Position_is_not_good | | | ✓ |
| NO | Say_no | | | ✓ |
| OK | Say_ok | | | ✓ |
| ⊞ Open_Airway | 07 | | | ✓ |
| START_explain_the_situation | present_the_situation_to_user | | | ✓ |

Decision Options

Automatic Action : 06_speak_to_injured_man

Decision Time-out : numeric 0

State Variables it influence : Chest_compression_Frequency

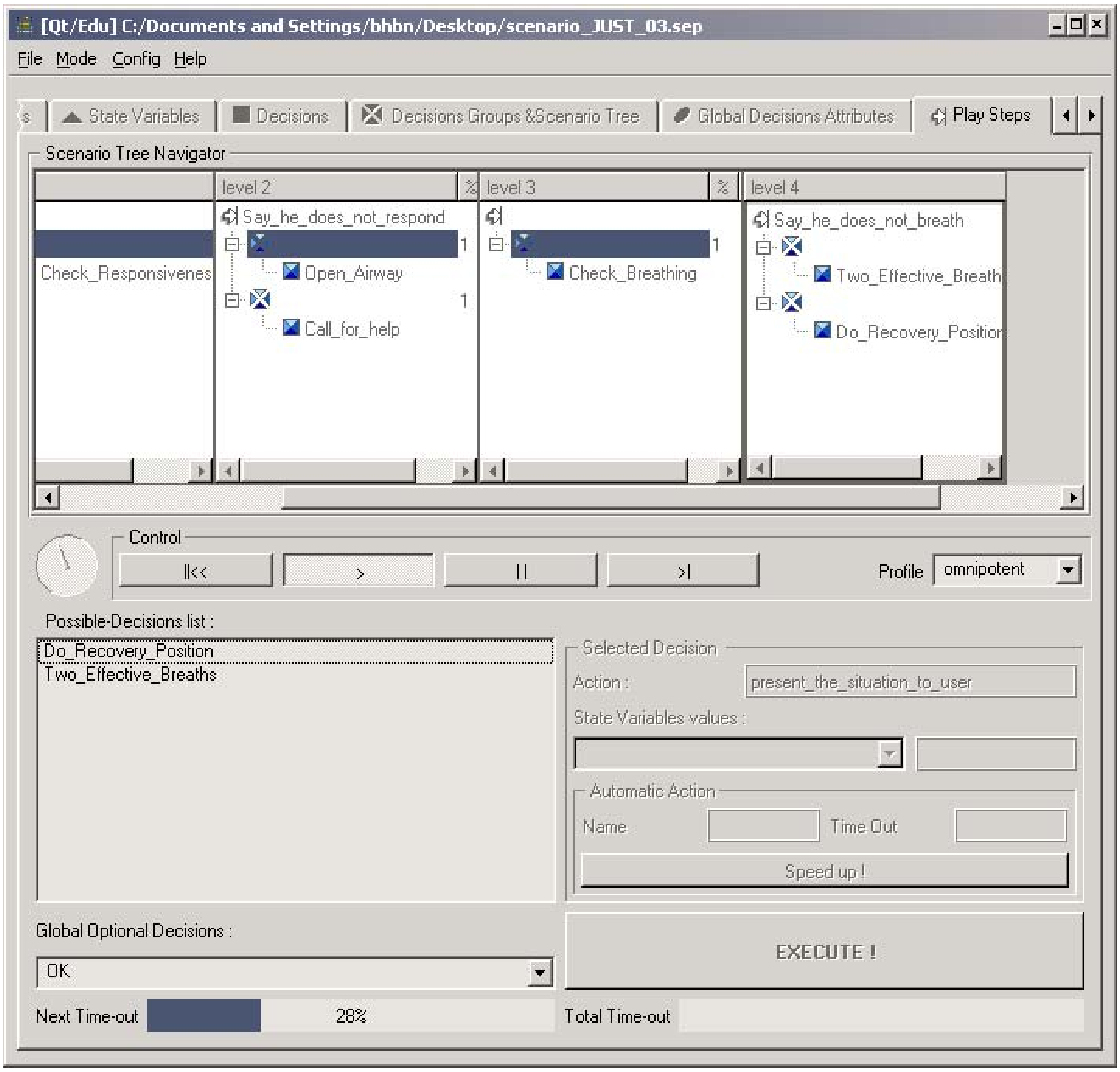
New Value: 0

Decision

Name : Compress_Chest

Select Action : 11_12

Figure 6.4: Interactive storytelling decisions authoring GUI.



Profile: omnipotent

Possible-Decisions list :
Do_Recovery_Position
Two_Effective_Breaths

Action : present_the_situation_to_user

Global Optional Decisions : OK

Next Time-out 28%

Figure 6.5: Interactive storytelling execution control GUI.

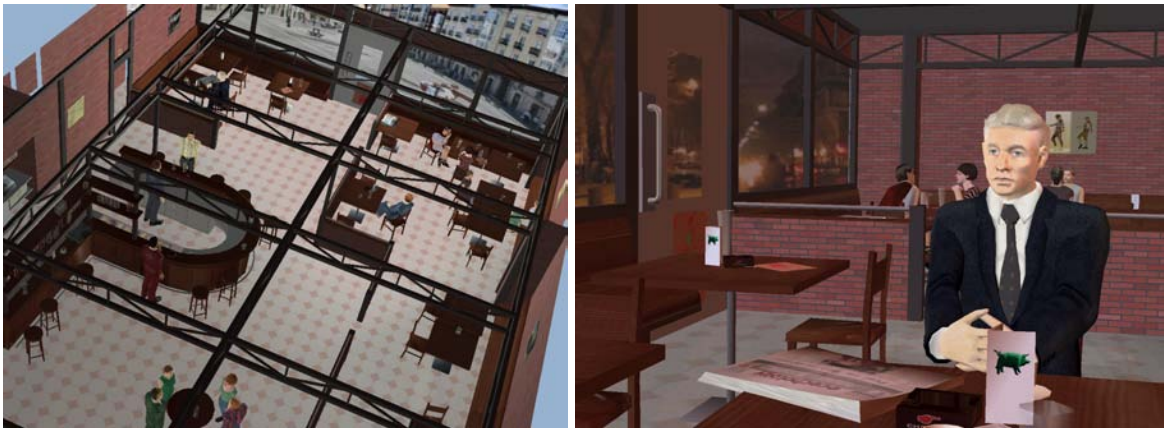


Figure 6.6: Bar environment.

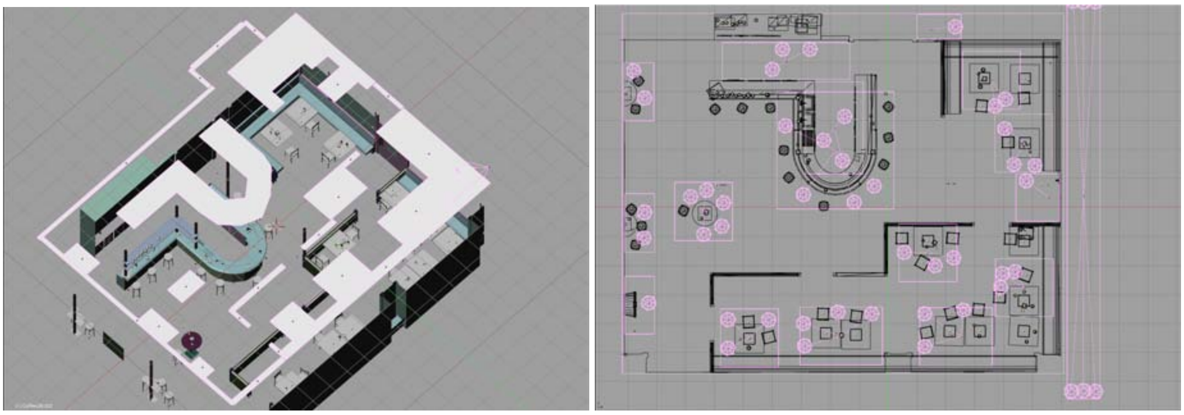
could author scenarios taking place in a given environment by themselves and then drive the interactive exposure in the direction they have planned.

### 6.3.3 Declarative scenario

For the needs of scenarios favoring the development of social skills, we have recreated the ambiance of a bar where the patient can meet people. The bar is populated with people sitting together or standing at the counter and having a drink. In a corner, a group of young people dance to the music of a jukebox. Therapists could then imagine scenarios in which a patient simulates an appointment with a stranger (fig.6.6).

However, the complexity of this scenario would quickly rise when simulating a normal life for numerous people in such social context; the barman taking orders and serving, some people discussing and having a drink, while other may dance or chat. Since the behavior of many agents in an environment cannot be driven in real time at a low level by a single user, the previously exposed control modes should anyway be combined with autonomous behaviors. This approach would of course lead to many discontinuities in the behaviors, if not some contradictions. Therefore, we proposed to consider the virtual population as a set of autonomous agents driven by artificial intelligence engines and to indirectly control them by giving high level goals. The achievement of the goal may not be immediate, but the AI should ensure it will be achieved when all the pre-requiresitesrequirements are fulfilled.

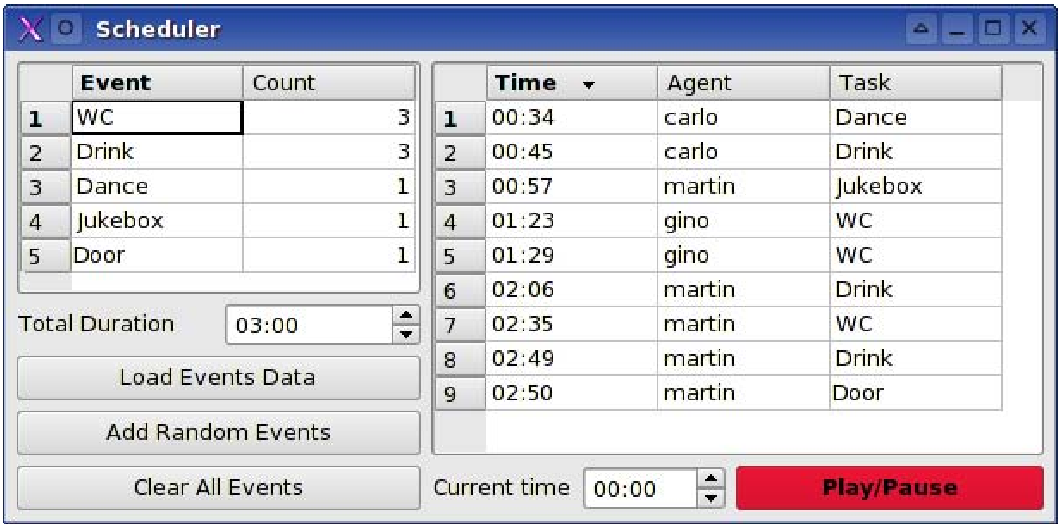
We relied on a facilitator based multi-agents simulation framework designed by Abaci *et al.* (2004) to support the planning of actions of our VHs. At a low level this system integrates the scripting of animations and the path planning for displacements. At a middle level, it infers some logical rules for the computation of action plans. Figure 6.7 shows the semantic descriptions required for the logical definition of the bar environment. Finally, at a high level and based on a declarative programming of agents behavioral rules, we obtained



(a) Path planning map

(b) Rooms and agents locations

Figure 6.7: Semantic information for the bar environment.



| | Event | Count |
|---|---|---|
| 1 | WC | 3 |
| 2 | Drink | 3 |
| 3 | Dance | 1 |
| 4 | Jukebox | 1 |
| 5 | Door | 1 |

Total Duration 03:00

| | Time | Agent | Task |
|---|---|---|---|
| 1 | 00:34 | carlo | Dance |
| 2 | 00:45 | carlo | Drink |
| 3 | 00:57 | martin | Jukebox |
| 4 | 01:23 | gino | WC |
| 5 | 01:29 | gino | WC |
| 6 | 02:06 | martin | Drink |
| 7 | 02:35 | martin | WC |
| 8 | 02:49 | martin | Drink |
| 9 | 02:50 | martin | Door |

Current time 00:00

Figure 6.8: Declarative scheduling of events for the bar scenario.

an autonomously running environment populated with several agents behaving in a parallel and coherent manner.

The next phase was to allow for indirect control on this scenario. The idea here was to rely on the ability of the action planner to handle all the intermediate actions to reach a goal. From a therapeutic point of view, the goal of an exposure session is to generate stress. For social phobic people, the stress is provoked by the gaze of others. We have therefore selected goals involving gaze actions toward the patient, such as someone passing next to him when going to the bathroom or the waiter coming to offer his services.

The added value of such a system compared to a basic scripting is better explained with an example. Imagine a therapist had the possibility to trigger an agent to walk to the bathroom as, on the way, it will cross the patient and look at him. In order to stress him, the therapist may repeat the operation several times. However, from the patient's perspective, the agent may be walking back and forth without reason. When using action planning, the agent needs to perform many actions before going to the bathroom; it will go there only when it needs to, it will need to only after having finished a drink, but to have a drink it must have ordered it —i.e. call and wait for the waiter, talk to him, wait again, and take the drink. This chain of events gives a sense to actions from the patient's point of view. Moreover, it is an automatic consequence of the rules describing the agents and is not hard-coded. As opposed to scripting and storytelling, the combination of multiple goals will provoke a complex chain of parallel interdependent actions and will bring life to the VE.

By declaring goals on a simple scheduling GUI (fig.6.8), therapists could control anxiety-provoking actions and global levels of stress generated within the bar scenario.

### 6.3.4 Discussions

Thanks to our collaboration with a therapist during the case studies, we have observed that there was also the need for changes in scenarios from a session to another. The VE may be the same, agents' behaviors not much different, but the behavioral exercises should evolve according to the patient's progress. In the office scenario, patients could just talk, talk while looking at everybody, talk while looking at people in the eyes, prepare a professional presentation, or finally give an impromptu job interview. From this global therapeutic point of view, the progression among exposure sessions also has to be organized and personalized for each patient with design and authoring tools for IVEs. As such, the management of VRE tools may soon be part of VR therapists' expertise.

Our experience with VRE systems and scenario authoring tools allows us to be quite optimistic on the possibilities to simplify the management of IVE contents. Choosing an environment, picking a few agents, placing them and connecting with pre-scripted behaviors, working on a story or selecting goals; this technology may soon be available thanks to the advance in applied research and in game engines development.

We have also observed that improvements in clinical assessment have a strong impact on the management of the therapeutic process. As therapists who could easily make behavioral and affective observations could compare throughout sessions and correlate with scenario content in order to determine the highest anxiety provoking situation for a particular patient. By targeting the stimuli and providing judicious cognitive feedback, we believe that therapy efficiency could highly be improved.

# Chapter 7

# Conclusion and future works

VR entered the field of mental health some years ago in order to determine if it could be a valid replacement to classical in-vivo exposure. Numerous studies on many phobias such as fear of flying, acrophobia, or arachnophobia have shown that the simulation of the anxiety provoking situation with VR was efficient in the course of a CBT. Among the anxiety disorders to be treated, SP is one of the most common. However, although optimistic regarding the potential of simulated public speaking situations, the previous experiments have shown that some innovations and improvements in the simulation of virtual humans may still be required.

## 7.1 Contributions and limitations

We have therefore proposed software solutions for the development of a generic VRE platform and implemented applications for the exposure to public speaking situations. Our system supports HMD and back-projection immersion setups and has proved to be clinically suitable and affordable in the case of exposure therapy. In addition to the basic requirements for 3D immersion, we have developed advanced VH simulations, with scriptable behaviors for social interaction and specific actions targeting visual contact. Our systems has been validated under experimental and clinical conditions, with subjects presenting no or mild social discomfort as well as with social phobic patients. The observations we have been able to perform during a preliminary study, three clinical case studies and a larger validation study were all coherent over more than 200 people; absence of anxiety during VRE with non-phobic subjects and significant correlation between social phobia tendencies and anxiety.

Our contribution regarding these questions not only confirms that VRE fulfills the conditions of SAD therapeutic exposure but also shows that VR can integrate the CBT clinical process without problem. In our case studies, the therapist have been able to efficiently take advantage of different immersion setups and has started to make use of the generic office IVE to vary the exposure scenarios. Further investigations on the potential variety of scenarios available with a generic VR platform has supplied us with new opportunities to improve the ecological validity of behavioral exposure therapy. Our studies have been limited in the amount of characters and the richness of their behaviors, but the upcoming improvements in computer graphics and animation shall certainly increase the application possibilities for VRE.

Moreover, we do not consider that VRE should be limited to ”reality on demand”, like a song on a CD or a movie on a DVD. In order to provide augmented functionality to the VRE tools, we have equally developed systems for monitoring anxiety responses during exposure and have been experimenting with emotional and behavioral assessment tools, respectively based on biofeedback and eye tracking devices. Our investigations with physiological measurements have leaded to the selection of computerized features based on two principal signals (heart rate and skin conductance level) allowing for a reliable assessment of the arousal of affective states, but have also shown the limitations in physiological signals interpretation and the limited validity across individuals. Our experimentation with eye tracking technology have shown that it was possible to integrate gaze tracking with VRE. We have developed a 3D gaze tracking method for the picking of moving articulated targets from arbitrary point of view and evaluated its precision under immersion conditions. Additional experiments on the tracking of gaze behaviors during exposure have then proved its usability for objective and reliable assessment of gaze safety behaviors linked to the pathology —i.e. eye-to-eye avoidance in the case of social phobia.

The innovations regarding these issues do not stand in the actual usage of these techniques, but rather in their integration into a VRE system. We have shown that the use of such assessment tools could be simplified and more adapted to a VR therapeutic process than to classical in-vivo exposure. We believe this could improve the quality and efficiency of therapy, for example by improving clinical evaluation or providing rating references for behavioral exercises. However, the validity of affective and behavioral assessment is still limited by the important variations among individuals. Only further experimental studies on very large cohorts would be able to provide statistically valid models for absolute and objective evaluations.

## 7.2 Influence of phobia on presence

We have seen that the collaboration between therapist and VR expert was fruitful for the former. In this section, we want to emphasize on the reciprocal benefit for the latter in the understanding of the underlying mental mechanisms supporting VRE during therapeutic exposure, in other words in the understanding of immersion, presence and sense of presence.

For our researches needs, we did not want to evaluate SoP directly with questionnaires — as anyway, their validity is highly controversial. However, we have had to deal with several indirect aspects of presence when analyzing human reactions during immersion — reaction to stimuli (Lee, 2004), physiological responses (Wiederhold *et al.*, 2002a), task efficiency or even relations to gaming Klimmt & Vorderer (2003). Then, we experimented with VR systems and have observed some differences in the way phobic and non-phobic react to VRE sessions. According to our experience, the VRE for social phobia has offered an appropriate experimental platform allowing for the explanation of some psychological factors of presence.

### 7.2.1 Stimulus, arousal, cognition

Before concluding on assumptions concerning presence or sense of presence, we prefer to analyze the objective elements of response we have:

- The correlation we have been able to make in all our experiments between social discomfort tendencies (scored with LSAS) and expressed or measured anxiety clearly shows that a phobic reaction due to a strong psychological stimuli can actually provoke anxiety in VR.
- We have been able to confirm that the most anxiety provoking stimuli related to social discomfort were the same in VR than in reality —i.e. gazes, feeling of being observed by VHs.
- We have observed that the cognitive impact of exposure sessions was effective in reality although the patients were conscious that their training/exposure was done in a virtual place with simulated humans.

Globally, we have observed that the usual triplet of anxiety related factors —stimulus, arousal and cognition— was involved in VRE in a similar, perhaps even in an identical, way as in reality. The phobic reaction is not very much affected by the artificiality of the experience.

We may find a logical explanation by considering the duality between rational and emotional reasoning. Emotional reasonings, which amplify the effects of other cognitive distortions, are indeed very frequent for people suffering from anxiety disorders. For social phobics, it may be expressed through thoughts such as “If I am nervous, then I must be performing terribly.” As this kind of reaction is aside (or above) any other logical pragmatic reasoning, the fact that the person knows the experience is not true does not affect the impact of the stimuli on these emotional cognitive functions. This could explain why cognitive readjustments made by therapists during VRE take effect in reality.

Moreover, as similar observations were made relatively to very different phobias such as acrophobia or arachnophobia (Krijn *et al.*, 2004; Renaud *et al.*, 2002a), we believe that our conclusions can be applied to many anxiety disorders.

### 7.2.2 Subjective risk and arousal

Pertaub *et al.* (2002) have observed that subjects felt greater anxiety in a 'negative audience' situation than with the other 'neutral' and 'positive' audiences and concluded that the subject's responses were "affected by the behavior of that audience even though they know it to be virtual" (p.76). We have also been able to confirm that the intensity and the nature of the stimuli affected the arousal of reactions during immersion, but we want to make nuances by correlating this with the social phobia tendencies.

We have seen that the distinction we established during our preliminary and validation studies between no-social-discomfort and mild-social-discomfort seems to be the limit for the emergence of a significant anxiety response to exposure in (virtual) social situation. Hence, we would consider that the impact of social stimuly is mainly dependent on the subject. If not in a clinical context, a minimum knowledge of subject's history may be necessary to estimate the expected level of reaction before concluding on a level of presence — experience with technologies, fears, phobias, etc.

In a more general point of view, we suggest that the ability of a stimulus to affect an individual is related to subjects' ability to control an emotional reasoning state: strong mental resistances or high emotional instabilities may lead to completely different reactions. In the case of mental health rehabilitation, patients' psychological vulnerability highly influences their reactions to VRE.

More specifically, we would even consider the subjective appreciation of danger as a key factor. For social phobics, gazes or people being highly frightening factors, their reproduction in a VE provokes an arousal proportional to the subjective feeling of risk it represents —even if it is a simplified or symbolic representation. The same applies to many fears and was confirmed in the VRET with other clearly identified factors (spiders, height, etc.). It could even be generalized by considering the enjoyment of people playing computer games would be augmented by the risks and the possibility to be virtually killed or injured.

### 7.2.3 Links between arousal, behavior and presence

We have experienced, within a large range of social anxiety levels, that safety behaviors usually observed in reality were indeed reproduced in VR with respect to the social phobia tendencies — e.g. social phobics have gaze avoidance toward virtual gazes whereas non-phobic face the 3D character when addressing them. This was confirmed by Garau *et al.*

(2005) who observed within simulated social situations that subjects “had respected some social norms despite the fact they knew the agents were computer-generated” (p114), and that this reaction seemed more flagrant with subjects expressing a higher social discomfort.

There is no reason to prevent the mental processes involved in the reaction of people suffering from anxiety disorders from being applied in VR, as anyway they are neither rational reasoning nor conscious reactions. Even though subjects are conscious of the artificiality of the situation, they react as they would in reality when they are performing in IVEs simply because they don’t have or know any other way to do it. They may be internally fighting with contradictory thoughts and feelings (“it’s not real, but it is still there!”) but will do ’only’ their best to face the situation anyway, often ending with the same schematics as in reality. Keeping only rational reasoning in the situation is not our natural way of behaving either. As such, keeping in mind that the environment is artificial requires mental efforts or a very particular state of mind (e.g. VR programmer).

From this perspective, the illusion of non-mediation would not be the cause of presence as suggested by Lombard & Ditton (1997), but only a vector for the stimulating efficiency of the content. This does not mean that the role of immersion technologies should be underestimated. Their function is to provide favorable conditions for the arousal rising. In VRE, there is an essential need for appropriate mediation of the stimuli to guarantee a high anxiety level; the media accessibility has to be direct and easy in order for the stimuli to affect the patient without going through high level cognitive processes.

Our feeling is that, from the mental process point of view, entering a virtual place is not different from behaving in reality. At the brain level, the additional stimuli related to VR mediation are of equal importance in comparison to the content perceived. When considering the perceptions processing, we can imagine how stimuli are continuously elected to be the most urgent/important to be treated according to their affective/cognitive impact. If the ’winning’ stimulus comes from the virtual environment (e.g. a spider), the reaction may be linked to virtuality —we may call that presence. If the ’winning’ stimulus comes from reality (e.g. someone talking outside), the reaction may be linked to reality —we may call that break in presence. Accordingly, VR hardware setups need to be natural and audiovisual contents rich and targeted to guarantee a successful impact in mind of simulated stimuli.

However, such a simplification of the mental process may only be considered as an aid for VR experts and, thanks to the collaboration with mental health professionals, the whole complexity of the human mind should be taken into account in further investigations.

## 7.3 Future works in virtual reality exposure therapy

From a general point of view, we have been able to observe that therapists are very demanding but also still very dependent on technologies experts to make use of the benefits of VRE. We believe that immersive 3D simulations, considered as an advantageous substitution to in-vivo exposure, may become quite common for psychiatrists in a near future.

However, the possibilities to provide rich session management and improved clinical assessment still need to be investigated in order to be fully validated and integrated into clinical usage. Besides the technological limitations in VR simulation (scenario management, crowds rendering, behavioral animation, etc.) and computerized assessment (affective computing, tracking), the main challenge will now be to better integrate the possibilities offered by these new tools from the therapeutic point of view. The first step toward this goal would be to validate in close collaboration with therapists the usability and efficiency of the scenario design and management tools we developed. To be more precise, we have determined at least three innovations in the therapeutic process which should be investigated more in depth;

- Adjustment of exposure intensity during a session (direct control interfaces, automatic biofeedback loop),
- Elaboration of progressive exposure programs through sessions (scenario authoring tools, behavioral exercises adapted to previous avoidance score),
- Adaptation to the specific anxiety provoking situation for each individual (e.g. assessment with various virtual stimuli, change of immersion solution).

In addition to these improvements related to the use of VRE by therapists, the possibilities to provide such systems for home exercises should be explored in parallel.

Finally, the particular case of social phobia could let us imagine how the presence of virtual humans in IVE could be more influent on the sense of presence in VR than the lack of peripheral/stereo vision or the limits in realism. However, general conclusions could not be drawn and further integration of VR in the field mental health may soon bring better psychologically valid interpretations of the mental mechanisms of presence.

# Bibliography

Abaci, T., Ciger, J., & Thalmann, D. 2004. Speculative planning with delegation. *In: Proceedings of the 3rd International Conference on Cyberworlds*. Tokyo, JAPAN: IEEE.

Baerlocher, P., & Boulic, R. 2004. An Inverse Kinematic Architecture Enforcing an Arbitrary Number of Strict Priority Levels. *The Visual Computer*, **20**(6), 402–417.

Beidel, D.C., Turner, S.M., & Jacob, R.G. 1989. Assessment of Social Phobia: Reliability of an Impromptu Speech Task. *Journal of Anxiety Disorders*, **3**, 149–158.

Biocca, F. 1997. The Cyborg's dilemna: Progressive embodiment in virtual environments. *Journal of Computer Mediated Communication*, **3**(2).

Boulic, R., Ulicny, B., & Thalmann, D. 2004. Versatile Walk Engine. *Journal Of Game Development*, **1**(1).

Bradley, M.M., & Lang, P.J. 1994. Measuring emotions: the Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, **25**, 49–59.

Cahen, O. 1990. *L'image en relief*. Masson.

Carlin, A.S., Hoffman, H.G., & Weghorst, S. 1997. Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behaviour Research and Therapy*, **35**(2), 153–8.

Cottraux, J., Berthoz, A., Jouvent, R., Pull, C., & Gueyffier, F. 2004 (Sept.). A comparative controlled study of VR and Cognitive Behaviour Therapy in Panic disorder with agoraphobia: methodology. *Pages 69–74 of: Proceedings of the Third International Workshop on Virtual Rehabilitation (IWVR04)*.

de Heras Ciechomski, P., Ulicny, B., Cetre, R., & Thalmann, D. 2004. A case study of a virtual audience in a reconstruction of an ancient Roman odeon in Aphrodisias. *In: The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*.

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

Dillon, C., Keogh, E., Freeman, J., & Davidoff, J. 2000 (March). Aroused and Immersed: The Psychophysiology of Presence. *In: Proceedings of Presence 2000 - the 3rd International Workshop on Presence*.

Dillon, C., Keogh, E., Freeman, J., & Davidoff, J. 2001 (May). Presence: Is Your Heart In It? *In: Proceeding of the fourth Annual International Workshop on Presence*.

DSM-IV. 1994. *Diagnostic and Statistical Manual of Mental Disorders*. 4 edn. American Psychiatric Association, Washington DC.

Egges, A., Molet, T., & Magnenat Thalmann, N. 2004 (Oct.). Personalised Real-time Idle Motion Synthesis. *Pages 121–130 of: Proceeding of Pacific Graphics 2004*.

Ekman, P., & Friesen, W.V. 1978. *Facial Action Coding System*. System Manual., Human Interaction Laboratory, Dept. of Psychiatry., University of California Medical Center, San Francisco.

Emering, L., Boulic, R., Molet, T., & Thalmann, D. 2000. Versatile Tuning of Humanoid Agent Activity. *Computer Graphics Forum*, **19**(4), 231–242.

Foa, E.B., Gilboa-Schechtman, E., Amir, N., & Freshman, M. 2000. Memory Bias in Generalized Social Phobia: Remembering Negative Emotional Expressions. *Journal of Anxiety Disorders*, **14**(5), 501–519.

Foley, J.D., van Dam, A., Feiner, S.K., & Hughes, J.F. 1990. *Computer graphics: principles and practice*. Second edn. Addison-Wesley Longman Publishing Co., Inc.

Gaggioli, A., & Breining, R. 2001. *Communications Through Virtual Technology: Identity Community and Technology in the Internet Age*. IOS Press Amsterdam. Chap. 5 Perception and cognition in immersive Virtual Reality.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, A.M. 2003. The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. *In: Proceedings of Conference on Human Factors in Computing Systems (CHI2003)*. Lauderdale, Florida: ACM.

Garau, M., Slater, M., Pertaud, D.P., & Razzaque, S. 2005. The Response of People to Virtual Humans in an Immersive Virtual Environment. *Presence, Teleoperators and Virtual Environments*, **14**(1), 104–116.

Garcia-Palacios, A., Hoffman, H.G., Carlin, A., Furness, T-A., & Botella, C. 2002. Virtual reality in the treatment of spider phobia: A controlled study. *Behaviour Research and Therapy*, **40**(9), 983–993.

Geiger, C., Oppermann, L., & Reimann, C. 2003. 3D-Registered Interaction-Surfaces in AR Space. *In: Proceedings of 2nd IEEE International Augmented Reality Toolkit Workshop.*

Glardon, P., Boulic, R., & Thalmann, D. 2004. PCA-based Walking Engine using Motion Capture Data. *In: Computer Graphics International (CGI'04).*

Hanjalic, A., & Xu, L.Q. 2001 (Dec.). User-oriented Affective Video Content Analysis. *In: Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'01).*

Harris, S.R., Kemmerling, R. L., & North, M.M. 2002. Brief Virtual Reality Therapy for Public Speaking Anxiety. *Cyberpsychology & behavior*, **5**(6), 543–550.

Healey, J., & Picard, R.W. 1998. *Digital Processing of Affective Signals*. Tech. rept. 444. Massachusetts Institute of Technology.

Heeter, C. 1992. Being there: the subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, **1**(2), 262–271.

Heeter, C. 2003. Reflexions on Real Presence by a Virtual Person. *Presence, Teleoperators and Virtual Environments*, **12**(4), 335–345.

Herbelin, B., Riquier, F., Vexo, F., & Thalmann, D. 2002. Virtual Reality in Cognitive Behavioral Therapy: a preliminary study on Social Anxiety Disorder. *In: Proceedings of the 8th International Conference on Virtual Systems and Multimedia (VSMM).*

Herbelin, B., Ponder, M., & Thalmann, D. 2005. Building exposure: synergy of interaction and narration through social channel. *Presence. Teleoperators and Virtual Environments*, **14**(2).

Hodges, L.F., Kooper R. Rothbaum B.O. Opdyke D. de Graaff D.D. Williford J.S. & North M.M. 1995. Virtual Environments for treating the fear of heights. *IEEE Computer*, **28**(7), 27–34.

Hodges, L.F., Rothbaum, B.O., Watson, B.A., Kesser, G.D., & Opdyke, D. 1996. Virtual Reality Exposure for Fear of Flying Therapy. *IEEE Computer Graphics Applications*, **16**(6), 42–49.

Hoffman, H.G., Garcia-Palacios, A., Carlin, A., & Botella-Arbona, C. 2003. Interfaces that heal: coupling real and virtual objects to treat spider phobia. *International Journal of Human Computer Interaction*, **16**(2), 283–300.

Hofmann, S.G., Gerlach, A.L., Wender, A., & Roth, W. 1997. Speech Disturbances and Gaze Behavior During Public Speaking in Subtypes of Social Phobia. *Journal of Anxiety Disorders*, **11**(6), 573–585.

Horley, H., Williams, L.M., Gonsalvez, C., & Gordon, E. 2003. Social phobic do not see eye to eye: A visual scanpath study of emotional expression processing. *Journal of Anxiety Disorders*, **17**(:), 33–44.

James, Laura K., Lin, Chien-Yu, Steed, Anthony, Swapp, David, & Slater, Mel. 2003. Social anxiety in virtual environments: Results of a pilot study. *Cyberpsychology & behavior*, **6**(3), 237–243.

Jang, D.P., Ku, J.H., Choi, Y.H., Wiederhold, B.K., Nam, S.W., Kim, I.Y., & Kim, S.I. 2002. The development of virtual reality therapy (VRT) system for the treatment of acrophobia and therapeutic case. *IEEE Transactions on Information Technology in Biomedicine.*, **6**(3), 213–217.

Jo, H.J., Ku, J.H., Jang, D.P., Shin, M.B., Ahn, H.B., Lee, J.M., Cho, B.H., & Kim, S.I. 2001. The development of the virtual reality system for the treatment of the fears of public speaking. *Studies in health technology and informatics*, **81**, 209–11.

Kaur, M., Tremaine, M., Huang, N., Wilder, J., Gacovski, Z., Flippo, F., & Mantravadi, C.S. 2003 (November). Where is “it”? Event Synchronization in Gaze-Speech Input Systems. *In: 5th international conference on Multimodal interfaces, ICMI2003*.

Kim, E.J. 2005. The effect of the decreased safety behaviors on anxiety and negative thoughts in social phobics. *Journal of Anxiety Disorders*, **19**, 69–86.

Klimmt, C., & Vorderer, P. 2003. Media Psychology "is not yet there": Introducing Theories on Media Entertainment to the Presence Debate. *Presence: Teleoperators and Virtual Environments*, **12**(4), 346–359.

Krapichler, C., Haubner, M., Engelbrecht, R., & Englmeier, K.H. 1998. VR interaction techniques for medical imaging applications. *Computer Methods and Programs in Biomedicine*, **56**(1), 65–74.

Krijn, M., Emmelkamp, P.M., Biemond, R., de-Wilde de Ligny, C., Schuemie, M.J., & van-der Mast, C.A. 2004. Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy*, **42**(2), 229–239.

Le Callennec, B., & Boulic, R. 2004. Interactive Motion Deformation with Prioritized Constraints. *In: Proceedings of ACM SIGGRAPH / Eurographics Symposium on Computer Animation.*

Lee, J.M., Ku, J.H., Jang, D.P., Kim, D.H., Choi, Y.H., Kim, I.Y., & Kim, S.I. 2002. Virtual Reality System for Treatment of the Fear of Public Speaking Using Image-Based Rendering and Moving Pictures. *Cyberpsychology & behavior*, **5**(3), 191–196.

Lee, K.M. 2004. Why Presence Occurs: Evolutionary Psychology, Media Equation, and Presence. *Presence: Teleoperators and Virtual Environments*, **13**(4), 494–505.

Levin, A.P., Saoud, J.B., Strauman, T., Gorman, J.M., Fyer, A.J., Crawford, R., & Liebowitz, M. 1993. Responses of "Generalized" and "Discrete" Social Phobics During Public Speaking. *Journal of Anxiery Disorders*, **7**, 207–221.

Liebowitz, M.R. 1987. Social Phobia. *Modern Problems in Phamacopsychiatry*, **22**, 141–173.

Lin, C.S., Huan, C.C., Chan, C.N., Yeh, M.S., & Chiu, C.C. 2004. Design of a computer game using an eye-tracking device for eye's activity rehabilitation. *Optics and Lasers in Engineering*, **42**(1), 91–108.

Lincoln, T.M. 2003 (Feb.). *Cognitive and Behavioural Treatment of Social Phobia*. Ph.D. thesis, Christoph Dornier Foundation for Clinical Psychology, Institute Marburg (GE).

Lombard, M., & Ditton, T. 1997. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, **3**(2).

Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., & Horvath, K. 2000. Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. *In:* W. IJsselsteijn, J. Freeman, & H. de Ridder (ed), *Proceedings of the Third International Workshop on Presence.*

Manganas, A., Tsiknakis, M., Leisch, E., Ponder, P., Molet, T., Herbelin, B., Magnenat-Thalmann, N., Thalmann, D., Fato, M, & Schenone, A. 2005. Just in Time Health Emergency Interventions: an innovative approach to training the citizen for emergency situations using virtual reality techniques and advanced it tools (the VR tool). *The Journal on Information Technology in Healthcare*, **2**(6).

Marks, I.M., & Mathews, A.M. 1979. Brief standard self-rating for phobic patients. *Behaviour Research and Therapy*, **17**, 263–267.

McCraty, R. 2002. Heart Rhythm Coherence-An Emerging Area of Biofeedback. *Biofeedback*, **Spring**, 23–25.

Meehan, M., Insko, B., Whitton, M., & Brooks, F.P. Jr. 2002. Physiological measures of presence in stressful virtual environments. *Pages 645–652 of: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. ACM Press.

Mennin, D.S., Fresco, D.M., Heimberg, R.G., Schneier, F.R., Davies, S.O., & Liebowitz, M.R. 2002. Screening for social anxiety disorder in the clinical setting: using the Liebowitz Social Anxiety Scale. *Journal of Anxiety Disorders*, **16**, 661–673.

Merinkangas, K.R, & Angst, J. 1995. Comorbidity and social phobia : evidence from clinical epidemiologic, and genetic studies. *European Archives of Psychiatry & Clinical Neuroscience*, **244**(6), 297–303.

Morris, J.D. 1995. SAM:The Self-Assessment Manikin. An Efficient Cross-Cultural Measurement of Emotional Response.(Observations). *Journal of Advertising Research*, Nov., 1–5.

Nasoz, F., Lisetti, C.L., Alvarez, K., & Finelstein, N. 2003 (June). Emotional Recognition from Physiological Signals for User Modeling of Affect. *In: Proceedings of the 3rd Workshop on Affective and Attitude User Modeling*.

North, M.M., North, S.M., & Coble, J.R. 1998. Virtual reality therapy: an effective treatment for the fear of public speaking. *International Journal of Virtual Reality*, **3**(3), 2–7.

North, M.M., Schoeneman, C.M., & Mathis, J.R. 2002. Virtual reality therapy: case study of fear of public speaking. *Pages 18–20 of: Medicine Meets Virtual Reality 02/10*.

Papagiannakis, G., Schertenleib, S., O'Kennedy, B., Poizat, M., Magnenat-Thalmann, N., Stoddart, A., & Thalmann, D. 2005. Mixing Virtual and Real scenes in the site of ancient Pompeii. *Computer Animation and Virtual Worlds*, **16**(1), 11–24.

Parrott, M., Bowman, D., & Ollendick, T. 2004. A methodology for designing specific animal phobia stimuli for virtual reality exposure therapy. *Cyberpsychology & behavior*, **7**(3), 300–301.

Paul, G. 1966. *Insights vs. desensitization in psychotherapy*. Stanford University Press.

Peinado, M., Herbelin, B., Wanderley, M., Le Callennec, B., Boulic, R., Thalmann, D., & Méziat, D. 2004 (Sept.). Towards Configurable Motion Capture with Prioritized Inverse Kinematics. *In: Proc. of the third International Workshop on Virtual Rehabilitation (IVWR'04)*.

Pertaub, D-P, Slater, M, & Barker, C. 2001. An experiment on fear of public speaking in virtual reality. *Studies in health technology and informatics*, **81**, 372–378.

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

Peruch, P., & Mestre, D.R. 1999. Between desktop and head immersion: Functional Dynamic Field during vehicle control and navigation in virtual environments. *Presence: Teleoperators and Virtual Environments*, **8**(1), 54–64.

Picard, R.W. 1997. *Affective Computing*. The MIT Press, Cambridge, MA.

Picard, R.W., Vyzas, E., & Healey, J. 2001. Towards Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(10), 1175–1191.

Ponder, M. 2004. *LinkComponent-based methodology and development framework for virtual and augmented reality systems*. Es Science, Ecole Polytechnique Federale de Lausanne, Lausanne.

Ponder, M., Herbelin, B., Molet, T., Schertenleib, S., Ulicny, B., Papagiannakis, G., Magnenat-Thalmann, N., & Thalmann, D. 2003a (May). Immersive VR decision training: telling interactive stories featuring advanced virtual humans. *In: Proceedings of the 9th Eurographics Workshop on Virtual Environments (EGVE)*.

Ponder, M., Papagiannakis, G., Molet, T., Magnenat-Thalmann, N., & Thalmann, D. 2003b (July). VHD++ Development Framework: Towards Extendible, Component Based VR/AR Simulation Engine Featuring Advanced Virtual Character Technologies. *Pages 96–104 of: Computer Graphics International (cgi 2003)*.

Prothero, J., Parker, D., Furness, T.A., & Wells, M. 1995. Towards a Robust, Quantitative Measure of Presence. *Pages 359–366 of: Proceedings of Conference on Experimental Analysis and Measurement of Situational Awareness*.

Rand, D., Kizony, R., Feintuch, U., Josman, N., Rizzo, A.A., & Weiss, P.L. 2003 (Sept.). Video projection versus HMD VR system for rehabilitation: Within and between comparisons. *Pages 105–112 of: Proceedings of the 2nd International Workshop on Virtual Rehabilitation*.

Rani, P., Sarkar, N., & Smith, C. 2003 (Sept.). Affect-Sensitive Human-Robot Cooperation: Theory and Experiments. *Pages 232–238 of: Proceedings of the IEEE International Conference on Robotics and Automation*.

Renaud, P., Bouchard, S., & Proulx, R. 2002a. Behavioral avoidance dynamics in the presence of a virtual spider. *IEEE Trans Inf Technol Biomed*, **6**(3), 235–43.

Renaud, P., Cusson, J.-F., Bernier, S., Décarie, J., Gourd, S.-P., & Bouchard, S. 2002b (Nov.). Extracting perceptual and motor invariants using eye-tracking technologies in virtual immersions. *Pages 73–78 of: Proceedings of the IEEE International Workshop on Haptic Virtual Environments and their Applications (HAVE'2002).*

Riquier, F., Stankovic, M., & Chevalley, A.F. 2002. Virtual gazes for social exposure: Margot and Snow White. *In: Proceedings of the 1st. International Workshop on Virtual Reality Rehabilitation.*

Riva, G., & Waterworth, J.A. 2003 (Apr.). *Presence and the Self: a congitive neuroscience approach*. Presence-connect.

Riva, G, Bacchetta, M, Baruffi, M, & Molinari, E. 2001. Virtual reality-based multidimensional therapy for the treatment of body image disturbances in obesity: a controlled study. *Cyberpsychology & behavior*, **4**(4), 511–526.

Riva, G, Bacchetta, M, Cesa, G, Conti, S, & Molinari, E. 2003a. Six-month follow-up of in-patient experiential cognitive therapy for binge eating disorders. *Cyberpsychology & behavior*, **6**(3), 251–258.

Riva, G., Alcaniz, M., Anolli, L., Bacchetta, M., Banos, R., Buselli, C., Beltrame, F., Botella, C., Castelnuovo, G., Cesa, G., Conti, S., Galimberti, C., Gamberini, L., Gaggioli, A., Klinger, E., Legeron, P., Mantovani, F., Mantovani, G., Molinari, E., Optale, G., Ricciardiello, L., Perpina, C., Roy, S., Spagnolli, A., Troiani, R., & Weddle, C. 2003b. The VEPSY UPDATED Project: Clinical Rationale and Technical Approach. *Cyberpsychology & behavior*, **6**(4), 433–9.

Riva, G., Botella, C., Légeron, P., & Optale, G. (eds). 2004. *Cybertherapy: Internet and Virtual Reality as Assessment and Rehabilitation Tools for Clinical Psychology and Neuroscience*. Amsterdam: Ios Press.

Rizzo, A.A., & Jounghyun Kim, G. 2005. A SWOT Analysis of the Field of Virtual Rehabilitation and Therapy. *Presence, Teleoperators and Virtual Environments*, **14**(2), 119–146.

Rizzo, A.A., Wiederhold, M., & Buckwalter, J.G. 1998. Basic issues in the use of Virtual Environments for Mental Health applicaitons. *Pages 21–42 of:* Riva G. Wiederhold B. Molinari E. (Eds.) (ed), *Virtual Environments in Clinical Psychology and Neuroscience: Methods and techniques in advanced patient-therapist interaction.* Ios Press: Amsterdam.

Rizzo, A.A., Schultheis, M., Kerns, K.A., & Meteer, C. 2004a. Analysis of Assets for Virtual Reality Applications in Neuropsychology. *Neuropsychological Rehabilitation*, **14**(1/2), 207–239.

Rizzo, A.A., Bowerly, T., Shahabi, C., Buckwalter, J. G., Klimchuk, D., & Mitura, R. 2004b. Diagnosing Attention Disorders in a Virtual Classroom. *IEEE Computer: Innovative Technology for Computer Professionals*, **37**(6), 87–89.

Rizzo, A.A., Pair, J., & McNerney, PJ. 2004c (Dec.). From training to toy to treatment: design and development of a post traumatic stress disorder virtual reality therapy application for Iraq War veterans. *In: Proceeding of the 24th Army Science Conference*.

Rizzo, M., Moon, J., Wilkinson, M., Bateman, K., Jermeland, J., & Schnell, T. 2002. Ocular search of simulated roadway displays in drivers with constricted visual fields. *Journal of Vision*, **2**(7), 162.

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

Rothbaum, B.O., Hodges, L., Alarcon, R., Ready, D., Shahar, F., Graap, K., Pair, J., Hebert, P., Gotz, D., Wills, B., & Baltzell, D. 1999. Virtual reality exposure therapy for PTSD Vietnam Veterans: a case study. *Journal of traumatic stress*, **12**(2), 263–271.

Rothbaum, B.O., Hodges, L., Anderson, P.L., Price, L., & Smith, S. 2002. 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.

Roy, S. 2003. State of the art of virtual reality therapy (VRT) in phobic disorders. *PsychNology Journal*, **1**(2), 176–183.

Roy, S., Klinger, E., Legeron, P., Lauer, F., Chemin, I., & Nugues, P. 2003. Definition of a VR-Based Protocol to Treat Social Phobia. *Cyberpsychology & behavior*, **6**(4), 411–420.

Schuemie, M. 2002. Effect of an external viewpoint on therapist performance in virtual reality exposure therapy. *Pages 834–835 of: CHI '02 extended abstracts on Human factors in computing systems*. ACM Press.

Sheffer, C.E., Penn, D.L., & Cassisi, J.E. 2001. The effect of impression management demands on heart rate, self-reported social anxiety, and social competence in undergraduate males. *Journal of Anxiety Disorders*, **15**, 171–182.

Sibert, L.E., & Jacob, R.J. 2000 (April). Evaluation of Eye Gaze Interaction. *In: CHI 2000 Conference on Human Factors in Computing Systems.*

Simons, R.F., Detenber, G.H., Roedema, T.M., & Reiss, J.E. 1999. Emotion-Processing in Three Systems: The Medium and the Message. *Psychophysiology*, **36**(5), 619–627.

Slater, M. 2003 (Jan.). *A note on Presence terminology.* Presence-Connect (on-line).

Slater, M. 2004. How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, **13**(4), 484–493.

Slater, M., & Usoh, M. 1993. Representations systems, perceptual position, and presence in immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, **2**(3), 221–233.

Slater, M., & Usoh, M. 1994. *Life and Virtual Reality.* John Wiley and Sons. Pages 125–148.

Slater, M., Usoh, M., & Steed, A. 1994. Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, **3**(2), 130–144.

Slater, M., Pertaud, D.P., Barker, C., & Clark, D. 2004 (Sept.). An Experimental Study on Fear of Public Speaking Using a Virtual Environment. *In: Proceedings of the third International Workshop on Virtual Rehabilitation (IWVR2004).*

Tsui, K.C., Azvine, B., Djian, D., Voudouris, C., & Xu, L.Q. 1998. Intelligent Multi-modal Systems. *BT Technology Journal*, **16**(3), 134–144.

Usoh, M., Catena, E., Arman, S., & Slater, M. 2000. Using Presence Questionnaires in Reality. *Presence, Teleoperators and Virtual Environments*, **9**(5), 497–503.

Velten, E. 1968. A laboratory task for induction of mood states. *Behavior Research and Therapy*, **6**, 473–482.

Vincelli, F., Anolli, L., Bouchard, S., Wiederhold, B.K., Zurloni, V., & Riva, G. 2003. Experiential cognitive therapy in the treatment of panic disorders with agoraphobia: a controlled study. *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society*, **6**(3), 321–328.

Vyzas, E. 1999 (June). *Recognition of Emotional and Cognitive States Using Physiological Data.* Department of Mechanical Engineering, Massachusetts Institute of Technology.

Wald, J. 2004. Efficacy of virtual reality exposure therapy for driving phobia: A multiple baseline across-subjects design. *Behaviour Research and Therapy*, **35**(3), 621–635.

Wang, H., Prendinger, H., & Igarashi, T. 2004. Communicating Emotions in Online Chat Using Physiological Sensors and Animated Text. *Pages 1171–1174 of: Proceedings of the 1st international conference for human-computer interaction.*

Waterworth, J.A., & Waterworth, E.L. 2003 (Feb.). *The meaning of Presence.* Presence-Connect (on-line).

Watson, D., & Friend, R. 1969. Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, **33**, 448–457.

Wiederhold, B.K., & Wiederhold, M.D. 2003. Three-year follow-up for virtual reality exposure for fear of flying. *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society*, **6**(4), 441–445.

Wiederhold, B.K., Jang, D.P., Kim, S.I., & Wiederhold, M.D. 2002a. Physiological monitoring as an objective tool in virtual reality therapy. *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society*, **5**(1), 77–82.

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

Witmer, B.G., & Singer, M.J. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence, Teleoperators and Virtual Environments*, **7**(3), 225–240.

Yang, G.Z., Dempere-Marco, L., Hu, X.P., & Rowe, A. 2002. Visual search: psychophysical models and practical applications. *Image and Vision Computing*, **20**(4), 273–287.

# Appendix A

# Liebowitz Social Anxiety Scale

## A.1 Questionnaire

**Liebowitz Social Anxiety Scale** **Liebowitz MR. Social Phobia. Mod Probl Pharmacopsychiatry 1987;22:141-173**

Pt Name: Pt ID #:
Date: Clinic #: Assessment point:

| Fear or Anxiety: | Avoidance: |
|---|---|
| 0 = None | 0 = Never (0%) |
| 1 = Mild | 1 = Occasionally (1—33%) |
| 2 = Moderate | 2 = Often (33—67%) |
| 3 = Severe | 3 = Usually (67—100%) |

| | Fear or Anxiety | Avoidance | |
|---|---|---|---|
| 1. Telephoning in public. (P) | | | 1. |
| 2. Participating in small groups. (P) | | | 2. |
| 3. Eating in public places. (P) | | | 3. |
| 4. Drinking with others in public places. (P) | | | 4. |
| 5. Talking to people in authority. (S) | | | 5. |
| 6. Acting, performing or giving a talk in front of an audience. (P) | | | 6. |
| 7. Going to a party. (S) | | | 7. |
| 8. Working while being observed. (P) | | | 8. |
| 9. Writing while being observed. (P) | | | 9. |
| 10. Calling someone you don’t know very well. (S) | | | 10. |
| 11. Talking with people you don’t know very well. (S) | | | 11. |
| 12. Meeting strangers. (S) | | | 12. |
| 13. Urinating in a public bathroom. (P) | | | 13. |
| 14. Entering a room when others are already seated. (P) | | | 14. |
| 15. Being the center of attention. (S) | | | *15.* |
| 16. Speaking up at a meeting. (P) | | | 16. |
| 17. Taking a test. (P) | | | 17. |
| 18. Expressing a disagreement or disapproval to people you don’t know very well. (S) | | | 18. |
| 19. Looking at people you don’t know very well in the eyes. (S) | | | 19. |
| 20. Giving a report to a group. (P) | | | 20. |
| 21. Trying to pick up someone. (P) | | | 21. |
| 22. Returning goods to a store. (S) | | | 22. |
| 23. Giving a party. (S) | | | 23. |
| 24. Resisting a high pressure salesperson. (S) | | | 24. |

## A.2 Score Analysis

According to the answers, the therapist can compute a score between 0 and 144, with 55 being the pathological threshold and above 95 a very severe social phobia.

| LSAS score | Diagnostic |
|---|---|
| Greater than 95 | Very severe social phobia |
| 80 – 95 | Severe social phobia |
| 65 – 80 | Marked social phobia |
| 55 – 65 | Moderate social phobia |

Moreover, Mennin *et al.* (2002) have demonstrate the capacity of LSAS to distinguish SP from generalized social phobia (GSP) thanks to the distinctions which can be made by computing sub-scores. The four sub-scores which can be computed are:

- **LSAS.F** ; total score of column *Fear or Anxiety,*
- **LSAS.A** ; total score of column *Avoidance*,
- **LSAS.P** ; anxiety + avoidance scores for situations concerning social *performance* (13 items marked with 'P'),
- **LSAS.S** ; anxiety + avoidance scores for situations concerning social *situations* (11 items marked with 'S').

# Appendix B

# *Science et Cité* study forms

Here are the forms used during the "Science et Cité" festival in Lausanne for the large scale validation sudy described in section 5.3 on page 77. The introduction text and the agreement form are on the following page. Questionnaires Q.I and Q.II are on page 129. Questionnaires Q.III and Q.IV are on page 130.

**Festival Science et Cité 2005**

# Installation “J’suis timide mais j’me soigne”

Laboratoire de Réalité Virtuelle
Ecole Polytechnique Fédérale de Lausanne

## Préambule

Cette expérience consiste en une brève immersion en environnement virtuel durant laquelle vous vous adresserez à une assemblée d'humains virtuels. Ce protocole est en cours de validation pour le traitement des phobies sociales. L'objectif de l'étude est d'observer le comportement du regard dans de telles circonstances afin d'établir une base de données de référence.

## Scenario

Dans la scène virtuelle, vous devez parler à l'auditoire sur un sujet de votre choix, de préférence en relation avec vos activités: soutenir un projet professionel ou scolaire, vous présenter pour un poste ou un concours, présenter votre hobby, proposer une décision politique, etc. Imaginez que parmis toutes les personnes se présentant, ce jury devrait vous choisir vous! Pour être convaincant, pensez à bien les regarder tous.

## Conditions de participation

L'expérience est ouverte à toute personne de plus de 15 ans.

Le laboratoire de Réalité Virtuelle de l'Ecole Polytechnique Fédérale de Lausanne s'engage à

- préserver l'anonymat des sujets,
- ne pas diffuser les formulaires originaux,
- utiliser les données recueillies uniquement à des fins scientifiques (publications dans des conférences et journaux internationaux).

L'utilisation d'images de synthèse et du visio-casque pouvant parfois provoquer de légers troubles perceptifs chez des sujets particulièrement sensibles, les personnes jugées 'à risque', par exemple les épyleptiques, ne peuvent pas prendre part à l'expérience.

Aucun résultat clinique ou thérapeutique ne sera fourni à la suite de l'expérience.

Pour leur part, les personnes acceptant ces conditions consentent à

- signer le formulaire de consentement ci-dessous,
- remplir en toute honnêteté les deux formulaires qui vous seront remis.

## Formulaire de consentement

Je certifie avoir été informé des conditions de participation à l'expérience “J’suis timide mais j’me soigne” organisée par le Laboratoire de Réalité Virtuelle de l'Ecole Polytechnique Fédérale de Lausanne, et me porte volontaire pour y participer.

Signature:

## Formulaire A. (AVANT IMMERSION)

### I. Profil

| | | | | | |
|---|---|---|---|---|---|
| 1 | Sexe | ☐ | Feminin | ☐ | Masculin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Age (>15 ans) | ☐ | < 25ans | ☐ | < 35 ans | ☐ | < 45 ans | ☐ | < 60 ans | ☐ | > 60 ans |

### II. Questionnaire de tendances à la phobie sociale

Vous trouverez ci-après une liste de situations et d'activités face auxquelles l'on peut avoir peur ou que l'on peut éviter. A la suite de chaque question, faites une croix en dessous de la réponse qui vous convient le mieux concernant votre degré d'anxiété et votre tendance à l'évitement.

| | | **Anxiété** | | | | **Evitement** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Aucune | Légère | Moyenne | Sévère | Jamais | Occasionnel | Fréquent | Habituel |
| 1 | Téléphoner en public | | | | | | | | |
| 2 | Participer au sein d'un petit groupe | | | | | | | | |
| 3 | Manger dans un lieu public | | | | | | | | |
| 4 | Boire en compagnie dans un lieu public | | | | | | | | |
| 5 | Parler à des gens qui détiennent une autorité | | | | | | | | |
| 6 | Jouer, donner une représentation ou une conférence | | | | | | | | |
| 7 | Aller à une soirée | | | | | | | | |
| 8 | Travailler en étant observé | | | | | | | | |
| 9 | Ecrire en étant observé | | | | | | | | |
| 10 | Contacter par téléphone quelqu'un que vous ne connaissez pas très bien | | | | | | | | |
| 11 | Parler à des gens que vous ne connaissez pas très bien | | | | | | | | |
| 12 | Rencontrer des inconnus | | | | | | | | |
| 13 | Uriner dans des toilettes publiques | | | | | | | | |
| 14 | Entrer dans une pièce alors que tout le monde est déjà assis | | | | | | | | |
| 15 | Etre le centre d'attention | | | | | | | | |
| 16 | Prendre la parole à une réunion | | | | | | | | |
| 17 | Passer un examen | | | | | | | | |
| 18 | Exprimer son désaccord ou sa désapprobation à des gens que vous ne connaissez pas très bien | | | | | | | | |
| 19 | Regarder dans les yeux des gens que vous ne connaissez pas très bien | | | | | | | | |
| 20 | Faire un compte-rendu à un groupe | | | | | | | | |
| 21 | Essayer de "draguer" quelqu'un | | | | | | | | |
| 22 | Rapporter des marchandises dans un magasin | | | | | | | | |
| 23 | Donner une soirée | | | | | | | | |
| 24 | Résister aux pressions d'un vendeur insistant | | | | | | | | |

## Formulaire B (APRES IMMERSION)

### III. Questionnaire d'impressions générales

| | | non | légèrement | modérement | fortement | très fortement |
|---|---|---|---|---|---|---|
| 1 | Avez-vous ressenti que vous étiez le centre d'attention ? | | | | | |
| 2 | Avez-vous été dérangé par le fait de prendre la parole ? | | | | | |
| 3 | Avez-vous regardé tout le monde ? | | | | | |
| 4 | Avez-vous regardé les personnes dans les yeux ? | | | | | |

### IV. Questionnaire de sensations corporelles

| | L'expérience a-t-elle provoqué en vous : | non | légèrement | modérement | fortement | très fortement |
|---|---|---|---|---|---|---|
| 1 | Palpitations cardiaques | | | | | |
| 2 | Pression dans la poitrine | | | | | |
| 3 | Faiblesse dans les bras ou les jambes | | | | | |
| 4 | Tension musculaire, muscles endoloris | | | | | |
| 5 | Fourmillements, picotements ou engourdissements | | | | | |
| 6 | Faiblesse générale | | | | | |
| 7 | Difficultés à respirer, souffle court | | | | | |
| 8 | Sensation de chaleur ou de froid | | | | | |
| 9 | Sensation que les choses tournent | | | | | |
| 10 | Points devant les yeux | | | | | |
| 11 | Vision trouble ou distordue | | | | | |
| 12 | Tremblements, frissonnements | | | | | |
| 13 | Douleur dans le bas du dos | | | | | |
| 14 | Transpiration excessive | | | | | |
| 15 | Sensation de pression dans les oreilles | | | | | |
| 16 | Vertiges | | | | | |
| 17 | Nausée | | | | | |
| 18 | Estomac dérangé | | | | | |
| 19 | Noeud dans l'estomac | | | | | |
| 20 | Boule dans la gorge | | | | | |
| 21 | Gorge sèche | | | | | |
| 22 | Maux de tête | | | | | |

# CurriculumVitae

## Bruno Herbelin

### Education

| | |
|---|---|
| 2000 | Diploma of Advanced Studies (DEA) in Computer Sciences,<br>Université Louis Pasteur, Strasbourg (France). |
| 1997 | Engineering diploma (equivalent to Master) in Computer Sciences<br>with a specialization in software engineering,<br>Université de Technologie de Belfort-Montbéliard (France). |

### Research Experience

| | |
|---|---|
| 07/01 – 08/05 | Reaserach assistant in the Virtual Reality Laboratory, EPFL, Lausanne.<br>*Involved in the European Network of Excellence ENACTIVE (2004-2007), in the European project JUST (Training in Health Emergency Situation by Virtual Reality, 2000-2003), and in the National project PHOBIA (immersive exposure to treat social phobia, 2001-2002).* |
| 02/00 – 06/00 | Laboratoire des Sciences de l'Image de l'Informatique et de la Télédétection and Laboratoire des systèmes photoniques, Université Louis Pasteur, Strasbourg.<br>*Hand gesture communication in virtual environments; a neural network based gesture recognition system.* |

### Publications

B. Herbelin, M. Ponder & D. Thalmann. 2005. Building Exposure: Synergy of Interaction and Narration through Social Channels. *Presence, Teleoperators and Virtual Environments*, **14**(2).

A.Manganas, M.Tsiknakis, E.Leisch, M. Ponder, T.Molet, B.Herbelin, N.Magnenat-Thalmann & D.Thalmann. 2004. JUST in Time Health Emergency Interventions: An innovative approach to training the citizen for emergency situations using Virtual Reality Techniques and Advanced IT Tools (The VR Tool). *The Journal on Information Technology in Healthcare*, **2**(6), 399-412.

L. Emering & B. Herbelin. Body Gesture Recognition and Action Response. 2004. *Handbook of Virtual Humans*. N.Magnenat-Thalmann and D.Thalmann eds. John Wiley & Sons, Ltd. Chap.12, 287-302.

### Presentations and Papers

J. Ciger, B. Herbelin & D. Thalmann. 2004. Evaluation of Gaze Tracking Technology for Social Interaction in Virtual Environments. *Second International Workshop on Modelling and Motion Capture Techniques for Virtual Environments (CAPTECH)*, Zermatt.

M. Peinado , B. Herbelin , M. Wanderley, B. Le Callennec, R. Boulic, D. Thalmann & D. Méziat. 2004. Towards Configurable Motion Capture with Prioritized Inverse Kinematics. *Third International Workshop on Virtual Rehabilitation (IWVR),* Lausanne.

B. Herbelin, P. Benzaki, O. Renault, F. Riquier & D. Thalmann. 2004. Using physiological measures for emotional assessment: a computer-aided tool for Cognitive and Behavioural Therapy. *5th International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT)*, Oxford.

M.Ponder, B.Herbelin, T.Molet, S.Schertenleib, B.Ulicny, G.Papagiannakis, N.Magnenat-Thalmann & D.Thalmann. 2003. Immersive VR Decision Training: Telling Interactive Stories Featuring Advanced Virtual Humans. *9th Eurographics Workshop on Virtual Environments (IPT/EGVE)*, Zurich.

B. Herbelin, F. Riquier, F. Vexo & D. Thalmann. 2002. Virtual Reality in Cognitive Behavioral Therapy: a preliminary study on Social Anxiety Disorder. *8th International Conference on Virtual Systems and Multimedia (VSMM)*, Gyeongju, Korea.

B. Herbelin, F. Vexo & D. Thalmann. 2002. Sense of Presence in Virtual Reality Exposures Therapy. *First International Workshop on Virtual Rehabilitation*, Lausanne.

M.Ponder, B.Herbelin, T.Molet, S.Schertenleib, B.Ulicny, G.Papagiannakis, N.Magnenat-Thalmann & D.Thalmann. 2002. Interactive Scenario Immersion: Health Emergency Decision Training in JUST Project. *First International Workshop on Virtual Rehabilitation*, Lausanne.

## Research interests

Virtual environments and the immersive process: sense of Presence, perception and sensory illusions, links between psychology, neuroscience and Presence.

Virtual Reality exposure therapy and rehabilitation systems: evaluation of immersive setups, experimentation with interaction paradigms, human activity monitoring.

## Performances and Exhibitions

Co-author and manager of the project "Flying Cities" by Elisa Zurlo, European Culture 2000 programme for the year 2003. *This artistic project involved three partners (France, Italy, Germany) and the creation of interactive 3D, spatialized music and immersive scenography.* Art Salon of Seloncourt (France), november 2003.

http://flyingcities.ulipo-land.net

Software engineer for the artistic event "TECHNO-Wedding" by Fred Forest and Sophie Lavaud. *This live media show involved an interactive symbolic virtual environment in parallel of the civil act.* Issy-les-moulinaux, march 1999.

http://www.fredforest.org/technomariage/default.htm

Developer for the artistic installation ICARE by Ivan Chabanaud. *Immersive poetic virtual environment*. National Art Festivals (Terres Blanches festival, Montbéliard, 1997, IRCAM open doors, Paris, 1998), TV show (MCM, Paris, march 1998).

http://icare.cicv.fr/Icare