Fernando José Teixeira Franco Gouveia Pais

Interactive Avatars for Triggering Emotional Responses

Dissertação submetida para obtenção do grau de Mestre em Engenharia Electrotécnica e de Computadores, no ramo de especialização em Automação e Robótica, apresentada ao Departamento de Engenharia Electrotécnica e de Computadores da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Tuesday 5th September, 2017

· U C ·

Universidade de Coimbra

Acknowledgements

This dissertation would not have been possible without the support of many people. I would like to thank my advisor, Prof. Paulo Menezes, whose guidance and availability made this dissertation not only a reality but a pleasant experience. To my co-advisor, Bruno Patrão, for his help and advice whenever things looked bleak. To Dr. Antero Abrunhosa, for the time taken from his vacation to provide me with much needed help. To my colleagues at the Mobile Robotics Laboratory for the great work environment they create. To the Instituto de Sistemas e Robótica for the resources and the working conditions provided for my work in this dissertation.

To my colleagues, Bruno Silva and Nuno Gouveia, for their company in the many all-nighters done in the laboratory. To the Titans for all the nights of brainstorming and motivation provided. To all my friends for the good times spent during this journey and the vague memories that will be with me all my life.

Last but not least I would like to give special thanks to my family. To my mother and father for their support all these years, making me who I am today. To my grandparents for always being there, helping me whenever they could, and always being a source of inspiration.

Abstract

Virtual reality technology has seen explosive growth in the last few years, continuously improving user immersion on virtual environments. The affordability and diversity of such technology has led on to a wide array of applications being developed and released. This thesis focus on a usually neglected part of these environments, the avatars that populate them. It proposes the use of an emotional model to control the avatars' behaviour, endowing them with some semblance of intelligence. This allows for the creation of applications where the avatars play a central role instead of just being there for decorative purposes. During the course of this dissertation three different scenarios were developed, exploring some of the many uses such avatars might have. The first and second scenarios were also evaluated by a group of participants in terms of usability and effectiveness respectively.

Acronyms

AI - Artificial Intelligence

ALMA - A Layered Model of Affect

ANN - Artificial Neural Network

EEC - Emotion Eliciting Condition

GUI - Graphical User Interface

HMD - Head-Mounted Display

KSA - Knowledge Skills or Aptitudes

PAD - Pleasure-Arousal-Dominance

PANA - Positive Activation Negative Activation

PTSD - Post-Traumatic Stress Disorder

VEC - Virtual Emotion Center

VR - Virtual Reality

VRET - Virtual Reality Exposure Therapy

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Emotional intelligence - the ability to feel, identify, and act on emotions, is currently recognized as one of the most important aspects of human intelligence [4]. When people interact emotions are always present. Recognizing if certain actions are causing an emotional response, positive or negative, is essential to life in society. It also has a deep impact in professional success, since people with higher emotional intelligence show greater focus. TalentSmart tested emotional intelligence alongside 33 other important workplace skills. The results showed that emotional intelligence is the strongest predictor of performance, explaining 58% of success in all job types.

In current times, everyday life is also filled with interactions with machines. The ones who's purpose is to interact with humans need do so effectively; they need to appear as they are intelligent participants on the interaction. Machines might not need all the abilities people need, but in order to appear intelligent when interacting with humans, some affective computational model is required [5]. This is also true for virtual characters that try to replicate human behaviour and interact with human users.

The versatility of VR technology and its potential for research, training, and therapeutic applications makes it an excellent area of study. With the recent developments in affective computing models and the increase of hardware capabilities we are now able to increase the complexity of the virtual environments, focusing on a usually neglected component, the avatars that populate those environments.

1.1 Objectives

This work explores the use of avatars, with a emotional driven decision making system, to increase user immersion in 3D virtual environments. This allows for the creation of scenarios where the avatar behaviour tries to mimic human responses during certain emotional states.

Depending on the purpose of the scenario the avatar mood can be variable (changing in response to the avatar perception of a certain stimulus) or relatively fixed (in order to elicit some response on the user). Three settings where proposed to test the avatars:

1. Tool for emotional elicitation
2. Method for exposure therapy sessions
3. Means to impart some skill on the user

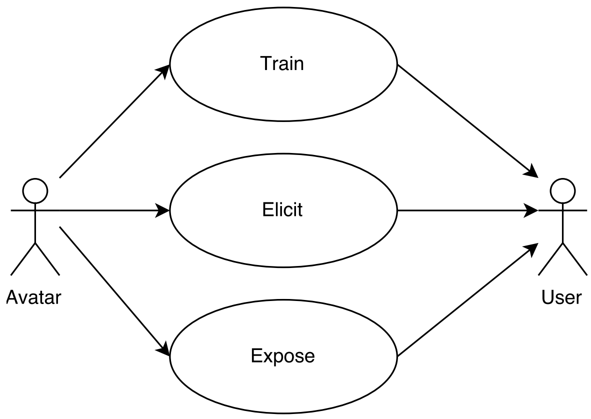


Fig. 1.1 Idealized system functionality

1.2 Related Work

For many years researchers have been trying to induce emotional states for scientific purposes. Methods involving hypnosis, images, music, interaction with trained confederates, the Velten self-statement technique, and even drugs have been tried in order to produce the desired response. Some of these methods bring with them ethical issues and/or problems of standardization [6]. Thanks to the developments done in the past few years on Virtual Reality technology, we are now able to explore new methods of emotional elicitation [7, 8]. Being able to control the elements in an elicitation experience is very desirable in order to get clear results and virtual environments allow for such control. It still has its flaws, since the imperfection of the virtual environments and the hardware complexity increase the likelihood of contamination [9], but the benefits trump its drawbacks.

On a more extreme side of emotion elicitation, virtual reality is also becoming an effective tool for exposure therapy. Virtual reality exposure therapy (VRET) immerses the user in a virtual environment where he is systematically exposed to a specific feared stimuli. Much research has been done in the last few years on the efficacy of VRET for the treatment of specific phobias (i.e. arachnophobia [10], agoraphobia [11], acrophobia [12], and aviophobia [13]. Early results point to an improved efficacy over imaginal exposure and as an attractive alternative to *in vivo* exposure [2].

Virtual reality also provides new ways to train important skills. Traditionally, training requires the trainee to not only learn the theoretical part but also do on-the-job training. This might be troublesome if the job requires either the expenditure of costly resources or the exposure to dangerous situations. Looking at this from a time management point of view, it is also harder to arrange time for on-site training than it is to train whenever time is available. Using virtual reality scenarios and environments for training addresses the previously mention problems. It provides a safe environment were the user can practice on-site skills, without the need of supervision or the expenditure of resources. Such scenarios are mainly being used for medical training [14, 15] but also in other areas such as welding [16], nuclear engineering [17], and tower cranes operating [18].

1.3 Developed Work

This work explored the use of avatars exhibiting some emotional behaviours, aiming at the induction of emotional responses on the users of the developed system. It is comprised by three different scenarios. The first scenario focus on territorial invasive situations by interacting with emotionally driven avatars. The avatars mood ranged from happy to angry, increasing in intensity as the distance toward the user is reduced.

A second scenario was developed during the course of this work, focusing on imparting the notion of personal space. This second scenario was populated by only one avatar whose mood depended only on the distance the user was from it. Using the concept of the four spaces proposed by Hall [19](i.e. Public, Social, Personal, and Intimate) the avatar behaviour changed, depending on which of the four spaces the user was. The action chosen by the avatar served as an indicator of how comfortable he was toward the user chosen distance.

The third and final scenario implemented on the course of this dissertation was one that combined both the training of an important skill and the exposure to one of the most common phobias, Glossophobia [20]. Glossophobia is the fear of public speaking, in extreme cases individuals completely freeze before their audience or have panic attacks characterized by sweating or trembling. This scenario consisted on a situation I will soon be confronted with, the presentation of a developed work to an audience. A virtual environment consisting of an auditorium populated with emotionally driven avatars was created, with their behaviour changing depending on the general quality of the user presentation.

1.4 Dissertation Structure

This Dissertation is divided in five chapters: Introduction, Theoretical Foundations, Developed Work, Experiments, and Conclusion. In the second chapter the literature and the reasoning that supports this work will be reviewed. It will start with a review on emotions and its different perspectives, some of the metrics created to quantify them, and emotion eliciting factors that were explored in this dissertation. After, a view on virtual reality as a therapy and training tool will be given, followed by a explanation on some of the available computational models of emotion and the one that was used. It ends with an overview of some of the most complex methods used in one of the developed scenarios: fuzzy systems and artificial neural networks. The third chapter introduces the developed work during the course of this dissertation. It begins by explaining the work done on the avatars, the developed modules, how they are interconnected, and ends with the constructed scenarios that were explored and tested. In the fourth chapter the experiments done with the working system will be presented. The results obtained from several participants will be exposed, followed by a discussion on them. Lastly the fifth chapter presents the conclusion and possible future work to be developed over this system.

Chapter 2

Theoretical Foundations

The concepts this dissertation addresses are introduced in this chapter and can be divided into five sections. It begins with an introduction to the concepts of emotion and personality, how they are represented, and some ways they are affected by social situations. The second part focuses on how virtual reality is being used in therapies and the new opportunities for training it provides. The third section centers around computational models of emotions, the different approaches taken, how it is relevant to both psychology and computer sciences, and the original emotional model that inspired the system used in the developed scenario. In the fourth and fifth sections an overview on fuzzy systems and artificial neural networks will be made, methods used on one of the developed scenarios.

2.1 Psychological Representation of Humans

As stated before, this section is dedicated to the concept of emotions and human behaviour. Since this thesis puts such focus on replicating human emotional states and human behaviour we need to first address the question of what emotions are. After that, the models used in this work to quantify abstract concepts like personalty and emotions are addressed. Lastly, some of the factors that influence human behaviour on a daily basis are discussed.

2.1.1 Theoretical Perspectives on Emotion

Defining what an emotion is seems like the obvious first step. Unfortunately its not such an easy thing to do since there is no singular or even preferred definition of emotion. In 1981 Kleinginna compiled 92 definitions, plus 9 skeptical statements and his own suggestions for a consensual definition [21]. So trying to explain emotion by defining it is not a viable approach. Instead of defining what an emotion is, lets look at the theoretical perspectives on emotions. Cornelius [22] identified four perspectives about how to define, study, and explain emotions: Darwinian, Jamesian, cognitive, and social constructivist.

The Darwinian perspective, introduced by Darwin in 1872 [23], sees emotion as evolved phenomena with a survival function, that endured because it helped us solve problems we faced as a species. According to Darwin, emotional expressions, and by extension, emotions, must be analyzed in terms of their functions and survival value. Plutchik [24] even stated that emotions exist in all organisms since he compares emotions in terms of their functional equivalent. This view allowed LeDoux to make inferences on human fear by analysis the brains of rats [25]. Using this perspective several so called fundamental/basic emotions, that fulfill some adaptive role, can be identified and other species can be studied to provide data on human emotions.

The Jamesian perspective, derived from the writings of William James [26], states that emotions derive from physiological changes. According to James, emotions exist because our bodies evolved in order to respond automatically to survival-related situations. So first there is a physiological response and what we consider emotions are just the experience of those changes. One important notion of followers of this perspective is that each emotion must be followed by a unique pattern of bodily response. As such, most of the research done has been on trying to prove this last proposition. After almost a century of research some conclusions were reached. First the importance of the body in emotion intensity has been somewhat validated, with research on people with spinal cord injuries showing a decrease of intensity of certain emotions proportional to the extent of the injuries [27]. Second Levenson and other researchers found evidence that a small number of emotions elicit specific autonomic responses [28].

The Cognitive perspective, popularized by Magda Arnold and said to be the dominant perspective, states that thought and emotion are inseparable and that all emotions are dependent on what she called appraisal [29]. Arnold describes appraisal as the process by which we judge events as either good or bad. For theorists that follow this perspective every emotion is associated with a different pattern of appraisals that depend on the person characteristics (i.e. temperament, personality, physiological state, etc). Critics of this perspective claim that Arnold's view of emotions is over-intellectualized and that emotions happen much faster than perception. This criticism however comes from a misunderstanding of Arnold definition of appraisals, since she defined them as "sense judgments" and being "direct, immediate, nonreflective, nonintellectual, [and] automatic".

Finally, the social constructivist perspective, the youngest of the four (in respect to the study of emotions in psychology), states that emotions are not evolved adaptations or primarily biological but rather cultural products from learned social rules [30]. Other perspectives don't dismiss the influence of culture on emotional display regulation, but social constructivists believe that to truly understand emotions we need to look at their social function. Doing so allows us to discover social-specific rules on how emotions should be experienced and expressed. Using anger as an example, Averill [31] states that it is elicited by the appraisal that one has been wronged in some way. But since in western society one cannot seek retribution by intentionally harm another person, anger allows one to violate the norm because it is something that takes control of a person and makes it not responsible for his/her actions.

All of these perspectives present some good points, sometimes even agreeing with each other, but they are still separate views on the working and purpose of emotions. This has to be taken into account when choosing how to model emotions from a computer science perspective and, as will be discussed in a following section, gave rise different approaches. The perspective chosen to guide this work and the approaches taken in it is the cognitive perspective. As such the central process that influences the avatars and their actions its their appraisal of the situation.

2.1.2 Big Five personality traits

Even if all previous perspectives on emotions clash on defining and study them they do agree on something, emotions, their impact, and the elicited reaction to them depend on the individual personality. Multiple definitions for the word personality exist, depending on the personality theorist theoretical position. Still, most agree that the field of personality is the study of how individuals differ from one another, disagreeing on how to conceptualize those differences. One approach to this problem is that of trait theory, that focus on measuring personality traits. According to this theory, personality traits (e.g conscientiousness) are relatively stable over time, differ across individuals, and give rise to characteristic patterns of behaviour. This theory and its instruments came under critique by W. Mischel [32] who believed that this view ignored a important factor, the situation the individual was in, starting a person versus situation debate on the field of psychology. He made four key critics:

1. Traits could not predict behavior with a correlation above 0.3.
2. The observed stability was a result of semantic similarity that biased the observer reports over time.
3. If the stability was there it's caused by the situation, not the person.
4. Behavior is not cross-situationally consistent.

Roberts [33] explains how the first three critics made by Mischel have already been refuted by later work on the subject. In fact, even Mischel [34] states that his intentions with his book was not to argue cross-situational consistency in personality functioning is low, so it should be ignored, but to warn that the inherent biased of researchers and ignoring the situation as a factor causes that low correlation. He attributes this misinterpretation to a "substitution problem", in that what was said about the book began to replace what was in fact in the book.

One popular approach to quantify personality is the five-factor model, also known as the Big Five personality traits. This model is not based on neuropsychological experiments but instead on language descriptors. It reduces the number of relevant personality factors to five domains and lists the trait facets that make up each them[35]:

1. Openness - which characterizes the degree of curiosity, creativity and a preference for novelty. The trait facets usually used to describe this domain are: fantasy, aesthetics, feelings, actions, ideas, and values.

2. Conscientiousness - which characterizes a preference for order, self-discipline, and planed rather than spontaneous behaviour.The trait facets usually used to describe this domain are: competence, order, dutifulness, achievement striving, self-discipline, deliberation.

3. Extraversion - which characterizes affection for social engagement, need for stimulating situations, and energy of the individual. The trait facets usually used to describe this domain are: warmth, gregariousness, assertiveness, activity, excitement seeking, positive emotions

4. Agreeableness - which characterizes a preference for social harmony and tendency to get along with other people. The trait facets usually used to describe this domain are: trust, straightforwardness, altruism, compliance, modesty, tender-mindedness.

5. Neuroticism - which characterizes a tendency to feel negative emotions and is sometimes refered to as emotional instability. The trait facets usually used to describe this domain are: anxiety, angry hostility, depression, self-consciousness, impulsiveness, vulnerability

Using these five factors we can describe a person personality or, in the case of this work, the personality of the avatars.

2.1.3 PAD Space

The Pleasure-Arousal-Dominance (PAD) emotional state model is a psychological model developed by A. Mehrabian and J. Russel [36] that describes emotional states using a three parameter representation. The need for such model comes from a essential requirement that any integrated science has: the availability of basic dimensions suitable for analyses of all its problems. MehrabiaN [37] stated that the study of emotions was crippled for decades by behaviorist bias, noting that this neglected area of human function could possibly yield the set of basic dimensions psychology required.

This model classifies emotional states in terms of pleasure-displeasure, arousal-nonarousal, and dominance-submissiveness values.

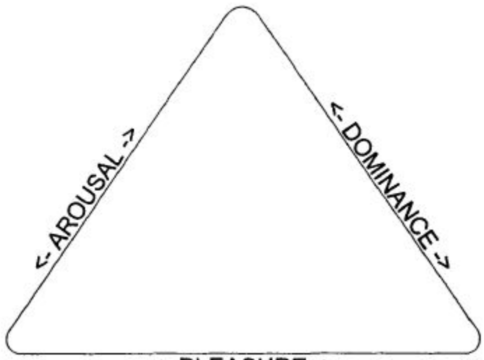


Fig. 2.1 Representation of the PAD Emotional State Mode

It was tested in relation to 42 verbal-report scales of emotional response and showed that all the reliable variances in those scales was explained in terms the PAD scales. In summary, these three dimensions provide a useful and reliable description of emotions.

Mehrabian also used PAD space in another model, the PAD Temperament Model [38]. Emotional states are transitory conditions that vary substantially over the course of a day. In contrast emotional traits are much more stable, remaining invariant over periods of years or even entire lifetimes. The same way as the Five Factor Model describes personality in terms of five variables, the PAD Temperament Model describes an individual emotional predisposition toward certain emotional states using PAD space. The PAD Temperament Model uses three measures: Trait Pleasure-displeasure (i.e. predominance of positive vs negative affective states), Trait Arousability (i.e. how easily a person is aroused and how slowly that the arousal returns to baseline levels), and Trait Dominance-submissiveness (i.e. a person feeling of control over his life vs feeling of being controlled). It should come as no surprise the existence of work relating the Big-five personality factors and the PAD Temperament Model [39].

So, not only does PAD space enables us to define emotions in terms of numerical values, it also allows us to determine how a personality described in the Five Factor Model affects emotional states. This model and its relation to the Big Five is applied in the emotional model.

2.1.4 Territoriality

The concept of territoriality exists in humans as well as animals. For animals its used to define behaviour associated with the defense from invaders of a certain geographical area, for which they have a sense of ownership. In the case of humans, territoriality concerns not only physical spaces, but also ideas, objects and even personal space. Territorial behaviours can be useful since they helps to coordinate activities, regulate population density, and provide a sense of well-being, but can also be a source of social conflict. Using the amount of ownership felt/warranted as a criterion, Altman [40] identified three different types of territories: primary, secondary, and public. Primary territories are the exclusive domain of the user and are the ones defended more intensely. Personal space belongs to this group, since the attributed ownership is extremely high. Secondary territories are not perceived as the exclusive domain of the user but instead are semi-public, this blurred boundary makes them the most propitious to conflicts. Public territories are available to anyone for temporary ownership, with conflicts arising if the intrusion is overstayed.

There are two methods used for territorial defense: prevention and reaction. Prevention consists on either staking the territory or marking it with objects so others recognize its occupied and move on. In the context of this work the focus will be given to reactions. But how do we react somebody invades our territory? What are the factors that influence the intensity of the reaction? According to Lyman and Scott [41], territorial encroachments are not all the same and can be divided into three types:

- Violation - which is the unwarranted use of another's territory. This does not necessarily needs to be a physical use of the territory but can also be done with the eyes (e.g. staring at somebody) or voice (e.g. talking loudly)
- Invasion - which is a attempt to take over another's territory. Invasions are a more permanent encroachment than violations and cause stronger reactions.
- Contamination - which is the pollution of another's territory, not by the presence of an invader, but by what was left behind by a previous occupant (e.g. when we go to a public bathroom and the previously user didn't flush).

The type of encroachment is not the only factor that affects the intensity of the reaction towards it. In fact, a large number of factors are taken into account such as:

- Identity of the intruder.
- Reason for the encroachment.
- Territory that was encroached.
- How the encroachment was done.
- How long it lasted.
- Chances of happening again.

All these factors, and probably some others that haven't been the subject of study yet, determine the intensity of the reaction. But what determines the reaction itself? According to McBride [42] and Finando [43] in face-to-face encounters with other people we are physiologically aroused, with an increase in galvanic skin response and heart rate. These responses are not limited to human interactions also extend to interactions with virtual characters. Slater [44] measured changes in electrodermal activity when the subjects were approached by groups of virtual characters. But since the arousal can be either positive or negative, we need to label it once it happens. According to Patterson's model [45], if the arousal state is positively labeled it encourages a reciprocal or enhancing reaction. On the other hand, if the state is negatively labeled it promotes a compensatory reaction.

In summary, we can see that encroachments on a person territories provokes a positive or negative reaction on that person. Virtual characters are also able to provoke these reactions, a fact we will explore on this work.

2.1.5 Proxemics

Maintaining personal space is one essential parts of our daily life. Every interaction we have, no mater how small, applies ”our” concept of personal space and interaction distances. Even when we are in crowded social spaces, like a bar or a bus, we try to preserve a ”comfortable” distance between ourselves and other people Usually people know that they should not come near each other closer than a certain distance, which tends to cause discomfort in the involved participants. But how do we define where other people must be when we interact with them?

Psychologists have researched proxemics, the study of human use of space and the factors that influence it, for many years. According to experts in the field, several factors influence the distances maintained during interactions. Some of the most important are age, sex, and culture of the people interacting. According to Hall [19] when the conversations are neutral or friendly, women tend to choose shorter distances when interacting than men do. On the other hand, if the conversation is threatening or alienating, men prefer shorter distanced than women. Age is also taken in consideration when interacting. If the distance taken relatively to another persons reflects comfort, it is a easy to conclude that interactions are closer between people of the same age group. However this notion does not apply to certain groups, such as the very old and the very young, who usually prefer to interact at very close distances. Culture and ethnic background is one of the most interesting and hardest to grasp concepts in interactions. Different cultural needs and norms result in very different interactional distances. Still, we can divide these distances into four interpersonal spaces:

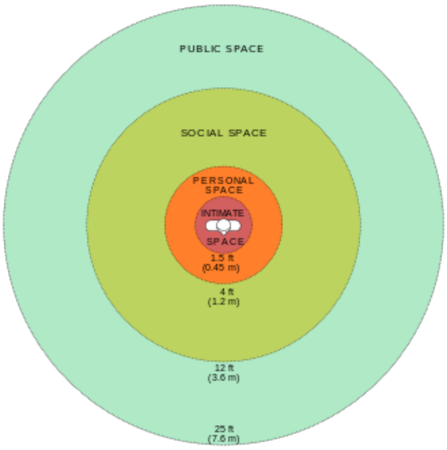


Fig. 2.2 Four spaces of interpersonal distance

- Public Space - In the beginning of public space it's easy for an alert subject to take evasive or defensive response if threatened. The end of public space is marked by a significant degradation of the information our sensory inputs can acquire on the other person
- Social Space - In this space impersonal business occurs, with an increase in distance usually reflecting an increase in the formal character of the interaction.
- Personal Space - The expression "keep one at arm's length" is the best way to describe the personal space. It can be thought of as a protective sphere that a person maintains between itself and others.
- Intimate Space -In the intimate space the presence of the other person is explicit, due to the amount of sensory inputs, and can sometimes be overwhelming.

The distances chosen for interaction gradually expand, beginning when a person is six years old and continuing to grow until adolescence when adult norms begin to reflect [46]. A way to practice and learn these distances was developed during this thesis.

2.2 Computational models of emotion

In the last few years we have witnessed a expansion in the research of affective computational models, driven by work done both from the side of psychology as well as computational sciences. This interdisciplinary partnership can be justified by the growing interest and need from both sides for such models.

Psychologists see it as a way to demonstrate their theoretical concepts on emotions, which usually stay on an abstract level. The need to explicitly define a model, so it can be computationally implemented, may reveal implicit assumptions or hidden complexities in it. By doing so, computational models become not only means to concretize theories but also frameworks to develop them. They also provide psychologists new ways to stimulate subjects during experiments. Virtual humans with such models have been used to show how behaviours and physical appearance affect social interactions such as conversations [47], willingness to cooperate [48], and level of social aggression [49].

The computational sciences interest in these models comes from two areas: artificial intelligence and human-machine interaction. Research on cognitive science, psychology, and neuroscience of emotion suggests that emotions have important social and cognitive functions that intelligent systems need. Simon [50] early work argued that emotions have a critical role in the ability of the organism to balance competing goals, and a wide range of studies point to emotions as the mean by which individuals establish values for decisions and their outcomes. This is not a new notion for AI [51] but, due to being treated as opposed to intelligence and rationality, has been practically ignored by AI research of late 20th century. But the increase in complexity of cognitive systems and the need to manage multiple goals as well oversee resources allocation has rekindle the interest in emotions. How emotions can address such control has been a theme of much research in the field of AI [52, 53].

Looking at human-human interactions, we can easily observe the roles that emotions and their expression have. Considerable information can be extracted about the mental state of an individual by observing his emotional displays. The correct interpretation of such information and how we react to it is crucial in human relations. Its not hard to conclude that machines need this ability to satisfactorily interact with humans. Citing R. Picard [5] "Machines may never need all of the emotional skills that people need: however, there is evidence that machines will require at least some of these skills to appear intelligent in interacting with people.". Much work has been done on the subject of how emotional displays in virtual characters affect user interactions. Thomas and Johnston [54] showed how such emotional displays make artificially generated characters more lifelike and human. Several studies substantiate this claim, e.g. Kramer [55] showed how such characters tend to make people act more politely, nervous and make socially correct choices. The effect on peoples trust [56] and empathy [57] on emotional agents have also been the subject of study. The impact emotional agents have in training scenarios has also been explored by using empathy and bonding between student and character [58] and by using praising and emotional displays has motivational tools in a tutoring application [59].

In summary, the expectations and interest on computational models of emotions depend on the interested party. Psychology researchers focus on the fidelity of the model with respect to human emotional processes, AI researchers emphasize the impact of emotions on reasoning processes and agent performance, and research on human-machine interaction interest is whether the model improves the interaction. Let's now look at some of the computational models available, their assumptions, behaviours and objectives.

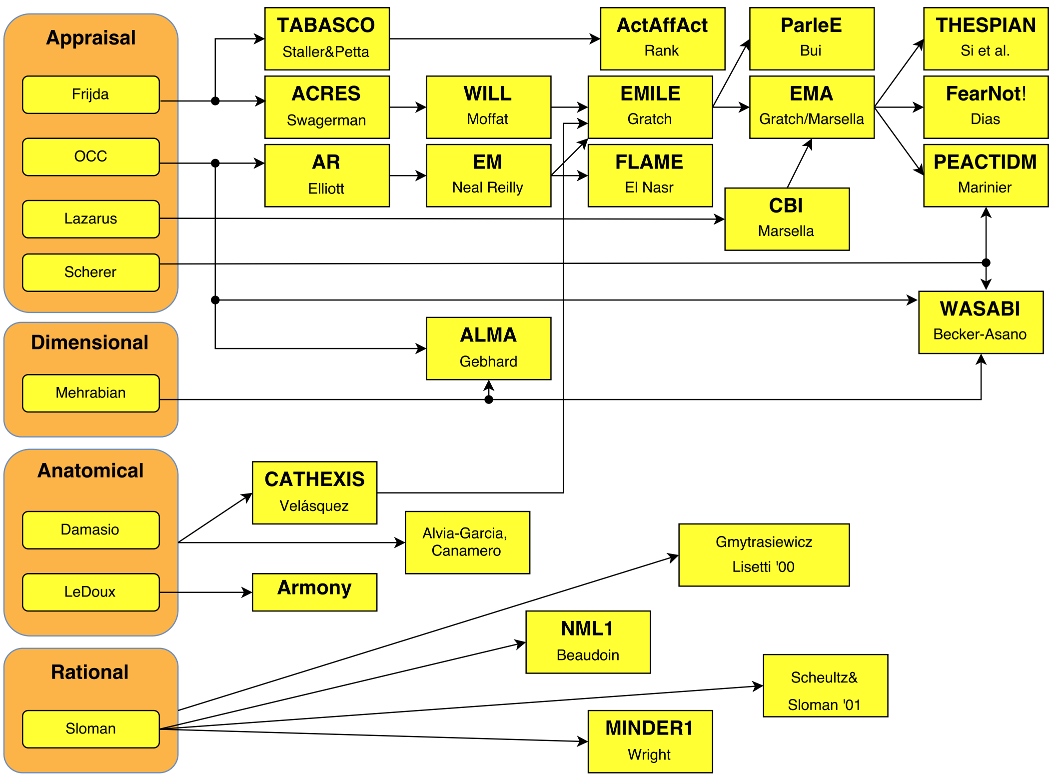


Fig. 2.3 Computational models of emotion [1]

The models listed in this figure are very independent entities, with inputs and outputs to suit their objectives and based on irreconcilable assumptions. This is manly due to the term "emotion" itself being open to different interpretations and contrasting definitions (as explained in 2.1.1). Since the work developed on this thesis is mostly based on the ALMA model the emphasis will be given to appraisal and dimensional theory.

Appraisal theory is a predominant theory among psychological perspectives on emotion. It emphasizes the connection between emotion and cognition, making it one of the most interesting theories for designing AI systems. In appraisal theory emotions are extracted from evaluations (appraisals) concerning the relationship between the event and the individual beliefs and desires, usually referred to as person-environment relationship [60]. Essentially, as the name implies, appraisal theory considers appraisal as the cause of emotions. The relationship between appraisal variables and specific emotions has been the subject of much discussion. Although most appraisal theorists agree that the same event may elicit multiple appraisals, they are unsure how they combine into an overall emotional state or how that state should be represented. Some believe the emotional state is best represent as a discrete component while others state that a continuous representation is necessary.

As we can expect, models derived from this theory consider appraisal as the main process, using complex systems for computing the necessary appraisal variables. Different approaches to this problem were taken, using decision-theoretic plans [61], reactive plans [62], detailed cognitive models [63], and Markov-decision processes [64]. The approach commonly taken for the emotion itself is much simpler, being usually represented as a discrete component. Some models use a generalized affective state or "mood" to represent the effect of the elicit emotions [3] while others treat emotions as labels to which behaviours can be associated to [65]. Recent models have begun to consider the impact emotions and mood have on the actual appraisal [66]. Such models have been mostly applied to computer applications, mainly in the creation of interactive characters in order to make them more realistic [67] and better understand humans [68].

The focus of dimensional theories is on the components of emotions rather than the appraisal itself. Dimensional theorists defend that instead of the term emotion, which is usually tightly linked to a appraisal, we should look at concepts such as mood, affect or core affect and their temporal and structural dynamic. Theorists also defend that emotions should not considered as discrete entities but rather as points in a continuous multidimensional space [69]. Dimensional models usually have two dimensions, like for instance the circumplex model [70] and the PANA model [71], or three dimensions, like the Lövheim cube of emotion [72] and the previously mentioned PAD space. Dimensional theories question the assumption that appraisal and emotion are has tightly linked as appraisal theorists presume.

Instead they idealize mood as not being about some object but a contribution of many other factors. Computational dimensional models are mostly used for generating behaviours for animated characters but have also been used has representational frameworks in systems that try to recognize emotional states in humans [73].

Other approaches are not as useful as these in the context of this work but are also interesting. The anatomic approach attempts to recreate the neural links and processes linked to emotions in living organisms [74]. Most researchers focus on one specific emotion instead of a dealing with a wide range of them. Rational approaches focus on the function emotions serve rather than the emotions themselves. Most of the research using this approach is done in the field of artificial intelligence, that conceptualizes emotions as being a set of processes and constraints that improve adaptive behaviour [75]. Communicative theories defend that emotions function as a communicative mechanism to convey mental states to other individuals in order to facilitate social interactions. Communicative theorists emphasize the result of the emotional display rather than the internal emotional model [76].

2.3 A Layered Model of Affect

The ALMA(A Layered Model of Affect) [3], has was previously stated, is an emotional model that combines appraisal and dimensional theories. It implements the OCC appraisal model [77], to deal with the emotion eliciting events, and combines it with the dimensional PAD emotional state model [36] and the big five model of personality [35]. To do that it simulates three different affective variables that interact with each other in order to replicate human behaviour. Those are:

- Emotions - reflect short-term affective states and are usually associated to appraisals of events. Emotions are time bound and decay over time until they eventually disappear.
- Mood - reflects a medium-term affective state and is not associated with specific appraisals. It's a more stable affective state and has an impact on the agent's decisions.
- Personality - reflects a long term, usually permanent, affective state and represents the agent mental characteristics.

Emotions are the lowest layer of affective state, elicited by a specific event, object or action. They derive from cognitive appraisals by the OCC model and their intensity is controlled by the personality. Emotions are mapped in the three dimensional PAD space and are given a pleasure, arousal, dominance values. The mapping of emotions into 3 dimensional affective space was done in [39], however since not all necessary 24 emotions types exist in that mapping the remaining values were extrapolated from similarities to comparable emotions. The following table shows the mapping the ALMA system relies on

Emotion	P	A	D	Mood Octant
Admiration	0.5	0.3	-0.2	+P+A-D Dependent
Anger	-0.51	0.59	0.25	-P+A+D Hostile
Disliking	-0.4	0.2	0.1	-P+A-D Hostile
Disappointment	-0.3	0.1	-0.4	-P+A-D Anxious
Distress	-0.4	-0.2	-0.5	-P-A-D Bored
Fear	-0.64	0.6	-0.43	-P+A-D Anxious
FearsConfirmed	-0.5	-0.3	-0.7	-P-A-D Bored
Gloating	0.3	-0.3	-0.1	+P-A-D Docile
Gratification	0.6	0.5	0.4	+P+A+D Exuberant
Gratitude	0.4	0.2	-0.3	+P+A-D Dependent
HappyFor	0.4	0.2	0.2	+P+A+D Exuberant
Hate	-0.6	0.6	0.3	-P+A-D Hostile
Joy	0.4	0.2	0.1	+P+A+D Exuberant
Liking	0.40	0.16	-0.24	+P+A-D Dependent
Love	0.3	0.1	0.2	+P+A+D Exuberant
Pity	-0.4	- 0.2	-0.5	-P-A-D Bored
Pride	0.4	0.3	0.3	+P+A+D Exuberant
Relief	0.2	-0.3	0.4	+P+A+D Relaxed
Remorse	-0.3	0.1	-0.6	-P+A+D Anxious
Reproach	-0.3	-0.1	0.4	-P-A+D Disdainful
Resentment	-0.2	-0.3	-0.2	-P-A-D Bored
Satisfaction	0.3	-0.2	0.4	+P-A+D Relaxed
Shame	-0.3	0.1	-0.6	-P-A-D Anxious

Table 2.1 OCC emotions into PAD space [3]

As said before, emotions decay over time until they eventually disappear, but their effect can still be felt in the mood variable. Mood can be seen as an "average" emotional state. Mehrabian also depicts mood in the three dimensional PAD space, classifying each of the eighth octants with a discrete description of mood.

+P+A+D Exuberant	-P-A-D Bored
+P+A-D Dependent	-P-A+D Disdainful
+P-A+D Relaxed	-P+A-D Anxious
+P-A-D Docile	-P+A+D Hostile

Table 2.2 Mood octants of the PAD space [3]

The strength of a mood is defined by its distance from the zero point. Mood can be further separated into two components, current mood and default mood. Current mood, has the name indicates, is the mood the agent currently has. Default mood is the mood the agent starts with and is given by the big five personality values and is computed using the following method.

$$P = 0.21 \times E + 0.59 \times A + 0.19 \times N \tag{2.1}$$

$$A = 0.15 \times O + 0.30 \times A - 0.57 \times N \tag{2.2}$$

$$D = 0.25 \times O + 0.17 \times C + 0.60 \times E - 0.32 \times A \tag{2.3}$$

The challenge of this model is in simulating human-like mood changes depending on the default mood, current mood and active emotions. For that purpose all active emotions are combined into a single point in PAD space called the virtual emotion center(VEC). Since emotions tend to result from distinct appraisals, depending on the situation the VEC position may vary widely. The VEC influences the current mood position in a function called *pull and push mood change function*. The pull phase of the function occurs when the current mood is between the PAD zero point and the VEC and pulls the current mood towards the VEC. The push phase occurs when the current mood is at or beyond the VEC and pushes the current mood further into the current octant. This push phase is the implementation of the concept that the longer a person is experiencing a mood the more intense it gets. If no emotional stimuli has happened for some time, the current mood slowly goes back to the default mood.

Before the mood can be computed, the *personality profile* and the global *computation parameters* for each agent need to defined. The computation parameters specify the PAD values for each emotion as well as how the personality affects each dimension of PAD space. The personality profile defines the personality of the agent as well as a set of subjective appraisal rules. The appraisal rules link specific events to emotion eliciting conditions (EECs), which are needed to compute the emotions.

In the VirtualHuman system [3], the system where ALMA was first implemented, this model was used to select the avatar visual appearance and speech output. Depending on the avatar emotions, mood, and role in the conversation, the character module selected the correct output. This emotion dependent behaviour was designed according to work developed in [78, 79].

2.4 Appraisal System Approaches

2.4.1 Fuzzy System

Since this work deals with subjective variables and situations it's necessary to use methods that can cope with it. One possible approach to this problem is to use a fuzzy system.

Fuzzy Logic

The term fuzzy logic, first introduced by L. Zadeh[80], refers to a logic that allows for sets to be represented by a continuous function between 0 and 1. Classical logic sets can only take value zero or one, not being able to deal with sets that don't have clear boundaries. This ambiguity between "true" or "false" is represented by a membership function, which defines how much a certain value belongs to a fuzzy set. Just as we use words to describe values, fuzzy logic also allows for the use of linguistic variables. Zadeh[81] defined linguistic variables as being characterized by (X, T, U, M), where:

- X - the name of the variable (i.e temperature, speed).
- T - the set of linguistic values that X takes (i.e {hot, cold},{fast, slow}).
- U - the universe of discourse.
- M - semantic rule that relates each linguistic value in T with a fuzzy set in U.

Sometimes the real world is too complex for precise description, approximation (fuzziness) must be introduced in order to create models. This type of logic allows us to formulate human knowledge in a systematic manner and use it to create fuzzy systems, which are defined thereafter.

Fuzzy Control System

Fuzzy control systems are rule-based systems that use fuzzy logic to solve engineering problems. The reason they are rule based is because at the heart of a fuzzy system are the so-called IF-THEN rules. Fuzzy IF-THEN rules are conditional statements in which some of the words are characterized by membership functions (i.g. IF speed is slow THEN apply more force to accelerator). But how do the rules combine to provide a proper answer? Three types of fuzzy systems are commonly used: pure fuzzy systems, Takagi-Sugeno-Kang fuzzy systems, and fuzzy systems with a fuzzifier and a defuzzifier. Since in pure a fuzzy system the inputs and outputs are fuzzy sets, and in a Takagi-Sugeno-Kang fuzzy system the versatility of fuzzy systems is not well represented we can ignore both of them. Let's focus instead on fuzzy systems with fuzzifier and defuzzifier, since they overcome the disadvantages of both those systems. The basic configuration of such fuzzy system is:

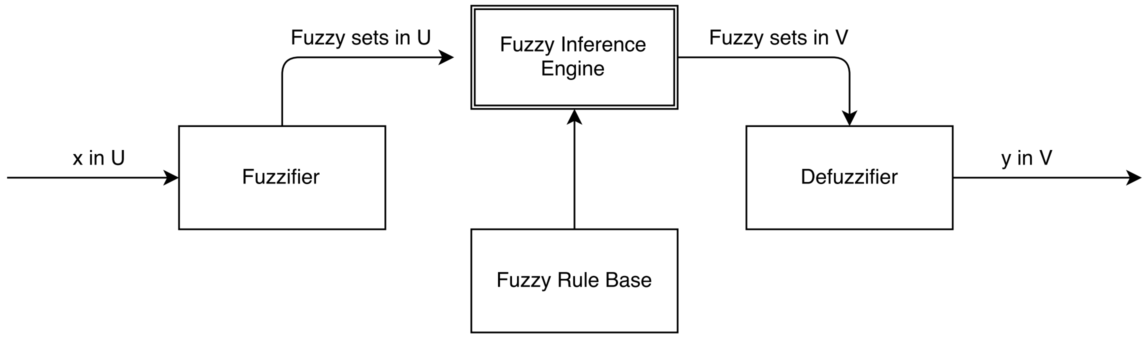


Fig. 2.4 Basic configuration of fuzzy systems with fuzzifier and defuzzifier.

As we can see, fuzzy systems with fuzzifier and defuzzifier are comprised by three key components: fuzzifier, fuzzy inference engine, and defuzifier. The fuzzifier maps a real-valued point $x^* \in U \subset R^n$ to a fuzzy set $A^{'}$ in U. Fuzzifiers are generally either sigleton fuzzifiers, which have membership value of 1 at x^* and zero in the remaining points, or represented by a continuous membership function(e.g. Gaussian, triangular, etc.). The fuzzy inference engine applies the defined IF-THEN rules to the output of the fuzzifier and combines their result into an output fuzzy set. There are two ways to infer with a set of rules: composition based inference and individual-rule based inference. Composition based inference combines all rules into a single fuzzy relation in $U \times V$, which is then viewed as a single IF-THEN rule. In individual-rule based inference for each rule its determined a output fuzzy set and the output of the fuzzy inference engine is the combination of the individual fuzzy sets. Two commonly used fuzzy inference engines are the product inference engine and the minimum inference engine. The product inference engine uses:

- individual rule based inference with union combination
- Mamdani's product implication
- algebraic product for the t-norm operators and max for the s-norm operators

and is defined as:

$$\mu_{B'}(y) = \max_{l=1}^{M}[\sup_{X \in U}(\mu_{A'}(x)\prod_{i=1}(\mu_{A_i^l}(x_i)\mu_{B^l}(y)] \tag{2.4}$$

The Minimum inference engine uses:

- individual rule based inference with union combination
- Mamdani's minimum implication
- min for the t-norm operators and max for the s-norm operators

and is defined as

$$\mu_{B'}(y) = \max_{l=1}^{M}[\sup \min(\mu_{A'}(x), \mu_{A_1^l}(x_1), ..., \mu_{A_n^l}(x_n), \mu_{B^l}(y))] \tag{2.5}$$

The defuzzifier maps from fuzzy set $B^{'}$ in $V \subset R$, which is the output of the inference engine, to a crisp point in $y^* \in V$. His duty is to specify which point in V better represent the fuzzy set $B^{'}$. When selecting a defuzzification scheme its important to consider the plausibility (ify^* represents $B^{'}$ from an intuitive point of view), continuity (a small change in $B^{'}$ should not result in a large change in y^*), and the computational simplicity the scheme provides. Three possible choices for defuzzifiers are: the center of gravity (y^* is the center of the area covered by the membership function), the center average (y^* is obtained from the weighted average of the centers of the M fuzzy sets), and the maximum defuzifier (y^* is the point in V where $\mu_B(y)$ is maximum).

By using these systems we can introduce an analog input, apply human experience and reasoning to it, and obtain a reasonable output in response. Since we are trying to model human behaviour mainly from observations and theories this seems like a interesting approach to explore.

2.4.2 Artificial Neural networks

Work on artificial neural networks (ANN) has been motivated by the ability that the human brain has to compute information in an entirely different way than computers do. The brain is a very complex, nonlinear, and parallel information-processing system. It is able to to organize its structural components (i.e. neurons) in order to perform certain computations (e.g. perception, recognition, and motor control) much faster than digital computers are.

ANN is a massively parallel distributed processor, made up of basic processing units (artificial neurons), whose structure can learn from data and use it. Knowledge is acquired by the ANN in the learning process and stored in the neurons by means of synaptic weights. This learning process can be performed in ways analogous to human learning: learning with a teacher, reinforcement learning, and unsupervised learning. In learning with a teacher (also known as supervised learning) the neural network has a set input–output examples to be trained with. The network learns by adjusting its synaptic weights, depending on the difference between the response computed and the one provided by the "teacher". In reinforcement learning the input-output mapping is performed through interactions with the environment in order to minimize a cost function. In unsupervised learning (also known as self-organized learning) no one oversees the learning process and instead the network is optimized in respect to a task-independent measure of the quality of representation.

The architecture of the neural network is firmly linked with the learning algorithm used to train it. In general, three classes of network architectures can be identified:

• Single-Layer Feedforward Networks - it has only the input layer (that receives the inputs) and the output layer (that returns the outputs), with the input layer projecting directly to the output layer (feeding forward).

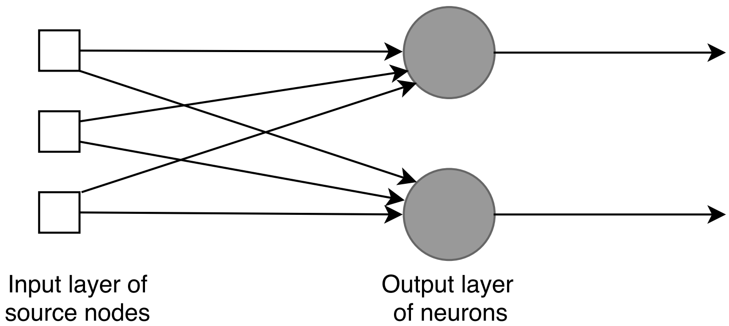


Fig. 2.5 Single-Layer Feedforward Network

• Multilayer Feedforward Networks - it has one or more hidden layers between the input and the output layers. Its called a hidden layer because its not directly seen from the output or the input perspective. Adding more layers allows for more complex models to be learned (with care needed since overfitting the network makes it lose its ability to generalize).

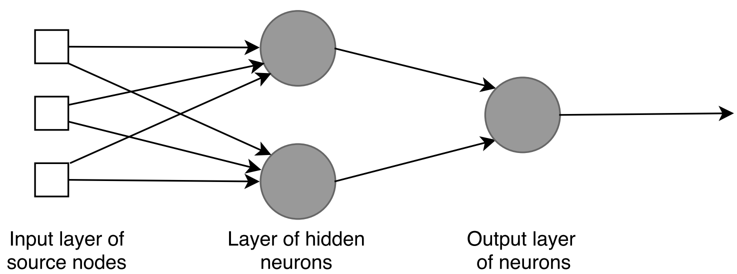


Fig. 2.6 Multilayer Feedforward Network

• Recurrent Networks - recurrent networks key feature in relation to the feedforward ones is that it has at least one feedback loop. This feedback loop involve the use of unit-time delay elements, which results into a nonlinear dynamic behaviour.

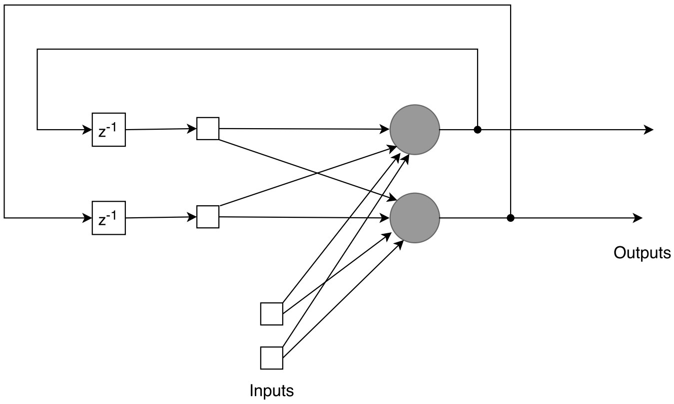


Fig. 2.7 Recurrent Network

There are many situations where its advantageous to apply artificial neural networks instead of conventional methods. Some of them are when:

- Human expertise is absent
- Humans are unable to explain their expertise
- Solution changes with time
- Solution needs to be adapted to particular cases
- The problem size is to vast for our limited reasoning capabilities

Neural networks and other machine learning methods are being used in a variety of areas (e.g. financial forecasting [82], computer vision [83], and medicine [84]), showing very promising results. Looking at the appraisal process described previously we can see how it fits almost perfectly with these situations. As such a neural network was trained and used in one of the developed scenarios, providing data to an appraisal.

2.5 Virtual reality

Although virtual characters can be integrated in different interactive platforms, we are mostly focused on their inclusion on virtual scenarios. As such, I could not end this chapter without a reference to virtual reality. In the last few years we have seen an unprecedented boom in Virtual Reality technology, with new applications being presented on a regular basis. But why did this sudden rise in interest happen and when did it started?

It has been fifty two years since I. Sutherland presented his vision of the Ultimate Display [85]. He envisioned a system that not only stimulated vision but provided immersion on all senses. But only in 1989 J. Lanier combined all the different concepts and actually coined the term Virtual Reality. This gave rise to a wave of interest by the academic community in general, with attempts to develop the technology and algorithms to fulfill Sutherland's vision. Unfortunately the initial euphoria died out after a few years, but the potential of such technology was recognized. It was only in 2012, with a Kickstarter project named Oculus Rift, that once again the interest in virtual reality technology was rekindle. Oculus Rift offered an affordable high-quality Head-Mounted Display (HMD) to the public and was the spark to the so-called second wave of VR that we are currently on. With the hardware being much more affordable than it was before, the financial difficulty for prototyping was greatly reduced. This resulted in the vast amount of applications we are currently seeing flooding the market. Affordability is not the only factor for this boom. The development of new input and output devices is also happening at an unprecedented speed. In this section two of the most promising applications for virtual reality will be presented.

2.5.1 Virtual reality exposure therapy

Though fear is a vital emotional response for survival [86] an excessive response can be debilitating or paralyzing, impairing normal life functioning. In order to deal with this problem several therapeutic strategies were developed, one of them being exposure therapy. Exposure therapy consists, as the name indicates, on exposing the patient to the feared object or context in order to reduce the negative affective response. Research shows that exposure therapy is an effective method for dealing with the symptoms associated with fear responses [87]. Exposure therapy can be done using three techniques: *in vivo* exposure, imaginal exposure, and interoceptive exposure. *In vivo* exposure consists on exposing the patient to actual fear-inducing situations, imaginal exposure relies on the patient imagining the situation they are afraid of, and in interoceptive exposure the patient is confronted with the bodily responses usually induced when in a fear state (i.e. Shortness of breath, elevated heart rate, etc) [88]. *In vivo* exposure therapy has been shown to have greater efficacy, when compared to imaginal exposure, especially in the treatment of specific phobias (Emmelkamp, 2003). Interoceptive exposure therapy is usually only used to treat specific disorders such as panic attacks or post-traumatic stress disorders.

Virtual reality exposure therapy is becoming increasingly more common on the treatment of anxiety and phobias. It consists on immersing the patient in a virtual environment where he is exposed to the feared stimuli in a relevant setting. A number of qualitative reviews has shown the potential for VRET in the treatment of specific phobias such as: arachnophobia [10], agoraphobia [11], acrophobia [12], and aviophobia [13]. This view derives from early evidence that VRET produces better results than imaginal exposure and results in equivalent outcomes to in vivo exposure. In terms of general effectiveness Thomas Parsons compiled the results of 52 studies, of which 21 met the inclusion criteria, and presented their effects on the patients [2]. The anxiety disorders were categorized into six affective domains (PTSD, Social phobia, Arachnophobia, Acrophobia, Panic disorder with agoraphobia, Aviophobia) and evaluated before and after the patient underwent VRET. The following table presents the average random effect sizes, including the variance and confidence limits for the mean effect sizes, for the affective domains and the anxiety total

Domain	Average random effect size	Effect size variance	95% CI		*R*	%
			Lower	Upper		
PTSD	0.87	0.01	0.64	1.10	0.40	0.16
Social phobia	0.96	0.10	0.34	1.59	0.43	0.19
Arachnophobia	0.92	0.12	0.25	1.59	0.42	0.18
Acrophobia	0.93	0.06	0.44	1.43	0.42	0.18
Panic disorder with agoraphobia	1.79	0.02	1.52	2.06	0.67	0.44
Aviophobia	1.59	0.05	1.16	2.01	0.62	0.39
Anxiety Total	0.95	0.02	0.69	1.21	0.43	0.18

Fig. 2.8 Effect overview of VRET [2]

Taking into account the available data we can conclude that VRET is a relatively effective tool for therapeutic treatments. Still, caution must be used when interpreting its clinical significance. Some of the studies on this subject don't have control groups or randomized trials, reducing their value as unbiased studies. Still, even if its efficacy is not as great as *in vivo* exposure, its an attractive alternative and cannot be dismissed.

2.5.2 Virtual reality as a training tool

Has stated before, virtual reality provides a new way to train skills. Traditional training requires not only a theoretical component but also a practical component. In certain cases allowing trainees to practice in the actual setting can be dangerous for them(e.g. dealing with dangerous materials), for others (e.g. doctors operating on real patients), or be prohibitively expensive (e.g. military airborne training). Virtual environments can reconstruct real world conditions and provide a safe environment for training with almost no costs (after the initial investment on the system).

Fig. 2.9 PeriopSim virtual environment

The most important question is if training in virtual environments leads to positive training transfer (i.e. improved performance), negative training transfer(i.e. degraded performance), or no effect at all. Training transfer is defined as something learned(knowledge, skills, or attitudes[KSAs]) under a training condition that is retained and applied in the operational condition [89]. Three aspects must be taken into account when trying to achieve positive training transfer with virtual environments. The first aspect is that the simulated task environment needs to have many elements in common with real-world tasks (identical element principle). The second aspect is that it needs to have high functional fidelity, the simulated equipment used by the trainees needs to emulate the real world equipment. The third aspect is the psychological fidelity, the trainee needs to perceive and act on the virtual environment the same way he would on the operational environment.

Several fields are already using virtual environments as training tools with relative success. In medical residents training several studies have researched the improvement in operating room performance in VR-trained groups versus non-VR-trained groups [90, 14]. Even though most studies focus on a specific scenario, laparoscopic cholecystectomy or other laparoscopic procedures, they verified that a positive training transfer occurs in training using VR surgical simulations. Other tasks, such as welding [16] and assembly [91], have also confirmed a positive training transfer in VR scenarios, with improved performance in VR trained students. Results in military training are harder to find, mainly due to the military unwillingness to disclosure information and participate in double-blinded studies . Nevertheless the usage of simulators as training tools, specifically in the air force, is widespread and their training efficacy [92] and cost effectiveness have been demonstrated.

The potential for virtual reality training tool is obvious, with clear validation on its positive training transfer. The fields that benefit from it are still limited but every year new scenarios are being created for training of specific skills. This work hopes to contribute to this area, with new scenarios focusing on neglected skills.

Chapter 3

Developed Work

The work developed for this thesis can be divided in several parts:

1. modifications to the used graphical engine (OpenAR, described below) for enabling the possibility to control the apparent gaze direction of animated characters.

2. Development of three different scenarios and respective applications, being each one suited for a specific purpose, which are: learning about proxemics, territoriality-based emotions, and which attitudes to take when doing an oral presentation.

3. Development of an emotional model for controlling the transitions between emotional behaviours of the avatars that are used in the different applications.

4. Implementation of a classifier for analyzing the user in terms of posture, focus of attention, and gestural pose and finally produce an output score that will be used by the emotional model-based controller. This combination of the classifier in tandem with the controller is used to make a crowd of avatars react positively or negatively during the presentation training application.

In the following, first an overview of the avatars implementation and their animations will be done. In the second section the developed scenarios that were build in order to demonstrate the system and the approaches taken will be described. The final section exposes the emotional model that was used in all scenarios, with focus on two of the most complex components.

3.1 Avatar Modeling

Before describing the emotional model and the scenarios developed on the course of this dissertation, a quick overview on the avatars must be given. All basic models for the avatars and the animations used by them were downloaded from the Mixamo website.

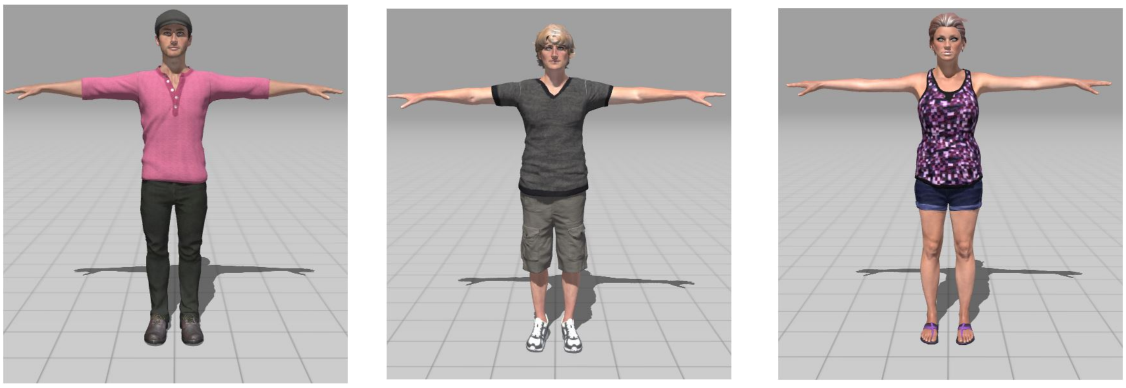


Fig. 3.1 Examples of Avatars

Mixamo is a 3D computer graphics technology company, funded in 2008 and acquired by Adobe in 2015, that started out as a cloud-based service offering animations and automatic character rigging. In their website dozens of 3D characters and hundreds of animations can be downloaded for free and used to develop all sort of applications.

3.1.1 Characters

Each character has three main components: (1) a skeleton which is a collection of bones connected by controllable joints, (2) a polygonal mesh whose shape is obtained by influence of the bones, and (3) a texture that creates the particular appearance. From all the available characters in the Mixamo website only five were selected and used in the developed scenarios. The reason for this was that, since the main use of these characters are in video games, and therefore most possess either a too unrealistic appearance or were designed to be used in fighting/shooting poses and not prone for other uses. In order to deal with this, specially in scenarios where many avatars are used, versions of those five original characters were created by modifying the models' textures, giving them new skin tones or clothes.

All the developments were done by making use and extending, as already stated above, OpenAR [93], which is a framework developed at ISR-Coimbra to enable the rapid creation of 3D-based graphical applications with support for different types of visualization and interaction devices.

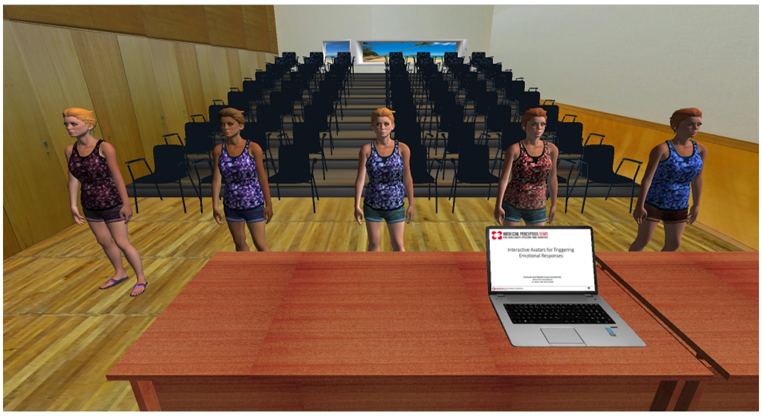


Fig. 3.2 Version of same base avatar

Using it, and after loading all the static components of a given scenario, an avatar can be loaded by setting its shader, scale, loading the character components, and setting its pose in the world. Then there are two possibilities, making use of the animation engine to have OpenAR generate the timed poses of the character that make it "alive", or control each if its joints directly by using some programmed or external information source. In this work both cases were used, the former for the autonomous avatars that appear in front of the user and will be explained hereafter, and the latter for mapping the user movements into its own avatar to increase the immersion sensation through embodiment manipulation.

3.1.2 Animations

The characters themselves are only stationary objects, they need sequences of poses in order to be animated. As the characters are based on skeletons, an animation is just a sequence of joint configurations that evolve along time. It should be noted that the skeleton of a character is, as expected, not visible, but each of its configurations define a specific pose for each of its bones, which in turn modify the position of the mesh vertices positions under their influence.

When an animation is run, the engine continuously interpolates joint values between predefined key-poses using time as the control (interpolating) parameter. In a realistic simulation, the characters must not repeat indefinitely the same movement, but instead should frequently make the transition between different movements (creating behaviours) in a smooth way. This is easily achieved by blending two animations via a blending factor that progressively increases the influence of the new animation and decreases that of the previous one. As the animations themselves contain specific information about the characters they run in, this means that its necessary to have animations for both female and male characters. All used 3D characters provided by Mixamo all have the same basic skeleton, with 52 interconnected joints that give shape to the character.

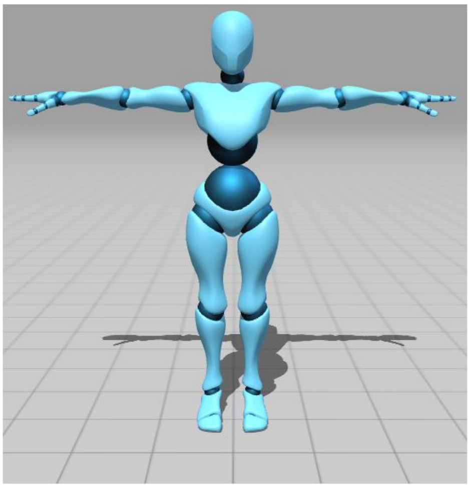


Fig. 3.3 Avatar Joint Structure

Even though Mixamo has hundreds of animations available, not all of them can be used in simulated real life scenarios. For that reason, and as the number of needed animations for each character depends on each particular scenario, new animations had to be created, combining aspects of several animations (e.g. an animation where the character was agreeing with one where it was seated). Still, all animations were assigned an emotional labels and interaction contexts, that will be used latter by an automated mechanism for chaining the appropriate animations, and thus generate behaviours that are adequate for each particular simulated situation. This mechanism and the use of the control variables will be explained in a section bellow.

3.1.3 Attention

The focus of attention (or gaze direction) is one type of information that is easily perceived by humans, and that plays an important role in human to human interaction and communication. Knowing where other person is looking at has become so natural that people that don't perform gaze saccades towards the focus of interest are perceived as having abnormal behaviour (or some handicap).

Thus, to explore this and make avatar's behaviours look even more human like, it should be possible to control this focus of attention along time. This can be also interesting for the training of people that have some kind of handicap, e.g. joint attention limitations in ASD.

Since the avatars are running animations with fixed joint positions, one of the first problems that needed to be solved was how to make them turn their head towards something.

In interactions with users or other avatars the realism of the interactions would be jeopardized if the avatar wasn't looking to the correct place. The difficulty in this is that the position and orientation of the head depends on the animation, so just turning the head with a fixed rotation is not effective. To tackle this, if the avatar is commanded to look at a certain point, the contribution(in rotation) of the animation of all joints preceding the neck is added into a rotation matrix. By applying the inverse to this rotation matrix and multiplying the neck rotation matrix by it, independently of the animation, the head is oriented toward the positive z-axis. Considering a unit vector $\vec{V}_1 = [0,0,1]$ as being the vector we have and the vector we want($\vec{V}_2$) being one starting at the head origin and directed towards the designated point. The final transformation needed to apply to the head is the rotation that transforms $\vec{V}_1$ into $\vec{V}_1$. This is calculated using

$$v = \vec{V}_1 \times \vec{V}_2$$

and

$$\theta = cos^{-1}\left(\frac{\vec{V}_1.\vec{V}_2}{\|\vec{V}_1\|.\|\vec{V}_2\|}\right)$$

where v is the vector product and θ is the angle, the rotation matrix is given by

$$R = I + [v]_\times + [v]_\times^2\left(\frac{1-cos(\theta)}{sin(\theta)}\right)$$

where I is the identity matrix 3×3 and

$$[v]_\times = \begin{bmatrix} 0 & -v[3] & -v[2] \\ v[3] & 0 & -v[1] \\ -v[2] & v[1] & 0 \end{bmatrix}$$

Doing so forces the avatar neck and head to be oriented towards the front, and on top of the character Z-axis. By obtaining a rotation matrix that converts a vector in the z-axis into a vector directed toward the point we defined we can make the avatar head turn in the desired direction.

3.2 Developed Scenarios

Three virtual reality scenarios were developed during this dissertation, each targeting a specific application for the developed system. The first enables the user to understand proxemics concepts and can be customized for understanding the differences according to age, sex and culture, by selecting the appropriate parameters. The second application is focused on territorial invasions and how they can trigger emotional responses. Finally a third applications demonstrates the use emotion-driven characters to create a training scenario, where the user can rehearse slide presentations in an auditorium with spectators reacting to the way it is being done.

3.2.1 Scenario 1: Proxemics

Personal space is one of the most essential components of our daily life. Every interaction we have with another person applies ”our” concept of personal space and interpersonal distances. The distances we chose to use when interacting gradually expand, since when we are around six years old until early adolescence, when adult norms begin to be reflected [46]. Another way to look at this is that if a child doesn’t correctly learn the social accepted distances from a early age, it can bring him/her problems in the future.

This first scenario was designed as a training tool to learn interpersonal distances. The virtual environment is comprised by an old town populated by the user and one avatar.

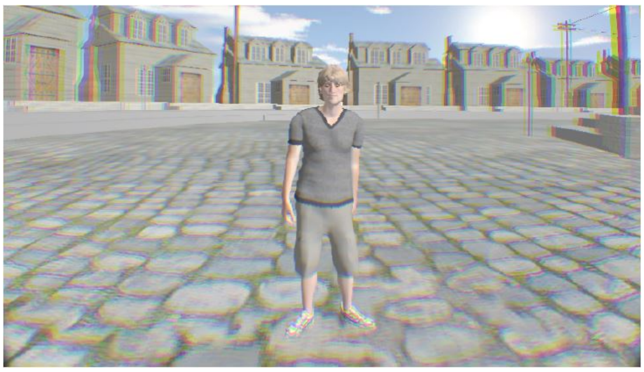


Fig. 3.4 Virtual environment of scenario 1

Different avatars can be selected depending on what training is intended since the distance used depends on who we are interacting with. Before the scenario is initialized the avatar four spaces [19] are specified, according to the objective of the simulation. As the user varies his distance towards the avatar his mood, and by association his behaviour changes. This distance is measured using data from a Kinect, which also allows a mapping of the user movements into his own avatar. In this scenario the user must interpret the avatar behaviour in order to ascertain in which of the four zones he/she is. The option to use visual help also exists, with each zone being marked on the floor.

Fig. 3.5 Virtual environment of scenario 1 with visual helping clues active.

If the user is outside the public zone the avatar doesn't acknowledge his presence. If he enters the public zone, the avatar acknowledges his presence by looking at him. In the social, personal and intimate zones the avatar is interacts with the user, showing discomfort if the user starts getting too close. If If the user try's to be aggressive and hit the avatar he runs away from the user and the scenario ends. After two minutes the scenario ends and a score is given, representing the user performance in the interaction.

This training scenario was presented as a demonstrator at Experiment@International Conference 2017 and published in IEEE Xplore [94].

3.2.2 Scenario 2: Territorial Invasions

As previously explained, virtual reality provides new ways for emotion elicitation. Emotional theorists are always looking for new tools to obtain experimental data from, and virtual reality provides well controlled environments to do just that.

This second scenario was designed to serve as such a tool. By creating territorial invasive situations with avatars, it elicits a response by the part of the user. Territorial encroachments are not all the same and the intensity of the reaction toward them depends on several factors. This scenario tests three variations, exploring how changes in the mood and number of intruders affects the appraisal. A state machine is used to reduce animation search space and control the interaction context, simulating a regular interaction with the user. In this scenario the user movements are also mapped using a Kinect, only this time it doesn't reflect the user distance. This forces the user to face the emotional avatar and let him invade his territory. The first variation has one avatar, with a mood in the hostile octant, approaching the user.

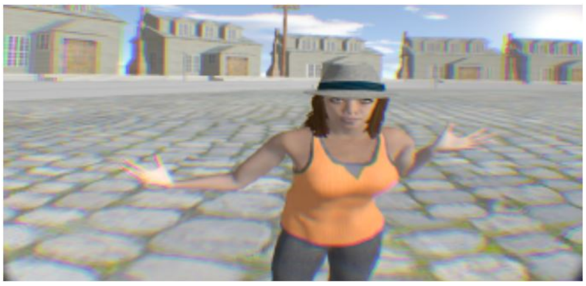


Fig. 3.6 Second scenario: one hostile avatar

The second variation also has only one avatar approaching the user but this time with the mood in the relaxed octant.

Fig. 3.7 Second scenario: one relaxed avatar

The final variation has several hostile avatars approaching the user, surrounding him from all sides.

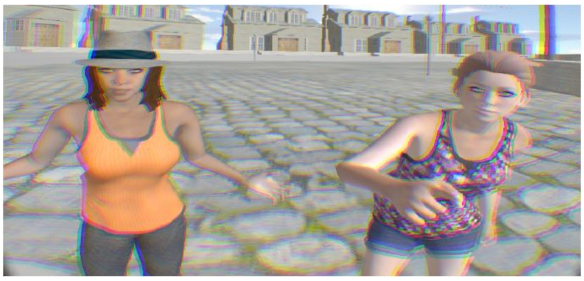


Fig. 3.8 Second scenario: several hostile avatars

In all of these variants the mood of the avatars gets pushed further into the starting octants as the distance from the user decreases. This scenario was presented as a demonstrator at Experiment@International Conference 2017 and published in IEEE Xplore [95].

3.2.3 Scenario 3: Presentation Trainer

This last scenario combines the functionality of the previous two, it serves as both a training tool and a emotion elicitation tool (in the form of exposure therapy). This scenario was designed to train something that many people have difficulty with, a oral presentation towards a audience. To do so a 3D model of an auditorium, replicating the ISR auditorium, was created and populated with emotive avatars.

Fig. 3.9 Third scenario: auditorium with interested crowd

Fig. 3.10 Third scenario: auditorium with bored crowd

The state machine for this scenario consists of only two states: ongoing (when the user is presenting), and concluded (when the user finishes the presentation). The mood of the avatars depends on their personality and their evaluation towards the user presentation. The evaluation of the presentation is done by a fuzzy evaluator, translating into an appraisal by

Fig. 3.11 Third scenario: projected slide as viewed by the presenter.

the avatars. The personality influences not only the the appraisal of the presentation but also default mood and the speed of the current mood. The auditorium can have any number of avatars and the difficulty of the evaluation is controllable.

3.3 Architecture

As explained before, this system uses emotional states to control the avatar behaviours and select animations. In this section a general view of how the system works, how it varies mood, and how it selects the "correct" animation is given.

3.3.1 Emotional Model

The emotional model used in this system was greatly inspired by the ALMA [3], using the Big Five to map personality and PAD space to map emotions and mood. The choice of the correct action depends on the current mood of the avatar, which is influenced by the active emotions (in the form of the VEC) and the personality (in the form of the default mood). Despite the ALMA considering that the default mood as only depending on the personality that is not the approach taken on this implementation. As W. Mischel [32] stated, the situation must also be taken into account when interpreting these factors. As such, the effect of the personality on the variables of this system depends on the scenario they are applied in. State machines are also implemented to better control and synchronize the interaction as well as to reduce search space. The action chosen during the interaction is not purely based on emotions but also on the current context. As such every animation is associated with a PAD emotional value and a state, associated with a specific context.

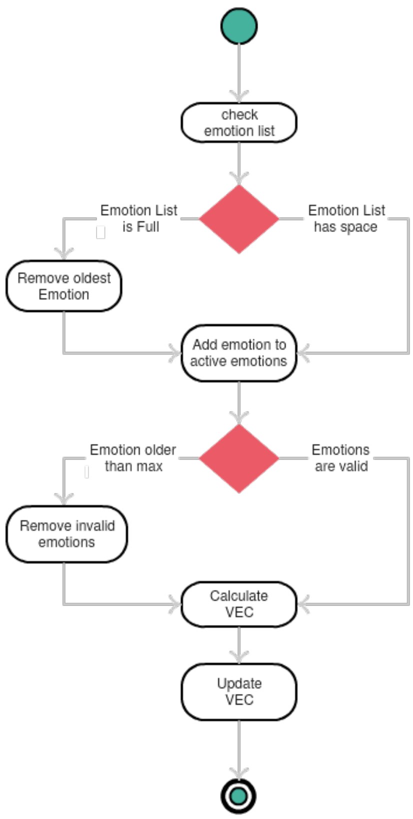


Fig. 3.12 Emotion management activity diagram

In order to better understand how it all works lets separate it into key stages: emotional appraisal, VEC calculation, current mood update, the context transition, and the animation selection. The appraisal of the situation, and subsequent elicited emotion, depend on the specific scenario. When a emotion is elicited in the avatar it is added to the active emotions list and the VEC is updated. The active emotions list is composed by all the emotions that have yet to expire. The behaviour in this stage can be seen in the activity diagram 3.12.

Every cycle the current mood is updated, influenced by the default mood and the VEC in the same pull and push way as the ALMA. The speed the current emotions moves depends on the personality, since some people change mood faster than others. The speed the current mood tends further into the octant is a fraction of that speed, because an elicitation is more effective than this effect. The context transition depends on the scenario since the state machine is different for each one of them. In the animation selection stage the animation that has the closest PAD value to avatar current mood and the same context the avatar is selected.

3.3.2 Scenario 2: State Machine

In the second scenario a more complex state machine than in the other scenarios was used. Special attention will be given to it, since it can be used not only in avatar-user interactions but also avatar-avatar interactions.

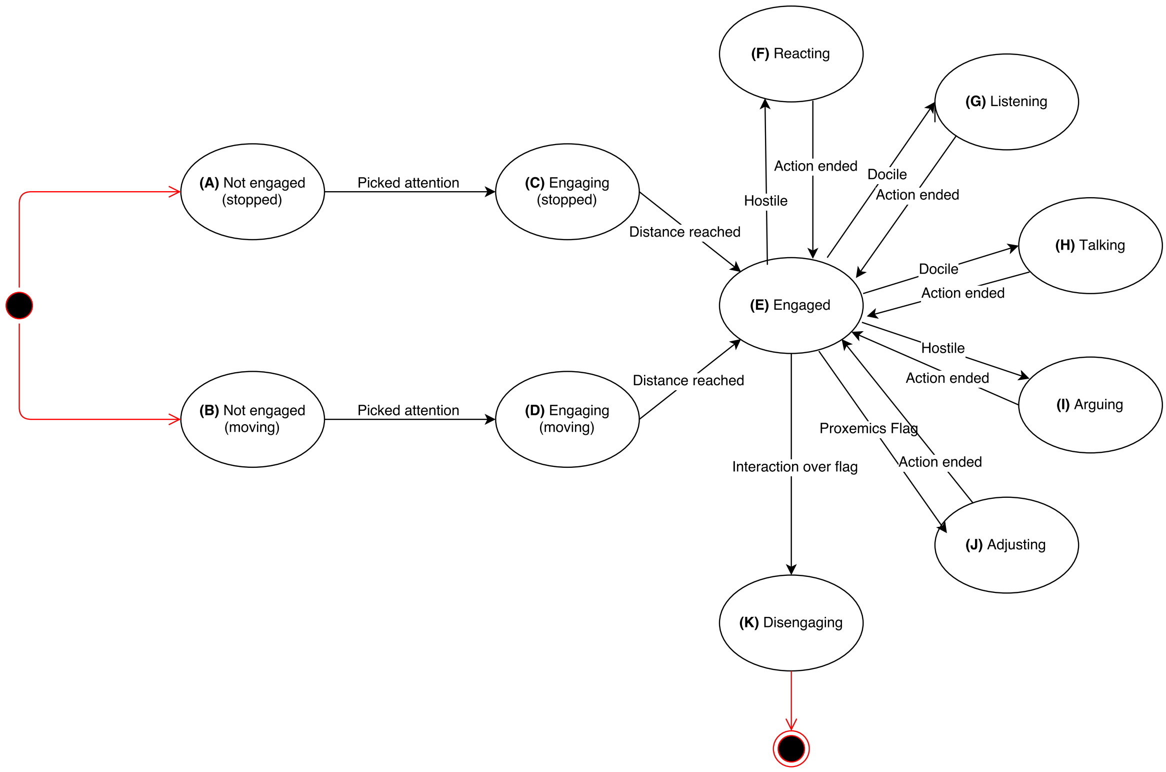


Fig. 3.13 Scenario 2 state machine

This state machine has ten states and each state has several possible animations assigned to it. Avatars either begin at A or B, either stopped in the same place or moving in some direction. From there, if something picked their attention, they move from A to C or from B to D. When the avatars are closer than a certain distance they go to state E. From E the avatar can go to all remaining states and almost all of them return to it. If its his turn to be the receiver the avatar can either go to F, if his mood is hostile, or to G, if his mood docile. The same way if its his turn to act he goes to H, if his mood is docile, or to I, if his mood is hostile. When the avatar needs to change its distance, either coming closer or moving away, it goes to state J. When the interaction time is over it goes to the final state K and ends the interaction.

3.3.3 Scenario 3: Fuzzy evaluator

The third scenario has a more complex appraisal mechanism than the other scenarios. It uses a fuzzy system that relies on data from several sources, including a neural network, to appraise a user presentation. Several components are important in a good presentation: eye contact with the audience, effective use of voice, and effective use of the body. Even though these three components alone don't determine the quality of the presentation, they are still good indicators of the quality of the speaker. The evaluation is done using a fuzzy system with fuzzifier and defuzzifier. It determines how the user is doing on those components and increments or decrements the current evaluation. Analyzing this fuzzy system by components:

Fuzzifier

For eye contact there are two variables that are important to the evaluation, time span and focus. Focus is if the user is looking at the audience or focusing on something else (Floor, Projected Presentation, etc.). For that, a crisp set was used, which indicates if the user is looking at the audience (Focused) or not (Unfocused). Time span means the duration of time the user has been focused or unfocused. For that a fuzzy set with three membership functions, Small Time Span, Medium Time Span, and Large Time Span, will be used.

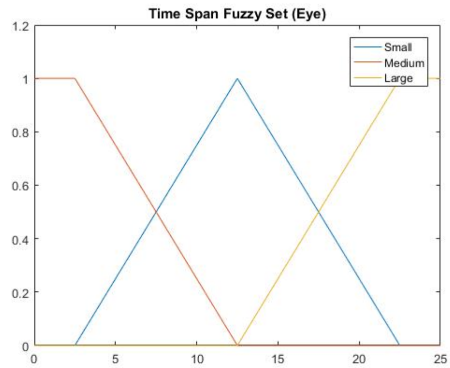


Fig. 3.14 Time Span Fuzzy Set Example

Once again, there are two variables that are important for the posture component, the posture and the time. For the posture a crisp set is once again used to describe Correct and Incorrect Posture. The same way as the previous components, we will describe time using a fuzzy set with three membership functions: Small Time Span, Medium Time Span, and Large Time Span.

Fuzzy inference engine

The implemented inference engine uses individual rule based inference by means of the Mamdani-minimum inference engine, described by the following equation

$$\mu_{B'}(y) = \max_{l=1}^{M}[\sup \min(\mu_{A'}(x), \mu_{A_1^l}(x_1), ..., \mu_{A_n^l}(x_n), \mu_{B^l}(y))] \tag{3.1}$$

With the following set of fuzzy IF-THEN rules:

```
1− IF Focused AND Small Time THEN Small Increment.
2− IF Focused AND Medium Time THEN Medium Increment.
3− IF Focused AND Large Time THEN Large Increment.
4− IF Unfocused AND Small Time THEN Small Decrement.
5− IF Unfocused AND Medium Time THEN Medium Decrement.
6− IF Unfocused AND Large Time THEN Large Decrement.
7− IF Correct Posture AND Small Time THEN Small Increment.
8− IF Correct Posture AND Medium Time THEN Medium Increment.
9− IF Correct Posture AND Large Time THEN Large Increment.
10− IF Incorrect Posture AND Small Time THEN Small Decrement.
11− IF Incorrect Posture AND Medium Time THEN Medium Decrement.
12− IF Incorrect Posture AND Large Time THEN Large Decrement.
```

Defuzzifier

The defuzzifier that was used was the MeOM (mean of maxima) defuzzifier, described by the following equation

$$y^* = \frac{y_0 + y_1}{2} \tag{3.2}$$

where y_0 is the first value of y with the largest degree of confidence and y_1 is the last value of y with the largest degree of confidence, so $y_0 < y^* < y_1$.

Variables

The focused/unfocused variable depends on the direction returned by the Oculos Rift. If the Oculos returns an orientation indicating that the user is facing the audience the variable takes the value 1, otherwise it takes the value 0. The posture variable is defined by the return from a artificial neural network, that receives as input the arms joints rotation matrices and returns if that pose is correct(1) or not(0). The arms joints rotation matrices are returned by the Kinect and the neural network is trained with a set of "correct" and "incorrect" poses.

3.3.4 GUI

Even though the system is able to change the avatars emotional state, and by consequence select their actions in a automatic way, its useful for therapeutic applications and validation to be a way to control it manually. With this in mind a graphical user interface(GUI) was developed using QT.

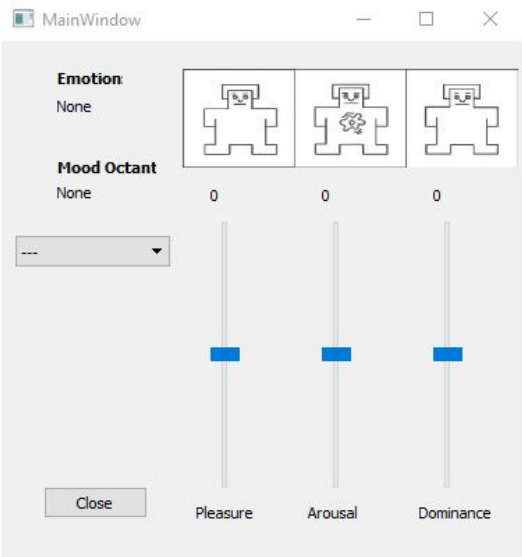


Fig. 3.15 Developed GUI

This application allows the user to define each value of PAD scale, using three sliders that adjust each variable between -1 and 1. Depending on the value an image from Self-Assessment Manikin (SAM) technique is displayed on top of the slider. The user can also select specific emotions, using the box on the left side. The current mood octant as well as closest emotion to those PAD values is displayed in the upper left side of the application. When the user changes the PAD value the GUI writes those values in a file and, if the user selected manual control, the system uses those values to define the avatars current mood.

Chapter 4

Experiments

In this section the results from each developed scenario are presented. Two of the developed scenarios were tested and evaluated by several participants and a questionnaire was filled by each of them. On the third scenario the results from the neural network training and the incorrect pose definition are also presented

4.1 Scenario 1

This scenario was created as a tool for learning interpersonal distances. It was evaluated in terms of usability and fidelity. To evaluate usability the System Usability Scale(SUS) [96] was used, with each user filing the questionnaire (shown in Annex 6.1) at the end of the scenario. In order to see if the system correctly portraits the interpersonal distances each user was asked to consider specific situations when approaching. A total of 25 participants tried the scenario and answered the questionnaire, obtaining the results bellow.

4.1.1 Results

Looking at the participants of this experience their age ranges from 55 to 16 years old with average of 29 years. From the 26 participants 12 had experience with virtual reality and 13 didn't. In terms of gender only 4 participants were female, with the remaining 21 being male. Analyzing the answers themselves, the average value for each question was:

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Avg	3,64	1,20	4,60	1,64	4,20	1,96	4.64	1.60	4.04	1.40

Table 4.1 Scenario 1 Questionnaire Answers

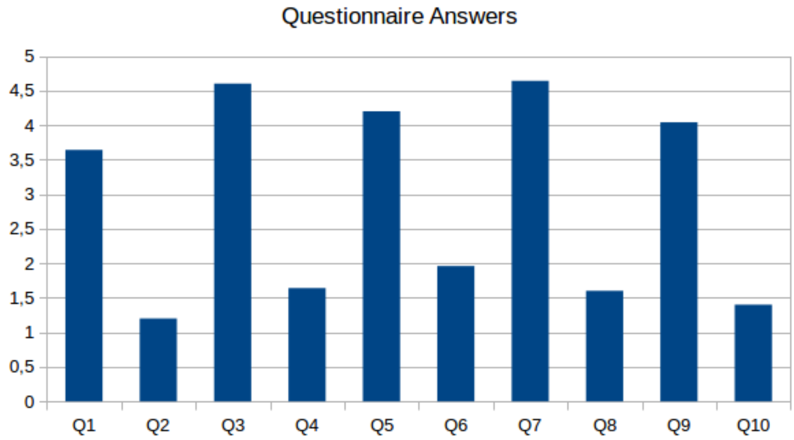


Fig. 4.1 Scenario 1 Questionnaire Answers

By scoring the answers according to the SUS method, subtract 1 from the positive questions and subtract from 5 the negative ones, we obtain:

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Avg	2,64	3,80	3,60	3,36	3,20	3,04	3.64	3.40	3.04	3.60

Table 4.2 Scenario 1 Questionnaire Results

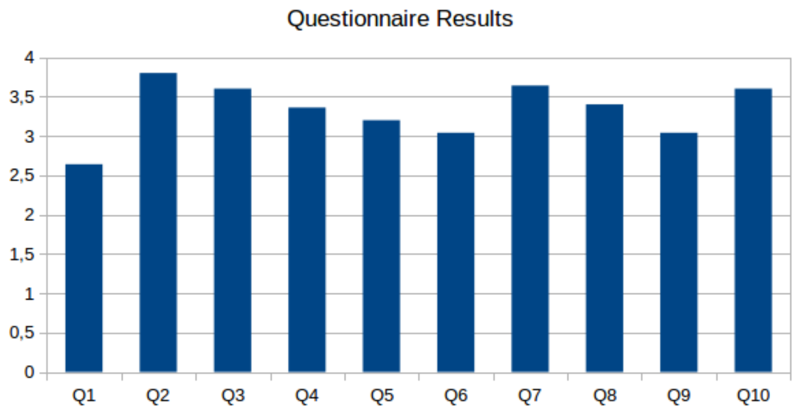


Fig. 4.2 Scenario 1 Questionnaire Results

By summing the answers and multiplying the result by 2.5 we obtain the SUS score, which in average was 83.3. Calculating the variance of each question we obtain:

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Var	0.990	0.166	0.333	0,823	0.416	0.623	0.406	0.916	0.873	0.583

Table 4.3 Scenario 1 Questionnaire Variance

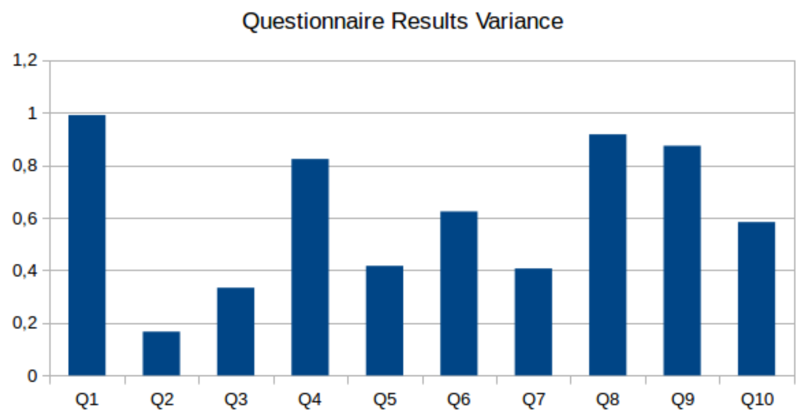


Fig. 4.3 Scenario 1 Questionnaire Variance

4.1.2 Discussion

Overall the participants evaluation of the scenario was positive, with an average score of 83,3. Analyzing the questions themselves, the ones with bigger variance are 1, 4, 8, and 9.

Question 1 asks if the user would like to use the system frequently, an answer that depends very much on the participant. Since this scenario is intended for training purposes the wiliness to use it frequently should be proportional to the participant need to train. Of the 25 participants 16 interacted at a correct distance in a formal situation and 19 in a friendly situation. For the distance chosen as "too close", 13 participants identified it inside the intimate zone and 7 as inside the close phase of the personal zone. Only 10 participants correctly identified all three distances and 1 participant was unable to identify any of them. Question 4 asked the user if it needed the support of a technical person to use the system, a question that once again was heavily dependent on the participant. The average score by participants that had experience with VR was 3,545 with variance 0,272 and from participants that didn't was 3,214 with variance 1,258. This difficulty in using the system is partially related to how familiar participants are with VR environments, since some participants with no experience needed more support. Question 8 asked if the user found the system very cumbersome to use. Like in the previous answer there was a significant difference between participants that had VR experience and those that didn't. The average score given by participants that had experience with VR was 3,909 with variance 0,091 and by those that didn't was 3 with variance 1,230. In fact, several participants mentioned how wearing the Oculus Rift was uncomfortable because of its weight. Question 9 asked how confident the user felt using the system. Once again there was a significant difference between participants that had VR experience and those that didn't. The average score given by participants that had was 3,090 with variance 0,490 and by those that didn't was 3 with variance 1,230.

We can say that this scenario was well received by the participants, with most of the differences in perception being traced back to their experience with VR or absence of it. This however doesn't validate its training potential, since all participants considered they already knew the correct distances to interact with and only tried the system once.

4.2 Scenario 2

This scenario was created as emotion elicitation tool, using territorial situations to do so. As such a questionnaire was created (shown in Annex 6.2) to determine the intensity of the reaction using the Self Assessment Mankin(SAM) [97] technique. This allows the quantification of the elicited reaction in PAD values. A total of 25 participants tried the three variants of the scenario and answered the questionnaire, obtaining the results bellow.

4.2.1 Results

The participants of this experience are basically the same as the previous one, less one 28 male participant with no VR experience being removed. The resulting octants from the answers of the questionnaires are:

	Exu	Dep	Rel	Doc	Bor	Dis	Anx	Hos
S1	4	3	4	0	2	6	2	4
S2	3	0	17	0	1	3	1	0
S3	0	3	1	0	3	1	16	1

Table 4.4 Scenario 2 questionnaire octants

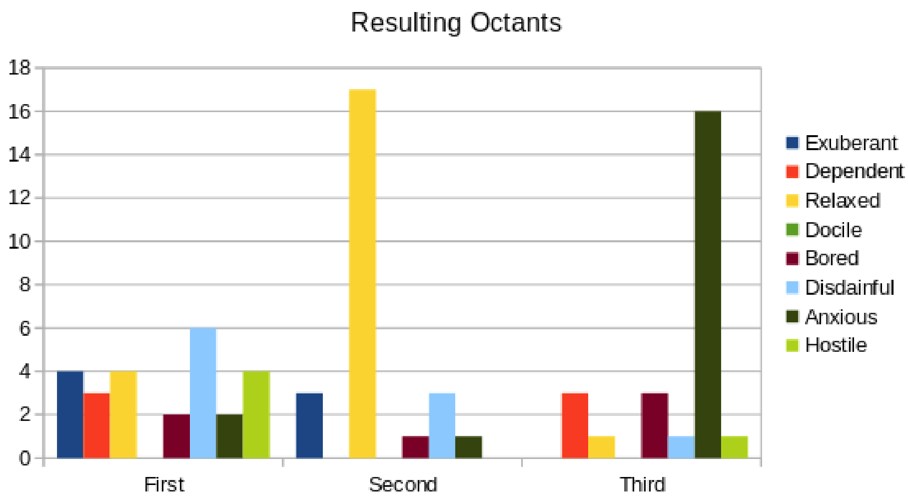


Fig. 4.4 Scenario 2 questionnaire octants

The average of each PAD component, divided by scenario, and its corresponding octant:

	P	A	D	Octant
S1	0.04	-0.1	0.27	3
S2	0.31	-0.36	0.53	3
S3	-0.46	0.33	-0.51	7

Table 4.5 Scenario 2 questionnaire average result

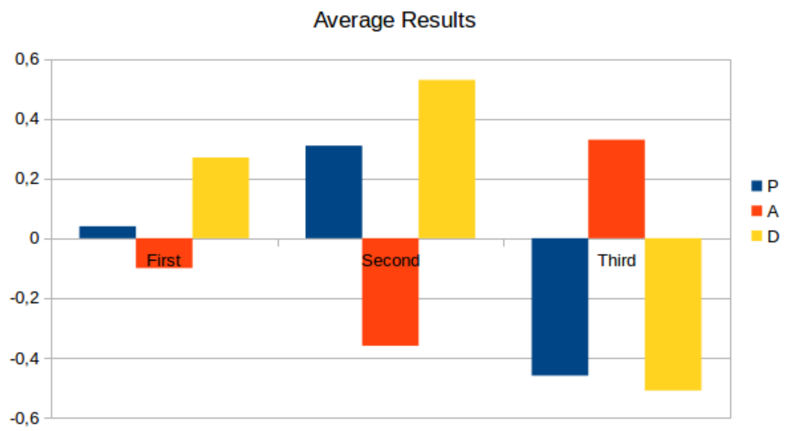


Fig. 4.5 Scenario 2 questionnaire average result

Looking now at the variance of each of the previous component we obtained:

	P	A	D
S1	0.306	0.390	0.364
S2	0.152	0.172	0.184
S3	0.342	0.467	0.357

Table 4.6 Scenario 2 questionnaire variance

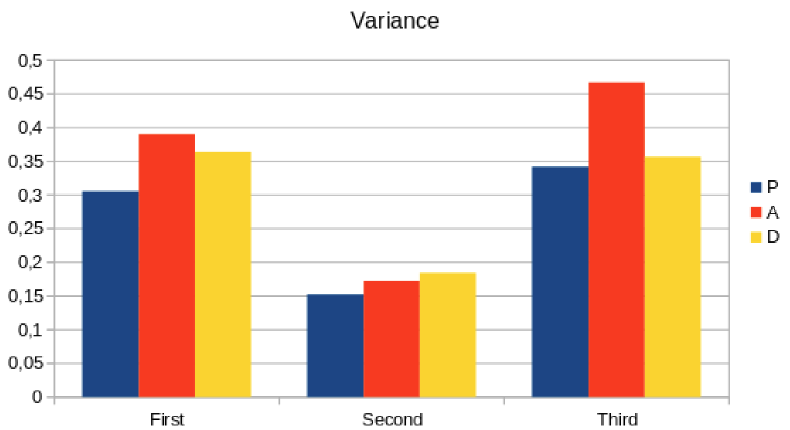


Fig. 4.6 Scenario 2 questionnaire variance

4.2.2 Discussion

Looking at the induced emotional states we can see some coherence in the second (one relaxed avatar) and third (several hostile avatars) variation. We can see that 64% of participants identified as being anxious in the third scenario and 68% as being relaxed in the second scenario. In comparison, the first scenario the most induced state was Disdainful for only 24% of the participants. Looking at the average value given as answers we once see the second and third variation pointing to relaxed and anxious mood respectively, with the first scenario not showing any considerable inclination.

Still, the second and third variation did not present the same coherence in elicited state. Looking at the variance, the third approach had the most spread out set of results of all variants. Even though the majority of participants identified as being anxious, some identified as strongly with different octants. In fact, even though most participants qualified their state as anxious, only a small number of them felt the need to act on that anxiety. In a discussion with participants after the experiment, some commented on the fact that the perception of being in a VR environment largely influenced their response toward it. The lack of actual danger is both a benefit and a drawback of this technology. Inducing negative effects without the element of surprise is much harder than to induce positive ones.

We can say that this scenario obtained moderate success. Even though the most territorial aggressive situation didn't produce the same coherency in the participants as we hoped, the majority of participants were induced into the expected emotional state.

4.3 Scenario 3

This scenario was created as both a training tool and an exposure therapy tool. As such a questionnaire using both the previous methods was created (shown in Annex 6.3). This allows the quantification of the elicited reaction to the presence or absence of the audience and the usability of the system. The participants of this scenario where asked to have a prepared presentation to use and the availability to be tested several times, in order to verify the training transfer. Unfortunately, due to lack of available participants, an evaluation of this system was not possible. Nevertheless the training of the neural network used for this scenario and the qualification of incorrect poses will is presented in this section.

4.3.1 Posture Definition

What was correct and incorrect to do in a presentation was defined by way of an inquiry involving five participants. They were asked to watch eight 30 seconds videos of presentations, five labelled as bad an three as good. During each video they were told to press space if they noticed a moment where the speaker did something bad. After they seen the video the participants where asked to rate the presentation in terms of the general quality of it, eye contact, posture, and vocal control (shown in Annex 6.4). As it was expected the videos labeled as bad received significantly lower grades than the ones labeled as good. Several movements and posture choices where identified as bad (hands in pockets, arms crossed, scratching the head, hands behind back) and as good (open gestures with the arms helping convey the message).

4.3.2 Neural net data collection and training

With that in mind a data colection from the Kinect on the identified postures was made, focusing on the both shoulders and elbows joints. The transformation matrices return by the Kitect from those four joints were then reduced to their equivalent Euler angles. Samples where collected from five participants on eight identified postures:

- scratching head right hand
- scratching head left hand
- arms crossed
- hands in pockets
- hands behind back
- left arm moving toward audience
- right arm moving toward audience
- both arms moving toward audience

For each posture 100 samples where collected from all participants, resulting in a total of 4000 total postures. The correct postures were labeled with the value 1 and the incorrect ones with the value 0. Using Matlab neural net toolbox 450 different artificial neural networks where trained with varying initial weights, activation functions, and number of neurons in the hidden layer. The table in 4.7 shows the results of the top ten networks trained.

Position	HiddenNodes	ActivationFunc	layerWeights	inputWeights	Correlation	MSE
1	24	'tansig'	'rands'	'rands'	0.98576	0.0065228
2	23	'logsig'	'rands'	'initlay'	0.98571	0.0065468
3	25	'tansig'	'initwb'	'initwb'	0.98522	0.0067766
4	25	'tansig'	'initwb'	'initlay'	0.98483	0.0069456
5	25	'logsig'	'rands'	'rands'	0.98481	0.006962
6	25	'logsig'	'initlay'	'initlay'	0.98468	0.0070161
7	25	'tansig'	'initlay'	'rands'	0.98436	0.0071603
8	22	'tansig'	'rands'	'initlay'	0.98425	0.0072105
9	22	'logsig'	'initlay'	'rands'	0.98406	0.0072965
10	19	'logsig'	'initlay'	'initwb'	0.98397	0.00734

Fig. 4.7 Top ten nets with maximum of 25

By considering the neural network result over 0.5 as correct and under as incorrect it is able to correctly label 99.70% of the poses. Still in order to optimize this component a less computationally expensive option was chosen. The table in 4.8 shows the results of the top ten networks with a maximum of 10 hidden neurons.

Position	HiddenNodes	ActivationFunc	layerWeights	inputWeights	Correlation	MSE
1	10	'logsig'	'initwb'	'rands'	0.97258	0.012481
2	9	'tansig'	'rands'	'initlay'	0.97191	0.012781
3	10	'logsig'	'initwb'	'initlay'	0.97151	0.012962
4	10	'logsig'	'initwb'	'initwb'	0.97119	0.013102
5	9	'logsig'	'initwb'	'rands'	0.97103	0.013181
6	10	'logsig'	'initlay'	'initlay'	0.9705	0.013412
7	9	'logsig'	'rands'	'initlay'	0.96967	0.013783
8	10	'tansig'	'initwb'	'initlay'	0.96871	0.014216
9	10	'logsig'	'rands'	'initwb'	0.96838	0.014361
10	9	'tansig'	'initlay'	'rands'	0.9678	0.014636

Fig. 4.8 Top ten nets with maximum of 10

Evaluating the results the same way as before the networks can still correctly label 99.32% of the poses. This neural network was then exported as Matlab function, converted into a c function, and used in this scenario.

Chapter 5

Conclusion and Future Work

Looking at the results obtained, we believe this approach toward avatar behaviour has great potential. In the first scenario, even though the imparting of skills was not tested, it was still positively evaluated in terms of usability by the participants. All were able to correctly identify when the avatar was enjoying the interaction and when he was uncomfortable with the distance taken. As such, a proposal for future work on this scenario revolves around its training validation, preferably with children that are still learning the correct distances. In the second scenario the experiments done also demonstrated the system potential for emotion elicitation. By varying the mood of the avatar, and therefore its behaviour, two of the scenario variants where able produce distinct emotional responses on the participants. Even though the scenario didn't elicited a response on the participants as strong as first thought it gives us a path to take. Future work on this scenario should explore the ability of the avatars to induce positive emotional responses, instead of using them to induce negative ones. The third scenario, the most complex of the three, needs to be carefully evaluated in order to prove it's effectiveness. Unfortunately, the particular type of participants needed to test this scenario makes it much harder to do than the other two. Still, it's the one with the largest potential for exposure therapy and the most interesting in terms of skill impartment.

Immersion is far from perfect, the user full notion of being on virtual environment affects his decision and behaviours. As such, the efficacy of these scenarios are heavily dependent on the user psychological disposition. In training scenarios a bad attitude toward them might weaken the training transfer or even cause a negative training transfer. In elicitation/exposure scenarios such predisposition might invalidate its benefits completely. Still results show that its worth to dwell further on this approach.

The work on this system is far from over with many other concepts and applications still to be implemented. Some more suggestion on possible future work besides the previously mentioned ones are:

- Physiological data analysis and use it for system auto-regulation.
- Validation by psychologists an psychiatrists of the benefits of the scenarios.
- Introducing sound to further improve immersion on the third scenario.
- Creation of more complex scenarios with more avatars.
- Exploring more complex ANN for avatar appraisals.

Chapter 6

Annex

INSTITUTE OF SYSTEMS AND ROBOTICS - UNIV. COIMBRA
PROXEMICS TRAINER - SYSTEM USABILITY SCALE

Gender: M ☐ F ☐ Age: ☐ Experience with VR: Y ☐ N ☐

Instructions: Mark the box that best describes your reaction to the virtual training tool you experienced today. The scale goes from strongly disagree(1) to strongly agree(5).

1- I think that I would like to use this system frequently.

1	2	3	4	5

2- I found the system unnecessarily complex.

1	2	3	4	5

3- I thought the system was easy to use.

1	2	3	4	5

4- I think that I would need the support of a technical person to be able to use this system.

1	2	3	4	5

5- I found the various functions in this system were well integrated.

1	2	3	4	5

6- I thought there was too much inconsistency in this system.

1	2	3	4	5

7- I would imagine that most people would learn to use this system very quickly.

1	2	3	4	5

8- I found the system very cumbersome to use.

1	2	3	4	5

9- I felt very confident using the system.

1	2	3	4	5

10- I needed to learn a lot of things before I could get going with this system.

1	2	3	4	5

Fig. 6.1 Proxemics trainer questionnaire

INSTITUTE OF SYSTEMS AND ROBOTICS - UNIV. COIMBRA

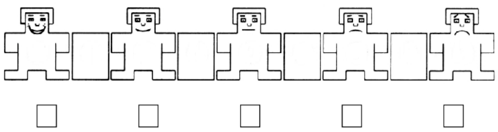
EMOTIONAL ELICITATION USING AVATARS

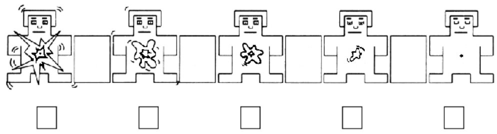
Gender: M ☐ F ☐ Age: ☐ Experience with VR: Y ☐ N ☐

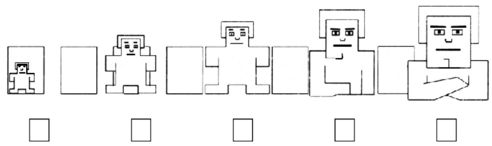
Instructions:

You will be asked to try three different interactions with emotional avatars. After each fill the corresponding section on how you felt during the interaction in three different factors: Pleasure, Arousal, and Dominance

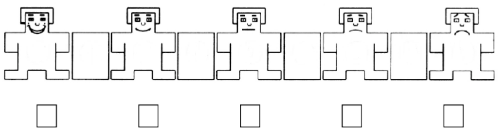
1 - First interaction

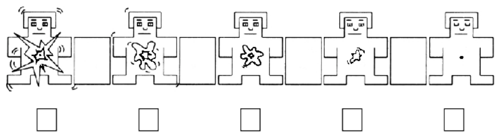


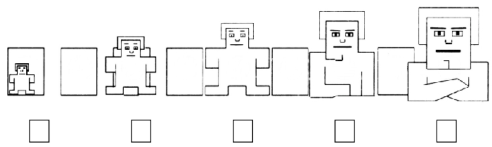




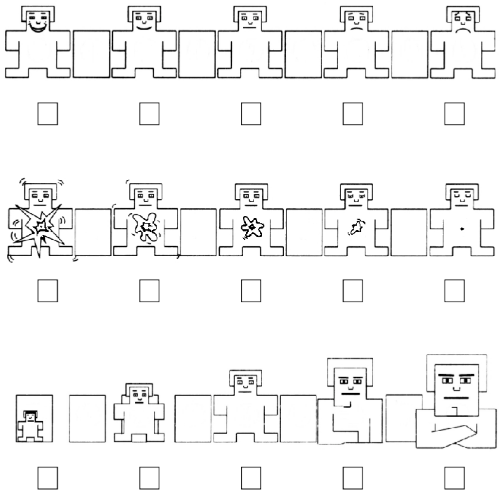
2 - Second interaction







3 - Third interaction



2

Fig. 6.2 Emotional elicitation questionnaire

INSTITUTE OF SYSTEMS AND ROBOTICS - UNIV. COIMBRA

PRESENTATION VIDEOS EVALUATION

Gender: M ☐ F ☐ Age: ☐ Experience with VR: Y ☐ N ☐

Instructions:

You will be shown eight videos, with a duration ranging from 30 seconds to 40 seconds, of excerpts of presentations. After viewing each one rate it, from a scale of 0 (Terrible) to 10 (Perfect), in terms of the general presentation, eye contact, posture, vocal control.

1 - During viewing

1 - Presentation ☐ Eye Contact ☐ Posture ☐ Vocal Control ☐

2 - Presentation ☐ Eye Contact ☐ Posture ☐ Vocal Control ☐

3 - Presentation ☐ Eye Contact ☐ Posture ☐ Vocal Control ☐

4 - Presentation ☐ Eye Contact ☐ Posture ☐ Vocal Control ☐

5 - Presentation ☐ Eye Contact ☐ Posture ☐ Vocal Control ☐

6 - Presentation ☐ Eye Contact ☐ Posture ☐ Vocal Control ☐

7 - Presentation ☐ Eye Contact ☐ Posture ☐ Vocal Control ☐

8 - Presentation ☐ Eye Contact ☐ Posture ☐ Vocal Control ☐

2 - After viewing

A - Rank the following items in terms of importance to a good presentation:

Eye Contact ☐ Posture ☐ Vocal Control ☐

Fig. 6.3 Presentation video questionnaire

INSTITUTE OF SYSTEMS AND ROBOTICS - UNIV. COIMBRA

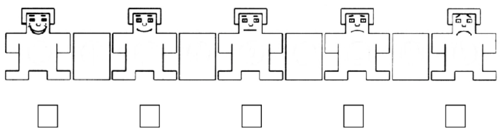
PRESENTATION TRAINING TOOL

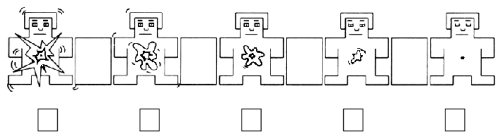
SUS AND ELICITED RESPONSE QUESTIONNAIRE

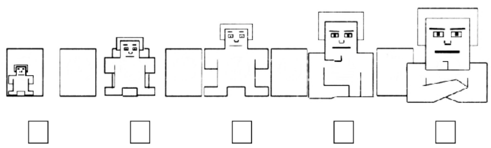
Gender: M ☐ F ☐ Age: ☐ Experience with VR: Y ☐ N ☐

Instructions: You will be asked to train a presentation with and without an audience. After each fill the corresponding section on how you felt during the interaction in three different factors: Pleasure, Arousal, and Dominance

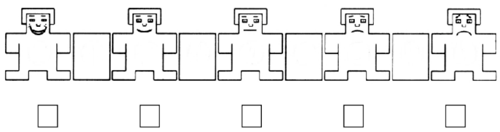
1 - With audience

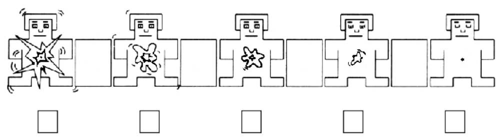


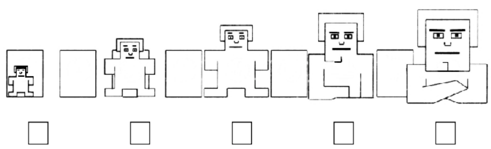




2 - Without audience







Instructions: Mark the box that best describes your reaction to the virtual training tool you experienced today. The scale goes from strongly disagree(1) to strongly agree(5).

1- I think that I would like to use this system frequently.

1	2	3	4	5

2- I found the system unnecessarily complex.

1	2	3	4	5

3- I thought the system was easy to use.

1	2	3	4	5

4- I think that I would need the support of a technical person to be able to use this system.

1	2	3	4	5

5- I found the various functions in this system were well integrated.

1	2	3	4	5

6- I thought there was too much inconsistency in this system.

1	2	3	4	5

7- I would imagine that most people would learn to use this system very quickly.

1	2	3	4	5

8- I found the system very cumbersome to use.

1	2	3	4	5

9- I felt very confident using the system.

1	2	3	4	5

10- I needed to learn a lot of things before I could get going with this system.

1	2	3	4	5

2

Fig. 6.4 Presentation training questionnaire

References

[1] Stacy Marsella. Computational models of emotion. *A Blueprint for Affective Computing-A sourcebook and manual*, 3:21–46, 2010. ISSN 2161-4393. doi: 10.1109/IJCNN.2005.1556117. URL http://ieeexplore.ieee.org/document/1556117/.

[2] Thomas D. Parsons and Albert A. Rizzo. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3):250–261, 2008. ISSN 00057916. doi: 10.1016/j.jbtep.2007.07.007.

[3] Patrick Gebhard. ALMA – A Layered Model of Affect. *AAMAS '05: Proceedings of the 4th international joint conference on Autonomous agents and multiagent systemsent systems*, pages 29–36, 2005. doi: 10.1145/1082473.1082478.

[4] Andrew Coleman. *A Dictionary of Psychology*. Oxford University Press, 2008.

[5] R.W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: analysis of affective\nphysiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, 2001. ISSN 0162-8828. doi: 10.1109/34.954607.

[6] Meike K Uhrig, Nadine Trautmann, Ulf Baumgärtner, Rolf-Detlef Treede, Florian Henrich, Wolfgang Hiller, and Susanne Marschall. Emotion Elicitation: A Comparison of Pictures and Films. *Frontiers in psychology*, 7(February):180, 2016. ISSN 1664-1078. doi: 10.3389/fpsyg.2016.00180. URL http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00180/abstract.

[7] Rosa Mar, Cristina Botella, Mariano Alca, Beatriz Rey, Víctor Liaño Cristina Botella Mariano Alcañiz Belén Guerrero Rosa María Baños, Beatriz Rey, Rosa Baños, Victor Liaño, Mariano Alcañiz, and Belén Guerrero. Changing induced moods via virtual reality. *Persuasive 2006*, pages 7–15, 2006. ISSN 03029743. doi: 10.1007/11755494_3. URL http://link.springer.com/chapter/10.1007/11755494{_}3{%}5Cnhttps://univpn.univie.ac.at/+CSCO+0h756767633A2F2F71626A6179626E712D69322E66636576617472652E70627A++/static/pdf/930/chp{%}3A10.1007{%}2F11755494{_}3.pdf?token2=exp=1428500284{~}acl=/static/pdf/930/ch.

[8] Bruno Patrão, Samuel Lago Pedro, Paula Castilho, and Paulo Menezes. Immersive Technologies as a Key Tool in Therapeutic Context. 2015.

[9] Jack M Loomis and James J Blascovich. Immersive virtual environment technology as a basic research tool in psychology. 31(4):557–564, 1999.

[10] Stéphane Bouchard, Sophie Côté, Julie St-Jacques, Genevì Eve Robillard, and Patrice Renaud. Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games. *Technology and Health Care*, 14:19–27, 2006. ISSN 0928-7329.

[11] Young-Hee Choi, Francesco Vincelli, Giuseppe Riva, Brenda K Wiederhold, Jung-Ho Lee, and Kee-Hwan Park. Effects of group experiential cognitive therapy for the treatment of panic disorder with agoraphobia. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, 8(4):387–93, 2005. ISSN 1094-9313. doi: 10.1089/cpb.2005.8.387. URL http://www.ncbi.nlm.nih.gov/pubmed/16092896.

[12] C.A.P.G. van der Mast P.M.G. Emmelkamp , M. Krijn , A.M. Hulsbosch , S. de Vries , M.J. Schuemie. Virtual Reality Treatment in Acrophobia : A Comparison with Exposure in Vivo Virtual reality treatment versus exposure in vivo : a comparative evaluation in acrophobia. 40(July 2001):509–516, 2016. doi: 10.1089/109493101300210222.

[13] Barbara Olasov Rothbaum, Larry Hodges, Samantha Smith, Jeong Hwan Lee, and Larry Price. A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology*, 68(6):1020–1026, 2000. ISSN 0022-006X. doi: 10.1037/0022-006x.68.6.1020.

[14] Dana K Andersen and Richard M Satava. Virtual Reality Training Improves Operating Room Performance Results of a Randomized , Double-Blinded Study. 236(4):458–464, 2002. doi: 10.1097/01.SLA.0000028969.51489.B4.

[15] Richard M Satava. Surgical Endoscopy T h e first steps. pages 203–205, 1993.

[16] Richard T Stone and Kristopher Patrick Watts. Virtual Reality Integrated Welder Training. 2011.

[17] Chen Xi and Stefano Markidis. 3D Virtual Reality for Education in Nuclear Engineering. 2009.

[18] Bruno Patrão and Paulo Menezes. A virtual reality system for training operators. 9(8): 53–55, 2013.

[19] E. T. Hall. *The hidden dimension*. Anchor Books, 1969. ISBN 0385084765.

[20] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. American Psychiatric Pub, 2013.

[21] Paul R Kleinginna and Anne M Kleinginna. A Categorized List of Emotion Definitions , with Suggestions for a Consensual Definition. 5(4), 1981.

[22] Randolph R Cornelius. *The science of emotion: Research and tradition in the psychology of emotions*. Prentice-Hall, Inc, 1996.

[23] Charles Darwin. *The expression of the emotions in man and animals*. Oxford University Press, 1998.

[24] R. Plutchik. A general psychoelvolutionary theory of emotion. In *R. Plutchik and H. Kellerman (Ed.), Emotion: Theory*, pages (3–33). 1980.

[25] Joseph Ledoux. Emotional networks and motor control : a fearful view. 107, 1996.

[26] William James. What is an Emotion ? 9(34):188–205, 2009.

[27] Kathleen Chwalisz, Ed Diener, and Dennis Gallagher. Autonomic Arousal Feedback and Emotional Experience : Evidence From the Spinal Cord Injured. *Journal of personality and social psychology*, 54(5):820–828, 1988.

[28] Robert W Levenson. Autonomic specificity and emotion. *Handbook of affective sciences*, 2:212–224, 2003.

[29] Magda B Arnold. *Emotion and personality*. 1960.

[30] J. R. Averill. A constructivist view of emotion. In *R. Plutchik and H. Kellerman (Ed.), Emotion: Theory*, chapter 12, pages 305–339. 1980.

[31] James R Averill. *Anger and aggression: An essay on emotion*. Springer Science & Business Media, 2012.

[32] W. Mischel. *Personality and assessment.* Wiley, New York, 1968.

[33] Brent W Roberts. Back to the future : Personality and Assessment and personality development. *Journal of Research in Personality*, 43(2):137–145, 2009. ISSN 0092-6566. doi: 10.1016/j.jrp.2008.12.015. URL http://dx.doi.org/10.1016/j.jrp.2008.12.015.

[34] Walter Mischel. From Personality and Assessment (1968) to Personality Science , 2009. *Journal of Research in Personality*, 43(2):282–290, 2009. ISSN 0092-6566. doi: 10.1016/j.jrp.2008.12.037. URL http://dx.doi.org/10.1016/j.jrp.2008.12.037.

[35] Gerald Matthews, Ian J. Deary, and Martha C. Whiteman. *Personality traits*. 2003. ISBN 9780521831079.

[36] Albert Mehrabian and James A. Russell. An approach to environmental psychology. 1974.

[37] Albert Mehrabian. Basic Dimensions for a General Psychological Theory Implications for Personality, Social, Environmental, and Developmental Studies. 1980.

[38] Albert Mehrabian. Pleasure-Arousal . Dominance : A General Framework for Describing and Measuring Individual Differences in Temperament. 14(4):261–292, 1996.

[39] Albert Mehrabian. Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model. 48(2):86–92, 1996.

[40] Irwin Altman. *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding*. 1975.

[41] Stanford Lyman and Marvin Scott. Territoriality: A neglected sociological dimension. 1967.

[42] G Mcbride, M G King, and J W James. The Journal of Psychology : Interdisciplinary and Applied Social Proximity Effects on Galvanic Skin Responses in Adult Humans. (November 2014):37–41. doi: 10.1080/00223980.1965.10544805.

[43] S. FINANDO. The effects of distance norm violation on heart rate and length of verbal response. 1973.

[44] Bernhard Spanlang and M E L Slater. Proxemics with Multiple Dynamic Characters in an Immersive Virtual Environment JOAN LLOBERA Starlab Barcelona – Universitat de Barcelona Universitat Politècnica de Catalunya and.

[45] Miles L Patterson. An Arousal Model of Interpersonal Intimacy. 83(3):235–245, 1976.

[46] John R Aiellol and Tyra De Carlo Aiello. The Development of Personal Space : Proxemic Behavior of Children 6 Through 16. 1974.

[47] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. Creating rapport with virtual agents. *Intelligent Virtual Agents*, pages 125–138, 2007. ISSN 03029743. doi: http://dx.doi.org/10.1007/978-3-540-74997-4_12. URL http://link.springer.com/chapter/10.1007/978-3-540-74997-4{_}12.

[48] Eva Krumhuber, Antony S.R. Manstead, Darren Cosker, Dave Marshall, Paul L Rosin, and Arvid Kappas. Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, 7(4):730–5, 2007. ISSN 1528-3542. doi: 10.1037/1528-3542.7.4.730. URL http://www.ncbi.nlm.nih.gov/pubmed/18039040.

[49] Cade McCall, Jim Blascovich, Ariana Young, and Susan Persky. Proxemic behaviors as predictors of aggression towards black (but not white) males in an immersive virtual environment. *Social Influence*, 4(2):138–154, 2009. ISSN 1553-4510. doi: 10.1080/15534510802517418. URL http://ezproxy.library.ewu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true{&}db=psyh{&}AN=2009-05720-004{&}site=ehost-live{&}scope=site{%}5Cnhttp://content.ebscohost.com.ezproxy.library.ewu.edu/ContentServer.asp?T=P{&}P=AN{&}K=36635392{&}S=L{&}D=sih{&}E.

[50] H A Simon. Motivational and emotional controls of cognition. *Psychological Review*, 74(1):29–39, 1967. ISSN 0033-295X. doi: 10.1037/h0024127. URL http://10.0.4.13/h0024127{%}5Cnhttp://search.ebscohost.com/login.aspx?direct=true{&}db=pdh{&}AN=rev-74-1-29{&}site=ehost-live.

[51] Marvin Minsky, Ray Kurzweil, and Steve Mann. *Society of mind*, volume 48. 1991. ISBN 0671657135. doi: 10.1016/0004-3702(91)90036-J.

[52] Matthias Scheutz and Aaron Sloman. Affect and Agent Control: Experiments With Simple Affective States. *Intelligent Agent Technology - Proceedings of the 2nd Asia-Pacific Conference on IAT*, (May):200–209, 2001. doi: 10.1142/9789812811042_0028. URL http://eproceedings.worldscinet.com/9789812811042/9789812811042{_}0028.html.

[53] Matthias Scheutz and Paul Schermerhorn. Affective Goal and Task Selection for Social Robots. *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*, page 74, 2009. doi: 10.4018/978-1-60566-354-8.ch005.

[54] F. Thomas and O. Johnston. *Disney animation: The illusion of life*. Abbeville Press, New York, 1981.

[55] N Krämer, B Tietz, and G Bente. Effects of embodied interface agents and their gestural activity. *Intelligent Virtual Agents*, (01):292–300, 2003. ISSN 03029743. doi: 10.1007/b12026.

[56] AJ Cowell and KM Stanney. Embodiment and interaction guidelines for designing credible, trustworthy embodied conversational agents. *Intelligent Virtual Agents*, pages 301–309, 2003. ISSN 03029743. doi: 10.1007/978-3-540-39396-2_50. URL http://link.springer.com/chapter/10.1007/978-3-540-39396-2{_}50.

[57] Ana Paiva, Daniel Sobral, Av Prof, Cavaco Silva, Sarah Woods, Ruth Aylett, Polly Sobreperez, and Lynne Hall. Caring for Agents and Agents that Care : Building Empathic Relations with Synthetic Agents. AAMAS '04:194–201, 2004.

[58] Ana Paiva, João Dias, Daniel Sobral, Ruth Aylett, Sarah Woods, Lynne Hall, and Carsten Zoll. Learning by Feeling: Evoking Empathy with Synthetic Characters. *Applied Artificial Intelligence*, 19(3-4):235–266, 2005. ISSN 0883-9514. doi: 10.1080/08839510590910165. URL http://hdl.handle.net/2299/514{%}5Cnhttp://www.tandfonline.com/doi/abs/10.1080/08839510590910165.

[59] James Lester, Stuart Towns, Charles Callaway, Jennifer Voerman, and Patrick FitzGerald. Deictic and emotive communication in animated pedagogical agents. *Embodied Conversational Agents*, (July):123–154, 2000. URL http://books.google.com/books?hl=en{&}lr={&}id=tHiKZGh9t7sC{&}pgis=1.

[60] Theodore D. Kemper and Richard S. Lazarus. Emotion and Adaptation. *Contemporary Sociology*, 21(4):522, 1992. ISSN 00943061. doi: 10.2307/2075902.

[61] Paul Thagard, Jonathan Gratch, and Stacy Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, 2004. ISSN 13890417. doi: 10.1016/j.cogsys.2004.02.002.

[62] Wsn Reilly. Modeling What Happens Between Emotional Antecedents and Emotional Consequents. *Ace*, (April), 2006. URL http://www.ofai.at/{~}paolo.petta/conf/ace2006/pres/Reilly-ModelingWhatHappensBetween-E0144-pres.pdf.

[63] Robert P. Marinier, John E. Laird, and Richard L. Lewis. A computational unification of cognitive behavior and emotion. *Cognitive Systems Research*, 10(1):48–69, 2009. ISSN 13890417. doi: 10.1016/j.cogsys.2008.03.004. URL http://dx.doi.org/10.1016/j.cogsys.2008.03.004.

[64] Magy Seif El-Nasr, John Yen, and Thomas R Ioerger. FLAME—Fuzzy Logic Adaptive Model of Emotions. *Autonomous Agents and Multi-Agent Systems*, 3(3):219–257, 2000. ISSN 13872532. doi: 10.1023/A:1010030809960. URL http://dx.doi.org/10.1023/A:1010030809960{%}5Cnhttp://link.springer.com/10.1023/A:1010030809960.

[65] Clark Davidson Elliott. The Affective Reasoner: A process model of emotions in a multi-agent system, 1992. URL http://portal.acm.org/citation.cfm?id=142741.

[66] Stacy C. Marsella and Jonathan Gratch. EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90, 2009. ISSN 13890417. doi: 10.1016/j.cogsys.2008.03.005. URL http://dx.doi.org/10.1016/j.cogsys.2008.03.005.

[67] Wenji Mao and Jonathan Gratch. Evaluating a Computational Model of Social Causality and Responsibility. *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 985–992, 2006. doi: 10.1145/1160633.1160809. URL http://doi.acm.org/10.1145/1160633.1160809.

[68] C Conati and H Maclare. Evaluating a probabilistic model of student affect. *Intelligent Tutoring Systems, Proceedings*, 3220:55–66, 2004. ISSN 03029743. doi: 10.1007/b100137.

[69] D Watson and A Tellegen. Toward a consensual structure of mood. *Psychological bulletin*, 98(2):219–235, 1985. ISSN 0033-2909. doi: 10.1037/0033-2909.98.2.219.

[70] J. A. Russell. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980. ISSN 0022-3514. doi: 10.1037/h0077714.

[71] David C Rubin and Jennifer M Talarico. NIH Public Access. 17(8):802–808, 2010. doi: 10.1080/09658210903130764.A.

[72] Hugo Lövheim. A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical Hypotheses*, 78(2):341–348, 2012. ISSN 03069877. doi: 10.1016/j.mehy.2011.11.016.

[73] Lisa Feldman Barrett. Are Emotions Natural Kinds? *Perspectives on Psychological Science*, 1(1):28–58, 2006. ISSN 17456924. doi: 10.1111/j.1745-6916.2006.00003.x.

[74] Jaak Pankseep. Affective Neuroscience: The Foundations of Human and Animal Emotions, 1998.

[75] John R Anderson and Christian Lebiere. The Newell Test for a theory of cognition. *The Behavioral and brain sciences*, 26(5):587–601; discussion 601–648, 2003. ISSN 0140-525X. doi: 10.1017/S0140525X0300013X. URL http://psych.stanford.edu/{~}jlm/papers/AndersonLebiere03.pdf.

[76] James C Lester, Stuart G Towns, Charles B Callaway, Jennifer L Voerman, and Patrick J FitzGerald. Embodied conversational agents. pages 123–154, 2000. doi: 10.1027/1864-9335.40.1.26. URL http://dl.acm.org/citation.cfm?id=371552.371560.

[77] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The cognitive structure of emotions.* Cambridge university press, 1990.

[78] Patrick Gebhard and Thomas Rist. Adding the Emotional Dimension to Scripting Character Dialogues. 2003.

[79] Patrick Gebhard, Martin Klesen, and Thomas Rist. Coloring Multi-Character Conversations through the Expression of Emotions. 2004.

[80] L A Zadeh, I Introduction, and U S Navy. Fuzzy Sets. *Information and Control*, pages 338–353, 1965.

[81] L A Zadeh. The Concept of a Linguistic Variable and its Application to Approximate Reasoning-I. 1975.

[82] Guoqiang Zhang, B Eddy Patuwo, and Michael Y Hu. Forecasting with artificial neural networks : The state of the art. 1998.

[83] Alex Krizhevsky and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. 2012.

[84] G Baxt. Application of artificial neural networks to clinical medicine. 1995.

[85] Ivan E Sutherland. The Ultimate Display. 1965.

[86] M. Fendt and M. S. Fanselow. The neuroanatomical and neurochemical basis of conditioned fear. *Neuroscience and Biobehavioral Reviews*, 23(5):743–760, 1999. ISSN 01497634. doi: 10.1016/S0149-7634(99)00016-0.

[87] BO Rothbaum and AC Schwartz. Exposure therapy for posttraumatic stress disorder, 2002.

[88] Jaye Wald and Steven Taylor. Efficacy of interoceptive exposure therapy combined with trauma-related exposure therapy for posttraumatic stress disorder: A pilot study. *Journal of Anxiety Disorders*, 21(8):1050–1060, 2007. ISSN 08876185. doi: 10.1016/j.janxdis.2006.10.010.

[89] Roberto K. Champney and Meredith Caroll. Conducting Training Transfer Studies in Virtual Environments. In *Handbook of virtual environments: Design, implementation, and applications*. 2007.

[90] Gunnar Ahlberg, D Ph, Lars Enochsson, D Ph, Anthony G Gallagher, D Ph, Leif Hedman, D Ph, Christian Hogman, M Sc, David A Mcclusky Iii, M D, Stig Ramel, D Ph, C Daniel Smith, F A C S, Dag Arvidsson, and D Ph. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. 193:797–804, 2007. doi: 10.1016/j.amjsurg.2006.06.050.

[91] A C Boud, D J Haniff, C Baber, and S J Steiner. Virtual Reality and Augmented Reality as a Training Tool for Assembly Tasks.

[92] Jacobs Hays, R. T. and Prince J. W. Flight simulator training effectiveness: A meta-analysis. 1992.

[93] Paulo Menezes. Openar - an augmented reality and animation library. URL http://orion.isr.uc.pt/OpenAR/.

[94] Fernando Pais, Bruno Patrão, and Paulo Menezes. Virtual Reality as a Training Tool For Human Interactions. *Exp.at'17 International Conference*, 2017.

[95] Fernando Pais, Bruno Patrão, and Paulo Menezes. Exploring Avatar Interactions to Trigger Social Identity-Related Responses. *Exp.at'17 International Conference*, 2017.

[96] John Brooke. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 1996.

[97] Margaret Bradley and Peter Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 1994.