Simulating Empathy in Virtual Humans

Dissertation
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

Hana Boukricha

Simulating Empathy in Virtual Humans

Hana Boukricha
AG Wissensbasierte Systeme (Künstliche Intelligenz)
Technische Fakultät
Universität Bielefeld
Postfach 10 01 31
D-33501 Bielefeld
Germany
E-Mail: hboukric@techfak.uni-bielefeld.de

Von der Technischen Fakultät der Universität Bielefeld zur Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat) genehmigte Dissertation.

Dekan der Fakultät:
Prof. Dr. Jens Stoye

Gutachter:
Prof. Dr. Ipke Wachsmuth
Prof. Dr. Eng. Ana Maria Paiva
Prof. Dr. Dirk Reichardt

Einreichung der Arbeit: 31. Oktober 2012
Tag der Verteidigung: 15. Februar 2013

Die offizielle Druckversion wurde auf alterungsbeständigem Papier gemäss DIN-ISO 9706 gedruckt.

To my family for their love and support
my daddy, my mommy, my brother, and my husband

Acknowledgment

This doctoral thesis was realized during my employment in the Artificial Intelligence Group headed by Prof. Ipke Wachsmuth at the Faculty of Technology in Bielefeld University. The research done in this thesis was kindly supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center 673 (CRC 673), Alignment in Communication.

First of all, I thank my professor Ipke Wachsmuth for giving me this opportunity and for believing in my qualification to succeed in this hard work. His wise advice and very helpful assistance were key contributors to the realization of the present thesis. I also thank Prof. Ana Maria Paiva and Prof. Dirk Reichardt for agreeing to review my thesis. I further thank Prof. Karl Grammer, Andrea Hofstätter, Prof. Pia Knoeferle, and Maria Nella Carminati for their collaboration and support.

I thank the former and current members of the Artificial Intelligence Group for the very nice and supportive working atmosphere in our 'M4 Flur'. Especially, I thank Christian Becker-Asano, Stefan Kopp, and Thies Pfeiffer for the fruitful discussions during crucial moments of this thesis.

Special thanks are directed to my friends and working colleagues, Nhung (my binom), Alexa, Maha, and Lars for their moral support and for keeping together. I also thank my Tunisian and German friends for allowing me to refrain from intense work and to enjoy other moments in life.

Last but not least, I owe my dear family, Abderrahman, Dorra, Heithem, and Tarak, deepest thanks for everything, especially my daddy and mommy who were fully present despite the geographical distance.

Contents

1 Introduction

> [..] it were self-evident that hunters and gatherers had to help each other to survive, so humans must have helping genes. (Hoffman [55], p. 1)

In our modern society, as people become more and more concerned with managing the increasing number of tasks arising in their everyday lives, helping each other is no longer as 'self-evident' as it should be. Some possible reasons include the demands of being mobile and flexible, along with the lack of home care possibilities for the elderly and children. These constraints have resulted in, for example, a lower birth rate, which is expected to have negative consequences for the economic and biological development of modern society. Thus, there is a growing need for people to be supported in managing their everyday lives.

Accordingly, by advancing research in **A**rtificial **I**ntelligence (AI), we can improve the ability of artificial agents to help us solve such problems. The possibility of being confronted with artificial agents increases with each passing year. They can be embedded in multiple settings, e.g., caring and therapy, teaching, internet, and smart-phone applications. To facilitate successful communication and social interaction between artificial agents and human partners, it is essential that aspects of **human social behavior** be considered when designing **human-computer interfaces**. A crucial aspect is **empathy**, often referred to as the ability to perceive and understand others' emotional states. Further, artificial agents are not only required to support and facilitate our everyday lives but also present a suitable testbed to help us understand psychological phenomena such as empathy. Hence, the goal of the present thesis is to provide a computational model of empathy to enhance an artificial agent's social behavior, and to provide an experimental tool for the psychological theories shaping the model.

Parts of this thesis were already published in [11], [18], [14], [16], [17], [15], and [13].

1.1 Motivation

In the following, we motivate the role of empathy in social behavior and in human-machine interaction. Further, we emphasize the role of computational modeling in providing an experimental tool to help understanding psychological phenomena such as empathy.

1.1.1 Empathy, social behavior, and human-machine interaction

Psychologists like Davis [34] and Hoffman [55] have emphasized the role of empathy in **cooperative and prosocial behaviors** like helping, caring, and justice, as well as in preventing **antisocial behaviors** like aggression. For example, students with high empathy scores on a paper-and-pencil measure dispensed more money to a Telethon community and volunteered more hours at shelters for homeless families than did students with lower scores (cf. [55]). In a study by Batson & Ahmad [4], participants were induced to feel empathy for their co-participants in a prisoner's dilemma scenario. As result half of the participants did not defect despite knowing that their partners had already defected. Thus, if the participant feels empathy for the other person, then the dilemma still exists. In a study by Mehrabian & Epstein [77], students were asked to act as teachers and to punish a learner for each incorrect answer using varying intensities of electric shock. As result they found that teachers with high empathy scores on a paper-and-pencil measure were less aggressive than teachers with low empathy scores. Further, neuropsychological evidence [35] substantiates the claim that humans empathize with each other to different degrees depending on several factors including, among others, their mood, personality, and social relationships.

Research in AI describes *social ability* as one of three important properties in an intelligent agent [120]. An **intelligent agent** is defined as: 'a computer system situated in some environment, and capable of autonomous actions in this environment in order to meet its design objectives.' (Wooldridge [120], p. 15). In this regard, **social ability** is defined as an intelligent agent's capacity to interact with other agents, e.g., to cooperate and negotiate [120]. According to the role of empathy in promoting humans' cooperative and prosocial behavior, endowing artificial agents with the ability to empathize enhances their communicative and cooperative social skills, thus improving their intelligent behavior.

Brooks [22] claims that intelligent behavior emerges from an agent's interaction with its environment. Therefore, in order for an agent to exploit the actual situation and to interact with its environment, it needs a physical body composed of sensors and actuators that couple the agent with the environment, and thus influences the agent's internal processing. This is referred to as **embodiment** [22]. Accordingly, an artificial agent's embodiment is crucial to perceiving and understanding an interaction partner's emotional states, and to expressing and communicating empathy through, for example, facial expressions and verbal utterances.

Research on artificial agents exhibiting empathic behavior such as virtual humans and robots, provides valuable results for the integration of empathy in human-machine interaction, and substantiates the role of empathy in enhancing artificial agents' social behavior. **Virtual humans** are 3D animated characters with human like appearance. Prendinger & Ishizuka [95] found that a virtual human that provides empathic feedback through textual expressions can reduce the stress levels of candidates during job interview tasks. Brave et al. [19] found that in a game scenario of casino-style blackjack, an artificial agent that empathizes with the player's game situation is perceived as more likable, trustworthy, and caring. Leite et al. [69] found that a robot's empathic behavior in a chess game scenario is perceived by children as more engaging and helpful. Further, not only does an artificial agents' ability to empathize have a positive impact on human-machine interaction, but also their ability to evoke empathy in humans. In this regard, Paiva et al. [91] show that empathic virtual humans can evoke empathy in children and thus can teach them to deal with bullying situations. A virtual human's empathic behavior also contributes to its ability to build and sustain long-term socio-emotional relationships with human partners as demonstrated by Bickmore & Picard [9].

However, in the context of a competitive card game scenario, Becker et al. [7] found that a virtual human's positive empathic emotions are significantly arousing and stress inducing and thus inadequate. Therefore, in line with humans' ability to empathize with each other to different degrees, we believe that a **modulation** of a virtual human's empathic behavior through factors such as its mood, personality, and relationship to its interaction partner, will allow for a more **adequate empathic behavior** in the agent across different interaction scenarios. Although much effort has gone to providing artificial agents with features such as mood, personality and social relationships, little attention has been devoted to the role of such features in modulating their empathic behavior. Altogether, endowing artificial agents with appropriate empathic behavior contributes in enhancing their social behavior.

1.1.2 Interdisciplinary partnership: AI and psychology

In the previous section, the role of empathy in enhancing artificial agents' social skills was motivated. In this regard, not only does psychological research contribute to the computational modeling of psychological phenomena, computational modeling can also contribute to clarifying the underlying theories. Marsella et al. [73] refer to this as an interdisciplinary partnership. The authors argue that **computational modeling** can concretize concepts dealt with in psychological theories where they are typically described in natural language at an informal and abstract level. The process of developing a computational model facilitates the systematic examination and deeper understanding of the different theoretical aspects and assumptions. Furthermore, a more flexible evaluation of a theory can be achieved by defining different parameters and modules within the model, thus allowing for the investigation of the effects and interactions of hidden aspects in the theory. For example, De Vignemont & Singer [35] claim that humans empathize with each other to different degrees depending on several factors, e.g., their relationships with each other. However, in their theory, they do not provide concrete values for their proposed factors, now explain how they systematically influence the degree of empathy. A computational model would make it possible to test these proposed factors and to, in turn, develop a systematic definition of how they interact to impact the degree of empathy. In sum, computational modeling provides an **experimental tool** for the underlying theories, and contributes to their verification and enhancement.

1.2 Thesis scope and objectives

As mentioned earlier in this chapter, the goal of the present thesis is to provide a **computational model of empathy** to enhance an artificial agent's social behavior, and to provide an experimental tool for the psychological theories shaping the model. The proposed empathy model is realized for the virtual humans MAX and EMMA, and is applied and evaluated in the context of two different interaction scenarios. The virtual humans MAX and EMMA are developed in the AI-Group headed by Prof. Dr. Ipke Wachsmuth at the Faculty of Technology of Bielefeld University. In the following, we formulate the main problems and requirements to achieve our objectives. At the end of this section, the research context of the present thesis will be outlined.

Problems and requirements

After careful consideration of the theoretical background on empathy and of related work on empathic artificial agents, (Chapters 2 and 3), the following problems and requirements emerged. In this regard, we define three requirements for building a computational model of empathy that addresses three central processes to empathy. We refer to these requirements as Empathy Mechanism, Empathy Modulation, and Expression of Empathy. Further, we define two requirements that a proposed model of empathy should satisfy, namely, Adequacy and Universality.

- **Empathy Mechanism**: empathy arises in response to others' emotions. Therefore, an approach to recognize others' emotional states and to generate an empathic emotion is required.
- **Empathy Modulation**: empathy emerges in different degrees depending on factors such as mood, personality, and social relationships. Therefore, to generate different degrees of empathy, an approach to determine the values of the modulation factors in question, and to modulate the empathic emotion through these factors is required.
- **Expression of Empathy**: empathy is communicated by different modalities. Therefore, in order to define an expression of empathy, a repertoire of at least one modality that represents an adequate expression of empathy is required.
- **Adequacy**: the proposed empathy model should provide adequate empathic behavior in line with underlying theories and hypotheses.
- **Universality**: the empathy model should be easily applied and adapted to different context scenarios.

Research context

This doctoral thesis is realized in an interdisciplinary research context supported by the **C**ollaborative **R**esearch **C**enter 673 (CRC 673), **Alignment in Communication**. In the project part **A1 Modelling Partners**, affective partner modeling as a function of internally representing others' perceived emotional states, and emotional alignment are addressed. In this regard, the computational model of empathy proposed here allows a virtual human to align to interaction partners' perceived emotional states. Further, the empirical evaluation of the model was performed in cooperation with Prof. Dr. Pia Knoeferle and Dr. Maria Nella Carminati as members of the psycholinguistics section of the project.

1.3 Thesis structure

- *Chapter 2*: this chapter discusses the **psychological and neuropsychological background** relevant to understand the theoretical concepts underlying the proposed computational model of empathy. Accordingly, the chapter begins by discussing different theoretical models of empathy. Further, we focus on the implications of the theoretical models of emotion, and of theories and findings on emotion expression and perception for the realization of several theoretical aspects of empathy.

- *Chapter 3*: this chapter begins by giving an overview on some of the **computational models of emotion**, and their application and evaluation in virtual human and robot scenarios. Subsequently, approaches to the simulation of **emotion expression**, and methods for **emotion recognition** are presented. Further, different **computational models of empathy** and their application and evaluation in virtual human and robot scenarios are discussed. Finally, the requirements formulated in the present chapter are concluded, and related work on empathy is classified with respect to these requirements.

- *Chapter 4*: this chapter introduces a **new virtual human** with a female-like appearance, called EMMA. In this chapter, the different steps taken in developing EMMA are described. A highlight is EMMA's ability to express a wide range of emotional states based on a large repertoire of emotional facial expressions, a property crucial to the generation and expression of empathic emotions.

- *Chapter 5*: this chapter introduces our **computational model of empathy** realized for the virtual humans MAX and EMMA. The chapter begins by presenting the overall structure of the model and its integration into an existing cognitive architecture. Further, the different processing steps of the model are introduced, and the implications of the theoretical background in shaping the model are detailed.

- *Chapter 6*: this chapter introduces the **application and evaluation** of the proposed empathy model. Two application scenarios are presented for the model. This is followed by the description of an empirical evaluation of the model, along with its results.

- *Chapter 7*: this chapter summarizes and discusses the **results and contribution** of the thesis in light of the goals and requirements formulated in the present chapter. The chapter ends with **future directions** for research.

2 Theoretical background

The development of a computational model of empathy is an interdisciplinary task involving several psychological and neuropsychological theories, and depending significantly on advances in these theories. Therefore, this chapter discusses the psychological and neuropsychological background relevant to understand the theoretical concepts underlying our computational model of empathy. Section 2.1 covers the debate on a universal definition of empathy and describes different theoretical models of empathy. Subsequently, Sections 2.2 and 2.3, respectively, elucidate the implications of theoretical models of emotion and of theories and findings on emotion expression and perception for the realization of several theoretical aspects of empathy.

2.1 Theoretical models of empathy

The same as other psychological phenomena such as emotion (see Section 2.2), empathy has no universally agreed upon definition. In the following, we introduce the debate on a universal definition of empathy, and discuss the different theoretical models of empathy. In Section 2.1.4, our working definition of empathy is given and three central processes to empathy are identified which are crucial to shaping the computational model of empathy proposed in the present thesis (see Chapter 5).

2.1.1 Definitions of empathy

The origin of the term empathy can be traced back to the **German term *Einfühlung*** which, literally translated, means '**feeling oneself into**' (Goldstein & Michaels [45], p. 4). In the nineteenth-century, Lipps introduced the term *Einfühlung* in the context of German aesthetics where he defined it as the tendency to project oneself 'into' an object of art during its contemplation so as to obtain a better understanding of the observed object [45]. In the early twentieth-century, Lipps introduced *Einfühlung* in a psychological context and extended its definition from understanding of art to the understanding of persons. In this regard, he states that the perception of another's emotional state leads to the imitation of the other's emotional expression. This results in a similar emotional reaction in the observer that allows a better understanding of the other [45]. In 1910, Titchener introduced the **English term *empathy*** as a translation of the German term *Einfühlung* [34].

A literal translation of the term *Einfühlung* means 'feeling oneself into' and refers more to an active process of projecting oneself into the other [34]. Consequently, there was a focus on empathy as an active process. This resulted in the emergence of a further definition of empathy as the cognitive understanding of the other by taking on the role or the perspective of the other [34]. This deviation from the original definition lead to a debate on a universal definition and to the emergence of different definitions. The multiple **definitions of empathy** can be subdivided into **three major categories** [87]: **(1)** empathy as an affective response to the other's emotion, **(2)** empathy as the cognitive understanding of the other's emotion, and **(3)** empathy as the combination of the above two definitional categories. The definitions of empathy as well as the processes underlying empathy are discussed next within different psychological and neuropsychological models.

2.1.2 Psychological models of empathy

In the following, we present two psychological models of empathy offering comprehensive and detailed models that unify different views and perspectives. Such models are crucial to the computational modeling of empathy, and as such have major implications for our computational model. These implications will be detailed in Chapter 5.

Hoffman's model of empathy

In his theory of prosocial moral behavior and development, Hoffman [55] emphasizes the important role of empathy in contributing to prosocial actions and moral principles. He examines empathy and defines it as

> [..] an affective response more appropriate to another's situation than one's own. (Hoffman [55], p. 4)

He argues that an empathic response need not to be a close match to the affect experienced by the other, but can be any emotional reaction compatible with the other's situation. Accordingly, Hoffman follows the definition of empathy as an affective response to the other's emotion (cf. Section 2.1.1). However, he accords a special importance to the cognitive processes underlying empathy, and claims that, in his perspective, empathy is defined in terms of the processes required for an empathic response. In this regard, he states that empathy is multidetermined, and introduces five processes by which an empathic response arises. He calls these processes **modes of empathic arousal**. Among these modes of empathic arousal, Hoffman distinguishes three that he defines as 'preverbal, automatic, and essentially involuntary' ([55], p. 5): *mimicry*, *classical conditioning*, and *direct association*. He states that these modes require lower-level cognitive processing. Further, he distinguishes two modes that he defines as requiring higher-level cognitive processing: *mediated association* and *role-taking*.

Mimicry Hoffman [55] defines mimicry as a process involving the imitation of another's facial expressions, voice, and posture that triggers an afferent feedback eliciting feelings in the observer that are similar to those of the observed other. Thus, he identifies two successive steps underlying the process of mimicry which he calls *imitation* and *feedback*. Hoffman discusses several empirical findings that provide supportive evidence for imitation and feedback.

With regard to **imitation**, Meltzoff & Moore [79] found that infants between two-and three-weeks of age can reproduce tongue protrusion, mouth opening, lip protrusion, and a sequential finger movement demonstrated by an adult. Thus, two-to three-week old infants are able to imitate adults' facial and manual gestures. Meltzoff & Moore not only provide evidence for infants' ability to imitate, but also provide a theoretical model explaining the mechanisms underlying infant's facial imitation [80]. Further, Termine & Izard [115] found that nine-month old infants reproduce their mother's facial and vocal expressions of joy and sadness. The same as Lipps, Bavelas et al. [6] refer to imitation as motor mimicry. In their attempt to study motor mimicry, they found that a broad class of incidents evoked motor mimicry in human observers, e.g., a painful situation, laughter, smiling, affection, embarrassment, discomfort, disgust, stuttering, facing a thrown projectile, ducking away from being hit, word-finding, reaching with effort, and succeeding and failing at a timed task.

Regarding **feedback**, Hoffman discusses supportive evidence provided by evaluation of the so called **F**acial **F**eedback **H**ypothesis (FFH) which refers to the assumption that changes in facial expressions elicit emotion (see Section 2.2).

Classical conditioning Hoffman [55] defines classical conditioning as the pairing of others' observed emotional cues with own simultaneous emotional experience. Thus, the others' observed emotional cues become conditioned and an emotional response is evoked as an empathic response when observing these emotional cues in the other. Hoffman [55] states that classical conditioning as an empathy arousing mode occurs during the preverbal years of childhood, e.g., through mother-child interactions. For example [55], when the mother's body stiffens as an anxious reaction, this negative feeling is transmitted to the child through body contact while holding the child. The child's negative feeling is paired with cues observed in the mother, which then become conditioned stimuli. Thus, the observation of these expressive cues evokes the negative feeling in the child, even in the absence of physical contact.

Direct association Hoffman [55] defines direct association as the process by which the observation of another's emotional cues, e.g., facial expressions or any other situational cue, elicits similar past experiences that evoke an emotional response in the observer compatible with the other's situation. Thus, the observer directly associates the other's emotional cues with similar past experiences. Hoffman also argues that direct association is a variant of classical conditioning. Compared to classical conditioning, direct association requires the previous experience of an emotion similar to that of the other.

Mediated association Hoffman [55] also refers to mediated association as verbal mediation. He defines it as the process by which the processing and decoding of another's verbal messages elicits similar past experiences that evoke an emotional response in the observer compatible with the other's verbally described situation. According to Hoffman, verbal mediation requires semantic processing and decoding of the meanings of words, and thus require higher-level cognitive processing and more mental effort than the previously introduced empathy arousing modes. Further, he claims that verbal mediation evoke an empathic response in the observer even in the absence of the other.

Role-taking Hoffman [55] defines role-taking as an observer's ability to put himself in the other's situation, and to imagine how the other feels. According to Hoffman, role-taking requires higher-level cognitive processing.

Stotland [113] found that participants' negative emotional response when exposed to a painful heat treatment procedure was higher when they imagined themselves being in this situation, than when they imagined how the other felt or when they simply attended to the other's movements. He also founds that participants in the first and second condition showed a delayed physiological response after the beginning of the heat treatment procedure than did participants in the last condition. Consequently, Hoffman states that the delay in participants' physiological response measured in Stotland's experiment was due to a higher-level of cognitive processing and the mental effort involved in role-taking.

Based on Stotland's results, Hoffman distinguishes two types of role-taking. *Self-focused role-taking* as referring to the 'imagine self' condition in Stotland's experiment. *Other-focused role-taking* as referring to the 'imagine him' condition in Stotland's experiment. Further, Hoffman also states that other-focused role-taking is more cognitively demanding than self-focused role-taking as it requires taking the other's 'inner' states into account.

Role-taking has also been defined by Higgins [53] as

> seeing the world through another's eyes or putting yourself in another's shoes. (Higgins [53], p. 119)

The same as Hoffman [55], Higgins [53] makes a distinction between two types of role-taking, *situational role-taking* and *individual role-taking*. In **situational role-taking**, an observer appraises another's situation by using the same appraisal mechanisms as if he were in the situation himself. In **individual role-taking**, an observer appraises another's situation by taking the other's viewpoint and characteristics into account. Hoffman also highlights the importance of many modes of empathic arousal.

Importance of many modes According to Hoffman [55], the many modes of empathic arousal allow observers to empathize with others in many different situations regardless of the emotional cues available. What determines which particular mode will operate is the nature of the situation, e.g., the perception of expressive cues will more likely foster mimicry. Furthermore, the different modes of empathic arousal are thought to commonly **operate in conjunction** with one another. For example, the perception of different emotional cues may activate different modes. In this case, mimicry, classical conditioning, and direct association precede mediated association and role-taking which involve higher-level cognitive processing. Moreover, role-taking may activate the other modes, e.g., it may activate mimicry through the imagination of the other's emotional expressions. The subsequent operation of the empathy arousing modes may intensify or fine-tune the empathic response. The many modes also contribute to getting an accurate empathic response, e.g., when the other's expressive cues belie his feelings in a certain situation, role-taking instead of mimicry will evoke a more accurate empathic response. Hoffman also addresses the development of empathy from childhood to maturity.

Development of empathy Hoffman [55] distinguishes four stages in the development of empathy with respect to self-other differentiation. These are *egocentric empathy*, *quasi-egocentric empathy*, *veridical empathy*, and *empathy beyond the situation*.

In the **egocentric empathy** stage, children toward the end of their first year have a confusion in their self-other differentiation. For example, a 10-month-old child who observed a friend falling down and crying, also started to cry and then put his thumb in his mouth and put his head in his mother's lap. The child's reaction was the same as when he himself was in this situation. At this stage, children are confused about the difference between what happens to others and what happens to themselves. Further, their empathy is limited to the preverbal empathy arousing modes: mimicry, classical conditioning, and direct association.

In the **quasi-egocentric empathy** stage, children early in their second year are aware of 'self and others as separate physical entities' (Hoffman [55], p. 64). For example, a 14-month-old child was comforting his crying friend by bringing him to his own mother despite the presence of the friend's mother. At this stage, children are confused about the difference between their own needs and the others' needs. Further, their empathy is extended to an early form of the empathy arousing mode self-focused role-taking.

In the **veridical empathy** stage, children late in their second year are aware of 'self and others as having independent internal states' (Hoffman [55], p. 64). Hoffman illustrates the transition from quasi-egocentric empathy stage to the veridical empathy stage with the following example. A two-year-old child comforted his friend by bringing him his own teddy bear, but when his friend did not stop crying, the child looked for his friend's teddy bear. At this stage, children can differentiate between their own needs and those of others, and are aware of others' inner states as distinct from their own. Further, they are now capable of self-focused and other-focused role-taking.

In the **empathy beyond the situation** stage, children become aware of 'self and others as having their own personal histories, identities, and live beyond the immediate situation' (Hoffman [55], p. 64). At this stage, empathy also occurs with respect to others' imagined life situations. Further, Hoffman states that observers' empathic response can have different degrees of intensity depending on different factors influencing their empathy bias.

Empathy bias According to Hoffman [55], an observer's empathic response can have different **degrees of intensity** depending on the salience and intensity of the other's emotional cues and on the relationship between the observer and the other. The higher the salience and intensity of the other's emotional cues, the more intense the observer's empathic response. With regard to the relationship between the observer and the other, Hoffman distinguishes *in-group bias*, *friendship bias*, *similarity bias*, and *here-and-know bias* as factors influencing the observer's empathy bias. That is, observers empathize to a higher degree with family members, primary group members, friends, people with similar concerns and needs, and with present rather than absent people. He refers to *in-group bias*, *friendship bias*, and *similarity bias* as *familiarity bias*. Further, he discusses empirical evidence for the proposed *familiarity bias*.

Costin & Jones [29] found that children watching a friend or an acquaintance in a hypothetical dilemma reported more empathy for their friends than for the acquaintance and were more motivated to help their friends than the acquaintance. Krebs [65] found that participants who believed to be similar in their personality profile to the observed person showed higher physiological response for the pleasure and pain experiences of the other person than for those believed to be dissimilar. In his theory of prosocial moral behavior and development, Hoffman claims that empathy plays a crucial role in the development and motivation of prosocial moral principles.

Empathy and prosocial moral principles Hoffman [55] emphasizes the important role of empathy in contributing to prosocial actions and moral principles. In this regard, he focuses on caring and justice as prosocial moral behaviors. Hoffman defines caring as the maximization of well being and happiness in others. Caring includes concern for others' welfare and helping others in need.

Bavelas et al. [5] argue for mimicry as a communicative act mediating solidarity and involvement with others. They state

> By immediately displaying a reaction appropriate to the other's situation (e.g., a wince for the other's pain), the observer conveys, precisely and eloquently, both awareness of and involvement with the other's situation. (Bavelas et al. [5], p. 278)

Accordingly, Hoffman states that 'mimicry-based empathy' (Hoffman [55], p. 45) not only provides a motive for prosocial behavior but is also a prosocial act.

Davis' organizational model of empathy

The same as Hoffman [55], Davis [34] attempts to unify the different perspectives on empathy. He proposes an organizational model of empathy which is based on a definition of empathy as

> [..] a set of constructs having to do with the responses of one individual to the experiences of another. (Davis [34], p. 12)

By means of such a broad definition, Davis' goal is to combine the different aspects of empathy that had been considered separately, i.e. empathy's affective and cognitive aspects. In this regard, Davis' definition of empathy falls into the third category of empathy definitions (cf. Section 2.1.1).

As compared to Hoffman, Davis not only considers different aspects of empathy but also defines their relationship to each other. Hence, he defines **four related constructs** that constitute the proposed organizational model of empathy, *antecedents*, *processes*, *intrapersonal outcomes*, and *interpersonal outcomes*. Figure 2.1 illustrates these constructs and their relationship to each other.

Antecedents Davis [34] defines antecedents as the characteristics of the observer and the situation which influence empathic processes and outcomes (see Figure 2.1). He listed the following as relevant characteristics of the observer: *biological capacities*, *individual differences*, and *learning history*. *Biological capacities* refer to the ability to engage in, e.g., higher-level cognitive processes such as role-taking. In this regard, Davis refers to Hoffman's four stages of empathy development with respect to self-other differentiation [55]. *Individual differences* refer to differences in the dispositional tendency to engage in certain empathic processes and to experience certain empathic outcomes. *Learning history* refers to acquired values and behaviors related to empathy.

With regard to the characteristics of the situation, Davis mentions *strength of situation* and *observer-target similarity*. *Strength of situation* refers to the strength of perceived emotional cues, e.g., the perception of a strong facial display of an emotion leads to a strong affective response as an empathic outcome. *Observer-target similarity* refers to

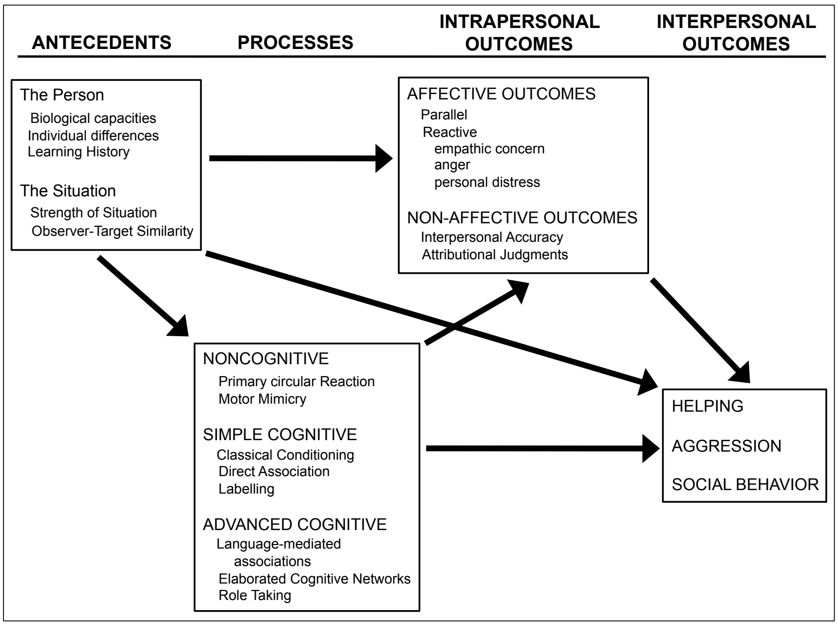


Figure 2.1: Davis' organizational model of empathy [34], p. 13.

the degree of similarity between the observer and the other. The higher the degree of similarity, the more intense or likely the empathic outcomes. The definition of *strength of situation* is in line with Hoffman's [55] assertion that the higher the salience and intensity of the other's emotional cues, the more intense the observer's empathic response. Furthermore, the definition of *observer-target similarity* is in line with Hoffman's introduced *similarity bias*.

Processes Davis [34] defines processes as the mechanisms by which empathic outcomes are generated (see Figure 2.1). As processes, Davis introduces *primary circular reaction*, *motor mimicry*, *classical conditioning*, *direct association*, *language mediated association*, and *role-taking* as Hoffman's modes of empathic arousal [54] [55]. Following Hoffman [54], Davis defines *primary circular reaction* as the newborn's reactive cry to another infant's cry. This process is not included in Hoffman's modes of empathic arousal according to his more recent work [55].

Further, Davis introduces *labeling* and *elaborated cognitive networks* as empathic processes as proposed by Eisenberg et al. [37]. Accordingly, he defines *labeling* as making inferences about the other's experience based on simple cues, e.g., knowing that college graduations are likely to produce happiness. He defines *elaborated cognitive networks* as the use of available cues to access stored knowledge and to use this knowledge to make inferences about the other. Davis states that labeling and elaborated cognitive networks are respectively similar to direct association and language mediated association.

Intrapersonal outcomes Davis [34] defines intrapersonal outcomes as affective and non-affective outcomes generated by the above introduced empathic processes. These are affected by the empathic antecedent conditions (see Figure 2.1), and are not manifested in overt behavior toward the other.

He defines affective outcomes as an observer's affective responses to observed experiences of the other, and distinguishes *parallel* from *reactive* outcomes. Parallel outcomes refer to affective responses that match or reproduce the affect observed in the other. Consequently, parallel outcomes are more likely to be produced through lower-level cognitive processes, e.g., motor mimicry. Furthermore, *observer-target similarity* increases the likelihood of parallel outcomes. Reactive outcomes refer to affective responses that are different from affect observed in the other. For example, witnessing someone being unfairly treated may lead an observer to react with *empathic anger*, or with *empathic concern* as a feeling of compassion or pity, or with *personal distress* as a feeling of discomfort or *anxiety*. Consequently, experiencing an affective response different from the affect observed in the other requires higher-level cognitive processing and interpretation of the other's state. Thus, reactive outcomes are more likely to be produced through higher-level cognitive processes such as role-taking.

With regard to non-affective outcomes, Davis defines *interpersonal accuracy* and *attributional judgments* as cognitive outcomes that are more likely to be produced through higher-level cognitive processes such as role-taking. *Interpersonal accuracy* refers to the adequate estimation of others' internal states such as their feelings and thoughts. *Attributional judgments* refer to how an observer explains what causes the other's behavior.

Interpersonal outcomes Davis [34] defines interpersonal outcomes as overt behaviors toward the other that mainly result from intrapersonal outcomes. Such interpersonal outcomes are helping, aggression, and social behavior (see Figure 2.1). Davis defines helping as

> [..] actions taken by one person which, at some cost to the self, improve the welfare of another by either reducing negative states and/or increasing positive states for that other. (Davis [34], p. 127)

Further, he defines aggression as antisocial behavior inhibited by empathy and negatively associated with empathy. Social behavior refers to all behaviors that occur in the context of establishing or maintaining social relationships, and that are affected by empathy. Helping and aggression are also included in these behaviors. According to Davis' mediational model [34] which addresses the role of empathy in social relationships, one's ability for empathy impacts one's social behavior which is perceived by others and which therefore influences one's social relationships with others.

Conclusion

Both Hoffman [55] and Davis [34] offer comprehensive and detailed models of empathy based on an elaborate discussion of empirical and theoretical work. However, they do not provide an explanation of the exact mechanisms underlying the different aspects of empathy considered in their models, e.g., the mechanisms required to decode and recognize others' emotional states. Further, in Davis's model some hypothesized relationships between the proposed empathy constructs remain difficult to understand, e.g., the relationship between the processes and the intrapersonal outcomes. In this regard, emotion theories, and theories and findings on emotion expression and perception may allow for a more precise explanation of the mechanisms in question (see Sections 2.2 and 2.3).

2.1.3 Neuropsychological models of empathy

Neuropsychological models of empathy investigate the **neural mechanisms** underlying empathy and thus offer further support to psychological models. They also allow for a more precise consideration of the processes underlying empathy. Accordingly, they provide additional support for the computational modeling of empathy, and as such have crucial implications for our computational model. Again, these implications will be detailed in Chapter 5.

The shared neural network hypothesis

In the context of neurobiological research on the neural correlates of motor action execution and of motor action observation, a set of neurons that fires during execution as well as during observation of goal-directed motor actions such as grasping, manipulating, or holding objects was discovered and referred to as **mirror neurons** by Rizzolatti et al.

[97] [96].

Subsequently, it has been suggested that the observation of an action causes the internal automatic simulation or reproduction of the same action in the observer and that this mechanism could be at the basis of action recognition and understanding [43]. In line with this hypothesis, neurobiological studies show that the activation of the human mirror neuron system is crucial for recognizing and understanding others' actions, and that it is also crucial to imitation and to intention understanding (see [44], [43], and [96] for a review and discussion).

Further neurobiological studies show that another similar 'mirror' mechanism underlies **emotion understanding**, and is thought to be fundamental to empathy [35]. That is, observing someone's emotional state activates brain areas involved in experiencing that same emotional state. Wicker et al. [118] found that exposure to disgusting stimuli (disgusting odorants) activated areas in the human brain responsible for viscero-motor responses, and that observing facial expressions of disgust activated the same brain areas. Botvinick et al. [10] found that exposure to a painful stimuli (thermal stimulation) engaged the human viscero-motor brain areas, as did observing faces expressing pain. Carr et al. [24] noticed an activation of overlapping brain areas responsible for viscero-motor responses both while imitating and while observing facial expressions of emotion (see [44], [96], and [43] for a review and discussion of these and of further studies).

Taken together, the results of neurobiological studies investigating humans' 'mirror mechanisms' suggest the existence of a *shared neural network* for the representation of actions and emotions. Consequently, humans are assumed to possess a **shared representational system** critical to understanding others' actions and emotions. Accordingly, Gallese [43] introduces **embodied simulation** as a functional mechanism at the basis of understanding others. It consists of internal and automatic simulative processes using a 'pre-existing body-model in the brain' (Gallese [43], p. 42), making embodied simulation similar to the empathy arousing mode mimicry (cf. [55]).

The empathic brain

In their attempt to narrow the definition of empathy, De Vignemont & Singer [35] state that there is empathy if:

> (i) one is in an affective state; (ii) this state is isomorphic to another person's affective state; (iii) this state is elicited by the observation or imag-

> ination of another person's affective state; (iv) one knows that the other person is the source of one's own affective state. (De Vignemont & Singer [35], p. 435)

Accordingly, they follow the definition of empathy as an affective response to the other's emotion (cf. Section 2.1.1). De Vignemont & Singer [35] claim that an empathic response is automatically elicited by the observation or imagination of another person's affective state based on the activation of a shared neural network. They point out that automatically empathizing with others when exposed to their emotions would result in being in a 'permanent emotional turmoil' (De Vignemont & Singer [35], p. 436). Consequently, they propose several **modulation factors** that may influence the amplitude of empathic brain responses. They also discuss neuropsychological evidence for the 'modulation of the empathic brain' (De Vignemont & Singer [35], p. 437).

Saarela et al. [103] found stronger brain responses when participants empathized with people displaying acute pain than when they empathized with people displaying chronic pain. Singer et al. [110] showed that men but not women had significantly weaker empathic brain responses when the observed person in pain was judged as unfair as when the observed person was judged as fair and likable. Lamm et al. [66] found that participants' empathic brain responses were reduced when they were told that painful treatment was beneficial. Gu & Han [49] showed that participants' empathic brain responses to pictures or cartoons of hands in painful situations were stronger when they focused their attention on the intensity of the pain than when they were asked to simply count the number of hands in the pictures or cartoons. Cheng et al. [25] found that participants who practice acupuncture had a reduced empathic brain response to pictures of needles penetrating different body parts as compared to naive participants (see [35] and [52], for a review and discussion of these and of further studies).

Based on these findings, De Vignemont & Singer [35] distinguish **four main categories of modulation factors**. These are *features of observed emotion*, *relationship between empathizer and observed other*, *situational context*, and *features of the empathizer*. As *features of observed emotion*, they propose valence, intensity, saliency, and primary versus secondary emotions. They state that it is may be easier to empathize with primary emotions such as *fear* and *happiness* than with secondary emotions such as *jealousy.* As *relationship between empathizer and observed other* they propose affective link, familiarity and similarity, and communicative intentions as one's intentions to communicate the desire to empathize. As *situational context* they propose appraisal of the situation and display of multiple emotions. They state that it becomes difficult to concurrently empathize with different others displaying different emotions. Finally, as *features of the empathizer* they propose mood, level of arousal, personality, gender, and age, emotional repertoire, and emotional regulation capacities.

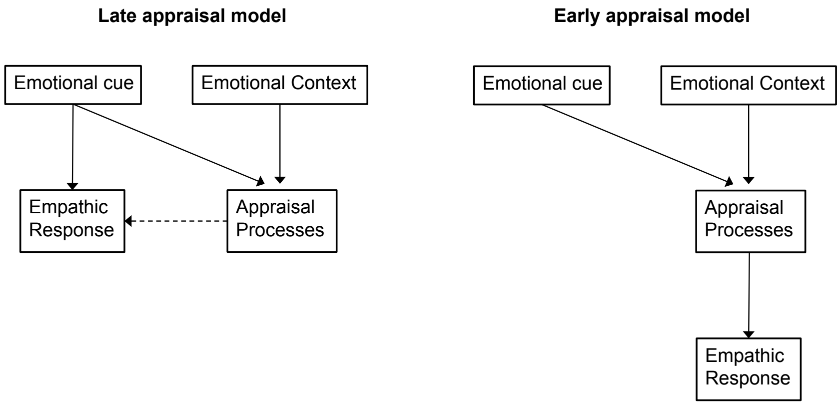


Figure 2.2: Late and early appraisal models [35], p. 438.

According to De Vignemont & Singer [35], an important question is at what stage modulation occurs during empathic processing. They propose two possible routes, and state that current neuropsychological research cannot yet clearly distinguish between them. First, in the **late appraisal model of empathy** (see Figure 2.2, left), the perception of an emotional cue directly and automatically activates an empathic response. Subsequently, the empathic response is modulated or inhibited through different modulation factors. Second, in the **early appraisal model of empathy** (see Figure 2.2, right), the perception of an emotional cue does not directly and automatically activate an empathic response. The emotional cue is first processed and evaluated in the context of different modulation factors. Whether an empathic response arises depends on the outcome of the evaluation.

Conclusion

Findings of neuropsychological research further substantiate and refine psychological theories on empathy. For example, the shared neural network hypothesis is crucial for the empathy arousing mode mimicry as defined by Hoffman [55]. Further support is also provided for the existence of empathy bias and for empathy antecedent conditions, as respectively defined by Hoffman [55] and Davis [34]. These are very similar to the empathy modulation factors defined by De Vignemont & Singer [35].

2.1.4 Conclusion and working definitions

Both Hoffman [55] and Davis [34] provide comprehensive and detailed models of empathy where several aspects of empathy and different perspectives on empathy are unified. As compared to Davis [34], Hoffman provides a more precise definition of empathy. Accordingly, in the present work, we follow **Hoffman's definition of empathy** as '[..] an affective response more appropriate to another's situation than one's own'. (Hoffman [55], p. 4)

Further, after consideration of the theoretical models of empathy introduced in the previous sections, we identify **three processes** that are central to human empathy and that have been addressed in most of the presented theories. First, the **Empathy Mechanism** which we define as the process by which an empathic emotion is generated. In this regard, Hoffman [55] proposes five empathy arousing modes, Davis [34] refers to Hoffman's modes as empathic processes, and De Vignemont & Singer [35] follow the shared neural network hypothesis. Second, the **Empathy Modulation** which we define as the process by which an empathic emotion is modulated. In this regard, Hoffman [55] defines factors influencing observers' empathy bias, Davis [34] defines empathy antecedent conditions, and De Vignemont & Singer [35] define several modulation factors. Third, the **Expression of Empathy** which we define as the process by which an empathic emotion is expressed. In this regard, Hoffman [55] and Davis [34], emphasizes that empathy manifests through prosocial behaviors such as helping, caring, justice, and mimicry. Davis [34] refers to such behaviors as interpersonal outcomes. These three identified processes of empathy are crucial in shaping the computational model of empathy proposed in the present thesis (see Chapter 5).

2.2 Empathy and theoretical models of emotion

In psychological research on emotion, there is no consensus on a general definition of emotion and thus no explicit general answer to the question 'what is emotion?'. Consequently, different theories analyzing and determining the nature of emotion emerged. In this section, the implication of theoretical models of emotion in further explaining the processes underlying empathy is emphasized. In this regard, theoretical models of emotion also provide further support for the computational modeling of empathy.

According to Hoffman [55], mimicry is defined as composed of two subsequent steps which he calls *imitation* and *feedback* (cf. Section 2.1.2). Regarding the second step in his definition of mimicry, Hoffman refers to the feedback theories and discusses supportive evidence provided by evaluation of the **F**acial **F**eedback **H**ypothesis (FFH). In the following the feedback theories are presented.

Feedback theories

The feedback theories are based on the thesis that bodily changes, e.g., changes in the functioning of glands and muscles, and changes in bodily expressions, follow directly from the perception of an exciting event, and that the feeling of these changes is the felt emotion. This thesis was introduced by James [58] who illustrated it in the following statement:

> [W]e feel sorry because we cry, angry because we strike, afraid because we tremble and not that we cry, strike, or tremble because we are sorry, angry, or fearful. (James [58], p. 190)

James substantiated his thesis by drawing on Darwin's studies [32] on emotion expression, where changes in the functioning of glands and muscles, and changes in the circulatory apparatus were discovered. In the 1960s, James' theory was reconsidered in the context of empirical studies. This resulted in the emergence of refined versions of the theory known as the **neo-Jamesian theories** of emotion.

The idea that bodily expressions influence emotion can be traced beyond James to Darwin's older work [32]. Darwin noticed that either the expression or the repression of the external signs of emotion, respectively, intensified or softened the intensity of that emotion. Further, Darwin focused on a more detailed and scientific description of the meaning of different facial expressions, as well as on the facial muscles accompanying them. He also underlined the specific and functional role of facial actions in expressing and communicating emotion. Inspired by Darwin's work, most of the neo-Jamesian theories attribute a primary role to facial expressions in the study of emotion, and hence investigate the assumption that facial expressions influence emotion. This assumption is commonly called **F**acial **F**eedback **H**ypothesis (FFH) [75]. In an elaborate review article, McIntosch [75] discusses research on the empirical evaluation of the FFH. The results show that facial actions modulate the intensity of emotion [114] and that they can initiate emotions [71]. These findings provide supportive evidence for the role of bodily changes in emotion elicitation, such as the role of facial actions in initiating and modulating emotion.

Further, McIntosch [75] refers to the possibility that representations of facial expressions in the central nervous system can produce emotion without the need to perform the related facial motion. A similar suggestion had already been made by Darwin who argued that 'Even a simulation of an emotion tends to arouse it in our minds.' (Darwin [32], p. 365) This suggestion is supported by findings provided in the context of evaluations of the **shared neural network hypothesis** which suggests the existence of a shared neural network for the representation and understanding of actions and emotions (cf. Section 2.1.3). As such it provides a refined explanation of feedback as the second step in Hoffman's definition of mimicry. Both mimicry as defined by Hoffman [55], and the shared neural network hypothesis, are crucial to the generation of an empathic emotion in our model (see Section 5.2.1).

According to Omdahl [87], while contemporary theorists such as Hoffman [55] and Davis [34] provide comprehensive models of empathy (cf. Section 2.1), they do not clearly explain the cognitive processing steps by which emotional states are decoded within their models. For example, in Hoffman's [55] empathy arousing mode role-taking, it is not clear what exact mechanisms allow an observer to decode the emotional state of the other. Therefore, Omdahl [87] claims that emotion theories such as **cognitive appraisal theories** may provide an explanation for how the other's emotional state is decoded. She took Scherer's cognitive appraisal theory of emotion [104] as an example, and empirically examined how Scherer's rules of cognitive appraisal of emotion explain emotion elicitation through role-taking. Role-taking is crucial to the generation of an empathic emotion in our computational model of empathy as based on the appraisal component of an existing computational model of emotion [8] (see Section 5.2.2). Thus, in the following, two example appraisal theories are presented.

Appraisal theories

Appraisal theories are based on the thesis that emotions are elicited by a continuous subjective evaluation of perceived stimuli against a number of criteria. These are criteria for the satisfaction of goals, motives, likes, dislikes, norms, and values. In the following, Scherer's appraisal theory [104] and Ortony, Clore, and Collins' (OCC) appraisal theory [88] are introduced.

The component process model In his **C**omponent **P**rocess **M**odel of emotion (CPM) [104], Scherer defines an appraisal module as one among three functionally defined modules. Only the appraisal module is presented here (see [104] for a description of the other modules). The appraisal module determines whether an emotion will be elicited. Scherer defines four necessary **appraisal objectives** required to adaptively react to a stimulus event. These are **(1)** determination of the relevance of the event; **(2)** determination of the implications or consequences of the event; **(3)** determination of the ability to cope with the event; **(4)** determination of the normative significance of the event. These appraisal objectives are pursued based on a subjective evaluation of the event using a number of appraisal criteria called **S**timulus **E**valuation **C**hecks (SECs). They represent a minimal set of criteria that differentiate emotions from one another.

To determine the relevance of a stimulus event, three SECs are required. A check for *novelty*, a check for *intrinsic pleasantness*, and a check for *relevance to goals and needs*. Further, to determine the implications or consequences of a stimulus event, five SECs are needed. A check for *cause*, a check for *probable outcomes*, a check for *failure to meet expectations*, a check for *conduciveness to goals and needs*, and a check for *urgency*, i.e. how urgently is an adaptive response to the event. To determine the ability to cope with a stimulus event, three SECs are required. A check for *control* as the extent to which one can influence or has control over the event, a check for *power* in terms of the ability to change the event's outcomes in line with one's goals and motives, and a check for *potential for adjustment* in terms of one's ability to adapt and live with the outcome of the event. To determine the normative significance of a stimulus event, two SECs are required. A check with *external standards* as the extent to which an event is compatible with one's values and rules as a member of a social group, and a check for *internal standards* as the extent to which an event is compatible with one's personal self-ideal and moral code.

The cognitive structure of emotions **O**rtony, **C**lore, and **C**ollins (OCC) [88] focus on the analysis of the cognitive aspects underlying the appraisal function of emotion, and propose a cognitive theory for its causal origins. In their theory, they address the question of what differentiates emotions from one another in terms of the cognitive structure underlying appraisal. They call their theory *The Cognitive Structure of Emotions*. The authors propose to arrange **22 emotion types** into six representative groups, which in turn are classified into three basic classes. These are emotions as valenced reaction to consequences of events, to actions of agents, and to aspects of objects.

Regarding emotions as valenced reactions to consequences of events, OCC define, among others, **fortunes-of-other emotions** as emotions that are elicited through appraisal of an event as desirable or not desirable for another person. They refer to the fortunes-of-other emotions also as empathetic emotions, and define four types: *happy-for*, *pity*, *resentment*, and *gloating*. *Happy-for* or *resentment* are elicited when the event is appraised as desirable for the other while *gloating* or *pity* are elicited when the event is appraised as not desirable for the other.

OCC define three types of variables to determine the **intensity of emotion types**. Central intensity variables are those that affect the intensity of all emotion types in a basic class of emotions. Global intensity variables are those that affect the intensity of all emotion types in all basic classes of emotions. Finally, local intensity variables are those that affect the intensity of emotion types in particular representative groups, and are thus specific to some emotion types but not to others. The appraisal of emotion inducing situations is based on evaluating the three central intensity variables, namely, *desirability*, *praiseworthiness*, and *appealingness*, relative to one's goals, standards, and attitudes.

For the fortunes-of-other emotions, OCC define three local variables that affect the intensity of these emotions, *desirability-for-other*, *deservingness*, and *liking*. The local variable *desirability-for-other* reflects the degree to which an event is evaluated as desirable or undesirable for the other person. The local variable ***deservingness*** reflects the degree to which one evaluates that the other person deserved or did not deserved the event. The local variable ***liking*** reflects the degree to which one likes or dislikes the other person. The higher the values of these variables, the higher the intensity of the fortunes-of-other emotions.

The more one believes that the event is desirable for the other person, the higher the intensities of the emotion types *happy-for* or *resentment*. The more one likes the other person and believes that the other deserves the desirable event, the higher the intensity of the emotion type *happy-for*. The more one dislikes the other person and believes that the other does not deserve the desirable event, the higher the intensity of the emotion type *resentment*. The more one believes the event is undesirable for the other person, the higher the intensities of the emotion types *gloating* or *pity*. The more one dislikes the other person and believes that the other deserves the undesirable event, the higher the intensity of the emotion type *gloating*. The more one likes the other person and believes that the other does not deserve the undesirable event, the higher the intensity of the emotion type *pity*.

Compared to existing emotion theories, OCC's theory is the only one that provides an explicit definition of empathic emotions and their corresponding intensity variables. The proposed variables influencing the intensity of the fortunes-of-other emotions are very similar to the empathy bias, to the empathy antecedent conditions, and to the empathy modulation factors as defined respectively by Hoffman [55], Davis [34], and De Vignemont & Singer [35]. However, OCC provide a more detailed explanation of how their defined factors influence one's empathy for others. Furthermore, they also consider two types of fortunes-of-other emotions that can be referred to as 'negative empathic emotions', namely *gloating* and *resentment*. Further, Scherer and OCC propose to validate their theories through their application in an **A**rtificial **I**ntelligence (AI) system. Thus, appraisal theories of emotion provide a suitable and clear explanation of the cognitive processing steps underlying role-taking.

So far, the temporal development of an empathic emotion has not been explicitly addressed within theoretical models of empathy, e.g., the temporal decay of an empathic emotion and the impact of several factors such as the empathizer's changing mood over time on modulating the empathic emotion (cf. [35], see Section 2.1.3) . Accordingly, for the simulation of the **time course of an empathic emotion**, we based our model on an existing computational model of emotion [8] (see Section 3.1.1) where emotions, mood, and their mutual interaction over time are simulated within a dimensional emotion model. Furthermore, primary and secondary emotions are defined and distinguished within the model. As primary emotions, a set of emotion categories similar to Ekman's set of basic emotion categories [38] is defined. In the following, dimensional emotion theories, basic emotion theories, and Damasios' primary and secondary emotions [31] are presented.

Dimensional theories

Dimensional theories are based on the thesis that emotions are variations along **basic dimensions** that represent fundamental elements of emotions. This thesis claims that emotions are related to one another in a systematic manner, and that their relationships can be represented in a dimensional model.

Scholsberg [105] proposes three basic dimensions along which emotions vary in their intensity and quality. These three basic dimensions are **level of activation**, **pleasantness-unpleasantness**, and **attention-rejection**. The level of activation dimension represents variations in the intensity of emotion. The pleasantness-unpleasantness, and attention-rejection dimensions represent variations in the quality of emotion. Based on ratings of pictures of emotional facial expressions, Scholsberg provides evidence for the three dimensions. He found that facial expressions can be arranged around a circular surface spanned by the pleasantness-unpleasantness and attention-rejection dimensions, and that they can also be arranged with respect to the third dimension of level of activation, thus forming a cone-like three-dimensional emotion space.

As Scholsberg, Russell & Mehrabian [102] propose a three-dimensional emotion model. They postulate three independent and bipolar dimensions of **pleasure-displeasure**, **arousal-nonarousal**, and **dominance-submissiveness**. The pleasure-displeasure and arousal-nonarousal dimensions are similar to the pleasantness-unpleasantness and level of activation dimensions proposed by Scholsberg. However, Russell & Mehrabian define arousal-nonarousal as a second dimension, and introduce a new and third dimension of dominance-submissiveness as the degree to which one feels in control of a situation. They argue that the emotional states of *anger* and *anxiety* have similar levels of displeasure and arousal, and differ only in their levels of dominance [101]. To test this hypothesis, subjects were asked to report their feelings after reading verbal descriptions of situations expected to make them angry or anxious, and imagining themselves in these situations. The results show a positive amount of dominance for *anger*, a negative amount of dominance for *anxiety*, a negative amount of pleasure and a positive amount of arousal for both *anger* and *anxiety*. Further evidence for the three dimensions was provided by Russell & Mehrabian in the context of two other empirical studies (see [102]).

In a further investigation of a dimensional emotion model, Russell [99] argues for the two dimensions of pleasure and arousal, and against the consideration of a third dimension. In his exploration of a third dimension, he found dimensions that were interpreted in terms of antecedents or consequences of an emotion, and thus do not refer to the emotion itself. Based on the results of studies of self-reports of emotion, and of judgments of emotion words, Russell provides evidence for a two-dimensional circular model of emotion, called the **Circumplex Model of Affect**, as proposed earlier by Scholsberg. Compared to the dimensional emotion model of Scholsberg, Russell defines activation (arousal) as a second dimension instead of attention-rejection.

The dimensional theories define basic dimensions along which emotions vary with respect to their fundamental elements. However, there is no agreement on the number and interpretation of these basic dimensions although there is more evidence for a first and second dimension of pleasure and arousal, and less for a third dimension.

Basic emotion theories

Basic emotion theories are based on the thesis that there is a predefined set of emotion categories that are basic, hard-wired in the brain, and universally recognized. This thesis was pioneered by Darwin [32] who observed that the expressions of some emotions are universal among cultures, and are displayed by people born blind. He concluded that the expressions of these emotions must be innate and part of our evolutionary endowment.

Research on universals in facial expression of emotion provides supportive evidence for a universally recognized set of emotion categories. Ekman [38] discusses the empirical data of seven cross-cultural studies on universals in facial expression of emotion. The results show a significant degree of agreement across different cultures in the recognition of the **six basic emotions**: *happy*, *anger*, *fear*, *sadness*, *disgust*, and *surprise*. As compared to the dimensional theories, basic emotion theories provide only a limited set of emotion categories.

Damasio's primary and secondary emotions

Damasio [31] defines **primary emotions** as emotions that are experienced early in life as preorganized responses to specific features of perceived stimuli, e.g., its size and motion. He argued that these features activate the amygdala, a part of the brain crucial to emotional processing, which, based on innate dispositional representations, triggers bodily changes specific to the emotion, and influences cognitive processing in line with that emotion. In the next developmental step, the connections between perceived stimuli and emotion specific bodily changes become conscious, allowing for flexible behaviors in given situations. For example, knowing that a bear causes the emotion fear allows for avoidance of this animal in a given environment. However, the mechanism for primary emotions does not cover the full space of emotional behavior, and is thus followed by another developmental step encompassing the mechanism of secondary emotions.

Secondary emotions are experienced by adults as responses to given situations based on a cognitive evaluation. Damasio [31] describes the process in three steps. First, the process begins with a cognitive evaluation of the situation through the deliberate consideration of the situation in the form of mental images. Second, the processing of the mental images leads to a neural response in the prefrontal cortex that arises based on acquired dispositional representations of the connections between certain situations and emotional responses. As described above, these dispositional representations are acquired based on the innate dispositional representations needed for primary emotions. Third, the neural response of the prefrontal cortex activates the amygdala, which, based on innate dispositional representations, triggers bodily changes specific to the emotion and influences cognitive processing in line with the emotion. Hence, secondary emotions are expressed using the mechanism of primary emotions.

Conclusion

In sum, as elucidated by this section, theoretical models of emotion are suitable to clarifying and explaining several processes underlying empathy, and are thus crucial to the computational modeling of several theoretical aspects of empathy. In the next section, we present theories and findings on emotion expression and recognition that also have crucial implications for the realization of empathy.

2.3 Emotion expression and perception

Theories and findings on emotion expression and perception provide further support for a deeper understanding of several processes underlying empathy, such as the generation and expression of an empathic emotion. In this section, we present research investigating the expression and perception of emotion.

Humans communicate their emotions through a wide range of multimodal cues. For example, facial expressions, speech prosody and content, body postures and movements as **outward expressions**, and skin conductance, muscle tension activity, and heart rate as **inward expressions**. To analyze and access the way emotions manifest through outward and inward expressions, several methods are used and developed (see [50] for a review). As illustrated in Section 2.2, there are different theories on emotion. Accordingly, research on emotion expression and perception is based on these theories. Since facial expressions are central to the present thesis, our scope will be limited to this literature.

In their research on universals in facial expressions of emotion (cf. Section 2.2), Ekman et al. [40] introduce the **F**acial **A**ction **C**oding **S**ystem (FACS) as a comprehensive system to describe all possible and visually distinguishable facial movements. Since every facial movement emerges from contracting different facial muscles, Ekman et al. [40] argue that a comprehensive expression coding system could be obtained by discovering the impact of muscle contraction on changing the visible appearance of the face. Thus, based on an anatomical analysis of facial muscle actions, they define **A**ction **U**nits (AUs) as the core elements of FACS describing the contraction of single or multiple facial muscles as well as eye and head movements (see Table A.1, p. 204). In order to code the time course of facial movements, Ekman et al. [40] define a method to score the intensity of AUs using a five point scale. Further, to code and analyze the facial expressions accompanying basic emotions, they define rules to link AUs to the six basic emotion categories (see Table A.2, p. 205).

According to Russell [100], facial expressions do not signal specific emotions such as basic emotions. Instead, they manifest primary information that can be automatically, easily, and universally understood. Russell [100] distinguishes two kinds of information, first, quasi-physical information such as muscle contractions and visual attention, and second, the overall level of pleasure (pleased vs. displeased) and arousal (agitated vs. sleepy). Researchers such as Scholsberg [105] posit that facial expressions convey primary information such as pleasure and arousal (activation) (see Section 2.2). Russell [100] argues that the combination of this primary information with a context allows for the attribution of a specific emotion. Further, he states that dimensions are primary, elemental, and universal while emotion categories are complex, derived, and variable with language and culture. Moreover, he emphasizes that spontaneous facial expressions displayed in daily settings are relatively mild and subtle as compared to the rather extreme configurations introduced by Ekman et al. [40].

While most of the researchers followed a **top-down approach** in that they investigated the emotional meaning of whole facial expressions, and then decomposing them into components (e.g., [40]), others followed a **bottom-up approach** in that they investigate the emotional meaning of individual facial actions and of their possible combinations (e.g., [111]). Thus, by investigating the meaning of single facial actions, it is possible to investigate the underlying properties of a whole facial expression and of the expressed emotion. For example, the presence of the same facial action in two different facial expressions of emotion assigns a shared affective meaning to both expressions (cf. [100]), e.g., raised eyebrows convey high arousal for both fear and surprise. Moreover, the intermixing of single facial actions would result in a wider range of emotional expressions rather than being limited to predefined expressions of emotion categories. Snodgrass [111] followed a bottom-up approach by combining FACS with the dimensional emotion space of pleasure and arousal [100]. Single AUs and AUs combinations were rated in terms of pleasure and arousal values in a first study, while in a second study they were rated in terms of emotion categories. Some representative results of the study by Snodgrass [111] are summarized in Figure 2.3 (left).

AU	Description	P	A
1	Inner brow raised	-1.92	-.85
2	Outer brow raised	1.85	2.11
4	Brow furrowed	-2.00	1.23
5	Upper eyelid raised	-1.19	3.04
6	Cheek raised	1.46	.73
7	Lower eyelid raised	-1.85	.58
9	Nose wrinkled	-2.33	2.33
10	Upper lip raised	-3.67	1.97
15	Lip corner depressed	-2.26	-.63
17	Chin raised	-.93	.60
1 + 2	Whole brow raised	.85	2.85
1 + 4	Inner brow raised and furrowed	-2.58	.12
1 + 2 + 4	Whole brow raised and furrowed	-1.39	.61
1 + 2 + 5	Whole brow raised + upper eyelid raised	.27	3.58
4 + 5	Brow furrowed + upper eyelid raised	-.66	1.16
5 + 7	Upper and lower eyelids raised	-2.04	2.73

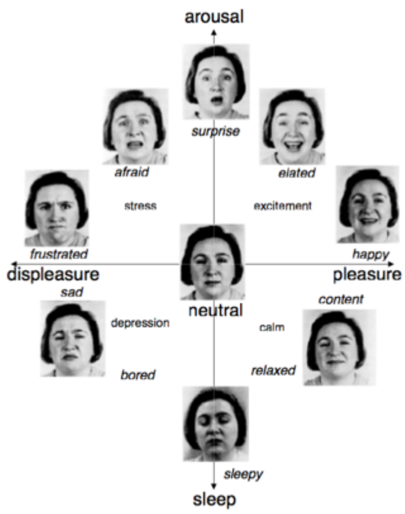


Figure 2.3: Left: some representative results of the study by Snodgrass [111]. Note that the AUs were presented to participants at their maximal intensity. **P** and **A**, respectively, denote pleasure and arousal. Right: facial expressions reconstructed in pleasure-arousal space by Russell [100] (figure taken from [20], p. 126).

The results show significant agreement within subjects and allow the association of emotion categories with values of pleasure and arousal. These findings support Russell's hypothesis [100] that the information signaled by a facial expression is present in its single components. Russell [100] illustrates Snodgrass's results [111] by placing the AUs within his dimensional emotion space of pleasure and arousal thus combining them into eight facial expressions (see Figure 2.3, right).

The results provided by Snodgrass's study and further discussed by Russell are valuable in supporting a dimensional approach to the facial expression of emotion. As compared to the basic emotion approach, which limits the set of facial expression of emotion, the dimensional approach allows for a wider range of expressions. Furthermore, a bottom-up approach allows for the combination of different facial actions, and their emotional meaning, to reconstruct new facial expressions. These findings are crucial for the generation and expression of an empathic emotion in our computational model of empathy (see Sections 4.3, 5.2.1, and 5.4).

2.4 Summary

In this chapter, we discussed the theoretical background relevant to understand the theoretical concepts underlying our computational model of empathy. In Section 2.1, we began by introducing the debate on a universal definition of empathy and the resulting three major categories of **definitions of empathy**. Further, we introduced Hoffman's [55] and Davis' [34] psychological models of empathy which offer comprehensive and detailed models of empathy, and as such unify the different views and perspectives on empathy. Within these models, different aspects of empathy are considered and discussed. For example, the **mechanisms eliciting empathy** as divided into lower-level cognitive mechanisms, such as mimicry, and higher-level cognitive mechanisms, such as role-taking, the emergence of **different degrees of empathy** based on several factors, and the motivation of prosocial and moral behaviors by empathy. While such comprehensive and detailed models of empathy are crucial to the computational modeling of empathy, they do not provide an explanation of the exact mechanisms underlying the aspects of empathy in question. For example, questions remain about the mechanisms required to decode and recognize others' emotional states within the empathy arousing modes as defined by Hoffman and the empathic processes as defined by Davis.

Neuropsychological models of empathy investigate the neural mechanisms underlying empathy thus offering further support for psychological models of empathy, and allowing for a more precise consideration of the processes underlying empathy. In this regard, we introduced the shared neural network hypothesis that suggests the existence of shared neural networks for understanding actions and emotions, and which is crucial for the empathy arousing mode mimicry as defined by Hoffman [55]. Further, we presented De Vignemont & Singer's [35] model of empathy as based on the shared neural network hypothesis and which also introduces different modulation factors of empathy leading to the emergence of different degrees of empathy. According to the discussed models of empathy, we introduced **Hoffman's definition of empathy** as our working definition of empathy and we identified **three central processes of empathy** key to our model (see Chapter 5). These processes are the *Empathy Mechanism* which we define as the process by which an empathic emotion is generated, the *Empathy Modulation* which we define as the process by which an empathic emotion is modulated, and the *Expression of Empathy* which we define as the process by which an empathic emotion is expressed.

Further in Sections 2.2 and 2.3, respectively, we elucidated the implications of different theoretical models of emotion and of theories and findings on emotion expression and perception in providing deeper explanations for several processes underlying empathy. For example, the feedback theories provide support for Hoffman's definition of mimicry, while the appraisal theories provide a suitable explanation for the cognitive processing steps underlying role-taking. Both mimicry and role-taking are crucial for the generation of an empathic emotion in our model (see Sections 5.2.1 and 5.2.2). An important aspect that was not explicitly addressed within theoretical models of empathy is the **temporal development of an empathic emotion** and its dynamic interaction with factors such as the empathizer's changing mood over time (cf. [35]). This issue is addressed in the proposed computational model of empathy based on an existing computational model of emotion [8] (see Section 3.1.1). Accordingly, we introduced the emotion theories shaping the model and crucial to the simulation of emotion dynamics within the model, such as dimensional emotion theories.

Findings within research on emotion expression and perception provide support for a dimensional approach to the facial expression of emotion which allows for the expression and perception of a wider range of emotions, as compared to the basic emotion approach. These findings are crucial to the generation and expression of an empathic emotion within the computational model of empathy proposed in the present thesis (see Sections 4.3, 5.2.1, and 5.4).

3 Related work

As emphasized in the previous chapter, besides theoretical models of empathy, theoretical models of emotion and theories and findings on emotion expression and perception are crucial to the computational modeling of empathy. Section 3.1 gives an overview of different approaches to the computational modeling of emotion as well as their application and evaluation in virtual human and robot scenarios. Subsequently, in Section 3.2, we present approaches to emotion expression simulation, and methods for emotion recognition. In Section 3.3, we introduce different computational models of empathy as well as their application and evaluation in virtual human and robot scenarios. This section concludes with the requirements formulated in Chapter 1 and with classifying the introduced empathy models with respect to these requirements.

3.1 Emotion in virtual humans and robots

In this section, we give an overview on different computational models of emotion and on their application and evaluation in virtual human and robot scenarios. In this regard, we present the computational model of emotion, WASABI [8], as well as its application and evaluation in scenarios involving the virtual human MAX [62]. This computational model of emotion underlies the computational model of empathy we propose (see Chapter 5). Furthermore, the virtual human MAX is also involved in the application and evaluation of our empathy model (see Chapter 6). We will therefore describe WASABI and the virtual human MAX in more details than other presented works.

3.1.1 Computational models of emotion

Marsella et al. [73] summarized the history of research on computational models of emotion over the last 15 years, based on several example models, and their theoretical foundations (see Figure 3.1). As depicted in Figure 3.1, **appraisal theories** influenced mostly the computational modeling of emotion. As introduced in Section 2.2, appraisal

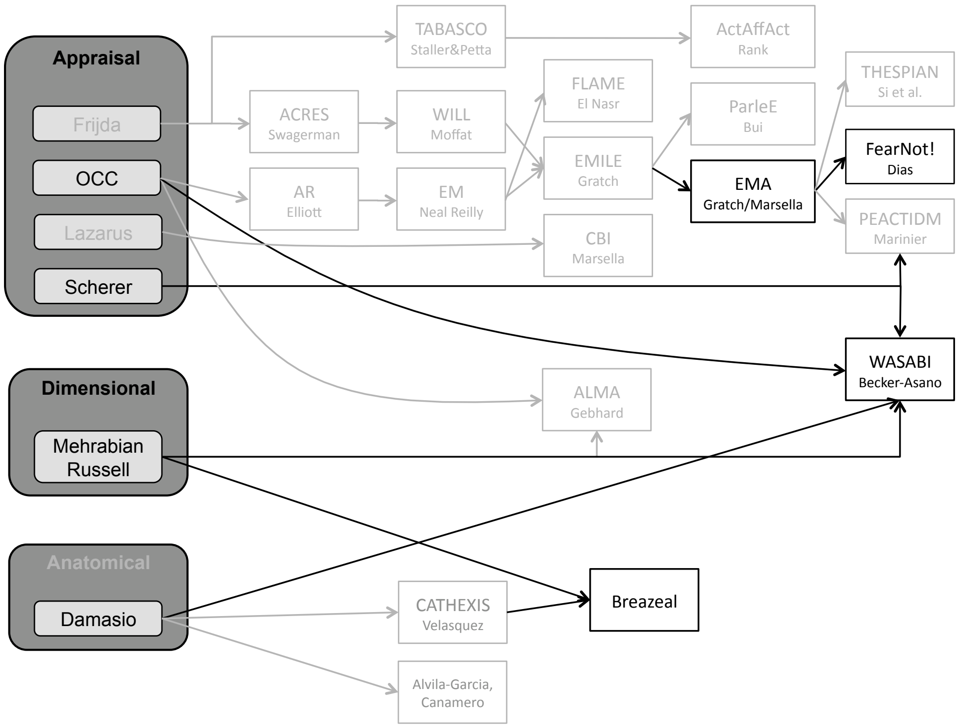


Figure 3.1: The history of computational models of emotion. Figure adapted from [73], p. 22. The highlighted models are presented in this Section.

theories provide explicit and clear models of the appraisal mechanisms that elicit emotions (appraisal criteria) and are thus believed to be more suitable to the computational modeling of emotion than other theories. In this regard, OCC [88] refer to their appraisal theory of emotion (see Section 2.2) as a 'computationally tractable model of emotion' (OCC [88], p. 181). Consequently, several computational models follow OCC's [88] appraisal theory. One of the first computational models based on their theory is the **A**ffective **R**easoner (AR) [42] (see Figure 3.1). In the following, some example models are presented to illustrate how the computational modeling of emotion is influenced by the different emotion theories, and how these models differ from each other. For a detailed review of existing computational models of emotion see [8].

EMA: EMotion and Adaptation

EMA (**EM**otion and **A**daptation) [47] is a domain independent computation framework for modeling the antecedents and consequences of emotions. EMA is based on domain independent decision theoretic plans and a **B**elief-**D**esire-**I**ntention (BDI) approach. To appraise the emotional significance of ongoing events, EMA uses a set of appraisal variables such as *desirability* and *likelihood*. The values of the appraisal variables are derived via **domain independent** inference rules over features of the agent's decision theoretic plans, and are organized within multiple appraisal frames. Each appraisal frame represents the appraisal of an event from the agent's perspective or from other agents' perspectives. This allows the agent to reason about how events or their own actions impact other agents' emotions. Each appraisal frame is mapped into an emotion with a specific type and intensity using rules for emotion eliciting conditions. The use of appraisal frames and the appraisal of a situation from the perspectives of other agents are notions taken from the AR [42]. Six emotion types are considered in EMA, *hope*, *fear*, *joy*, *distress*, *anger*, and *guilt*. Generated emotions are aggregated into a current emotional state and an overall mood state. Given the case that different emotions with equal intensities are generated (e.g., *fear* and *hope*), the overall mood state biases the intensities of these emotions thus increasing the intensity of the mood congruent emotion.

In conclusion, EMA provides a powerful and comprehensive model of domain independent 'plan-based appraisal' that explains the dynamics of appraisal over time. In the following, we present a computational model of emotion that builds upon the notions adopted by the EMA framework (see Figure 3.1, FearNot! [36]).

A computational model for emotional characters

Dias & Paiva [36] propose a computational model of emotion where the appraisal component comprises of a reactive layer and a deliberative layer. The appraisal component is based on the appraisal theory of OCC [88] (see Section 2.2) and its deliberative layer builds upon the notion of domain independent 'plan-based appraisal' adopted in the above presented EMA framework. The reactive layer is based on a set of domain dependent emotional reaction rules. Every generated emotion has a type, an intensity value, and a positive or negative value of valence. The intensity of an emotion is influenced by the character's values of arousal and mood. Mood is represented by a variable whose value increases with the valence of positive emotions and decreases with the valence of negative ones.

While in AR [42] and EMA [47] an agent's personality is related to the agent's goals, standards, and preferences. In Dias & Paiva 's model, further **personality related factors** are included, e.g., different activation thresholds for each of the 22 emotion types. For example, a character who is resistant toward the emotion *anger* has a high activation threshold for this emotion. Furthermore, Dias & Paiva's model integrates a generic computational model of empathy as will be presented in Section 3.3.1.

Breazeal's computational model of emotion

Breazeal [20] proposes a computational model of emotion that is based on a three-dimensional emotion space (cf. Section 2.2) of **A**rousal, **V**alence, and **S**tance (AVS) (see Figure 3.1). 14 emotion categories are located within AVS space and are surrounded by so called **emotion regions**. Perceived stimuli are appraised in terms of their intensity, relevance, intrinsic affect, and goal directedness, which are mapped to AVS values. Compared to the dominance dimension defined by Russell & Mehrabian [102] (see Section 2.2), the third dimension, stance, represents how approachable a percept to an agent is and makes it to distinguish between the emotion categories *anger* and *fear*. The AVS values are associated to emotion categories and are averaged, thus resulting in a net AVS value for each emotion category. The position of the net AVS value within the emotion region increases the intensity of the emotion category. The emotion category with the highest value of intensity together with the corresponding AVS value represent the agent's current emotional state. Based on a previous work by Velásquez [117], primary and secondary emotions are distinguished following Damasio [31] (see Section 2.2). **Primary emotions** correspond to the elicitation of emotions via innate and hard-wired releasers (emotion eliciting stimuli) while **secondary emotions** correspond to the elicitation of emotions via learned releasers using emotion memories.

Compared to the previously presented models, Breazeal uses a dimensional approach for the computational modeling of emotion that allows for smooth and continuous trajectories of emotion dynamics. In the following, we present WASABI [8], a computational model of emotion where emotions, moods, and their mutual interaction as well as primary and secondary emotions are simulated within a dimensional emotion model (see Figure 3.1).

WASABI: Affect Simulation for Agents with Believable Interactivity

WASABI ((**W**) **A**ffect **S**imulation for **A**gents with **B**elievable **I**nteractivity) [8] is the computational model of emotion that underlies the computational model of empathy developed in the present thesis (see Chapter 5).

WASABI is composed of two interconnected modules, a cognition module and an emotion module. The cognition module is composed of a reactive layer and a reasoning layer where respectively primary and secondary emotions, as distinguished by Damasio [31] (see Section 2.2), are triggered. Within the reactive layer, **primary emotions** are triggered as hard-wired reactions to perceived stimuli while in the reasoning layer, **secondary emotions** are triggered by BDI-based deliberation using memories and expectations. The appraisal processes within both layers result in negative or positive values of emotional valences, also called *emotional impulses*, that drive an agent's emotion dynamics over time.

The emotion module is comprised of two components, a **dynamics/mood** component and a **P**leasure-**A**rousal-**D**ominance (PAD) space. Within the dynamics/mood component, the time course of emotions and moods and their mutual interaction are calculated in an orthogonal space of their respective valence components (see Figure 3.2, left). The influence of emotions on mood is modeled by interpreting the valence of emotion as a gradient with respect to which the valence of mood increases or decreases according to Equation 3.1.

$$\frac{\Delta y}{\Delta x} = a \cdot x \tag{3.1}$$

The values of valences of emotions and moods decay over time based on two independently simulated spring-mass systems for each axis, creating two reset forces F_x and F_y that are proportional to the valences x and y. The spiral springs for each axis are virtually anchored to the origin and attached to the point of reference (see Figure 3.2, left). Furthermore, the concept of boredom is also considered within the dynamics/mood component and is represented by a third orthogonal z-axis. **Boredom** emerges in consequence to the absence of emotional stimuli. Once the point of reference lies in an epsilon neighborhood of absolute zero (ϵ_x and ϵ_y), the degree of boredom starts to increase linearly as per Equation 3.2.

$$z(t+1) = z(t) - b \tag{3.2}$$

Once the point of reference is outside the epsilon region, the value of boredom is reset to zero. The variable a in Equation 3.1, the spring constants d_x and d_y, the inertial mass m of the point of reference (see Figure 3.2, left), and the parameter b can be considered as **personality related factors**. Smaller values of a and greater values of the spring constants and the inertial mass result in a more lethargic agent, while greater values of a and smaller values of the spring constants and the inertial mass result in a more

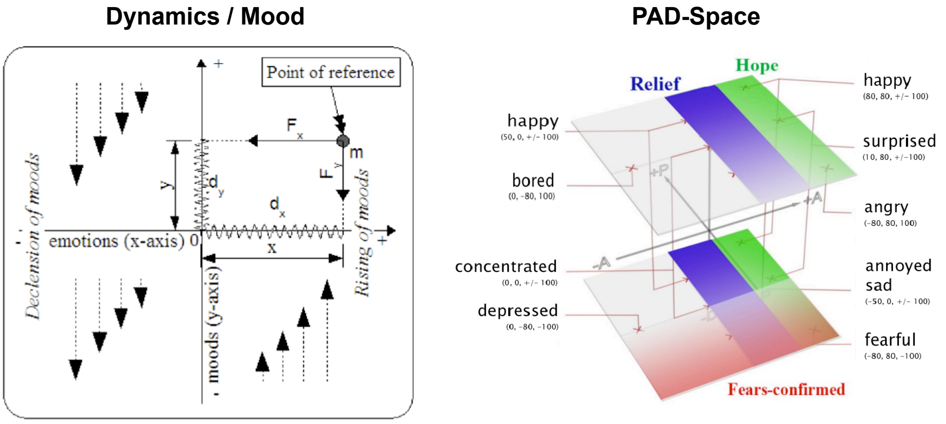


Figure 3.2: Left: the dynamics/mood space with the simulated spring-mass systems for each axis [8], p. 90. Right: the PAD space with primary and secondary emotions [8], p. 94.

temperamental agent. Greater values of b result in an agent that becomes bored easily in the absence of emotional stimuli.

Within PAD space (see Figure 3.2, right), primary and secondary emotions are represented and their awareness likelihood values are calculated. As primary emotions, *anger*, *happy*, *surprised*, *sad*, and *fearful* as five of the six basic emotions proposed by Ekman [38] (see Section 2.2), and the emotions *bored*, *annoyed*, and *depressed* as well as the neutral emotional state *concentrated* are simulated. These primary emotions are located in PAD space according to Russell & Mehrabian [102] (see Section 2.2). As secondary emotion, *relief*, *hope*, and *fears-confirmed* proposed by OCC [88] as prospect-based emotions are simulated. These secondary emotions are represented in PAD space by weighted polygon areas as depicted in Figure 3.2 (right). For more details on the location of primary and secondary emotions within PAD space, see [8].

For the calculation of the awareness likelihood values of primary and secondary emotions, the valence of emotion x, the valence of mood y, and the degree of boredom z provided by the dynamics/mood component over time are mapped into PAD space according to Equation 3.3.

$$
\begin{aligned}
&PAD(x_t, y_t, z_t) = (p(x_t, y_t), a(x_t, z_t), d(t)), \text{ with} \\
&p(x_t, y_t) = \frac{1}{2} \cdot (x_t + y_t) \text{ and } a(x_t, z_t) = |x_t| + z_t
\end{aligned}
\tag{3.3}
$$

The value of dominance $d(t)$ represents an agent's feeling of control over a situation and is derived from the cognition module. Only **two extreme values of dominance** are considered, dominant vs. submissive. The values of dominance allow the distinction between *fear* and *anger* as well as between *sad* and *annoyed* (cf. [102], see Section 2.2).

Each primary emotion within PAD space is surrounded by **two circular regions** representing its activation and saturation thresholds. The closer the point of reference within PAD space gets to the location of a primary emotion, the more likely an agent becomes aware of that emotion. The awareness likelihood value of a primary emotion is maximal within the saturation region and is equal to null outside the activation region. The awareness likelihood values of secondary emotions are calculated based on a linear interpolation within their polygon areas in PAD space once they are triggered by the reasoning layer. For example, the more pleasurable and aroused an agent feels, the more likely it is that he becomes aware of the secondary emotion *relief* once it is triggered by the reasoning layer. The emotion module outputs values of pleasure, arousal, and one of two possible values of dominance, as well as the awareness likelihood values of primary and secondary emotions.

Discussion

Gratch & Marsella [47] and Dias & Paiva [36] focus mainly on emotional appraisal as based on domain independent 'plan-based appraisal' that explains the dynamics of appraisal over time. Compared to this, Breazeal [20] and Becker-Asano [8] primarily focus on emotion as variations along basic dimensions and on the simulation of the **time course of emotions** independent of any elaborate cognitive appraisal (cf. [8]). Furthermore, also **primary and secondary emotions** as defined by Damasio [31] are distinguished in both of these models. However, compared to Breazeal's computational model of emotion, a more elaborate consideration of **cognitive appraisal** is realized within WASABI. In this regard, similarly to Dias & Paiva's model, a reactive and a reasoning/deliberative layer are considered within WASABI. The same as Gratch & Marsella's model, WASABI's reasoning layer is based on a BDI-approach. Furthermore, **personality factors** as considered within WASABI are very similar to those proposed by Dias & Paiva. Thus, within the WASABI computational model of emotion, several aspects of emotion are addressed and unified (cf. Figure 3.1).

3.1.2 Application and evaluation

There are several works on the application and evaluation of computational models of emotion within virtual human and robot scenarios. In the following, some example scenarios are presented just to give a small overview of how emotions are integrated into virtual humans and robots.

The Mission Rehearsal Exercise (MRE) system

The **M**ission **R**ehearsal **E**xercise (MRE) system [59] is a virtual reality training environment to teach decision-making skills in critical situations that can face members of the US Army. The MRE system includes virtual humans acting as friendly and hostile forces capable of task-oriented reasoning, and of natural language and non-verbal communication with the trainees. Figure 3.3 (left) depicts a situation where the unit of a sergeant has a collision with a civilian vehicle injuring a boy. The trainee acting as the leader of the sergeant can communicate with his sergeant, a doctor, and the boy's mother as virtual humans. According to Gratch & Marsella [47], central to such training is the recognition of subordinates' emotions to judge their behavior accordingly and to asses the impact of orders on them. In this regard, they propose to incorporate **EMA** (see Section 3.1.1) into the MRE system to simulate the virtual humans' emotions. Figure 3.3 (right) shows how, based on EMA, a trainee's order impacts a sergeant's verbal and non-verbal behavior.

An evaluation of EMA within the MRE system demonstrated that the virtual humans' emotional behavior has been rated as quite natural. A further evaluation of EMA by comparing its behavior to human behavior based on an interactive questionnaire as a standard clinical instrument to asses human emotion, showed that EMA mimics human emotional behavior quite well [48].

The social robot Kismet

Kismet is an expressive anthropomorphic robot [20] (see Figure 3.4) that engages persons in a face-to-face interaction through his expressive and emotional behavior. Kismet perceives interacting persons through video and audio sensors, and recognizes their communicated affective intent such as praise, soothing, or prohibition through their tone of voice. Based on Breazeal's computational model of emotion (see Section 3.1.1), a person's affective intent impacts the robot's affective state, which triggers a corresponding emotional expression in the robot.

An evaluation of Kismet's affective behavior shows that subjects made use of Kismet's emotional expression as expressive feedback to check on the success of their communication of affective intent to Kismet. Kismet's emotional expressions are linked to the dimensional AVS space of his underlying emotion system, thus allowing him to reflect the trajectories of his emotion dynamics within the dimensional emotion space (see Section 3.2.1).

Lt:	Sergeant, send two squads forward
Sgt:	(interprets utterance, exhibits brief fearful facial expression)
Sgt:	That is a bad idea. We shouldn't split our forces
Sgt:	(gazes at Lt, exhibits brief hopeful expression)
Sgt:	Instead we should send one squad to recon forward along our route to Celic
Lt:	Send two squads forward!
Sgt:	(exhibits brief angry expression)
Sgt:	Against my recommendation sir (averts gaze and shakes head sideways)

Figure 3.3: Left: example scenario of MRE system [59]. Right: example emotional dialog between a trainee (Lt) and a virtual sergeant (Sgt) [47].

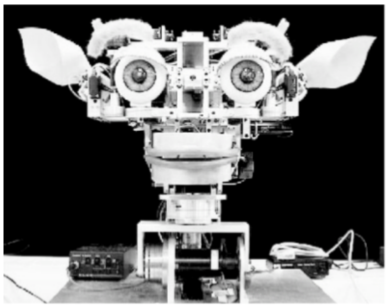


Figure 3.4: The social robot Kismet [20], p. 123.

The Multimodal Assembly eXpert (MAX)

The **M**ultimodal **A**ssembly e**X**pert (MAX) [62] is an embodied conversational agent embedded in a 3D computer graphics environment and able to engage with human partners in multimodal interaction using speech, gestures, and facial expressions. MAX was developed in the context of the Collaborative Research Center SFB 360 'Situated Artificial Communicators' and is still acquiring new skills through works in the Artificial Intelligence Group (AI-Group) at Bielefeld University. In a first scenario, MAX is visualized in a CAVE-like virtual reality environment, and guides his human partner through an interactive construction task [63]. In a further scenario, MAX acts as a guide in a public computer museum [62] (see Figure 3.5, right) where he engages in multimodal small talk dialogs with visitors, and provides them with information about the museum or the exhibitions. Furthermore, MAX also engages people in a small talk dialog in the corridor of the AI-Group at Bielefeld University.

MAX's behavior is controlled by a **cognitive agent architecture** [70] that builds upon the classical perceive-reason-act triad (see Figure 3.5, left) where the processes in the triad's components run concurrently. MAX's reactive behavior consists of an immediate response to perceived events and is realized by a direct mapping from perception to action (see Figure 3.5, left). Example reactive behaviors of MAX are gaze tracking, focusing the interaction partner in response to prompting signals, and performing permanent secondary behaviors such as eye blinks and breathing in order to appear more lifelike. MAX's deliberative behavior consists of reasoning about which actions to perform in response to perceived events, and is realized by adopting a BDI-approach [70] (see Figure 3.5, left). In order to increase the believability and likability of MAX (cf. [8]), his cognitive architecture is extended to an **emotion simulation module** based on **WASABI** (see Section 3.1.1); (see Figure 3.5, left). Accordingly, the primary or secondary emotion with the highest value of awareness likelihood is expressed by one of seven emotional facial expressions of MAX (see Section 3.2.1). The value of awareness likelihood of the displayed emotion modulates the intensity of its corresponding expression.

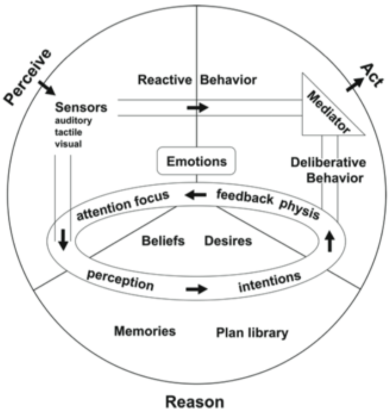


Figure 3.5: Left: MAX's cognitive architecture [70]. Right: MAX acting as a museum guide [62], p. 2.

An application of MAX's emotion simulation module is the museum guide scenario [8]. Within this scenario, MAX's primary emotions are simulated. For example, the perception of persons in MAX's visual field triggers his emotions positively, and the interpretation of the museum visitors' utterances as a compliment or as politically incorrect, respectively, triggers MAX's emotions positively or negatively. MAX's emotions not only impact his facial behavior, but also further behaviors such as leaving the screen when in a very angry emotional state. An evaluation of MAX's behavior within this scenario revealed that he engages the visitors in natural communication and that they ascribe him with a certain degree of sociality (cf. [62]).

A further application of MAX's emotion simulation module is a non-conversational 'Skip-Bo' card game scenario where MAX's goal is to win the game against a human opponent [8]. In a first version of this scenario, MAX's primary emotions are simulated. For example, playing a joker card triggers MAX's emotions positively and may result in MAX feeling happiness. An evaluation of MAX's primary emotions within this scenario revealed the appropriateness of his simulated emotions [7]. In another version of the gaming scenario, MAX's secondary emotions are simulated. For example, the expectation that the opponent would play a joker card triggers his emotions negatively and results in a secondary emotion of *fears-confirmed* when the opponent actually plays that card. An evaluation of MAX's secondary emotions showed, as expected, that MAX was judged to be significantly younger when expressing only primary emotions without secondary ones [8]. Within this scenario version, MAX also expresses empathy toward his opponent [7] (see Section 3.3.2).

Around 10 years after the development of MAX, a new virtual human called EMMA – **E**mpathic **M**ulti**M**odal **A**gent – is developed in our AI-Group and is introduced in the present thesis (see Chapter 4).

Discussion

The successful application and evaluation of the computational models of emotion within virtual human and robot scenarios substantiates the appropriateness of the proposed models and the role of emotions in enhancing the behavior of artificial agents. The application and evaluation of Dias & Paiva's model [36] is realized in the context of empathy and is thus presented in Section 3.3.2. Compared to the application and evaluation scenarios presented above, WASABI is successfully applied and evaluated within two different kinds of scenarios. A **conversational agent scenario** where emotions modulate the virtual human MAX's conversational and non-verbal behaviors, and a **gaming scenario** where they only modulate MAX's non-verbal behavior. A highlight is the application of WASABI in a **permanently running scenario** setting, specifically, the museum guide scenario. Thus, the successful application and evaluation of WASABI within these different scenarios is demonstrative for the feasibility, flexibility, and robustness of the model.

3.1.3 Conclusion

As emphasized in this section, the WASABI computational model of emotion addresses and unifies different aspects of emotion, and is successfully applied and evaluated within several scenarios involving the virtual human MAX. This highlights the appropriateness and feasibility of WASABI for the realization of several theoretical aspects of empathy (cf. Chapter 2). Accordingly, the computational model of empathy proposed in the present thesis is based on WASABI (see Chapter 5). In this regard, WASABI's **appraisal component** is crucial to the cognitive evaluation of others' situation, and thus for the generation of an empathic emotion. Furthermore, the **simulation of emotion dynamics** within WASABI is crucial to the simulation of the time course of an empathic emotion and to its modulation through factors such as the empathizer's changing mood over time (cf. Section 2.1).

3.2 Emotion expression and recognition

Humans' expression of emotion is multimodal and conveys different kinds of emotional information (cf. Section 2.3). Furthermore, humans possess elaborate and refined mechanisms to perceive others' emotions. Consequently, there are several works on the simulation of emotion expression and on developing methods for emotion recognition. In his book *The Expression of Emotion in Man and Animals*, Darwin [32] emphasized the specific and functional role of facial expressions in expressing and communicating emotion. Ekman's findings [38] on universals in facial expressions of emotion further substantiate their role in expressing and communicating emotion. Accordingly, **facial expressions** of emotion are central to the present thesis, and the related work presented in the next sections is mainly on the facial expression of emotion.

3.2.1 Emotion expression

Following Grammer & Oberzaucher [46], an **expression simulation system** is composed of two components: a control architecture for linking expressions to emotions and an expressive output component for animating expression patterns. Works on an expression simulation system usually considered the facial expression of emotion. In this regard, there are two standards mainly used for an expressive output component, FACS [40] (see Section 2.3) and the MPEG-4 standard [61] [90]. Within **MPEG-4**, **F**acial **A**nimation **P**arameters (FAPs) are specified that correspond to facial actions such as pulling the lip corners. FAPs are defined by means of 84 specified facial feature points that should be located on the face model in question. The MPEG-4 standard offers a large number of complex parameters to manipulate a facial expression on the basis of the specified facial feature points. Thus, conflicting or exaggerated facial deformations may occur and result in unnatural expressions. In contrast, **FACS** [40] offers a more controlled and elaborate method that accounts for humans' anatomical constraints to define facial expressions.

There are several works on an expression simulation system (see [57] for a detailed review). In a number of works, the control architecture is primarily based on linking emotion expressions to a set of discrete emotion categories such as Ekman's six basic emotion categories [38] (see Sections 2.2 and 2.3). New expressions are usually generated on the basis of this set of discrete emotion categories by means of fading or blending.

For example, the virtual human MAX (see Section 3.1.2) can display different intensities of seven emotional facial expressions [8] corresponding to his primary emotion categories (see Figure 3.6). MAX's face is comprised of 21 facial muscles that can be mapped to **A**ction **U**nits (AUs) from the FACS [40]. MAX's emotions also influence his

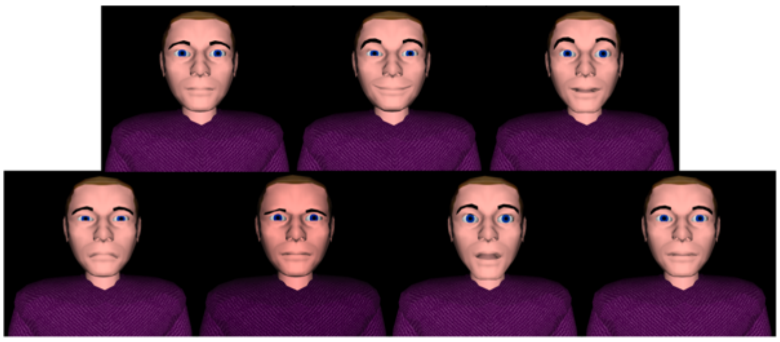


Figure 3.6: MAX's seven facial expressions of emotion. From left to right: *concentrated*, *happy*, *surprised*, *annoyed/sad/depressed*, *angry*, *fearful*, and *bored*.

speech prosody. MAX has discrete values of emotional pitch that correspond to his primary emotion categories. The frequencies of his secondary behaviors, breathing and eye blinking, are influenced by his value of arousal. The successful evaluation of the WASABI emotion model [8] (see Section 3.1.1) within a non-conversational 'Skip-Bo' card game scenario, where MAX expresses his emotions by means of his facial expressions, suggests the appropriate modeling of MAX's facial expressions.

Niewiadomski et al. [84] propose an algorithm to generate so-called 'complex' facial expressions for their embodied conversational agent **Greta**. As 'complex' facial expressions they define expressions that display emotion blends (e.g., superposed expressions) or expressions that are modified with respect to some socio-cultural rules (e.g., faked, masked, or inhibited expressions). 'Complex' facial expressions are computed as a composition of eight facial areas that can display different emotions (e.g., see Figure 3.7). For each type of 'complex' facial expression, a set of rules associated to Ekman's six basic emotion categories [38] is defined. These rules describe the facial areas of each basic emotion that are displayed within a type of 'complex' facial expressions as well as the compositions of facial areas of different basic emotions. Greta's face is animated using MPEG-4 FAPs.

Figure 3.7: From left to right: Greta's facial expression of *anger*, superposition of *sadness* and *anger*, *sadness* masked by *anger*, and Greta's facial expression of *sadness* [85].

In further works on an expression simulation system, the control architecture is based on linking emotion expressions to a dimensional emotion model such as Russell's *Circumplex Model of Affect* [99] (see Section 2.2). As compared to basic emotion models (see Section 2.2), dimensional emotion models present dense and continuous emotion spaces that are more inclined to characterize the continuity and subtlety of emotion expressions, and that allow for the expression of a wide range of emotions. Furthermore, Russell [100] argues that facial expressions do not signal specific emotions such as basic emotions, but instead primary information such as pleasure and arousal, and that the combination of this primary information with a context allows for the attribution of a specific emotion (cf. Section 2.3).

In their attempt to link facial expressions to a dimensional emotion model, Albrecht et al. [1] propose an algorithm to generate so-called 'intermediate' facial expressions based on a disc-shaped evaluation-activation space. Intermediate facial expressions are generated by means of interpolation between facial expressions of Ekman's basic emotions [38] mapped into the disc-shaped evaluation-activation space. To animate the facial expressions, they used a physics-based facial animation system modeled after human anatomy.

The same as Albrecht et al. [1], Courgeon et al. [30] propose an algorithm to generate intermediate facial expressions by means of interpolation between expressions of eight emotion categories (*fear*, *admiration*, *anger*, *joy*, *reproach*, *relief*, *distress*, *satisfaction*) placed at the extreme points of PAD space. To animate the facial expressions, they used MPEG-4 FAPs.

Breazeal [20] proposes an approach for linking emotional expressions to the three-dimensions of AVS space. In this approach, Breazeal defines basis postures for the social robot Kismet (see Section 3.1.2) that are associated to emotion categories lying at the extremes of AVS space (see Figure 3.8). New expressions related to each point in AVS space are generated by means of interpolation between the basis postures.

Altogether, these models use the dimensional emotion model to generate a wide range of emotion expressions including shades and mixtures of emotions. However, the generation of new expressions by means of **interpolation** between predefined expressions of emotion categories may generate unnatural and perceptually invalid expressions. In order to obtain perceptually valid expressions, current research seeks to link emotion expressions to a dimensional emotion model based on the **analysis of real expressions**, e.g., through image processing methods or perception studies. Within these models, low-level expressive features such as facial muscles are directly linked to the dimensional emotion model rather than predefined expressions. Following Russell [100], it is more appropriate to link low-level expressive features, such as single facial actions to primary information such as pleasure and arousal, rather than to emotion categories (cf. Section 2.3). In this regard, he postulates that the information signaled by a facial expression is present in its single components and substantiates his assumption through Snodgrass's findings [111].

Stoiber et al. [112] propose an approach to generate a two-dimensional space to control the facial expressions of a synthetic character based on analyzing physical deformations of a human actor's face displaying emotions. Their approach is based on detecting geometric and textural variation patterns (e.g., deformations of mouth, eyes, skin) and their reorganization in a two-dimensional control space. While they do not rely on any theoretical emotion space, their resulting two-dimensional control space is very similar to Plutchik's emotion wheel [94] (see Figure 3.9).

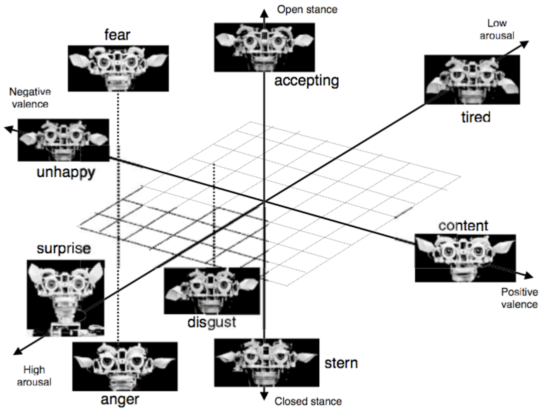


Figure 3.8: Basis postures of the social robot Kismet in AVS space [20], p. 141.

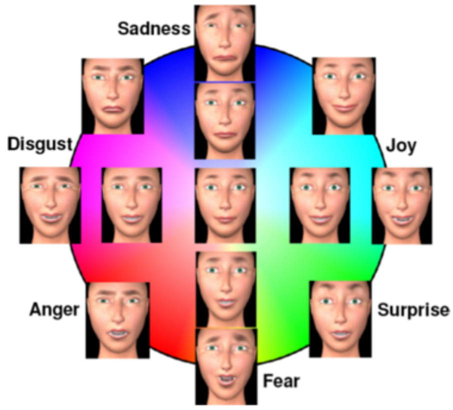


Figure 3.9: The facial expressions generated by Stoiber et al.'s approach [112].

Zhang et al. [121] propose a method to link MPEG-4 FAPs of a synthetic character to PAD values. Their approach is based on generating a database of synthetic facial expressions based on a mapping from a database of human faces displaying emotion. The database of synthetic facial expressions was annotated with PAD values using bipolar adjective pairs from [78]. The database was used to train a model to map PAD values to the synthetic character's MPEG-4 FAPs. The rating of the resulting expressions in terms of PAD values and emotion categories showed that the emotion underlying a generated facial expression was consistent with human participants' ratings of that expression.

Grammer & Oberzaucher [46] propose an approach to link AUs from FACS [40] to PAD values. In the context of an empirical study, a facial expression repertoire is reconstructed according to what they call a *reverse engineering* approach. In their study, human participants rated randomly generated facial expressions displayed by a 3D face using bipolar adjective pairs from [78]. The 3D face can express 25 AUs from FACS [40]. Based on a **P**rincipal **C**omponent **A**nalysis (PCA) and a **two-dimensional regression analysis**, the meaning of each AU within PA space was extracted. Note that the dominance dimension was not further considered in the regression analysis. Three example regression planes showing the meaning of AU1, AU12, and AU5 (cf. Table A.1, p. 204) within PA space are listed in Figure 3.10. By combining all regression planes for all AUs, a facial expression repertoire is reconstructed.

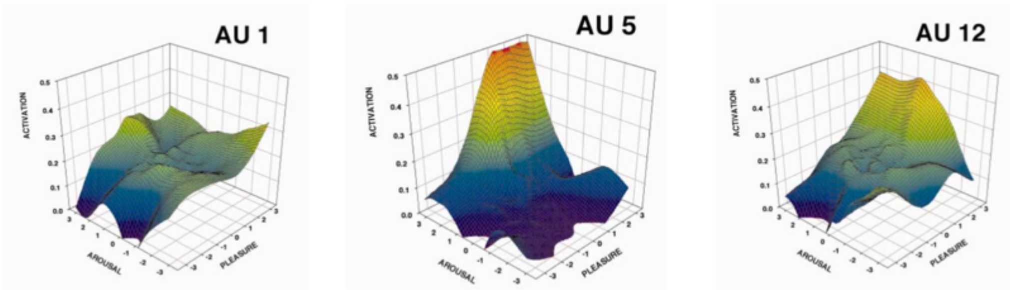


Figure 3.10: The regression planes representing the meaning of AU1, AU5, and AU12 in PA space [46], p. 10. The vertical axis (activation) represents the values of intensities of the AUs.

An interesting aspect in the expression of emotion is behavior expressivity, i.e. the velocity or abruptness of an expressive behavior. Accordingly, Chi et al. [26] propose the EMOTE (**E**xpressive **MOT**ion **E**ngine) model for defining the effort and shape qualities of expressive movements. The effort quality of a movement can be manipulated in terms of, e.g., weight as the impact of the movement (light vs. strong) or time as the urgency of the movement (sustained vs. sudden). The shape quality of a movement can be manipulated in terms of, e.g., direction of movement.

Discussion and conclusion

Ekman et al. [40] provide a set of clear and straightforward rules that define universal facial expressions of emotion. However, the simulation of emotional facial expressions on the basis of these rules is limited to this predefined set. In contrast, a dimensional emotion model is believed to be more convenient to characterizing the facial expression of emotion. However, the generation of new expressions by means of interpolation between predefined expressions located within a dimensional emotion space may produce unnatural expressions of emotion. Further approaches propose to link low-level expressive features of facial expressions to a dimensional emotion space to obtain perceptually valid expressions.

As such, we propose an approach to **link AUs to PAD emotion space** based on the work by Grammer & Oberzaucher. [46]. This approach is realized for a new virtual human, called EMMA (see Chapter 4), where an important aspect is to provide her with a large repertoire of facial expressions (see Section 4.3). EMMA's facial expression repertoire is crucial to the generation and expression of an empathic emotion in the proposed computational model of empathy (see Sections 5.2.1 and 5.4).

3.2.2 Emotion recognition

Similar to an expression simulation system [46] (see previous section), an **emotion recognition system** is composed of two components: feature recognition and mapping to emotion. Feature recognition consists of recognizing the features that are representative for emotion from expressive and contextual cues. Mapping to emotion consists of mapping the recognized features to emotion.

Feature recognition from expressive cues such as facial expressions, body gestures, audio signals, and bio-potential signals is based on several methods such as image and signal processing methods (see [51] for a review of existing methods). New approaches to facial feature recognition attempt to **recognize AUs** from FACS [40] since their detection allows for a more flexible and versatile interpretation of facial expressions as argued by Bartlett et al. [3], Valster & Pantic [116], and El Kaliouby & Robinson [41].

While much effort is invested in developing techniques for feature recognition from different expressive cues, several works on an emotion recognition system usually map the wide range of recognized features to a limited set of discrete emotion categories such as Ekman's six basic emotion categories [38] (see [51] for a review). Therefore, as for an expression simulation system (see Sections 3.2), dimensional emotion models present dense and continuous emotion spaces that are more inclined to characterize the continuity and subtlety of emotion expressions and that allow for the recognition of a wide range of emotions. However, little attention has been devoted to emotion recognition following a dimensional rather than a categorical approach, in particular regarding emotion recognition from facial expressions (see [50] for a review). As such, this field is currently in its pioneering stage.

In the works by Karpouzis et al. [60], Shin [108], Cao et al. [23], and Nicolaou et al. [83], different approaches are proposed for mapping recognized facial features to a dimensional emotion space. The proposed approaches are based on machine learning and classification methods, and on facial expression data sets annotated in terms of affective dimensions. In the works by Karpouzis et al. [60] and Nicolaou et al. [83], other modalities are also considered such as vocal and bodily expressions.

Since facial feature recognition in terms of AUs allows for a more flexible and versatile interpretation of facial expressions, Shugrina et al. [109] proposed an approach to recognize facial features in terms of AUs, and to **map AUs to PA values** based on the empirical findings of Snodgrass [111] (see Section 2.3). Accordingly, Shugrina et al. [109] associated each AU with a corresponding vector in PA space. The likelihood values of each recognized AU are used to calculate a scalar weighting of the AU. These values are used in a weighted summation of corresponding PA vectors to calculate the PA value of a facial expression.

With regard to feature recognition from contextual cues, several works are based on reasoning about other's cognitive states and their link to affective states based on, e.g., a BDI-approach, and on the computational modeling of emotion (see Section 3.1.1).

Discussion and conclusion

As for emotion expression, dimensional emotion models are believed to be more inclined to characterize and thus to analyze the expression of emotion. In this regard, the approaches proposed by Karpouzis et al. [60], Shin [108], and Cao et al. [23] rely on machine learning and classification methods, and on facial expressions data sets annotated in terms of the emotion dimensions. On the other hand, the approach proposed by Shugrina et al. [109] is based on the interpretation of facial expressions in terms of the PA values related to each recognized AU. Hence, emotion recognition is based on the extracted meaning of each AU in terms of PA values, e.g., [46], thus allowing for the interpretation of facial expressions in terms of the affective meaning of the different facial actions. Furthermore, it also makes it possible to investigate how individual facial actions contribute to the affective meaning of a whole expression.

As such, we propose an approach to map facial expressions displaying emotions to PAD values based on the meaning of each AU within PAD space, as defined in the virtual human EMMA's facial expression repertoire (see Section 4.3). Here we assume a system of automatic recognition of AUs as proposed by Bartlett et al. [3], Valster & Pantic [116], and El Kaliouby & Robinson [41], and focus on **mapping AUs to PAD values**. The proposed approach is crucial to the generation of an empathic emotion in our model (see Section 5.2.1).

3.2.3 Conclusion

According to Sections 3.2.1 and 3.2.2, dimensional emotion models are believed to be more convenient for emotion expression and recognition. With regard to **emotion ex-**

pression, proposed approaches can be classified into approaches that link whole facial expressions to emotion dimensions and approaches that link lower-level facial features to emotion dimensions. With regard to **emotion recognition**, proposed approaches can be classified into approaches that are based on data sets of whole facial expressions annotated in terms of emotion dimensions or into approaches that are based on data sets of lower-level facial features, e.g., facial actions, annotated in terms of emotion dimensions. According to Section 2.3, approaches that consider the emotional meaning of a whole facial expression can be referred to as top-down approaches while those that consider the emotional meaning of single facial actions can be referred to as bottom-up approaches. In this regard, our proposed approaches to emotion expression and to emotion recognition from facial expression can be classified as **bottom-up approaches**.

3.3 Empathy in virtual humans and robots

In this section, we present and discuss different computational models of empathy, as well as their application and evaluation in virtual human and robot scenarios.

3.3.1 Computational models of empathy

In this section, we focus on presenting works that mainly address the computational modeling of several theoretical aspects underlying empathy (cf. Section 2.1). One of the first models that provides a computational model of empathic emotions based on OCC's fortunes-of-other emotions [88] (see Section 2.2) is the **A**ffective **R**easoner (AR) [42].

CARE: Companion Assisted Reactive Empathizer

Following Davis [34] (cf. Section 2.1.2), McQuiggan et al. [76] propose an inductive framework for modeling **parallel and reactive empathy**. They called their framework CARE (**C**ompanion **A**ssisted **R**eactive **E**mpathizer) and based it on learning empirically informed models of empathy during human-agent social interactions. In a learning phase, users' situation data, such as their actions and intentions, users' affective states, bio-potential signals, and other characteristics such as their age and gender are gathered while they interact with virtual characters. The virtual characters respond to the user's situation with either parallel or reactive empathy. During interaction with the characters, users are able to evaluate their empathic responses using a 4 point Likert scale. Naive Bayes classifiers, decision trees, and support vector machines are used to learn models of empathy from 'good examples'. The induced models of empathy are used at runtime in a test phase to drive virtual characters' empathic responses.

A computational model of empathic emotions

Based on an empirical and theoretical approach, Ochs et al. [86] propose a computational model of empathic emotions. In order to precisely determine the conditions of elicitation of users' emotions during human-machine interaction, they empirically analyzed human-machine dialog situations to identify the characteristics of dialog situations that may elicit users' emotions. The results of this **empirical analysis** were combined with a **theoretical model** of emotion to provide a model of empathic emotions. The theoretical model of emotion is based on Scherer's appraisal theory of emotion [104] (cf. Section 2.2) where a description of the conditions of elicitation of emotions are defined. These are, in turn, based on a number of appraisal variables to determine the type and intensity of an elicited emotion. Thus, once the user's potential emotion is determined, the agent's empathic emotion from the same type is triggered toward the user. The elicitation and intensity of an empathic emotion depend on several factors. Accordingly, they define a **degree of empathy** as a value that affects the base intensity of the empathic emotion depending on the liking relationship between the user and the agent, and on the degree to which a user deserves or not his immediate situation (cf. [88], see Section 2.2). For their computational model of empathic emotions, they propose a formal representation using a BDI-like approach.

A generic computational model of empathy

Rodrigues et al. [98] propose a generic computational model of empathy following neuropsychological research on empathy [35] (cf. Section 2.1.3). Their model is integrated into an existing affective agent architecture [36] (see Section 3.1.1) and is comprised of two major components, an empathic appraisal component and an empathic response component (see Figure 3.11).

Accordingly, a perceived event by an agent that evokes an emotional cue in another agent is input to the empathic appraisal component together with the emotional cue. Rodrigues et al. [98] define an emotional cue as any observable signal indicating the

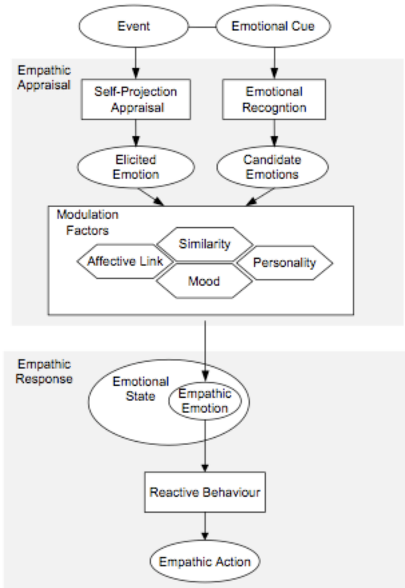


Figure 3.11: The generic computational model of empathy proposed by Rodrigues et al. [98].

presence of an emotion, e.g., facial expressions. Within the empathic appraisal, the emotional cue is input to an **emotion recognition** module, and the event is input to a **self-projection appraisal** module. The emotion recognition module provides several candidate emotions where the one with the highest potential is set as default. The self-projection appraisal module provides an elicited emotion as follows. The empathic agent assumes the other agent's situation by means of self-projection and appraises this situation using its own appraisal rules. The outputs of both modules are combined to determine an empathic emotion as the output of the empathic appraisal. If the elicited emotion is included in the set of candidate emotions, then it is set as the empathic emotion. Otherwise, the default emotion included in the candidate emotions is set as the empathic emotion. An important aspect addressed here is the **modulation of the empathic emotion** by several factors following De Vignemont & Singer [35]. Accordingly, the intensity of the empathic emotion is affected by the following modulation factors: *similarity*, *affective link*, *mood*, and *personality*. Similarity is determined as the degree of congruence of the elicited emotion with the candidate emotions. Affective link is determined as the value of liking between the agents. The average value of both of these factors affects the base intensity of the empathic emotion. That is, the higher the values of similarity and affective link, the higher the value of intensity of the empathic emotion. Mood refers to the empathizing agent's mood which then affects the intensity of the empathic emotion as it affects that of other emotions [36] (see Section 3.1.1). Personality refers to the empathizing agent's resistance to feel particular emotions. The more the agent is resistant to feel the empathic emotion, the less likely the emotion is to be output by the appraisal component.

With regard to the empathic response component, the empathic emotion generated by the empathic appraisal is added to the agent's emotional state, and triggers a situation appropriate action as a reactive behavior. The empathic appraisal process works concurrently to the reactive and the deliberative appraisal of the affective agent architecture [36] (see Section 3.1.1).

Breazeal et al.' s computational model of facial mimicry

According to the theoretical background on empathy introduced in Section 2.1, facial mimicry is crucial as a mechanism by which an empathic emotion is produced. In this section, we present Breazeal et al.'s computational model of facial mimicry for a social robot [21]. In this regard, Breazeal et al. [21] propose a biologically inspired approach to **imitation learning** based on findings by Meltzoff & Moore [80] (see Section 2.1.2). Their approach is based on two phases, one where a human interaction partner imitates the robot, and another where the robot imitates the human interaction partner. In the first phase, the robot starts a *motor babbling* action by displaying a random facial expression. Subsequently, the human partner tries to imitate the robot's displayed expression. Using separate neural networks for three different facial regions (right eye, left eye, and mouth) the robot learns a correspondence between its own facial display, and the human's recognized facial display during imitation. The neural networks are successfully trained with example input-output pairs. In the second phase, the robot imitates new facial expressions displayed by the human partner on the basis of the learned input-output pairs. Further, the imitated facial expression induces the corresponding emotional state in the robot based on the robot's emotion model [20] (see Section 3.1.1). However, they do not explicate how this issue is realized.

Discussion

As mentioned at the beginning of this section, one of the first models that provided a computational model of OCC's fortunes-of-other emotions [88] (cf. Section 2.2) is the

Affective **R**easoner (AR) [42]. Computational models of empathy that are based on OCC's fortunes-of-other emotions tend to be restricted to only two types of empathic emotions, *happy-for* and *pity*. However, an empathic emotion can be any emotion that is compatible with the other's situation (cf. Section 2.1).

McQuiggan et al. [76] provided a model of parallel and reactive empathy following Davis [34] (cf. Section 2.1.2). However, their approach is based on learning empirically informed models of empathy which hence constrain the flexibility of their model, and thus the easiness to adapt it to other contexts. Ochs et al. [86] provided a formal model of empathic emotions that is based on an empirical analysis of human-machine dialog situations. Similarly to McQuiggan et al.'s model [76], their model is tied to the considered context, and is also not easy to adapt to other contexts. Compared to these works, Rodrigues et al. [98] propose a generic computational model of empathy, and focus on an appropriate combination of different theoretical aspects of empathy within the model. In their model, two different mechanisms are integrated to generate an empathic emotion. Their definition of self-projection is similar to situational role-taking or self-focused role taking as defined respectively by Higgins [53] and Hoffman [55] (cf. Section 2.1.2). Further, their model is very similar to the early appraisal model of empathy [35] (cf. Figure 2.2, p. 22).

According to Hoffman's definition of mimicry [55] as composed of two subsequent steps, imitation and feedback (see Section 2.1.2), Breazeal et al. [21] focus mainly on the realization of the imitation step in their approach to facial mimicry.

3.3.2 Application and evaluation

In this section, we present different works on the application and evaluation of computational models of empathy in virtual human and robot scenarios. Accordingly, we focus on presenting works that mainly address the evaluation of the **impact of empathy** on human-agent interaction.

The Empathic Companion

The Empathic Companion is an embodied character developed in the context of a web-based job interview scenario to address users' recognized emotional state [95] by providing them with empathic feedback (see Figure 3.12). The goal of developing such a character is to investigate the effectiveness of empathic feedback to reduce users' stress in frustrating situations. In this scenario, an interviewing agent interacts with a user by asking him various questions. Bio-potential signals and context-based information are used to recognize users' affective states. Depending on a user's recognized affective state, empathic feedback consists of three options. The character shows empathic concern for a user whose recognized affective state indicates high arousal and negative valence, or encourages a user who is not aroused, or congratulates a user whose recognized emotional state indicates high arousal and positive valence. In order to evaluate the impact of the Empathic Companion on users, bio-potential signals are also used to measure users' perception of the embodied character. In this regard, two versions of the Empathic Companion were considered, a *not empathic* version where the character is defined as not supportive, and an *empathic* version where it is defined as supportive. The results show that the presence of the Empathic Companion significantly reduces users' value of arousal and stress when getting an interview question.

Empathic MAX playing Skip-Bo

Within the Skip-Bo card game scenario (cf. Section 3.1.2), the virtual human MAX provides the interaction partner with empathic feedback [7] based on the work by Prendinger & Ishizuka [95]. MAX's empathic feedback consists of displaying facial expressions reflecting the partner's recognized emotional state, thus overriding the emotional behavior triggered by MAX's own emotional state (see Figure 3.13). In order to evaluate the impact of MAX's emotional behavior on the interaction partner, bio-potential signals are also used to measure participants' perception of MAX's behavior. For this purpose, four different conditions were considered: a *non-emotional* condition where MAX does not show emotional behavior; a *self-centered* condition where MAX displays his own emotions according to his own game moves; a *negative empathic* condition where MAX displays self-centered emotional behavior and a non-congruent emotional state to that of the partner, e.g., showing positive emotions in response to the partner's negative emotions; a *positive empathic* condition where MAX displays self-centered emotional behavior and a congruent emotional state to that of the partner. The results show that displaying **positive empathic emotions** in a competitive card game scenario is arousing and stress inducing.

In a further work by Boukricha et al. [12], **situational role-taking** as defined by Higgins [53] (see Section 2.1.2) was realized for MAX within the Skip-Bo card game scenario. Thus, MAX infers the partner's emotional state by appraising the partner's game situation using own appraisal mechanisms. The partner's inferred emotional state is simulated within MAX's emotion simulation module besides the virtual human's own emotional state [8]. A first evaluation of this approach was realized with respect to three from a total of 32 game sessions recorded in the previous work by Becker et al. [7]. The results show that the arousal course provided by situational role-taking is quite similar to that provided by the emotion recognition system used in [7]. The proposed approach to situational role-taking is further considered and refined in the context of the present thesis (see Sections 5.2.2 and 6.1.2).

Figure 3.12: The empathic companion supporting a human candidate in a job interview task [95].

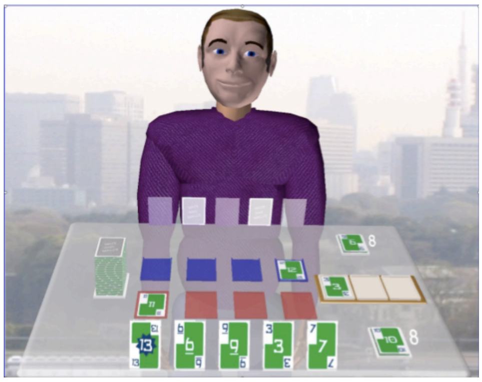


Figure 3.13: The virtual human MAX playing Skip-Bo against a human opponent [8], p. 104.

Casino-style blackjack

In the context of a casino-style blackjack scenario, Brave et al. [19] investigated the psychological impact of emotional and empathic agents on human users. Within this scenario, the agent and a user are playing against a disembodied dealer. The agent was represented by a photograph of a human face. Both interaction partners can observe their respective performances during the game. Based on a very simple emotion model, the agent displays positive emotions if he or the user won, and negative ones if he or the user lost. The agent's emotions are expressed by facial and textual expressions. The impact of the agent's emotional behavior on users was evaluated by asking participants to complete a questionnaire at the end of the game. In this regard, two conditions were considered, a *self-oriented emotion* and *other-oriented empathic emotion*. The results show that the agent is perceived as more likable, trustworthy, and caring in the *other-oriented empathic emotion* condition than in the *self-oriented emotion* condition.

A virtual training environment: Crystal Island

The application and evaluation of CARE [76] (see Section 3.3.1) is realized in the context of a virtual training environment called Crystal Island designed to teach middle school students microbiology and genetics. Within this environment, a student is assigned a particular task and a virtual character. Through his virtual character, the student can interact with other characters in the world to ask questions and form hypotheses in order to solve the assigned task. Within this scenario, data is collected for the students' ratings of the characters empathic behavior as defined in the CARE framework (see Section 3.3.1). The collected data was organized into training and test data. The evaluation of the characters' empathic behaviors according to these data sets shows that the induced empathy models within the CARE framework produce appropriate empathic behaviors.

EDAMS: An Empathic Dialog Agent in a Mail System

The computational model of empathic emotions [86] presented in Section 3.3.1 is applied and evaluated in the scenario of a dialog interaction in a mail system. In this scenario, a user interacts with an empathic agent, called **E**mpathic **D**ialog **A**gent in a **M**ail **S**ystem (EDAMS) to get information about his mails using predefined sentences. A 3D talking head is used to display empathic emotions by facial expressions. In order to calculate the intensity of an empathic emotion, some values are fixed prior to the interaction, e.g., the agent's degree of empathy. The empathic behavior of the agent is evaluated by asking participants to complete a questionnaire after interaction with the EDAMS. Three conditions are considered, similarly to those in [7]: a *non-emotional* condition, an *empathic* condition, and a *non-congruent emotional* condition. The results show that the agent is perceived more positively in the *empathic* condition and more negatively in the *non-congruent emotional* condition.

FearNot!

FearNot! is a computer application designed to help address the problem of bullying in schools [91]. FearNot! enables children between 8 and 12 years old to observe simulated bullying situations from a **third person perspective** and to experience this situation through role playing. The simulated bullying situations are realized with synthetic characters in a 3D virtual environment (see Figure 3.14). In an observed bullying situation, a child interacts with the victim (the bullied character) and advises him what to do by imagining being his friend. The child's interaction with the victim influences the victim's subsequent behavior in the observed situation. In this regard, Paiva et al. [91] focus on building characters with whom observers can feel a degree of proximity/familiarity, and that therefore evoke empathic responses in the observer. Accordingly, certain aspects of bullying situation were researched and taken into account in the design of such situations. The characters' emotional behavior and expression are controlled based on the emotion model presented in Section 3.1.1. Cartoon-like characters were chosen because they are preferred by children. The results of an empirical evaluation of the system with children and adults show that the system, as expected, evoked empathic responses in children but not in adults.

The generic computational model of empathy [98] presented in Section 3.3.1 was applied and evaluated in a scenario of four characters as designed within FearNot!. The characters were assigned different roles and values of liking relationships between each other. An empirical study investigating the perceived values of relationship and empathy between the characters is conducted with respect to two conditions, one with the empathy model and one without. The results show that the perceived values of empathy and relationship between empathizer and target are significantly higher in the empathy condition. The results are in line with the theoretical assumptions underlying the proposed model of empathy (see Section 3.3.1).

The relational agent Laura

Bickmore & Picard [9] introduce an approach to relational agents which they define as computer agents able to build and sustain long-term socio-emotional relationships with users. Their relational agent, called Laura, is a virtual agent that acts as advisor in the context of a fitness program. Laura interacts with users based on designed multimodal dialogs. In the dialogs, the agent responds multimodally using speech and synchronized non-verbal behavior. The user interacts with Laura by choosing items from a menu of predefined textual phrases. One important **socio-emotional phenomena** that contributes to the proposed relational agent approach is empathy. In the considered scenario, Laura's empathic behavior is realized through scripting.

The iCat as a chess companion

Leite et al. [69] proposed an approach to endow the iCat robot with empathic behavior in a chess game with children to improve their chess performance (see Figure 3.15). In their work they focus on providing the robot with a model of the children's affective states using a multimodal approach that combines visually recognized features such as facial and gaze behaviors, with contextual information about the game progress. Since prototypical facial expressions of emotion were not often observed in the children, they decided to measure the children's value of valence. Contextual information about the game progress is obtained based on self-projection (cf. [98], see Section 3.3.1). Based on this multimodal information, the probability of the children having a positive or negative value of valence is calculated. This is based on support vector machines trained with a data corpus containing the same kind of multimodal information obtained from recorded videos of children playing chess with the robot.

Figure 3.14: The FearNot! synthetic Characters in a 3D virtual environment [91].

Figure 3.15: The iCat as a Chess companion [68].

Once the children's affective state is recognized, an **adaptive empathic response** in the robot is triggered. The goal of the robot's empathic response is to maximize the children's value of positive valence. By trial and error, the robot also learns to choose the empathic response that effectively contributes to maximizing a particular child's positive value of valence. An empirical evaluation of the children's perception of the robot's empathic behavior in the game was conducted. For this purpose, three conditions were considered, a *neutral* condition where the robot does not display any empathic behavior, but instead a negative empathic behavior (cf. [7] and [86]); a *random empathic* condition where the robot selects an empathic response randomly; an *adaptive empathic* condition where the robot selects a learned empathic response. The results show that the robot was perceived by children as significantly more engaging and helpful in the *empathic* conditions, than in the *neutral* condition. No significant differences were found between the two *empathic* conditions.

Discussion

The application and evaluation of computational models of empathy within human-agent and agent-agent interaction scenarios provided valuable findings about the impact of empathy on enhancing artificial agents' social behavior, as well as support for the underlying theories. For example, empathic agents appeared to reduce stress levels during job interview tasks [95] and to teach children to deal with bullying situations [91]. However, the evaluation of the impact of MAX's empathic behavior within a competitive card game scenario shows that displaying **positive empathic emotions** is significantly arousing and stress inducing and thus inappropriate [7]. Therefore, we believe that a modulation of the virtual human's empathic emotion through factors such as mood, liking, or deservingness (cf. Sections 2.1 and 2.2) will result in a more appropriate behavior that further enhances an artificial agent's social behavior. For instance, consider a teaching scenario where an artificial agent is confronted with more than one student. Modulating the agent's empathic emotion through the factor deservingness, for example, would allow the agent to empathize more with the student who does not deserved his current annoying situation, e.g., a student who is working hard but who failed the exam vs. a student who is lazy and also failed the exam. In this regard, most of the computational models of empathy have focused on the generation and expression of an empathic emotion, while calculating **different degrees** of empathy has received little attention.

While this issue is addressed in the model by Ochs et al. [86], the perception of the agent's different degrees of empathy and of the different values of the modulation factors in question were not evaluated. In contrast, Rodrigues et al. [98] evaluated the perception of the values of relationship between empathizer and target in two considered conditions, *non-empathic* condition and *empathic* condition with their results supporting their model.

3.3.3 Conclusion

According to the three central processes of empathy that we identified in Section 2.1.4, *Empathy Mechanism*, *Empathy Modulation*, and *Expression of Empathy*, and to the discussion of related work on the computational modeling of empathy, the **requirements** for a computational model of empathy formulated in Section 1.2 arise. Table 3.1 shows the classification of previous work on the computational modeling of empathy discussed in this section with respect to its contribution to these requirements.

As compared to the previous work presented in this section, within our computational model of empathy, we consider **facial mimicry** and **situational role-taking** as mechanisms by which an empathic emotion is generated (see Sections 5.2.1 and 5.2.2). However, as compared to the model by Rodrigues et al. [98], we consider the mechanisms separately. Further, facial mimicry as defined by Hoffman [55] has not received much attention in the previously presented works with the exception of the work by Breazeal et al. [21]. However, whereas in their model of facial mimicry, Breazeal et al. [21] focused on the imitation step, in our approach, we focus on the **feedback** step. Further, we also attribute high importance to the calculation of different **degrees of empathy**, as an aspect that has received little attention in the computational modeling of empathy and which is crucial in further enhancing an artificial agent's social behavior. While in the models by Rodrigues et al. [98] and Ochs et al. [86], only the intensity of an empathic emotion is modulated, in our model, we also modulate its related type (emotion category). In this regard, we follow Hoffman's [55] emphasis that an empathic response need not to be a close match to the affect experienced by the other, but can be any emotional reaction compatible with the other's situation. Furthermore, we evaluated our proposed model based on three different conditions that distinguished different degrees of empathy, thus allowing for a more **fine-grained evaluation** of the model and its underlying parameters.

3.4 Summary

As emphasized in the previous chapter, theoretical models of empathy, together with theoretical models of emotion and theories and findings on emotion expression and perception, are crucial to the computational modeling of empathy. Therefore, Section 3.1 provided an overview of different approaches to the computational modeling of emotion

Related Work	Empathy Mechanism	Empathy Modulation	Expression of Empathy	Adequacy	Universality
McQuiggan et al. 2008	Mult.EmoRec	modulated type of emp.emo.	Mult.EmoExp	appropriate empathic reactions	partly
Ochs et al. 2012	emotional appraisal	modulated intensity of emp.emo.	Mult.EmoExp	positive impact of empathic agent on users	partly
Rodrigues et al. 2009	self-projection Automatic EmoRec	modulated intensity of emp.emo.	Mult.EmoExp	perceived values of empathy and liking are higher in the empathic condition	universal
Breazeal 2003	facial mimicry (imitation)	-	facial expression	appropriate imitation of perceived expressions	universal
Prendinger & Ishizuka 2005	Mult.EmoRec	-	textual expression	empathic agent lowers users' stress level	partly
Becker et al. 2005	Mult.EmoRec	-	facial expression	positive empathic emotions in a competitive game are arousing and stress inducing	partly
Boukricha et al. 2007	situational role-taking	-	facial expression	arousal course similar to the one provided by EmoRec for three games recorded in Becker et al. 2005	partly
Brave et al. 2005	very simple emotional appraisal	-	Mult.EmoExp	agent more likable, trustworthy, and caring in empathic condition	-
Bickmore & Picard 2005	emotional appraisal	-	Mult.EmoExp	empathic behavior contributes to relational agents	-
Leite et al. 2012	self-projection automatic EmoRec	-	Mult.EmoExp adaptive behavior	agent more engaging and helpful in empathic condition	partly

Table 3.1: Related work on the computational modeling of empathy classified with respect to its contribution to the requirements formulated in Section 1.2. 'Mult. EmoRec' refers to multimodal emotion recognition. 'Mult. EmoExp' refers to multimodal emotion expression. 'emp. emo.' refers to empathic emotion.

and their application and evaluation in different virtual human and robot scenarios. In this regard, the **WASABI** computational model of emotion unifies several theoretical aspects of emotion, and is successfully applied and evaluated within different scenarios involving the virtual human MAX. This underlines the appropriateness and feasibility of WASABI for the realization of different theoretical aspects of empathy, and as such is at the basis of our model (see Chapter 5).

Section 3.2 presented different approaches to emotion expression simulation as well as methods for emotion recognition. In line with findings within research on emotion expression and perception (see Section 2.3), recent works on emotion expression simulation and on emotion recognition follow a dimensional approach. In this regard, a **dimensional emotion model** (cf. Section 2.2) appears to better characterize the continuity and subtlety of emotion expression, and to allow for the expression and recognition of a wider range of emotions than other models (cf. Section 2.2). Furthermore, the discussed works were classified into those that consider the emotional meaning of a whole facial expression, thus relying on a top-down approach and those that consider the emotional meaning of single facial muscles, thus relying on a bottom-up approach (cf. Section 2.3). In this regard, a **bottom-up approach** allows for the combination of different facial actions and their emotional meaning to reconstruct new facial expressions. Accordingly, we propose an approach to link AUs from FACS [40] to PAD space [102] thus providing a large repertoire of facial expressions of emotion (see Section 4.3). We also propose an approach to map facial expressions displaying emotions to PAD values based on the provided meaning of each AU within PAD space. The proposed approaches are crucial for the generation and expression of an empathic emotion within our computational model of empathy (see Sections 5.2.1 and 5.4).

Finally, Section 3.3 introduced different approaches to the computational modeling of empathy as well as their application and evaluation within virtual human and robot scenarios. The models discussed were divided into those that focused mainly on addressing the theoretical aspects of empathy, and those that focused on evaluating the impact of empathy on human-agent interaction. With regard to the former, we identified models that are empirically informed and thus tied to the considered context and models that provide a more generic model of empathy. Further, the application and evaluation of several computational models of empathy within different virtual human and robot scenarios provided valuable results for the role of empathy in enhancing an artificial agent's social behavior, and for the role of computational models as an experimental tools for underlying theories. However, most of the computational models of empathy have focused on the generation and expression of an empathic emotion, while the consideration of different **degrees of empathy** has received little attention despite being crucial to further enhancing an artificial agent's social behavior.

The related works on the computational modeling of empathy presented in this section were classified with respect to their contribution to the **requirements** for a computational model of empathy as formulated in Section 1.2. As compared to these works, in our computational model of empathy, facial mimicry and situational role-taking are considered as two separate empathy mechanisms. Further, we also consider the calculation of different degrees of empathy in our model. In line with Hoffman's emphasis that an empathic response need not to be a close match to the affect experienced by the other but can be any emotional reaction compatible with the other's situation [55], in our model, not only the intensity of an empathic emotion is modulated, but also its related type (emotion category). Furthermore, we evaluated our model based on three conditions that differentiated different degrees of empathy, thus allowing for a more fine-grained evaluation of the model and its underlying parameters.

4 The virtual human EMMA

This chapter introduces EMMA, a new virtual human with a female-like appearance. Besides the virtual human MAX, EMMA is developed at the Artificial Intelligence Group (AI-Group) at Bielefeld University. Section 4.1 begins with an overview of the history of developing a new virtual human. In Section 4.2, we present the name EMMA as an acronym for **E**mpathic **M**ulti**M**odal **A**gent. Section 4.3 introduces an empirically based approach to link EMMA's facial expressions to the dimensional emotion space of pleasure, arousal, and dominance.

4.1 A new virtual human: The history in a nutshell

This section gives an overview of the motivation and background behind developing EMMA as well as of the different steps taken in the design of her appearance.

4.1.1 Motivation and background

Besides the successful work on the virtual human MAX [62] (cf. Section 3.1.2), the idea was brought up in the AI-Group at Bielefeld University to create a new virtual human. This new virtual human is named EMMA and is the female counterpart to MAX (see Figure 4.1).

As compared to MAX whose face is comprised of 21 facial muscles that express seven emotion categories (cf. Section 3.2.1), an important aspect in developing EMMA is to provide her with a large repertoire of facial expressions. Hence, EMMA's face is modeled in line with the **F**acial **A**ction **C**oding **S**ystem (FACS) [40] and expresses a wide range of emotions along the **P**leasure-**A**rousal-**D**ominance (PAD) dimensions [102] (see Section 4.3). Further, the creation of a new virtual human besides MAX allows for the consideration of another type of interaction in addition to human-agent interaction, namely, agent-agent interaction. For example, the conversational agent scenario realized in the

Figure 4.1: The virtual humans MAX & EMMA in a 3D computer graphics simulation of a biosphere.

present thesis which involves the virtual humans MAX and EMMA and a human interaction partner (see Section 6.1.1). Furthermore, the creation of a virtual human with a female-like appearance allows for the computational modeling of gender differences as well as for their analysis and evaluation. However, this aspect has not yet been considered, and could be the subject of future work. In summary, there are **three important aspects** in creating EMMA:

- providing an elaborate model of facial expressions (see Section 4.3).
- allowing for the consideration of agent-agent interaction (e.g., see Section 6.1.1).
- modeling and evaluating gender differences.

4.1.2 Design and appearance

The design of EMMA's appearance was initiated by Christian Becker-Asano and Kerstin Hasse as former members of the AI-Group at Bielefeld University under the supervision of Prof. Ipke Wachsmuth. The process was finalized by Andrea Hofstätter in **cooperation** with the department of anthropology at the University of Vienna [56]. In the following, we describe the different steps taken to design EMMA's appearance.

As mentioned before, one important aspect in creating EMMA is to provide an elaborate model of facial expressions. Therefore, as compared to MAX's face, EMMA's face consists of a higher-resolution polygon mesh that allows for a refined modeling of **A**ction **U**nits (AUs) in line with FACS [40] (see Section 4.3.1). Further, EMMA's face mesh is designed using front and side photographs of a real woman's face (see Figure 4.2).

Figure 4.2: The real woman's face underlying EMMA's face model.

In addition to the face mesh, there are polygon mesh models for hair, ears, eyes, lacrimals, teeth, and a palate, which together form EMMA's entire head mesh (see Figure 4.3).

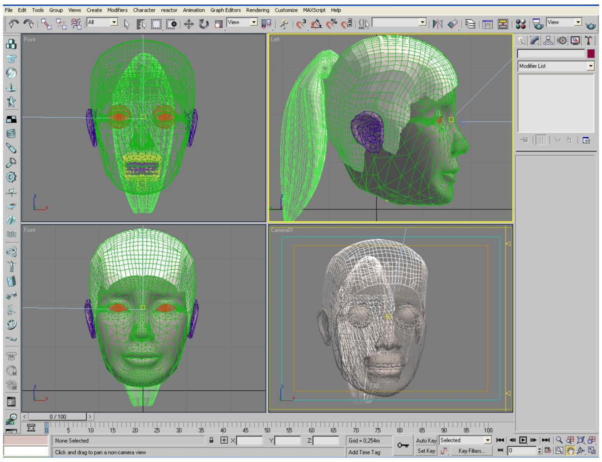


Figure 4.3: The polygon mesh models forming EMMA's head mesh [56], p. 18.

Also the texture of EMMA's face is designed based on a photograph of a real woman's face (see Figure 4.4).

Further, while MAX's body is based on a number of rigid object geometries enclosing his limbs and connected at the level of his joints, EMMA's body, the same as her head, consists of a polygon mesh model. Hence, EMMA's body allows for smoother, and

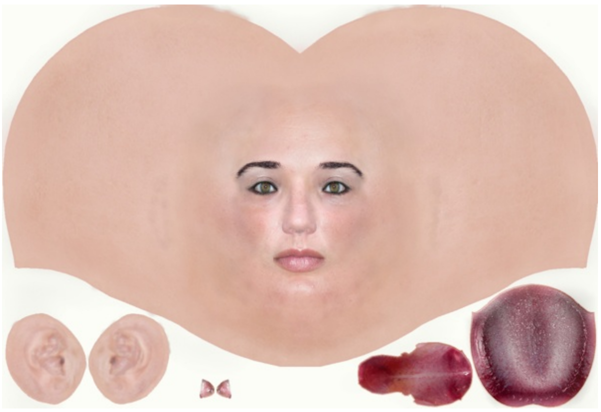


Figure 4.4: The textures for EMMA's face, ears, lacrimals, and palate [56], p. 19.

thus more natural body movements (see Section 4.2.2). Furthermore, the proportions of EMMA's body correspond to those of a real woman's body. As emphasized in the previous section, EMMA and MAX can appear together in the context of an agent-agent interaction (e.g., Section 6.1.1). Accordingly, it was decided to choose textures for EMMA's garments that are in line with those of MAX in order to obtain a coherent appearance for both virtual humans (see Figure 4.1).

4.2 EMMA: An Empathic MultiModal Agent

As mentioned in Section 4.1.1, EMMA's face is modeled in line with FACS [40] and expresses a wide range of emotions along the PAD dimensions [102] (see Section 4.3). Hence, the idea was brought up that EMMA could be suitable to realize empathy on the basis of the empathy mechanism **facial mimicry** (see Section 5.2.1). Effectively, a first interaction scenario of EMMA is carried out in the context of empathy in this thesis (see Section 6.1.1). To this end, the components of MAX's cognitive architecture (see Figure 4.5) are carried over, adapted, and extended for the new virtual human EMMA. In accordance with her capabilities within her first interaction scenario (see Section 6.1.1), we chose the name 'EMMA' to refer to an **E**mpathic **M**ulti**M**odal **A**gent.

In this section, we present the cognitive architecture by emphasizing the components that EMMA 'inherited' from MAX and those that are specific to EMMA.

4.2.1 The cognitive architecture

The virtual human MAX [62] has a cognitive architecture [70] that builds upon the perceive-reason-act triad (see Figure 4.5), and that is successfully applied within several interaction scenarios (see Section 3.1.2). Highlights include the permanently running

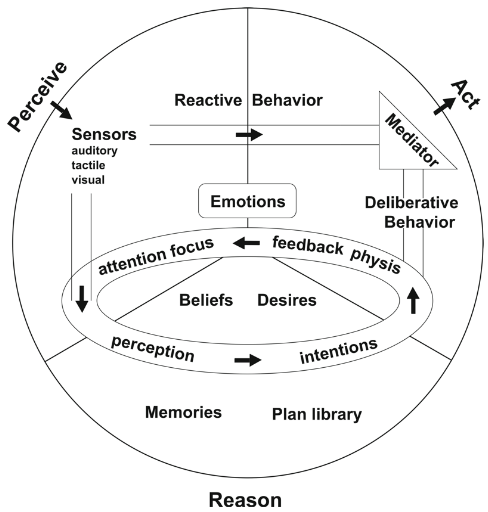


Figure 4.5: The cognitive architecture of the virtual human MAX carried over to the new virtual human EMMA [70].

scenarios of MAX acting as a guide in a computer museum, and of MAX engaging people in a small-talk dialog in the corridor of the AI-Group at Bielefeld University. The successful application of MAX's cognitive architecture within these and other scenarios is demonstrative of the feasibility, flexibility, and robustness of the architecture. Accordingly, MAX's cognitive architecture is carried over to the virtual human EMMA, and is then further adapted and extended.

The act section of the architecture triad (see Figure 4.5) includes a multimodal behavior generation module, called by Kopp & Wachsmuth [64] the **A**rticulated **C**ommunicator **E**ngine (ACE). Within this module, adaptations and extensions are performed for the facial expression component, the speech component, and the component that generates and controls body movements, e.g., gestures.

Further, a **new architecture component**, particular to EMMA and crucial to the computational model of empathy proposed in this thesis is for facial mimicry as introduced in Section 5.2.1. The adaptations and extensions carried out in the multimodal behavior generation module of the architecture are described in the next section.

4.2.2 The multimodal behavior

Similarly to the virtual human MAX, EMMA is capable of exhibiting multimodal behavior based on facial expressions, speech, and body movements. The multimodal behavior is triggered by the reactive and deliberative components of the architecture (see Figure 4.5) and is generated within the **A**rticulated **C**ommunicator **E**ngine (ACE) [64]. In the following, we present the adaptations and extensions carried out for EMMA.

Facial expression

FACS [40] is based on an anatomical analysis of facial muscle actions, and represents a comprehensive system to describe all possible and visually distinguishable facial movements (cf. Section 2.3). Furthermore, FACS is the most widely used expression coding system in the behavioral sciences, and has also been established as an eligible standard for the adequate and convenient modeling of artificial facial expressions (cf. Section 3.2.1). Further, **dimensional emotion models** present dense and continuous emotion spaces that are more inclined to characterize the continuity and subtlety of emotion expressions, and that allow for the generation of a wide range of expressions. Furthermore, it seems to be more appropriate to link low-level expressive features such as single facial actions to a dimensional emotion space rather than predefined facial expressions (cf. Section 3.2.1). Hence, **AUs** are modeled for EMMA in line with FACS [40], and are linked to the dimensional emotion space of pleasure, arousal, and dominance [102] (see Section 4.3).

In this regard, the facial expression component is extended to the specification and generation of facial behaviors on the basis of AUs. For example, the specification and generation of visemes as the visual representation of speech phonemes, of key frame animations of facial expressions, and of facial behaviors such as eye blinking as an incessant secondary behavior and eye brow raising as a conversational behavior (cf. [62]).

Further, the facial expression component is also extended to a **facial expression repertoire** linking PAD values to AUs. Thus, to each PAD value output by the emotion simulation module [8] of the cognitive architecture [70] (see Section 3.1.1), a corresponding facial expression animation is generated. Furthermore, similarly to MAX (see Section 3.2.1), the frequency of EMMA's eye blinking, as a secondary behavior, is influenced by her value of arousal. The facial behaviors generated within the facial expression component are synchronized with other triggered modalities such as speech and body movements.

Speech

As for facial expressions, the parameters of EMMA's speech prosody are also linked to the dimensional emotion space of pleasure, arousal, and dominance [102], thus allowing for a continuous and subtle modulation of EMMA's emotional speech prosody. In this regard, a new speech component based on the work by Schröder & Touvain [107] and Schröder [106] is added. In their work, Schröder & Touvain introduce a **German text-to-speech synthesis system** called MARY (**M**odular **A**rchitecture for **R**esearch on speech s**Y**nthesis) where PAD based emotional prosody rules are specified within a so called *EmoSpeak* module. Since the emotional prosody generated with regard to extreme PAD values was judged by members of our AI-Group as exaggerated, the values of some speech prosody parameters within the specified emotional prosody rules were adapted accordingly.

Body movements

To allow for the appropriate generation and control of EMMA's body movements such as her gestures, some adaptations are performed within the corresponding component. These adaptations involved the specification of the lengths of EMMA's limbs, number of her joints, and the transformations within EMMA's body-centered coordinate system.

In order to animate EMMA's body mesh (cf. Section 4.1.2), an underlying bone structure forming EMMA's skeleton is defined. The animation of EMMA's skeleton – her limbs and joints – further animates her body mesh based on a defined mesh deformation algorithm. Hence, as compared to MAX whose body is based on a number of rigid object geometries, EMMA is capable of smoother and thus more **natural body movements** especially in the torso region (cf. Figure 4.1).

Similarly to MAX (see Section 3.2.1), in addition to her incessant secondary behavior of eye blinking, EMMA has a breathing behavior defined as a body movement whose frequency is influenced by her value of arousal. Furthermore, EMMA is also able to synchronize speech and gestures as was previously realized for the virtual human MAX [64].

4.2.3 Conclusion

As emphasized in Section 4.2.2, similarly to MAX, EMMA is capable of exhibiting **multimodal behavior** based on facial expressions, speech, and body movements. However, EMMA can express a wider range of emotions along PAD dimensions using her face and speech prosody, and is capable of smoother and thus more natural body movements. Furthermore, EMMA's **large repertoire of facial expressions** is crucial to the realization of facial mimicry as an empathy mechanism (see Section 5.2.1), and for the expression of empathy (see Section 5.4) as requirements for building a computational model of empathy (cf. Section 1.2). In the following, the steps taken for the development of EMMA's facial expression repertoire are introduced.

4.3 EMMA's facial expressions in pleasure-arousal-dominance space

According to Grammer & Oberzaucher [46], an expression simulation system is composed of two components: a control architecture for linking expressions to emotions and an expressive output component for animating expression patterns (cf. Section 3.2.1). In this section, we present EMMA's **facial expression simulation system**. The expressive output component is based on modeling AUs for EMMA in line with FACS [40] (see next section). The control architecture is based on linking EMMA's AUs to the dimensional emotion space of pleasure, arousal, and dominance [102] using an empirically based approach (see Sections 4.3.2 to 4.3.5).

4.3.1 Implementing the facial action coding system for EMMA

In cooperation with the department of anthropology at the University of Vienna, EMMA's face is modeled based on FACS [40] with the help of experienced FACS coders [56]. Overall **44 AUs** (cf. Table A.1, p. 204) are implemented for EMMA. These include 34 AUs for nine upper face units and 25 lower face units. The remaining AUs represent head and eye units. Additionally, body turn left and body turn right are introduced and referred to as 'AU90' and 'AU91' (see Appendix A, p. 211).

The face units are implemented at their maximal contraction directly on the head mesh by means of so called *morph targets* (cf. [56]). Variations in the skin such as permanent **wrinkles** are also modeled for the relevant AUs, e.g., horizontal wrinkles for AU1 and AU2 and vertical wrinkles for AU4 (see Figure 4.6). Note that the visualization of the wrinkles in EMMA's current 3D computer graphics environment is still under development.

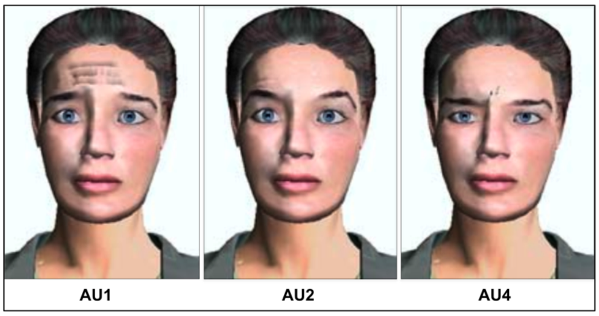


Figure 4.6: Horizontal wrinkles for AU1 and AU2 and vertical wrinkles for AU4 [56].

Figure 4.7 shows the **graphical user interface** for the visualization of different values for EMMA's AUs. For testing purposes, different configurations of AUs can be saved and loaded. The field 'SetPAD' is explained later in Section 4.3.3.

Ekman and colleagues [40] define rules to link AUs to the six basic emotion categories (see Table A.2, p. 205). Accordingly, Figure 4.8 depicts example facial expressions for EMMA showing the six basic emotions (cf. [40]).

Further, EMMA's **visemes**, as the visual representation of speech-phonemes, are specified based on her AUs following the work of Aschenberner & Weiss [2] who introduce a German viseme set (see Table A.3, p. 212). Three example visemes of EMMA are depicted by Figure 4.9. For a list of figures showing the remaining visemes, see Appendix A, p. 212. EMMA's visemes are modeled by Andrea Hofstätter in cooperation with the department of anthropology at the University of Vienna.

FACS allows for the modeling of a wide range of facial expressions based on a large number of AUs. However, Ekman and colleagues [40] reduce this wide range of possibilities to a limited set of discrete emotion categories (cf. Table A.2, p. 205). As mentioned before, EMMA's control architecture is based on linking her AUs to the dimensional emotion space of pleasure, arousal, and dominance thus allowing for a wider range of emotional facial expressions (cf. Section 4.2.2). EMMA's control architecture is presented in the following sections.

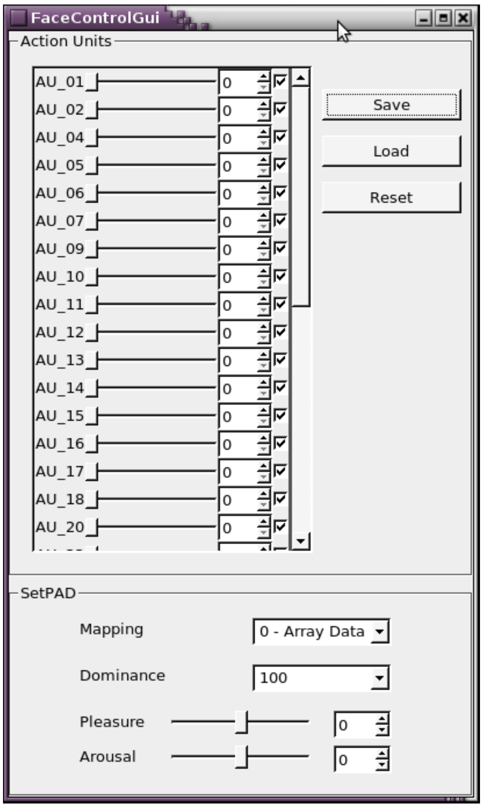


Figure 4.7: The graphical user interface for the visualization of different values for EMMA's AUs.

4.3.2 Empirical study

Based on the results of an empirical study, a facial expression repertoire comprised of facial expressions arranged in PAD space [102] is constructed for the virtual human EMMA. The facial expression repertoire is based on linking AUs [40] to PAD space. The empirical study was performed in **cooperation** with the department of anthropology at the University of Vienna, and relies on previous work by Grammer & Oberzaucher [46] (see Section 3.2.1). For more details on the empirical study, see [56].

The empirical study was conducted at the Biozentrum of the University of Vienna in Austria with **353 adult participants** between 18 and 65 years of age. A total number of 3517 randomly generated facial expressions were rated by the participants. Of these, 2099 (59.7%) were rated by 211 female participants and 1418 (40.3%) were rated by 142 male participants. Each participant rated 10 randomly generated facial expressions with a German translation of 18 bipolar adjectives from the 'Semantic Differential Measures

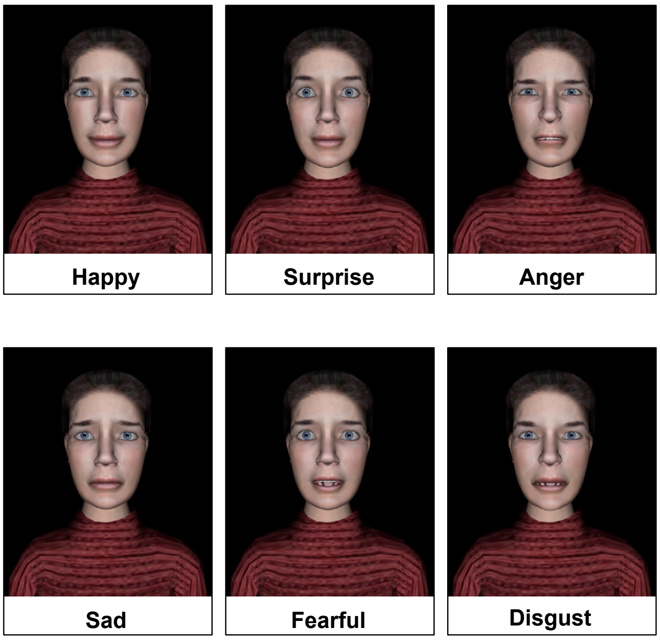


Figure 4.8: Example basic emotions modeled by EMMA's AUs (cf. Table A.2, p. 205).

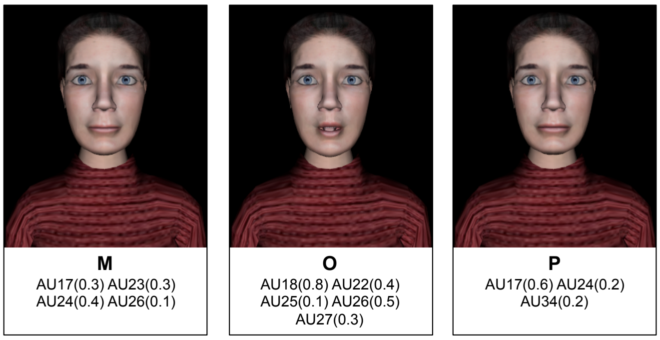


Figure 4.9: Three example visemes for EMMA, M,O, and P with corresponding values of AUs.

of Emotional State or Characteristic (Trait) Emotions' [78] (see Table 4.1). Each facial expression was rated with the 18 bipolar adjectives on a **1 to 7 Likert scale**. Following Osgood et al. [89], each group of six bipolar adjectives was associated with one of the dimensions of pleasure, arousal, and dominance.

Each facial expression rated by the participants was generated by means of a **facial expression randomizer program** that allows a random combination of AUs as well as the altering of their respective intensities. Accordingly, the following rule to generate random facial expressions on the basis of AUs was applied: *among 44 AUs, choose* *randomly 10 AUs with random intensities ranging between* 0% *and* 50% *of maximal intensity*. In this way, the generation of unnaturally exaggerated facial expressions is avoided. No facial expression was generated twice by the facial expression randomizer program. Furthermore, a number of predefined **co-occurrence rules for AUs** were applied to avoid the generation of anatomically impossible facial expressions. These rules are introduced by Ekman & Friesen [39] to describe the appropriate combination of AUs. A fuzzy-logical implementation of three co-occurrence rules is proposed by Wojdel et al. [119] and was applied within the facial expression randomizer program. For more details on the considered rules, see [56].

The randomly generated facial expressions were presented to participants in POSER 6 (Curious Labs, Santa Cruz, CA). Figure 4.10 shows the facial expression evaluation interface. The neutral facial expression was shown to the participants as a translation between two generated facial expressions and was also rated by five female and five male subjects between 20 and 32 years of age.

Figure 4.10: Facial expression evaluation interface with the rating adjective 'in control-cared for' [56], p. 25.

4.3.3 Methods and results

After a successful collection of data, a statistical analysis was performed (see [56] for more details on the used statistical methods). As mentioned in the previous section, the neutral facial expression was rated by five female and five male participants. The calculation of the median for each of the bipolar adjectives over the 10 participants showed that the **neutral facial expression** was rated as aroused, displeased, and submissive (see Table 4.1).

Emo. Dim.	Rating Adjectives	Value of Median (1-7 Likert scale)
Pleasure	happy-unhappy	6
	pleased-annoyed	4
	satisfied-unsatisfied	5
	contented-melancholic	5.5
	hopeful-despairing	6
	relaxed-bored	5
Arousal	stimulated-relaxed	2
	excited-calm	2
	frenzied-sluggish	4
	jittery-dull	2
	wide-awake-asleep	2
	aroused-unaroused	2
Dominance	controlling-controlled	5
	influential-influenced	5
	in control-cared for	4.5
	important-awed	5
	dominant-submissive	5
	autonomous-guided	5

Table 4.1: Values of median for the ratings of the neutral facial expression [56]. 'Emo. Dim.' refers to emotion dimensions.

Therefore, in order to consider the influence of the neutral facial expression on the ratings, its evaluation was taken into account in the statistical analysis of the data. Accordingly, for each of the bipolar adjectives, the mean value over the 10 participants as well as the difference of this mean value to the value 4 on the 1 to 7 Likert scale was calculated. The calculated value of difference for each of the bipolar adjectives was subtracted from the ratings of the randomly generated facial expressions (cf. [100]).

Further, each generated facial expression consisted of 10 randomly activated AUs with random intensities while the remaining AUs were not activated and thus had intensity values equal to 0. It should be noted that not only the presence of an AU in a facial expression contributes to its meaning, but also its absence. Hence, the values of 0 intensities for the non-activated AUs were also included in the statistical analysis. As statistical analysis, a principal component analysis and a regression analysis were carried out.

Principal component analysis

A principal component analysis with varimax-rotation was conducted on the data. For this purpose, the data were organized as a set of data points (vectors) ($N = 3482$) within an 18 dimensional space spanned by the 18 bipolar adjectives. That is, each data point represented a vector of the rating values of each randomly generated facial expression with the 18 bipolar adjectives. The principal component analysis yielded, as expected (cf. Section 4.3.2), a **three factor solution** that accounts for most of the variance in the data, namely, 66.7% of variance. The first factor is dominance which accounts for 25% of variance, the second factor is pleasure and accounts for 22.5% of variance, and the third factor is arousal and accounts for 19.2% of variance. Afterward, the data were standardized to a mean value of 0 and a standard deviation of 1. For more details on the results of the principal component analysis for each of the three factors, see [56]. Henceforth, only the factors pleasure, arousal, and dominance are considered in a subsequent regression analysis.

Regression analysis

In order to examine the correlation of the intensity values of AUs with the pleasure, arousal, and dominance factors, a two-dimensional non-linear regression of each AU with two of the three factors was carried out using a LOESS regression method (locally weighted regression) [27] [28]. This method uses a second-degree polynomial function and provides a smoothed curve over the considered data points called a *Loess Curve*. Accordingly, a two-dimensional non-linear regression of each AU with pleasure and arousal was calculated for those facial expressions that were rated as showing positive dominance, and those that were rated as showing negative dominance. As a result, **three-dimensional non-linear regression planes** for each AU in PA spaces corresponding to two values of dominance (positive vs. negative) were obtained. The resulting regression planes show the meaning of each AU within PAD space. Figure 4.11 shows two example regression planes for AU12 (Lip Corner Puller) and AU43 (Eyes Closed) in PA space of positive dominance. The regression plane for AU12 shows that AU12 has a maximal intensity value with respect to positive pleasure independently of the arousal value, while the regression plane for AU43 shows that AU43 has a maximal intensity value with respect to negative arousal independently of the pleasure value. Thus, AU12 mainly expresses positive pleasure while AU43 mainly expresses negative arousal.

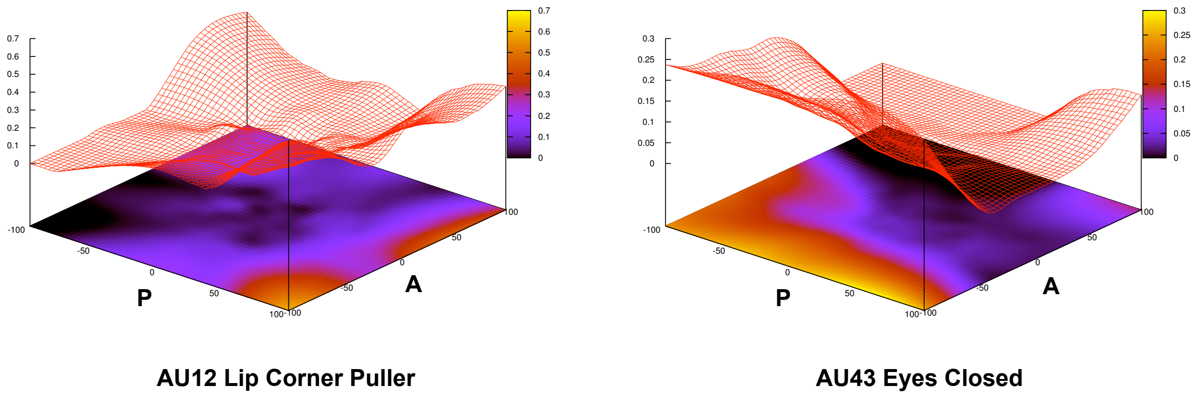


Figure 4.11: The non-linear regression planes corresponding to AU12 (left) and AU43 (right) in PA space of positive dominance. The *vertical axis* represents the AUs intensity values.

According to the face randomizer program (see Section 4.3.2), the values of AUs intensities range between 0% and 50% of maximal intensity. However, the regression function generated intensity values lower than 0% and higher than 50% of maximal intensity. Since an AU intensity value cannot be negative, except for AUs such as AU30 (Jaw Sideways) where negative values mean 'jaw moved to the left' and positive values mean 'jaw moved to the right', the negative values generated by the regression method are set to 0 and the regression planes are plotted in the range $[0, 1]$. The range $[0, 1]$ represents values between 0% and 100% of maximal intensity. The three-dimensional regression planes for each AU in PA spaces of positive and negative dominance are listed in Appendix A, p. 215. Note that these regression planes are provided by Andrea Hofstätter for the purpose of the present thesis and are not presented and discussed in [56]. Further, the regression planes for eye, head, and body units are not considered here as we decided to first focus on the regression planes for the face units. In the following, we report the **meanings of AUs within PAD space** in terms of their maximal intensity values as shown by their corresponding regression planes.

According to the three-dimensional regression planes for the AUs (see Appendix A, p. 215), AU10, AU12, AU14, AU18, AU26, AU27, AU30 right, AU31, AU34, and AU38, have their maximal intensity values in PA space of positive dominance while AU1, AU2, AU5, AU6, AU11, AU13, AU23, and AU39 have their maximal intensity values in PA space of negative dominance. The regression planes for AU4, AU9, AU15, AU17, AU25, AU29, AU30 left, AU33, AU35, and AU43 respectively do not show a significant difference with respect to their values of maximal intensity in PA spaces of negative vs. positive dominance.

Regarding the **PA space of positive dominance**, AU12 and AU18 have their maximal intensity values with respect to positive pleasure independently of the arousal value. AU25 and AU38 have their maximal intensity values with respect to positive pleasure and positive arousal while AU26, AU29, and AU30 right have their maximal intensity values with respect to positive pleasure and negative arousal. AU4 and AU17 have their maximal intensity values with respect to negative pleasure independently of the arousal value. AU9, AU10, AU15, AU27, AU31, and AU33 have their maximal intensity values with respect to negative pleasure and positive arousal. AU43 has its maximal intensity value with respect to negative arousal independently of the pleasure value.

Regarding the **PA space of negative dominance**, AU25 has its maximal intensity value with respect to positive arousal and positive pleasure. AU1, AU2, AU4, AU5, AU6, AU11, AU13, AU15, AU17, AU23, AU29, AU30 left, AU33, and AU39 have their maximal intensity values with respect to negative pleasure and positive arousal. AU35 has its maximal intensity value with respect to negative pleasure independently of the arousal value. AU9 has its maximal intensity value with respect to negative pleasure and negative arousal. As in the PA space of positive dominance, AU43 has its maximal intensity value with respect to negative arousal independently of the pleasure value. AUs that are not mentioned here have very low intensity values that do not appear to significantly influence the results.

By combining all regression planes obtained by the two-dimensional non-linear regression analysis, a facial expression corresponding to each point in PAD space is recomposed and a **facial expression repertoire** is reconstructed. The facial expression repertoire is comprised of facial expressions arranged in PAD space with respect to two values of dominance (positive vs. negative); (see Figure 4.12).

The restriction of the facial expression repertoire to two values of dominance was done in order to link the resulting facial expressions (see Section 4.3.5) to EMMA's

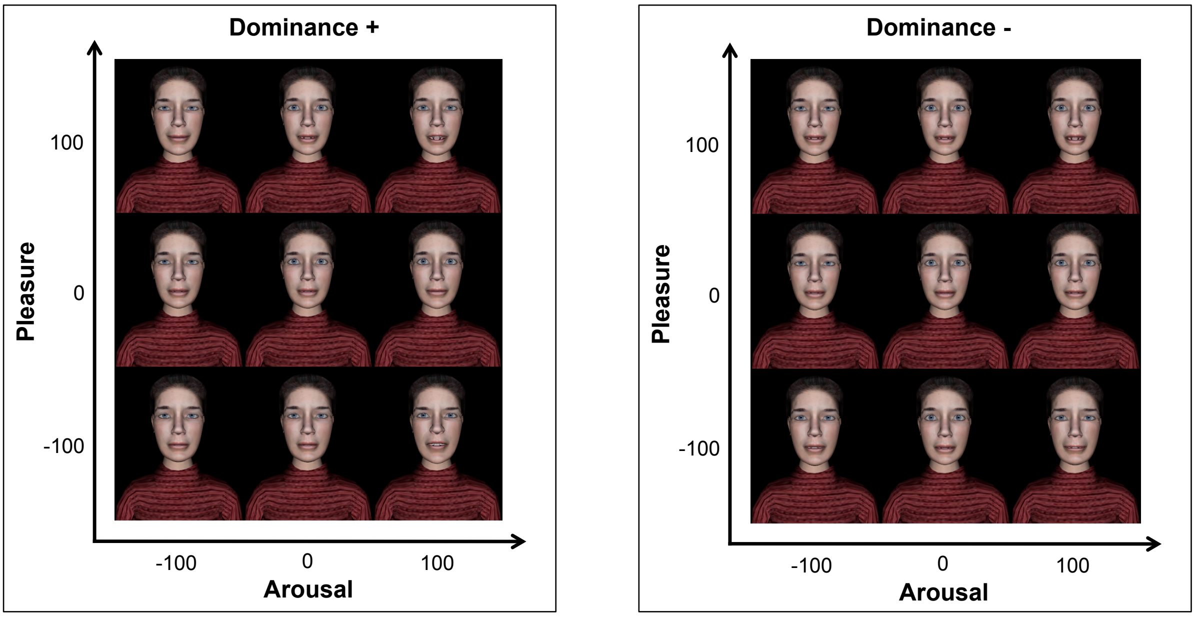


Figure 4.12: Facial expression repertoire based on the non-linear regression planes. Left: facial expression repertoire corresponding to PA space of positive dominance. Right: facial expression repertoire corresponding to PA space of negative dominance.

emotion simulation module [8], where only two values of dominance are considered (see Section 3.1.1). The 'SetPAD' field in the graphical user interface depicted by Figure 4.7 is used to visualize the resulting facial expression repertoire. By setting different values of pleasure, arousal, and dominance, the corresponding values of AUs are visualized in the field 'Action Units'.

4.3.4 Discussion and consequences

In this section, the results of the empirical study are discussed and the consequences for EMMA's facial expression repertoire are concluded. Note that the discussion of the results of the principal component analysis and of the regression analysis as well as the conclusion of their consequences for EMMA's facial expression repertoire are performed in the context of the present thesis.

Discussion

The results of evaluating EMMA's neutral facial expression are in line with those reported by Lee et al. [67] in the context of another empirical study. In the study by Lee et al. [67], the Extrinsic Affective Simon Task (EAST) used to measure the emotional value of target stimuli, was performed by 21 participants. As stimuli, two juxtaposed faces taken from pictures were presented. The juxtaposed faces showed *happy-happy*, *fearful-fearful*, *neutral-neutral*, and *happy-fearful* expressions. The subjects were asked to press a button labeled as positive or negative to rate the valence of the facial expressions. The reaction times of the participants and their emotional evaluations had been measured according to the EAST procedure. The idea behind this procedure is that the reaction time is shorter when the button's label corresponds to the presented facial expression. As a result, they found that the neutral facial expression was similarly rated as the fearful expression. This is in line with the evaluation of **EMMA's neutral facial expression** as aroused, displeased, and submissive (see Table 4.1), which corresponds to the emotional state *fearful* (cf. [101]).

Evidence for the existence of basic dimensions of emotions is provided by several dimensional theories (see Section 2.2). These basic dimensions are usually called pleasure-displeasure as the first, most agreed upon dimension, arousal-nonarousal as the second agreed upon dimension, and dominance-submissiveness as the third, less agreed upon dimension. However, in the present study, the principal component analysis provided a three factor solution with dominance as the first factor which accounts for 25% of variance in the data, pleasure as the second factor which accounts for 22.5% of variance, and arousal as the third factor which accounts for 19.2% of variance. In a previous study by Grammer & Oberzaucher [46] (see Section 3.2.1), the principal component analysis provided a solution with pleasure as the first factor which accounts for 27% of variance, dominance as the second factor which accounts for $24,5\%$ of variance, and arousal as the third factor which accounts for $16,7\%$ of variance. Compared to these findings, in the present study, the arousal factor also represents the third and last factor provided by the principal component analysis. This is further in line with the findings by Scholsberg [105] that arousal represents the third and last dimension after a second attention-rejection dimension (see Section 2.2).

Altogether, findings from the principal component analysis provide further support for the existence of **three basic dimensions** of emotions, and for the reliability of the 'Semantic Differential Measures of Emotional State or Characteristic (Trait) Emotions' [78]. Furthermore, the disagreement on the order of importance of the three basic dimensions highlights the importance of all three dimensions.

Further, Grammer & Oberzaucher [46] published three-dimensional non-linear regression planes for **AU1**, **AU5**, and **AU12** in PA space without considering the dominance dimension (see Section 3.2.1). According to these regression planes (see Figure 3.10, p. 53), **AU1** has its maximal intensity value with respect to positive pleasure independently of the arousal value. **AU5** has its maximal intensity value with respect to positive pleasure and positive arousal. **AU12** has its maximal intensity value with respect to positive pleasure and positive arousal. Quite similarly to these findings, the results of the present study (see Section 4.3.3 and Appendix A, p. 215) show that in PA space of positive dominance, **AU1** has its maximal intensity value with respect to positive pleasure and positive arousal, and **AU12** has its maximal intensity value with respect to positive pleasure independently of the arousal value. However, in PA space of negative dominance, **AU1** and even **AU12** have their maximal intensity values with respect to negative pleasure. With regard to **AU5**, the regression plane in PA space of positive dominance shows a maximal intensity value with respect to negative pleasure independently of the arousal value. The regression plane in PA space of negative dominance shows a maximal intensity value with respect to negative pleasure and positive arousal. Thus, both regression planes for AU5 provided by the present study are very different to that found by Grammer & Oberzaucher [46].

In a previous study by Snodgrass [111], **AUs and AUs combinations** were rated in terms of PA values. Some representative results of the study by Snodgrass [111] are summarized in Figure 2.3 (left, p. 33). Compared to these findings, in the present study (see Section 4.3.3 and Appendix A, p. 215), the regression planes for **AU1** show a maximal intensity value with respect to negative pleasure although accompanied by positive arousal in PA space of positive dominance. The regression planes of **AU2** show a maximal intensity value with respect to positive pleasure and positive arousal in PA space of positive dominance. The regression planes for both **AU4** and **AU5** show a maximal intensity value with respect to negative pleasure and positive arousal. The regression planes for **AU6** show a maximal intensity value with respect to positive pleasure and positive arousal in the PA space of positive dominance. The regression planes for **AU7**, **AU9**, and **AU10** show a maximal intensity value with respect to negative pleasure and positive arousal in PA space of positive dominance. Both regression planes of **AU15** show a maximal intensity value with respect to negative pleasure but accompanied by positive arousal. The regression planes for **AU17** show a maximal intensity value with respect to negative pleasure and positive arousal in PA space of negative dominance.

With regard to AUs combinations (cf. Figure 2.3, left, p. 33), the results of the present study show that a combination of the regression planes of **AU1 and AU2** results in a maximal intensity value of both AUs with respect to positive pleasure and positive arousal in PA space of positive dominance. A combination of the regression planes for **AU1 and AU4**, for **AU1, AU2, and AU4**, and for **AU1, AU2, and AU5**, respectively, results in a maximal intensity value of the AUs with respect to negative pleasure and positive arousal in PA space of negative dominance. A combination of the regression planes of **AU4 and AU5** results in a maximal intensity value of both AUs with respect to negative pleasure and positive arousal in PA spaces of both positive and negative dominance. A combination of the regression planes of **AU5 and AU7** results in a maximal intensity value of both AUs in PA spaces of both positive and negative dominance.

In sum, the results provided by the present study replicate only partly those provided by Snodgrass's study [111]. A crucial reason could be the evaluation of AUs combinations only, and the additional consideration of dominance in the present study. However, both the present study and Snodgrass's study [111] are very similar in their objectives which are to investigate the meanings of AUs in PA(D) space.

The visualization of the **facial expression repertoire** shows quite accurate facial expressions (see Figure 4.12). With regard to pleasure and arousal, the organization of the facial expressions in the facial expression repertoire has some similarities to that in the facial expression repertoire reconstructed by Russell [100] (see Figure 2.3, right, p. 33). One example is, the expression of positive pleasure and positive/neutral arousal in PA space of positive dominance. Further, in Russell's [100] facial expression repertoire there is no differentiation between positive and negative dominance. Accordingly, a **highlight** of the present study is that the facial expressions corresponding to negative pleasure, positive arousal, and positive dominance show more angry-like expressions while the facial expressions corresponding to negative pleasure, positive arousal, and negative dominance show more fearful-like expressions (see Figure 4.12). This finding further substantiates those by Russell & Mehrabian [102] that *anger* and *anxiety* have respectively a positive and a negative amount of dominance, while having similar amounts of negative pleasure and positive arousal (see Section 2.2).

Although, some of the facial expressions in the repertoire are difficult to interpret in terms of their corresponding PAD values, e.g., the facial expression of positive pleasure, positive arousal, and negative dominance. Another example is, the facial expression of negative pleasure, negative arousal, and negative dominance where AU12 can be observed. Furthermore, most of the regression planes show very low intensity values across PA space. This results in **inexpressive faces** across PA space, e.g., the facial expression of neutral pleasure, positive arousal, and positive dominance is very similar to the neutral facial expression (see Figure 4.12). A scaling of the regression planes with respect to the *vertical axis* (intensity values) does not significantly increase the expressiveness of these expressions, and results in **exaggerated expressions** at the extremes of PAD space. While the aim is to link the facial expression repertoire to EMMA's PAD space of her emotion simulation module [8] (see Section 3.1.1), some of the facial expressions in the repertoire **do not correspond** to the primary emotion categories located in PAD space as following Russell & Mehrabian [102]. For example, the facial expression that is more likely to express the emotion category *sad* corresponds to negative pleasure, negative arousal, and positive dominance (see Figure 4.12) while this emotion category is located in EMMA's PA space of negative dominance (see Figure 3.2, right, p. 42).

Further, the present study has a general caveat. By generating random facial expressions, no rule was followed that defined which AUs really occur together in spontaneous human facial expressions. This could result in the activation of AUs that show contradictory effects for the observer and that thus might provide inappropriate ratings. Furthermore, the facial expressions was presented as stills and without contextual information possibly making the rating process difficult and conflicting.

All in all, the findings of the present study corroborate the existence of three basic dimensions of emotions and provide further support for Russell's [100] assumption that facial expressions convey primary information and that the information signaled by a facial expression is also present in its single components (cf. Section 2.3).

Consequences

As already discussed in the previous section, some of the facial expressions in the facial expression repertoire provided by the above outlined study are either difficult to interpret in terms of their corresponding PAD values, or are inexpressive across PAD space, or do not correspond to the defined primary emotion categories in EMMA's emotion simulation module [8]. Consequently, we decided to **adjust the AUs regression planes** so that they produce facial expressions that are more expressive, and that fit the PAD space of EMMA's emotion simulation module. To this end, **two steps** are carried out to amend the AUs regression planes provided by the above outlined study.

First, some of the AUs are difficult to interpret in terms of PAD values, e.g., AU29 and AU30 (see Appendix A, p. 215). In their FACS manual, Ekman et al. [40] defined a set of AUs that are linked to the six basic emotion categories (see Table A.2, p. 205). Further, Snodgrass [111] provided a set of AUs expressing pleasure, arousal, as well as corresponding emotion categories as arranged in Russell's circumplex model of emotion [100] (see Figure 2.3, p. 33). Following these studies, only AUs are considered that clearly characterize pleasure, arousal, dominance, as well as primary emotion categories as arranged within PAD space of EMMA's emotion simulation module [8] (cf. Figure 3.2, right, p. 42). Further, based on the results provided by these previous works, Table 4.2 shows the **target set of AUs** together with their link to EMMA's emotion categories and their respective PAD values.

EmoCat - PAD	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU15	AU16	AU17	AU20	AU24	AU25	AU26	AU27	AU38	AU39	AU43
Happy (80, 80, +/-100) (50, 0, +/-100)	-	X	-	-	X	-	-	-	X	-	-	-	-	-	X	X	X	-	-	-
Surprised (10, 80, +/-100)	-	X	-	X	-	-	-	-	-	-	X	-	-	-	X	X	X	-	-	-
Angry (-80, 80, 100)	-	-	X	X	-	X	X	X	-	X	X	X	X	X	X	X	X	X	-	-
Concentrated (0, 0, +/-100)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Annoyed (-50, 0, 100)	-	-	X	-	-	X	-	-	-	X	-	X	-	-	-	-	-	-	-	X
Sad (-50, 0, -100)	X	-	X	-	-	X	-	-	-	X	-	X	-	-	-	-	-	-	-	X
Fearful (-80, 80, -100)	X	X	X	X	-	X	-	-	-	X	X	X	X	-	X	X	X	-	X	-
Bored (0, -80, 100)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X
Depressed (0, -80, -100)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X

Table 4.2: The target set of AUs and their link to EMMA's emotion categories as well as their respective PAD values.

Second, the LOESS regression method [27] [28] (see Section 4.3.3) does not provide a regression function that can be easily represented by a mathematical equation. Therefore, the regression planes provided by this method are difficult to adapt. Consequently, a **linear regression** over the non-linear regression planes of the considered AUs was calculated by Prof. Grammer in the context of our cooperation with the department of anthropology at the University of Vienna. As a result, mathematical equations of the form $ax + by$ are obtained for each AU regression plane that allow for the intended adaptation of these planes. The values x and y, respectively, represent pleasure and arousal values. Thus, the regression planes obtained by the linear regression can be adapted by simply adjusting the a and b parameters of the equations. Note that negative values provided by the linear regression are set to 0 (cf. Section 4.3.3); (see Figure 4.13). For a list of the new linear regression planes, see Appendix A, p. 227.

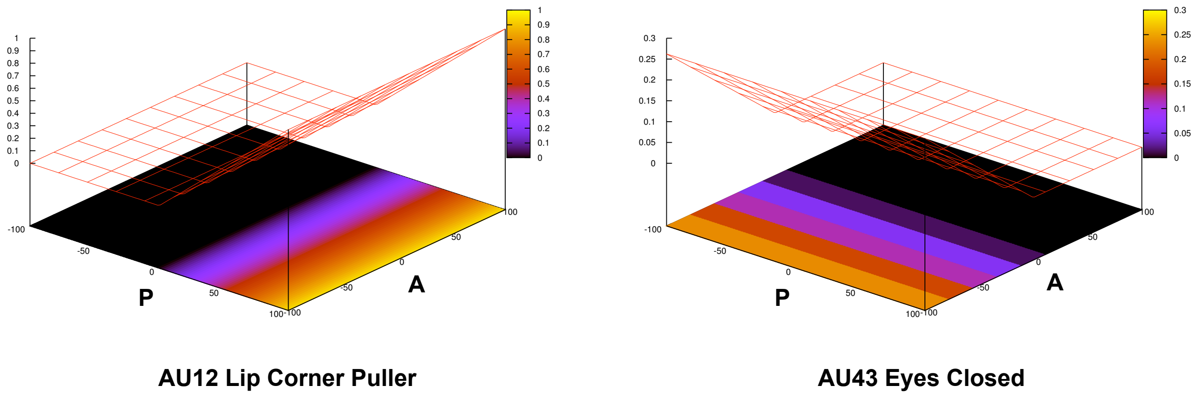


Figure 4.13: The linear regression planes corresponding to AU12 (Left) and AU43 (right) in PA space of positive dominance. The *vertical axis* represents the AUs intensity values.

According to Table 4.2, **AU1** mainly expresses negative pleasure and negative dominance. Thus, the regression plane for **AU1** (see Appendix A, p. 227) shows a maximal intensity value with respect to negative pleasure and negative dominance independently of the arousal value. Further, **AU2** mainly expresses positive arousal. Thus, in PA space of negative dominance, the regression plane for **AU2** shows a maximal intensity value with respect to positive arousal independently of the pleasure value. However, this is not replicated within PA space of positive dominance since **AU2** does not occur in a facial expression showing *anger*. Thus, the corresponding regression plane shows a maximal intensity value with respect to both positive pleasure and arousal.

AU4, **AU7**, **AU15**, and **AU17** mainly express negative pleasure. Thus, their regression planes show a maximal intensity value with respect to negative pleasure independently of the arousal value. **AU5**, **AU16**, and **AU20** mainly express negative pleasure and positive arousal. Thus, their regression planes show a maximal intensity value with respect to negative pleasure and positive arousal. **AU6** and **AU12** mainly express positive pleasure. Thus, their corresponding regression planes show a maximal intensity value with respect to positive pleasure independently of the arousal value. **AU9**,

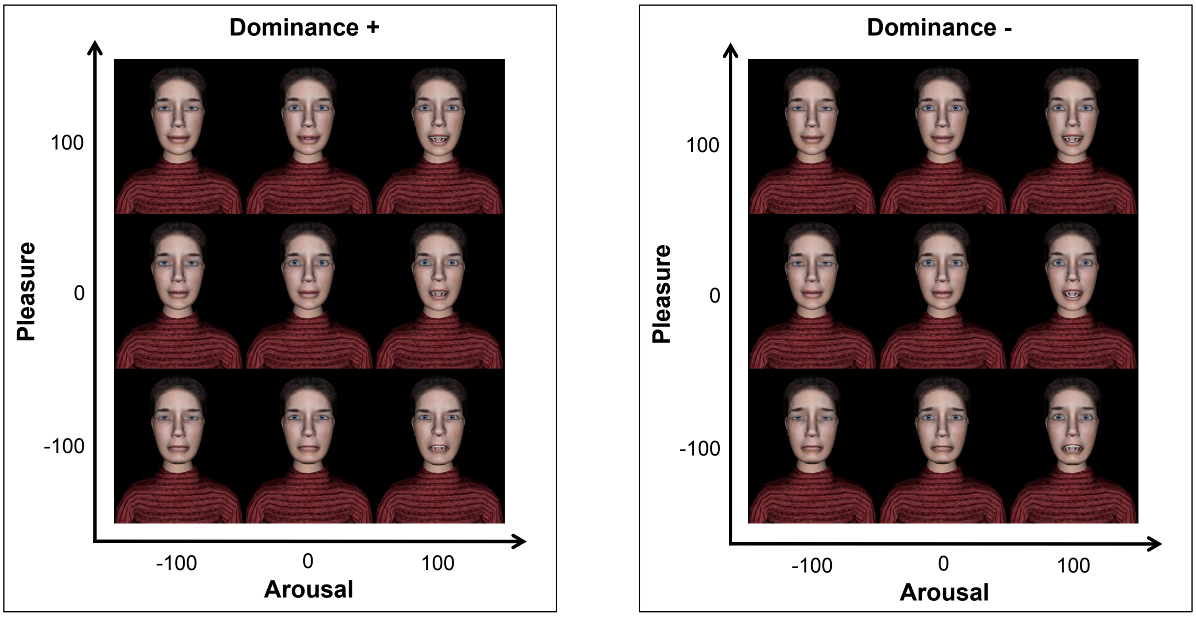


Figure 4.14: Facial expression repertoire based on the linear regression planes. Left: the new facial expression repertoire corresponding to PA space of positive dominance. Right: the new facial expression repertoire corresponding to PA space of negative dominance.

AU10, **AU24**, and **AU38** mainly express negative pleasure, positive arousal, and positive dominance. Therefore, their regression planes show a maximal intensity value with respect to negative pleasure, positive arousal, and positive dominance. **AU25**, **AU27**, and **AU26** mainly express positive arousal. Thus, their respective regression planes show a maximal intensity value with respect to positive arousal independently of the pleasure values. **AU39** mainly expresses negative pleasure, positive arousal, and negative dominance. Therefore, its regression plane shows a maximal intensity value with respect to negative pleasure, positive arousal, and negative dominance. **AU43** mainly expresses negative arousal and thus both of its regression planes show a maximal intensity value with respect to negative arousal independently of the pleasure value. Note, that the values of the maximal intensities for the AUs are fine tuned on the basis of the visualization of the resulting facial expressions with EMMA's face.

Again, by combining all of the regression planes, a **repertoire of facial expressions** arranged within PAD space with respect to two values of dominance (positive vs. negative) is reconstructed. The resulting facial expression repertoire is depicted in Figure 4.14.

As compared to the facial expression repertoire depicted in Figure 4.12, the new facial expression repertoire has more similarities to that reconstructed by Russell [100] (see Figure 2.3, right, p. 33) and has more **expressive** facial expressions. Some examples are the facial expressions of positive pleasure and positive arousal, of positive pleasure and neutral arousal, and of neutral pleasure and positive arousal. Furthermore, the generated facial expressions **correspond** to the emotion categories located in PAD space of EMMA's emotion simulation module [8] as following Russell & Mehrabian [102] (see Figure 3.2, right, p. 42). An **evaluation** of some of EMMA's facial expressions is carried out within an empathy context scenario and provided promising results (see Section 6.2).

While the linear regression provides a simple method to adapt the AUs regression planes with respect to PAD values, it results in a major loss of information and in a non-flexible way to define the meaning of AUs in PAD space. For example, according to FACS [40], AU6 occurs in facial expressions showing happiness and sadness. Thus, as compared to the non-linear regression planes, the linear regression planes do not allow AU6 to have a maximal intensity value with respect to positive pleasure and positive arousal, and also with respect to negative pleasure and negative arousal. In future work, one could try a non-linear polynomial function to re-establish the complete meaning of such AUs within PAD space and to get an appropriate smoother course of AUs intensity values. A major interest of the above outlined study and of further adaptation of its results is to explore the meaning of single AUs by trying to unify the current results with those provided by previous works, and to propose an approach to infer PAD values from facial expressions of emotions based on the provided AUs regression planes (see Section 5.2.1). In the following section, the dynamic link of EMMA's facial expression repertoire to PAD space of her emotion simulation module [8] is presented.

4.3.5 Facial expression dynamics

As presented in Section 3.1.1, EMMA's emotion simulation module [8] has as input, values of emotional valences also called *emotional impulses*. These values are either positive or negative and are triggered by the virtual human's, in this case EMMA, reactive or reasoning layer of her cognition module (see Section 4.2.1). The values of emotional impulses drive EMMA's emotion dynamics over time. At each point in time, the emotion module outputs values of pleasure, arousal, and one of two possible values of dominance as well as awareness likelihood values of primary and secondary emotions. Accordingly, EMMA has a repertoire of facial expressions arranged in PAD space with respect to two values of dominance (positive vs. negative); (see previous section). Therefore, each PAD value output by EMMA's emotion simulation module over time is expressed by its corresponding facial expression in the facial expression repertoire. Thus, EMMA's emotion dynamics drive her **facial expression dynamics** over time. Figure 4.15 shows the intensity values of EMMA's AU12 and of MAX's corresponding facial muscle (muscle smile), as well as the corresponding pleasure course in PA space of positive dominance over time. As compared to MAX, the intensity values of EMMA's AU12 over time reflect

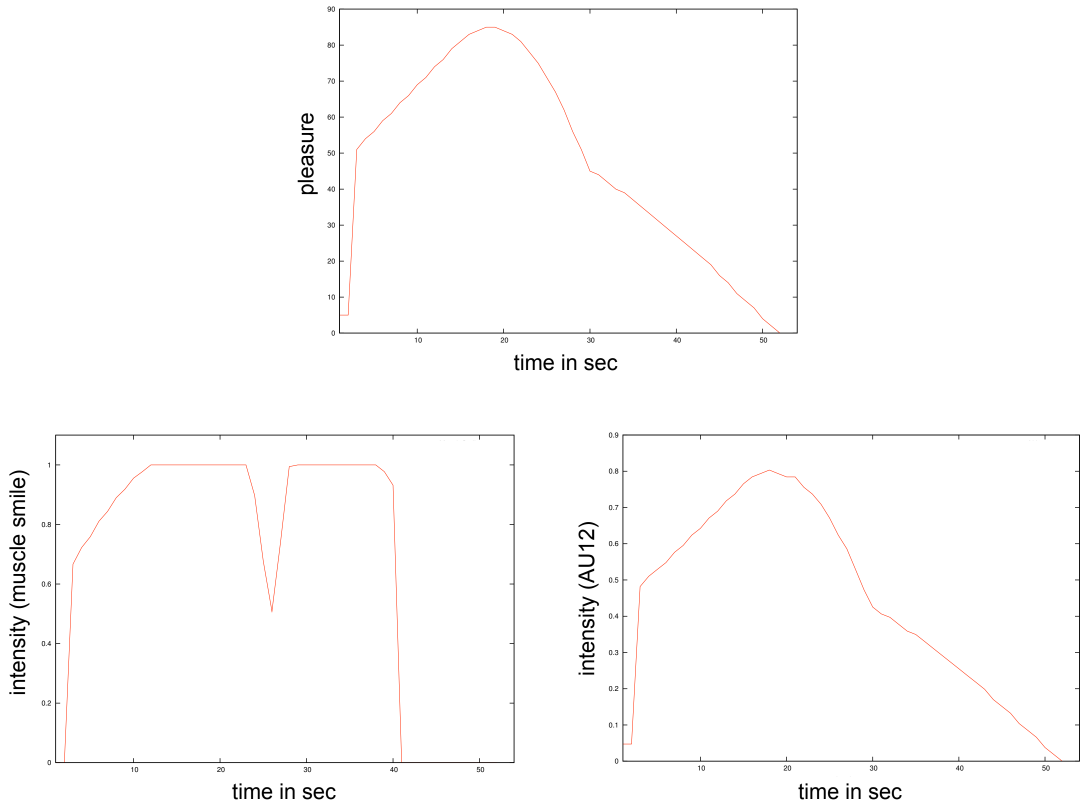


Figure 4.15: Top: pleasure course over time. Left: intensity values of MAX's smile muscle over time. Right: intensity values of EMMA's AU12 over time.

more the course of her values of pleasure over time. Figure 4.16 shows the intensity values of EMMA's AU27 and of MAX's corresponding facial muscle (muscle mouth open), as well as the corresponding arousal course in PA space of negative dominance over time. Also the intensity values of EMMA's AU27 over time better mimic the arousal course than those of MAX's corresponding facial muscle.

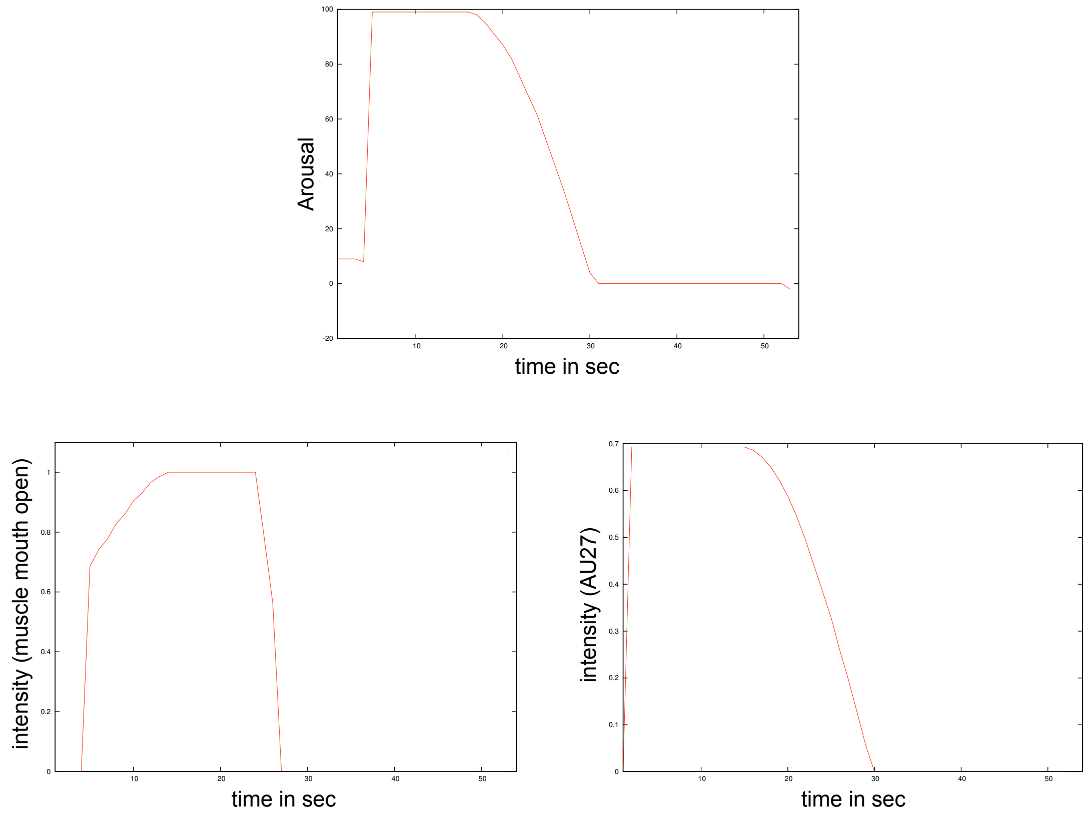


Figure 4.16: Top: arousal course over time. Left: intensity values of MAX's 'mouth open' muscle over time. Right: intensity values of EMMA's AU27 over time.

In sum, as compared to MAX, EMMA's facial expressions **reflect the trajectories** of her emotion dynamics within the PAD emotion space. Note that the space of negative arousal is not accessible to the virtual human's emotion dynamics, except during boredom in combination with neutral pleasure (cf. [8]). Unfortunately, EMMA's corresponding facial expressions cannot be displayed.

4.3.6 Conclusion

Based on the results of an empirical evaluation of EMMA's facial expressions, and on careful consideration of the results from previous research on the facial expression of emotion, we provided an **elaborate model of facial expressions** for the new virtual human EMMA. In this regard, using a **bottom-up approach** (cf. Section 2.3), we investigated the meaning of single AUs within the dimensional emotion space of pleasure, arousal, and dominance, and how it contributes to the meaning of a whole facial expression of emotion. Thus, as emphasized by Russell [100] for Snodgrass's work [111], we attempted to 'compile a **dictionary of the meaning** of elemental facial movements' (Russell [100], p. 301). The proposed regression planes for AUs were combined to result in a large repertoire of facial expressions which, regarding PA values, is quite similar to that reconstructed by Russell on the basis of Snodgrass's results (see Figure 2.3). Furthermore, the resulting facial expressions also correspond to emotion categories as located in PAD space following findings by Russell & Mehrabian [102] (cf. [8]) and reflect the trajectories of the time course of emotions within PAD space. Accordingly, the proposed AUs regression planes in PAD space can be further used and **easily adapted** for virtual humans other than EMMA whose faces, as EMMA's, are modeled following the FACS standard [40].

4.4 Summary

This chapter introduced the **new virtual human** EMMA developed at the AI-Group at Bielefeld University as the female counterpart to the virtual human MAX [62]. In Section 4.1, we gave an overview of the history and motivation for developing a new virtual human and distinguished **three important aspects** in creating EMMA. First, providing an elaborate model of facial expressions (see Section 4.3) based on linking AUs from FACS [40] to the dimensional emotion space of pleasure, arousal, and dominance [102]. Second, allowing for the consideration of another kind of interaction in addition to human-agent interaction, namely, agent-agent interaction (e.g., see Section 6.1.1). Third, allowing for the modeling and evaluation of gender differences. While the first two aspects are crucial to the present thesis, the third important aspect in creating EMMA could be the objective of future work. Further, we described the different steps taken in the design of EMMA's appearance, which was initialized by former members of our group and finalized in cooperation with the department of anthropology at the University of Vienna [56].

In Section 4.2, we introduced the name 'EMMA' as an acronym that refers to an **E**mpathic **M**ulti**M**odal **A**gent in accordance with EMMA's capabilities within her first interaction scenario (see Section 6.1.1). In this regard, EMMA's model of facial expressions (see Section 4.3) is crucial for the realization of the empathy mechanism facial mimicry (see Section 5.2.1) and for the expression of empathy (see Section 5.4). To this end, the virtual human MAX's cognitive architecture [70] was carried over, adapted, and extended for the new virtual human EMMA. Adaptations and extensions were carried out for the facial expression component, for the speech component, and for the component that generates and controls body movements. Based on **FACS** [40] as an eligible standard for the adequate modeling of facial expressions (cf. Section 3.2.1), the facial expression component was extended to the specification and generation of facial behaviors on the basis of AUs. Since dimensional emotion models are considered as more inclined to characterize the continuity and subtlety of emotion expression than other emotions models (cf. Section 3.2.1), the facial expression component was also extended to a facial expression repertoire linking AUs to the **dimensional emotion space** of pleasure, arousal, and dominance [102]. Also the speech component was extended to the simulation of emotional speech prosody on the basis of PAD values (cf. [107] [106]). The component to generate body movements was adapted to the specification of the lengths of EMMA's limbs, of the number of her joints, and of the transformations within EMMA's body-centered coordinate system. Thus, similar to the virtual human MAX, EMMA is capable of exhibiting multimodal behavior based on facial expressions, speech, and body movements. However, the new virtual human can express a wider range of emotions along the PAD dimensions using her face and speech prosody.

In Section 4.3, we introduced the steps taken for the development of EMMA's **facial expression repertoire**. In cooperation with the department of anthropology at the University of Vienna, EMMA's AUs were modeled with the help of experienced FACS coders [56] and EMMA's visemes were specified based on AUs. A total of 44 AUs were implemented for EMMA. Based on the results of an empirical study, a facial expression repertoire was provided by linking AUs to PAD values. The empirical evaluation was performed in cooperation with the department of anthropology at the University of Vienna [56]. In this study, human subjects rated randomly generated facial expressions of EMMA in terms of PAD values. As a result, three-dimensional **non-linear regression planes** were provided for each AU showing its meaning in PAD space. A facial expression repertoire was reconstructed by combining all regression planes for all AUs, and includes quite accurate facial expressions. As such, the results provide further support for Russell's [100] assumption that facial expressions convey primary information and that the information signaled by a facial expression is also present in its single components (cf. Section 2.3).

However, some of the facial expressions provided by the empirical study are either difficult to interpret in terms of PAD values, or are inexpressive, or do not correspond to EMMA's emotion categories as arranged within her PAD emotion space following [102] (cf. [8]). Accordingly, we decided to adapt the AUs regression planes by unifying the results provided by the present study, and those of previous work on the facial expression of emotion. For this purpose, **linear regression planes** for the AUs were used for an easy adaptation of the planes by simply adjusting the plane coefficients. By combining the new regression planes, a new facial expression repertoire was reconstructed comprised of facial expressions that are more expressive, and quite similar to those defined in Russell's proposed repertoire [100] (see Figure 2.3, p. 33). Furthermore, the resulting facial expressions also correspond to emotion categories as located in PAD space following Russell & Mehrabian [102] (cf. [8]), and reflect the trajectories of the time course of emotions within PAD space. Accordingly, the proposed AUs regression planes in PAD space can be used and easily adapted for virtual humans other than EMMA whose faces, like EMMA's, are modeled following the FACS standard [40]. However, the linear model results in a major loss of information and in a non-flexible way to define the meaning of AUs within PAD space. All in all, based on the results of an empirical evaluation of EMMA's facial expressions and on careful consideration of the results of previous works on the facial expression of emotion, we provided an **elaborate model of facial expressions** for the new virtual human EMMA. This model of facial expressions is crucial to the realization of facial mimicry as an empathy mechanism (see Section 5.2.1) and for the expression of empathy (see Section 5.4) as requirements for building a computational model of empathy (cf. Section 1.2).

5 A computational model of empathy

This chapter introduces a computational model of empathy realized for the virtual humans MAX and EMMA. Section 5.1 gives an overview of the structure of the proposed model and its integration into an existing cognitive architecture. Subsequently, Section 5.2 introduces two separate empathy mechanisms by which an empathic emotion is produced. In Section 5.3, an approach to modulate the empathic emotion produced by the empathy mechanisms is proposed. Further, Section 5.4 presents, the different modalities by which empathy is expressed.

5.1 Model overview

In this section, we give an overview of the structure of the proposed computational model of empathy as well as of its integration into an existing cognitive architecture.

5.1.1 Model structure

According to the requirements **Empathy Mechanism**, **Empathy Modulation**, and **Expression of Empathy**, formulated in Section 1.2 and further discussed in Section 3.3.3, the computational model of empathy proposed here is based on three processing steps:

1. The empathy mechanism as the process by which an empathic emotion is generated. Two separate mechanisms are considered, facial mimicry and situational role-taking.

2. The empathy modulation as the process by which the empathic emotion produced in step 1 is modulated and a degree of empathy is calculated.

3. The expression of empathy as the process by which the empathic emotion modulated in step 2 is expressed.

Furthermore, our model of empathy follows the **late appraisal model of empathy** [35] (cf. Section 2.1) in that the empathic emotion is produced first and than modulated. Therefore, in our model, the empathy mechanism takes place before the empathy modulation. Figure 5.1 depicts the structure of the computational model of empathy in terms of its three processing steps.

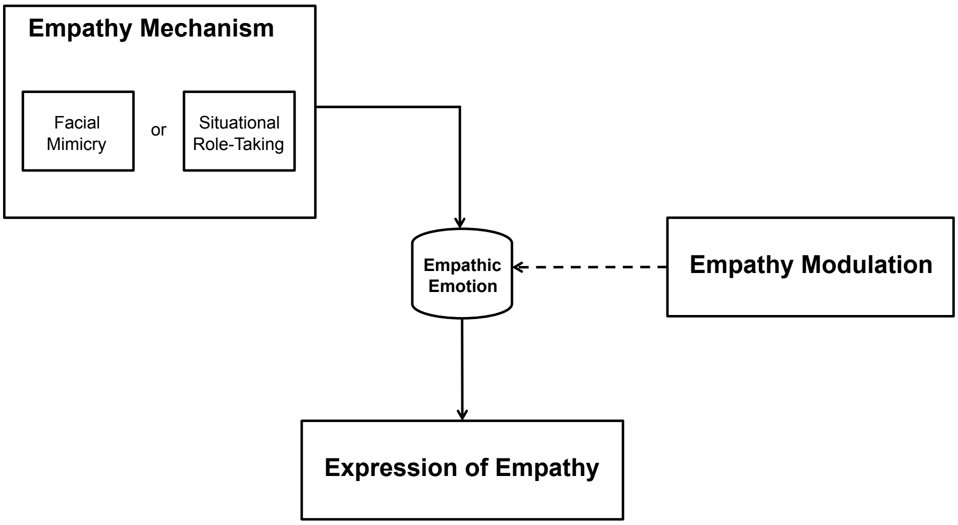


Figure 5.1: The structure of the proposed computational model of empathy in terms of its three processing steps.

The three processing steps underlying our model are introduced respectively in Sections 5.2, 5.3, and 5.4. In the following, we discuss the integration of the proposed model into an existing cognitive architecture.

5.1.2 Integration into a cognitive architecture

According to Section 3.1, WASABI [8] (see Section 3.1.1) presents a suitable computational model of emotion for the realization of several theoretical aspects of empathy (cf. Chapter 2). In this regard, WASABI's appraisal component is crucial to the cognitive evaluation of others' situations and thus for the generation of an empathic emotion. Furthermore, the simulation of emotion dynamics within WASABI is crucial to the simulation of the time course of an empathic emotion, and to its modulation through the empathizer's changing mood over time [35] (cf. Section 2.1.3). Therefore, we integrated our model into **WASABI** which represents the emotion simulation module of the virtual humans MAX's and EMMA's cognitive architecture [70] (see Sections 3.1.2 and 4.2.1). Figure 5.2 depicts the cognitive architecture of MAX and EMMA with the empathy model as integrated into the emotion simulation module of the architecture.

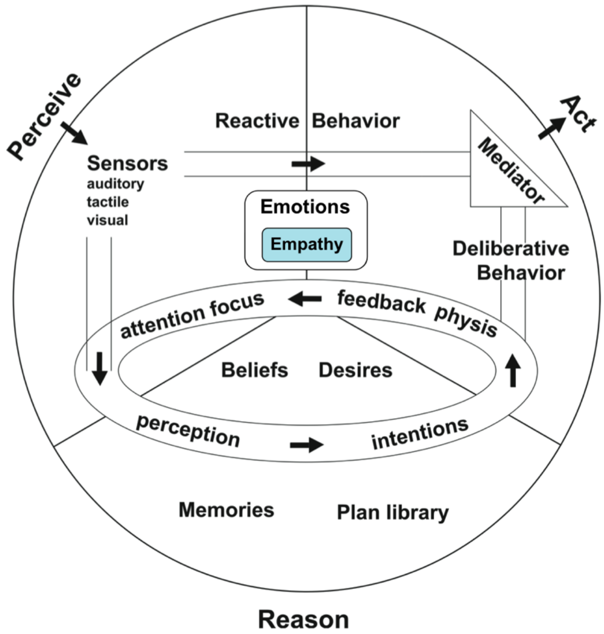


Figure 5.2: The cognitive architecture [70] of the virtual humans MAX and EMMA with the proposed computational model of empathy integrated into the emotion simulation module [8].

In the following sections, the three processing steps underlying our model are introduced.

5.2 Empathy mechanisms

According to the theoretical background on empathy introduced in Section 2.1, there are several mechanisms by which an empathic emotion is produced. As already emphasized in Section 3.2, facial expressions are crucial to expressing and communicating emotions [32] [38] and are thus central to the present thesis. Therefore, in our model, one considered empathy mechanism is **facial mimicry**. While expressive cues such as facial expressions may present a valid source of information about others' emotional states, contextual cues are another possible source of information based on the appraisal of others' situations (cf. Section 3.2.2). In this regard, in our model, we also consider the empathy mechanism **situational role-taking**. Facial mimicry and situational role-taking are considered separately and are applied to two different interaction scenarios (see Section 6.1).

5.2.1 Facial mimicry

In their theoretical models of empathy (see Section 2.1.2), both Hoffman [55] and Davis [34] introduce *mimicry* as an empathy mechanism. Hoffman defines mimicry as a process involving the imitation of another's facial expressions, voice, and posture that triggers an afferent feedback, eliciting feelings in the observer that are similar to those of the observed other. Thus, he identifies two successive steps underlying the process of mimicry which he calls *imitation* and *feedback*. Accordingly, **facial mimicry** is defined by Hoffman as the process involving the imitation of another's facial expressions that triggers an afferent feedback, eliciting the same feeling in oneself as that of the observed other. Following Hoffman, in our model, facial mimicry consists of an **internal imitation** of perceived emotional facial expressions, which results in an **emotional feedback** that represents the perceived emotional state.

Further, according to the *shared neural network hypothesis* (cf. Section 2.1.3), the observation of another's emotional state activates brain areas involved in experiencing that same state. This suggests the existence of a *shared representational system* involved in understanding others' emotions. Following this hypothesis, in our model, facial mimicry is based on the use of a **shared representational system**. That is, the same facial expression repertoire is used to express one's own emotions as well as to understand emotions from others' perceived facial expressions.

Damasio [31] distinguishes between primary and secondary emotions (see Section 2.2). While primary emotions are experienced early in life as preorganized reactive responses to specific features of perceived stimuli, secondary emotions are experienced by adults as responses to given situations based on a cognitive evaluation. Following Hoffman [55] (see Section 2.1.2), facial mimicry or mimicry in general is preverbal, automatic, and essentially involuntary. Thus, the emotion produced by this mechanism – the empathic emotion – corresponds to a **primary emotion**. Accordingly, in our model, facial mimicry generates a primary emotion as defined by the underlying emotion simulation module [8] where primary and secondary emotions are distinguished (see Section 3.1.1). The empathic emotion is thus produced within the **reactive layer** of the cognitive architecture

[70] (see Section 5.1.2).

As mentioned above, in our model, facial mimicry is based on the use of a shared representational system. Accordingly, using her own **A**ction **U**nits (AUs) and their intensity functions (regression planes) within **P**leasure-**A**rousal-**D**ominance (PAD) space (cf. Section 4.3), facial mimicry is realized for EMMA as follows:

1. **Internal imitation**: EMMA internally imitates a perceived facial expression by mapping perceived facial features to her own AUs.

2. **Emotional feedback**: the AUs resulting from *internal imitation* step are mapped to a PAD value representing the perceived emotional state.

Figure 5.3 depicts our proposed approach to facial mimicry as based on internal imitation and emotional feedback, and on using a shared representational system.

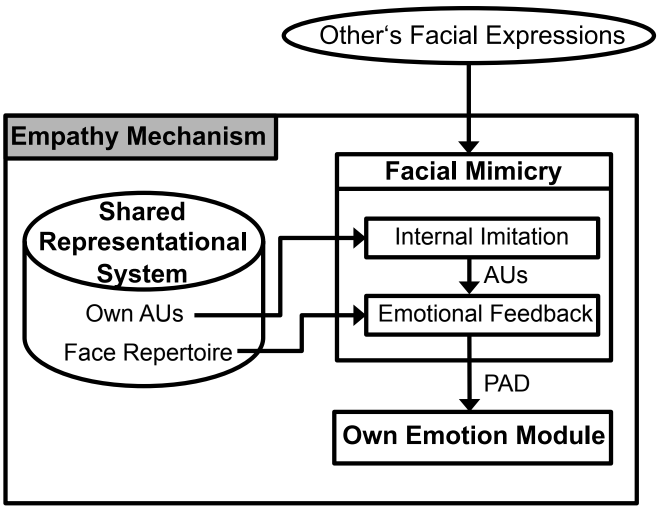


Figure 5.3: The empathy mechanism **facial mimicry** as based on internal imitation and emotional feedback, and on using a shared representational system.

In our realization of facial mimicry, we mainly focus on **emotional feedback**. In this regard, for internal imitation, we assume having a system of automatic AU recognition as those presented in Section 3.2.2. Accordingly, in order to determine a PAD value from AUs intensity values representing a perceived facial expression, the following two equations are used:

$$\begin{aligned}\sum_{i=1}^{n} AU_i(p,a,d) \cdot w_i(t) = AU_{\sum}(p,a,d,t) \\ (p,a,d)_{t,hint} = AU_{max}(t)\end{aligned} \tag{5.1}$$

$$\begin{aligned}\sum_{i=1}^{n} |AU_i(p,a,d) - w_i(t)| = AU_{\sum}(p,a,d,t) \\ (p,a,d)_{t,final} = AU_{min}(t)\end{aligned} \tag{5.2}$$

First, in **Equation 5.1**, the AUs intensity functions corresponding to each input AU are weighted with the corresponding input intensity values. The weights $w_i(t)$ represent the AUs input intensity values at each point in time t. Subsequently, a **sum** of the weighted AUs intensity functions is calculated. The PAD value corresponding to the **maximum** of the weighted summation is returned and is called $(p,a,d)_{t,hint}$. Since most of the AUs intensity functions have their maximum values at the boundaries of PAD space (see Appendix A, p. 227), the resulting PAD value is thus likely to lie at the boundaries of the space. This PAD value is further used to determine in which quadrant of PAD space the sought PAD value could lie. Finally, the domain of each considered AU intensity function is restricted to that quadrant of PAD space. The idea behind this method is to determine where the considered AUs have a **common meaning** within PAD space.

For example, consider a perceived facial expression composed of **AU12** and **AU43** with respective intensity values equal to 0.4 and 0.2 at time-stamp t. Now, consider the respective intensity functions for AU12 and AU43 in PA space of positive dominance (see Figure 5.4). According to these intensity functions, AU12 has a maximal intensity value with respect to positive pleasure independently of the arousal value while AU43 has a maximal intensity value with respect to negative arousal independently of the pleasure value. Thus, both AUs are expected to have a common meaning with respect to positive pleasure and negative arousal. By applying Equation 5.1, the intensity functions for AU12 and AU43 are respectively weighted with the values 0.4 and 0.2 and a sum of the weighted functions is calculated. As a result, $(p,a,d)_{t,hint}$ is equal to $(100,-100,100)$ (see Figure 5.5).

Second, in **Equation 5.2**, the AUs intensity functions with restricted domain are

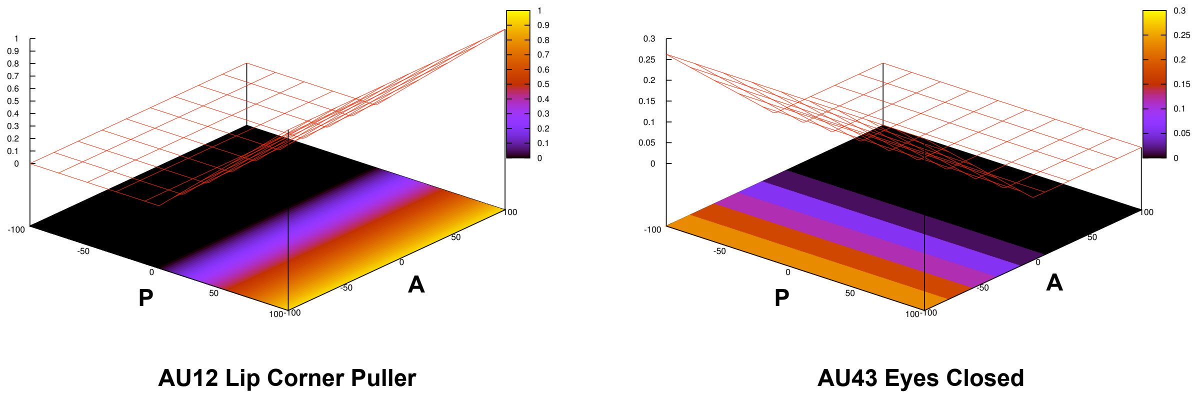


Figure 5.4: The linear regression planes corresponding to AU12 (Left) and AU43 (right) in PA space of positive dominance. The *vertical axis* represents the AUs intensity values.

translated with respect to the input intensity values – $w_i(t)$ – of their corresponding AUs. Subsequently, a **sum** of their absolute values is calculated. The PAD value corresponding to the **minimum** of the calculated sum denoted by $(p, a, d)_{t,final}$ is returned as the sought PAD value at time-stamp t.

Again, as an example, consider a perceived facial expression composed of **AU12** and **AU43** with respective intensity values equal to 0.4 and 0.2 at time-stamp t. The domains of the intensity functions for AU12 and AU43 are restricted to $[0, 100]$ for the pleasure dimension, and to $[-100, 0]$ for the arousal dimension. By applying Equation 5.2, the intensity functions for AU12 and AU43 are respectively translated to -0.4 and -0.2 and a sum of their respective absolute values is calculated. As a result, $(p, a, d)_{t,final}$ is equal to $(40, -77, 100)$; (see Figure 5.6). Increasing the input intensity value of AU12 would increase the value of positive pleasure while increasing the input intensity value of AU43 would increase the value of negative arousal. That is, smiling (AU12) with eyes closed (AU43) is interpreted as expressing positive pleasure and negative arousal.

Note that the calculation of $(p, a, d)_{t,hint}$ may provide more than one solution. Accordingly, the AUs in question have a common meaning in more than one quadrant of PAD space. Thus, the domain of each considered AUs intensity function is restricted to those quadrants of PAD space. It is also possible that the calculation of $(p, a, d)_{t,final}$ provides more than one solution. In this case, a possible way to get an **explicit value** of $(p, a, d)_{t,final}$ is to calculate a mean value of all possible solutions.

For example, consider only one AU, e.g., **AU12** with an input intensity value of 0.4.

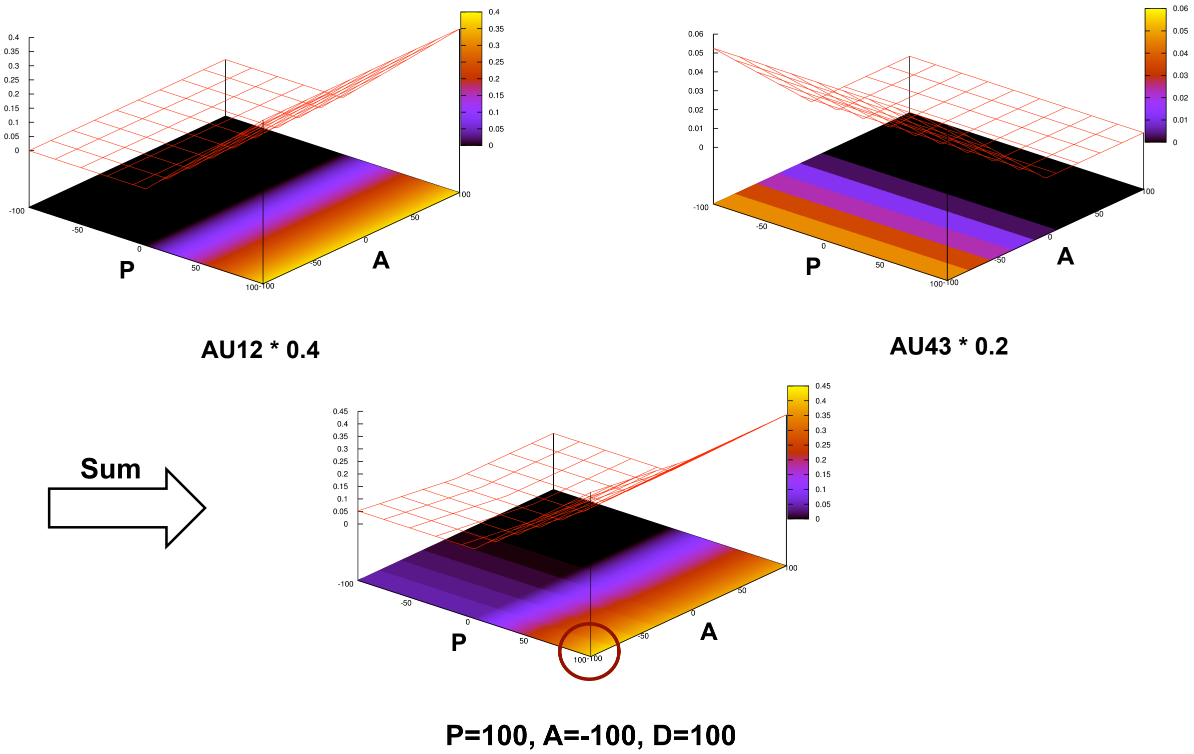


Figure 5.5: Top: the intensity functions of AU12 and AU43, respectively, weighted with the values 0.4 and 0.2. Bottom: the sum of the weighted intensity functions of AU12 and AU43 showing the value of $(p, a, d)_{t,hint}$.

The calculation of $(p, a, d)_{t,final}$ provides one solution for pleasure, many solutions for arousal, and two solutions for dominance (positive vs. negative) (see Appendix A, p. 227). By calculating the mean value of all possible solutions, a PAD value is provided that is equal to $(40, 0, 0)$ where the values 0 can be interpreted as: 'AU12 alone neither expresses positive nor negative arousal and dominance'.

However, this approach becomes less appropriate when, e.g., considering a facial expression with AUs displaying similar amounts of negative and positive pleasure, arousal, and dominance. Such facial expression is then interpreted as displaying neutral pleasure, neutral arousal, and neutral dominance. This result could also be provided by the approach proposed by Shugrina et al. [109] regarding pleasure and arousal (see Section 3.2.2). In this case, further involvement of **context related information**, e.g., through situational role-taking or self projection as realized in the model by Rodrigues et al. [98]

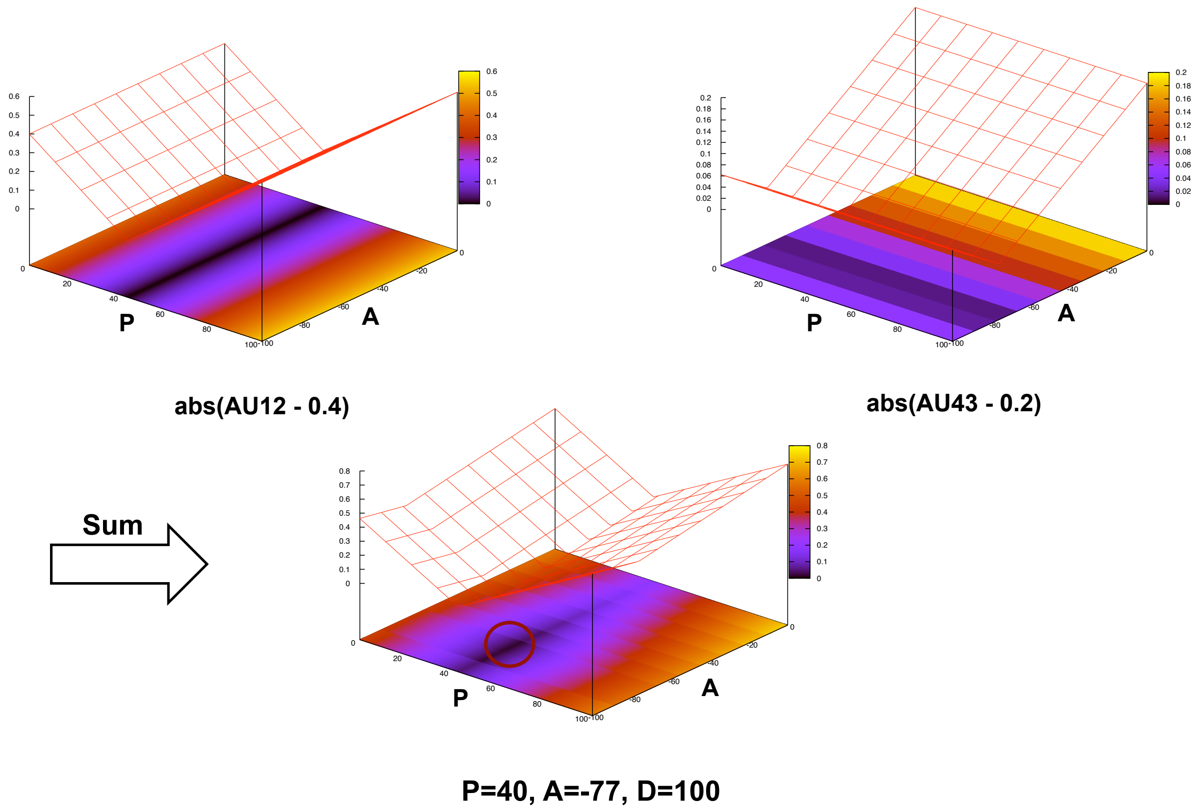


Figure 5.6: Top: the absolute values of the intensity functions of AU12 and AU43, respectively, translated to the values -0.4 and -0.2. Bottom: the sum of the above functions showing $(p, a, d)_{t,final}$. '**abs**' refers to the absolute value.

(see Section 3.3.1), may provide a more appropriate solution. Accordingly, in our approach, context related information is considered to get an explicit value of dominance (see Sections 5.2.2 and 6.1.1). Regarding pleasure and arousal, we rely on the calculation of the mean value of all possible solutions while the consideration of context related information to get an explicit solution should be considered in future work.

Further, as mentioned in Section 4.3.3, not only does the presence of an AU in a facial expression contribute to the meaning of that expression but also its absence. Hence, the AUs that are not activated in a perceived facial expression are considered in Equations 5.1 and 5.2 as having intensity values equal to 0. In order to easily explain these equations, non-activated AUs were not included in the above given examples.

The **appropriateness** of this approach was tested by applying it to the facial expression repertoire reconstructed on the basis of the linear AUs intensity functions and by using these functions. Further, it was also tested by applying it to the facial expression repertoire reconstructed on the basis of the non-linear AUs intensity functions and by using these functions (see Appendix A, p. 215). In both cases, the determined PAD values are exactly the same as those related to the input facial expressions.

Once a PAD value is determined by this approach, it is represented within PAD space of EMMA's emotion simulation module [8] (see Figure 5.3). The determined PAD value represents a **hypothesis** about the emotional state related to a perceived facial expression.

First results of this approach are investigated using the virtual human **MAX's facial expressions of emotion**. As mentioned in Section 3.2.1, the virtual human MAX has 21 facial muscles and can express different intensities of seven facial expressions of emotion. By applying facial mimicry, EMMA first internally imitates MAX's facial expression by getting the values of his facial muscles at each point in time and by mapping them to her own AUs with corresponding intensity values. Figure 5.7 depicts the mapping of MAX's facial muscles involved in his facial expressions of emotion to EMMA's AUs.

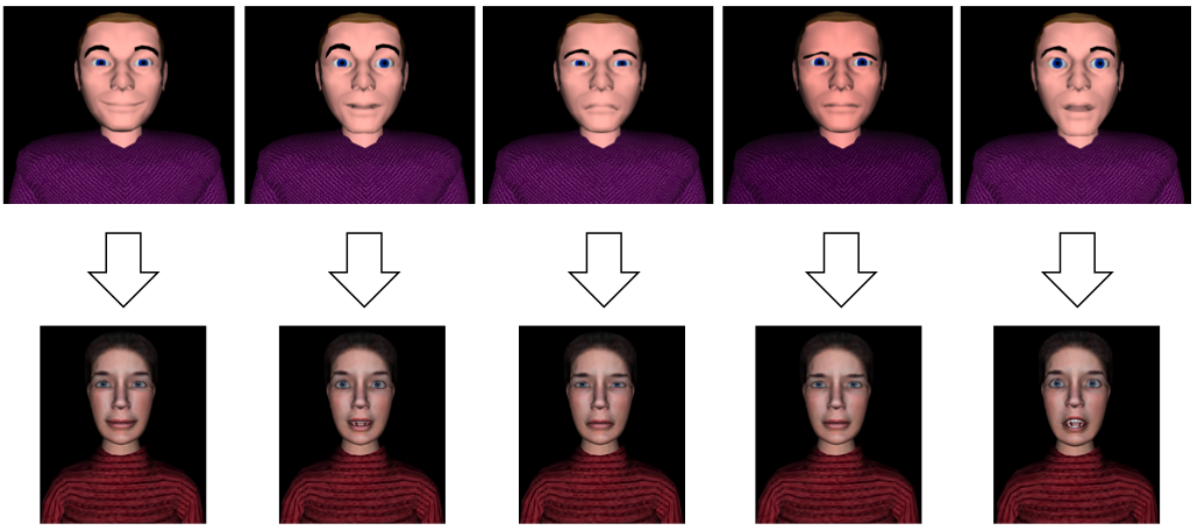


Figure 5.7: Top: MAX's facial expressions for *happy*, *surprised*, *annoyed/sad*, *angry*, and *fearful*. Bottom: EMMA's facial expressions resulting from internal imitation of those of MAX.

Subsequently, a PAD value is determined from the values of AUs provided by internal imitation of MAX's facial expression. The PAD values determined from MAX's facial expressions of emotion at their maximal intensity values are summarized in Table 5.1.

Note that MAX's facial expression corresponding to the neutral emotional state *concentrated* has some slightly activated facial muscles. As mentioned above, MAX's inactive facial muscles in a facial expression are considered in Equations 5.1 and 5.2 as having intensity values equal to 0.

MAX's Emo. in PAD space	$(p, a, d)_{t,final}$
Happy: (80, 80, +/-100), (50, 0, +/-100)	(96, -5, +/-100)
Surprised: (10, 80, +/-100)	(21, 66, +/-100)
Annoyed: (50, 0, 100)	(-96, -44, +/-100)
Sad: (50, 0, -100)	(-96, -44, +/-100)
Angry: (-80, 80, 100)	(-77, 8, 100)
Fearful: (-80, 80, -100)	(-75, 98, -100)
Concentrated: (0, 0, +/-100)	(-14, 7, +/-100)

Table 5.1: MAX's expressed emotions in PAD space and the determined values of $(p, a, d)_{t,final}$.

According to Table 5.1, the **PAD values** determined by the proposed approach are quite accurate as compared to those of MAX's expressed emotion categories. For each of the facial expressions corresponding to the emotion categories *happy*, *surprised*, *annoyed/sad*, and *concentrated*, two possible values of $(p, a, d)_{t,final}$ are determined that differ in their values of dominance. In this regard, context related information is needed to decide for an explicit value of dominance (e.g., see Section 6.1.1). Except for the facial expressions corresponding to *angry* and *fearful* where respectively one explicit PAD value is determined. As compared to MAX, EMMA can express positive pleasure with different values of arousal (see Figure 4.14, p. 94). Accordingly, MAX's facial expression of happiness is interpreted as expressing near neutral value of arousal. Further, MAX's facial expression of *annoyed/sad* is interpreted as expressing negative arousal while MAX's expression of *angry* is interpreted as expressing near neutral arousal. While the arousal values determined for these facial expressions are not similar to those of their corresponding emotion categories, the difference in arousal between the determined arousal values for these facial expressions is similar to the difference between those of their corresponding emotion categories.

Figure 5.8 (Top) depicts the determined **PA time course** from MAX's expression of happiness during activation of the emotion category *happy* after receiving a sequence of positive emotional impulses. The figure shows that the determined PA course is similar to

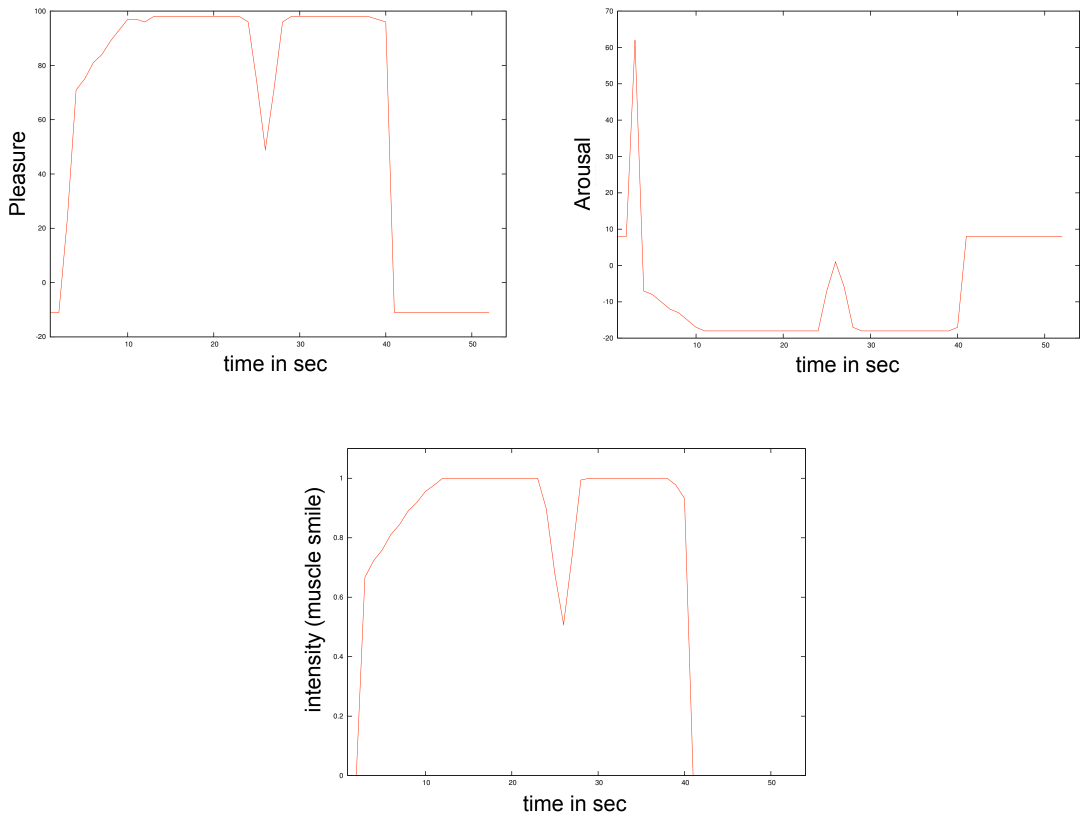


Figure 5.8: Top: PA time course determined by our proposed approach. Bottom: intensity values of MAX's smile muscle over time.

the course of the intensity values of MAX's smile muscle during activation of the emotion category *happy* (see Figure 5.8, bottom). Thus, the determined PA course reflects the changes in MAX's smile muscle during activation of his emotional state of happiness over time.

In sum, a first investigation of our approach to determine PAD values by combining the meaning of different AUs within PAD space shows **promising and interesting results**. However, this approach should be further investigated on the basis of human data provided by an automatic AU recognition system, or by a database of facial expressions annotated with AUs and their intensity values. This field is in its pioneering stage and thus currently, such human data are not easy to obtain.

5.2.2 Situational role-taking

Hoffman [55] and Davis [34] introduce *role-taking* as an empathy mechanism (see Section 2.1.2). Hoffman defines role-taking as an observer's ability to put himself in the other's situation and to imagine how the other feels. He distinguishes two types of role-taking, *self-focused role-taking* and *other-focused role-taking*. Self-focused role-taking refers to the ability to imagine oneself being in the other's situation while other-focused role-taking refers to the ability to imagine how the other feels in a considered situation. Similar to Hoffman, Higgins [53] also defines role-taking as seeing the world through the other's eyes or putting oneself in the other's shoes (cf. Section 2.1.2). He also distinguishes two types of role-taking that are similar to those introduced by Hoffman, *situational role-taking* and *individual role-taking*. In situational role-taking, an observer appraises the other's situation by using the same appraisal mechanisms as if he were in the same situation himself, whereas in individual role-taking, the additional implication of the other's viewpoint and characteristics are taken into account.

Accordingly, in our model, **situational role-taking** refers to the ability to generate a hypothesis about the other's emotional state by appraising the other's situation using one's own appraisal mechanisms.

Following Hoffman (cf. Section 2.1.2), role-taking, as compared to, e.g., mimicry, requires a higher-level cognitive processing. In this regard, individual role-taking is considered as being more cognitively demanding than situational role-taking, since it requires the additional consideration of the other's viewpoint and characteristics. Further, Damasio [31] defines secondary emotions as emotions that are experienced by adults in response to given situations based on a cognitive evaluation (cf. Section 2.2). According to Damasio, the process of secondary emotions starts with a cognitive evaluation of the situation through a deliberate consideration of the situation in the form of mental images. Similarly, role-taking is also a process that can be defined as starting with a **cognitive evaluation** of the other's situation through a deliberate consideration of the other's situation. This may further trigger reactive processes that generate the empathic emotion or it may generate an empathic emotion by mere cognitive evaluation. Thus, as empathic emotions, primary and secondary emotions can be generated through role-taking.

However, in our model, only the generation of **primary emotions** by means of situational role-taking is considered. Therefore, in our model, the empathy mechanism situational role-taking is triggered within the deliberative layer of the cognitive architecture [70] (see Section 5.1.2) and generates the empathic emotion on the basis of reactive processes within the reactive layer of the architecture.

Omdahl [87] underlines that contemporary theorists such as Hoffman and Davis provide comprehensive models of empathy where they discuss different empathy arousing modes but do not clearly explain the cognitive processing steps by which emotional states, such as *happy*, *sad*, or *angry* are decoded within these modes (cf. Section 2.2). Omdahl mainly considers the empathy mechanism role-taking and claims that **cognitive appraisal theories** (see Section 2.2) may provide an explanation for how the other's emotional state is decoded by role-taking.

Accordingly, as in our previous work [12] (see Section 3.3.1), situational role-taking is realized based on **WASABI**'s appraisal component [8] (see Section 3.1.1). Within WASABI, primary emotions are triggered as hard-wired reactions to perceived stimuli in the reactive layer of the cognitive architecture [70] (see Section 5.1.2). Secondary emotions are triggered by BDI-based deliberation within the deliberative layer of the architecture. The appraisal processes within both layers result in negative or positive values of emotional impulses and in one of two possible values of dominance that drive an agent's emotion dynamics over time. As mentioned earlier in this section, only primary emotions are generated by situational role-taking as realized in our model. Accordingly, by means of a **BDI**-based deliberation, a cognitive evaluation of the other's perceived situation is realized based on one's **own appraisal mechanisms** as defined within the deliberative layer. These appraisal mechanisms trigger further appraisal mechanisms defined within the reactive layer, which result in negative or positive emotional impulses and in one of two possible values of dominance (see Figure 5.9). The emotional impulses and a determined value of dominance are input to the virtual human's own emotion simulation module [8] (see Section 5.2.3) and represent a **hypothesis** about the other's emotional state.

The realization of situational role-taking requires a context scenario. While facial mimicry is realized for the virtual human EMMA, situational role-taking is realized for the virtual human MAX in cooperation with the doctoral thesis project of Nhung Nguyen [81]. A more detailed description of the realization of this empathy mechanism

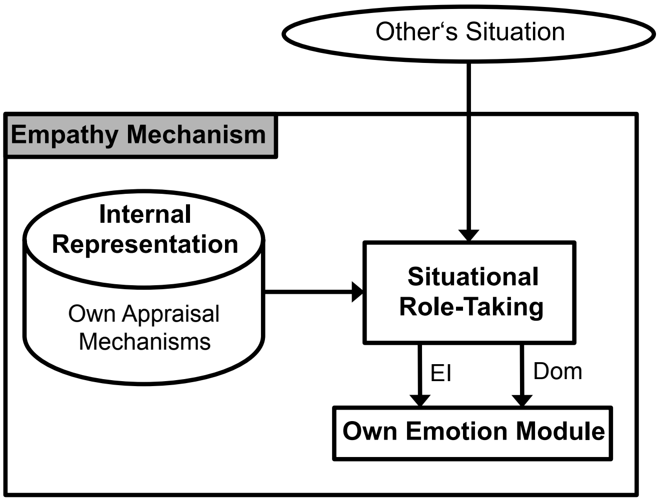


Figure 5.9: The empathy mechanism **situational role-taking** based on own internally represented appraisal mechanisms. '**EI**' refers to an emotional impulse, '**Dom**' refers to a value of dominance.

and of the underlying appraisal mechanisms are presented in Section 6.1.2.

5.2.3 The empathic emotion

According to OCC's emotion theory [88], fortunes-of-other emotions, also called empathetic emotions, are elicited through appraisal of an event as desirable or not desirable for another person (cf. Section 2.2). As such, in our model, an **empathic emotion** is defined as an emotion that is elicited with respect to the other's perceived emotional state. That is, an empathic emotion is elicited after detecting a rapid and, at the same time, salient change in the other's perceived emotional state, or if the other's emotional state is perceived as salient.

At each point in time, the emotional values resulting from facial mimicry (cf. Section 5.2.1) or from situational role-taking (cf. Sections 5.2.2) are input to the virtual human's emotion simulation module (cf. Section 3.1.1). These values drive the simulated dynamics of the other's **perceived emotional state** over time. Thus, the time course of the perceived emotional state is represented and simulated within the virtual human's emotion simulation module. Accordingly, a related primary emotion as well as a corresponding value of awareness likelihood are inferred.

The other's perceived emotional state is represented by an additional reference point within the virtual human's emotion simulation module. Thus, the virtual human always distinguishes between his own and others' perceived emotional states. Further, the empathy mechanisms facial mimicry and situational role-taking, as defined in our model, are respectively based on one's own facial expression repertoire and on ones' own appraisal mechanisms to generate a hypothesis about the other's emotional state. Thus, in line with Hoffman's [55] defined four stages in the development of empathy with respect to self-other differentiation (see Section 2.1.2), our model allows for **quasi-egocentric empathy**. That is, while the virtual human always distinguishes between his own and others' perceived emotional states, he evaluates others' situations from his own perspective.

The simulation of the dynamics of further emotional states in addition to the virtual human's emotional state has been realized in the context of our previous work on situational role-taking [12]. The dynamics of these emotional states are simulated in exactly the same way as the dynamics of the virtual human's emotional state (cf. Section 3.1.1). Note that negative values of arousal resulting from facial mimicry are mapped to neutral values of arousal since the negative arousal space is not accessible to the emotion dynamics as defined in [8] except during boredom (see Section 3.1.1).

Once the perception of the other's emotional state is interrupted, the perceived emotional state simulated within the virtual human's emotion simulation module **decays over time**, and reaches the state of boredom. Thus, the virtual human assumes that the other is no longer receiving emotional stimuli and that, as for himself, the other becomes bored in consequence to the absence of emotional stimuli. This can be interpreted as the virtual human's assumed default state for the other as long as no perceived emotional state of the other is available. Further, the perceived emotional state is also asserted as **belief** about the other's emotional state within the BDI-based deliberative layer of the cognitive architecture (see Section 5.1.2).

Once the other's perceived emotional state is represented within the virtual human's emotion simulation module, a defined **condition of elicitation** of an empathic emotion is checked. This condition of elicitation is checked with respect to the PAD values of the perceived emotional state over time. That is, with respect to a predefined short time interval T, the difference between perceived PAD values corresponding to the time-stamps t_{k-1} and t_k, with $t_k - t_{k-1} <= T$, is calculated as $\|PAD_{t_k} - PAD_{t_{k-1}}\|$. If the difference in PAD values exceeds a predefined saliency threshold $TH1$, or if $\|PAD_{t_k}\|$ exceeds a predefined saliency threshold $TH2$, then the emotional state PAD_{t_k} and its related primary emotion represent the empathic emotion (see Figure 5.10). The predefined thresholds can be interpreted as representing the virtual human's **responsiveness** to the others' perceived emotional states.

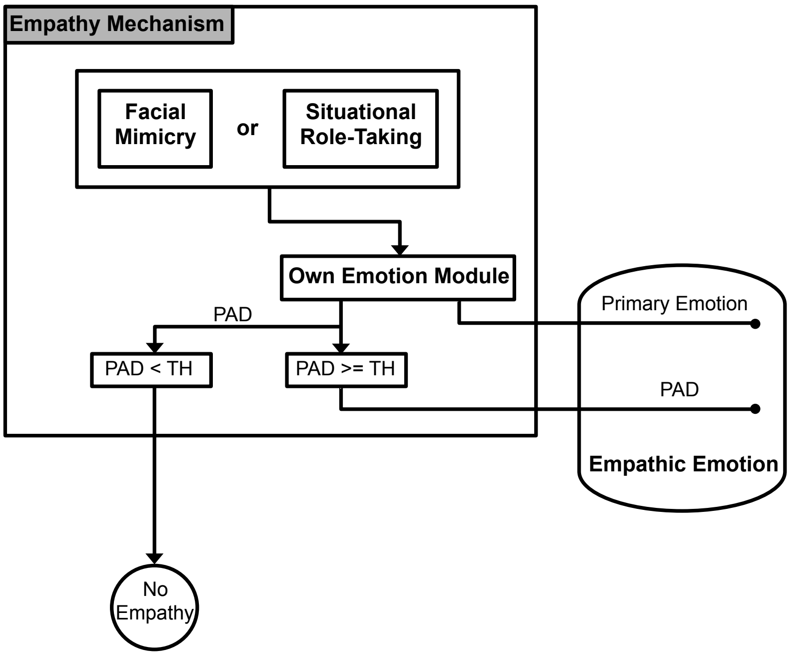


Figure 5.10: The **empathic emotion** as output by the empathy mechanism. **PAD** refers to $\|PAD_{t_k} - PAD_{t_{k-1}}\|$ and $\|PAD_{t_k}\|$. **TH** refers to $TH1$ and $TH2$.

By calculating the difference $\|PAD_{t_k} - PAD_{t_{k-1}}\|$, empathy is elicited with respect to the **dynamic change** in the perceived emotional state. In this regard, a rapid and salient negative change from positive emotional state toward a neutral emotional state, and a rapid and salient positive change from a negative emotional state toward a neutral emotional state elicit an empathic emotion. Thus, an empathic emotion is elicited in response to a neutral emotional state resulting from a rapid and salient change in emotional state. Accordingly, a perceived neutral facial expression that comes shortly after a negative or positive facial expression of emotion elicits empathy. In contrast, in previous work by Rodrigues et al. [98], an empathic emotion is produced in response to an emotional cue, e.g., facial expressions, which is defined as any signal indicating the arousal of an emotion.

Once an empathic emotion is elicited, the next processing step in our model is triggered, namely, empathy modulation.

5.3 Empathy modulation

Hoffman [55], Davis [34], OCC [88], and De Vignemont & Singer [35] (see Sections 2.1 and 2.2), introduce several factors that modulate an observer's empathic response. Following De Vignemont & Singer (cf. Section2.1.3), we group these factors into **four categories**: **(a)** features of observed emotion, **(b)** relationship between empathizer and the observed other, **(c)** situational context, and **(d)** features of the empathizer (see Table 5.2).

Authors	Features of observed emotion	Relationship between empathizer and the observed other	Situational context	Features of the empathizer
Hoffman (2000)	salience and intensity of observed emotion	familiarity bias: in-group bias friendship bias similarity bias	here-and-know bias	-
Davis (1994)	strength of situation	observer-target similarity	-	biological capacities individual differences learning history
OCC (1988)	desirability-for-other	liking	deservingness	desirability-for-self
De Vignemont & Singer (2006)	valence, intensity, saliency, primary vs. secondary emotions	affective link familiarity and similarity communicative intentions	appraisal of the situation display of multiple emotions	mood arousal personality, gender and age emotional repertoire emotional regulation capacities

Table 5.2: Factors modulating an observer's empathic response classified into four categories following De Vignemont & Singer [35].

Davis distinguishes between factors that directly modulate the empathic outcomes and those that influence the empathic processes thus modulating their empathic outcomes (cf. Section 2.1.2). Such factors are, e.g., *biological capacities*, *individual differences*, and *learning history*. For example, in our model, the empathy mechanism facial mimicry is based on using one's own facial expression repertoire. Thus, facial mimicry may, for example, produce an empathic emotion from type *annoyed* in response to the other's emotional state of *anger* (cf. Section 5.2.1). This can be explained as due to the factor *individual differences* since one's own facial expression repertoire may be different from the other's. While such factors modulate the empathic emotion, they do not influence an observer's **felt degree of empathy**. That is, an observer may experience a maximum degree of empathy even if the generated empathic emotion does not match the other's emotion in consequence to such factors. Therefore, in our model, we consider **empathy**

modulation as the process by which an empathic emotion is modulated by factors that affect the felt degree of empathy. Accordingly, we define the degree of empathy as the **degree of similarity** between the empathic emotion as generated by the empathy mechanism and the empathic emotion after the process of empathy modulation. That is, the more similar one's empathic emotion to the other's perceived emotion, the higher the degree of empathy.

In the following, the modulation process and the calculation of a degree of empathy are introduced.

5.3.1 Modulation process

In our model, the modulation process is realized within **PAD space** of the virtual human's emotion simulation module [8] (see Section 3.1.1). Accordingly, at each point in time when an empathic emotion is elicited, the following equation is applied:

$$\begin{aligned} empEmo_{t,mod} = ownEmo_t + \\ (empEmo_t - ownEmo_t) \cdot (\sum_{i=1}^{n} p_{i,t} \cdot w_i)/(\sum_{i=1}^{n} w_i) \end{aligned} \tag{5.3}$$

The value $empEmo_{t,mod}$ represents the modulated empathic emotion. Further, the value $ownEmo_t$ represents the virtual human's own emotional state as the modulation factor *mood*. The value $empEmo_t$ represents the non-modulated empathic emotion resulting from the previous processing step, the empathy mechanism. The values $p_{i,t}$ represent modulation factors in addition to the factor *mood*. The values w_i represent assigned values of weights for the modulation factors $p_{i,t}$.

The values $p_{i,t}$ represent arbitrary predefined modulation factors that can have values ranging in $[0, 1]$. Such modulation factors are for example, *liking* and *familiarity* (see Table 5.2). As proposed by OCC [88], *liking* can be represented by values ranging in $[-1, 1]$ from maximum dislike to maximum like where the value 0 represents neither like nor dislike. Following this, *familiarity* can be represented by values ranging in $[0, 1]$ from non-familiar to most-familiar. Only the impact of positive values of $p_{i,t}$ is considered in the present approach for empathy modulation. The values of $p_{i,t}$ are determined with respect to the **social context** and are asserted as **beliefs** within the BDI-module (see Section 6.1). The values w_i allow for the assignment of different values of weights to the modulation factors $p_{i,t}$. For example, *liking* may be assigned a higher value of weight than *familiarity* thus having a stronger effect on the modulation of the empathic emotion than *familiarity*.

By applying **Equation 5.3**, the modulated empathic emotion $empEmo_{t,mod}$ lies on the straight line spanned by the virtual human's own emotional state $ownEmo_t$ and the non-modulated empathic emotion $empEmo_t$ (see Figure 5.11, left).

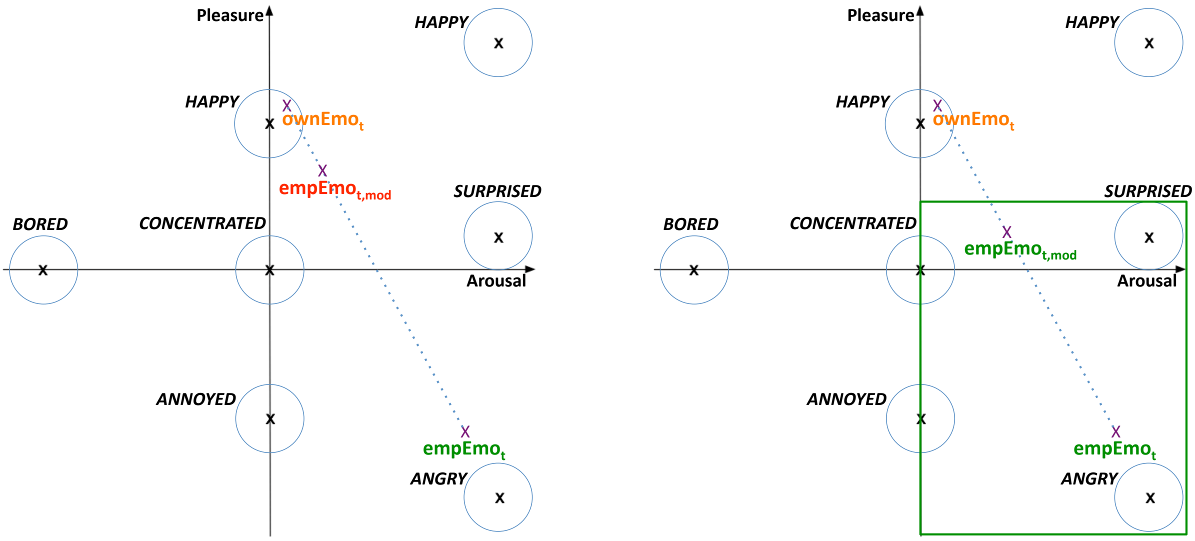


Figure 5.11: The PA space of positive dominance of the emotion simulation module [8]. Left: $empEmo_{t,mod}$ as lying on the straight line spanned by $ownEmo_t$ and $empEmo_t$. Right: the *empathy facilitation region* defined for the primary emotion category *angry*.

As mentioned in the previous section, we define the degree of empathy as the degree of similarity between the modulated empathic emotion and the non-modulated one. Thus, the degree of empathy is represented by the **distance** between $empEmo_{t,mod}$ and $empEmo_t$ within PAD space. That is, the closer $empEmo_{t,mod}$ to $empEmo_t$, the higher the degree of empathy. The less close $empEmo_{t,mod}$ to $empEmo_t$, the lower the degree of empathy. At each point in time t, the degree of empathy is impacted by the values of the modulation factors. The impact of the modulation factors $p_{i,t}$ is calculated as a weighted mean of their current values at each point in time t.

In previous work by Rodrigues et al. [98], the impact of the modulation factor *mood* is defined as follows. A negative mood increases the potential of a negative empathic emotion and decreases the potential of a positive one while a positive mood increases the potential of a positive empathic emotion and decreases the potential of a negative one. Accordingly, in our model, the virtual human is more **sensitive** to the empathic emotion when his emotional state is more similar to the empathic emotion. The virtual human is more **resistant** to the empathic emotion when his emotional state is less similar to the empathic emotion. That is, the closer the virtual human's own emotional state $ownEmo_t$ to the empathic emotion $empEmo_t$ the higher the degree of empathy and the less the modulation factors $p_{i,t}$ can impact the degree of empathy. The less close the virtual human's own emotional state $ownEmo_t$ to the empathic emotion $empEmo_t$ the lower the degree of empathy and the more the modulation factors $p_{i,t}$ can impact the degree of empathy.

With regard to the impact of the modulation factors $p_{i,t}$, the higher their value of weighted mean, the closer the modulated empathic emotion $empEmo_{t,mod}$ to the non-modulated empathic emotion $empEmo_t$ and the higher the degree of empathy. The lower their value of weighted mean, the less close the modulated empathic emotion $empEmo_{t,mod}$ to the non-modulated empathic emotion $empEmo_t$ and the lower the degree of empathy.

As compared to previous work, e.g., Rodrigues et al. [98] (see Section 3.3.1), our approach to empathy modulation allows for a **maximal degree of empathy** even if the virtual human's own emotional state $ownEmo_t$ is maximally different to the non-modulated empathic emotion $empEmo_t$. This occurs by maximal values of $p_{i,t}$. For example, consider the modulation factors *liking* and *familiarity*. The virtual human would experience a maximal degree of empathy with a maximally liked and familiar other independently of his own emotional state. The idea is that one's emotional state has less influence on his degree of empathy with strongly liked and familiar others. Further, the proposed approach also allows for higher degrees of empathy even if the values of $p_{i,t}$ are very low. This occurs when the virtual human's own emotional state $ownEmo_t$ is very similar to the non-modulated empathic emotion $empEmo_t$. For example, consider again the modulation factors *liking* and *familiarity*. The virtual human would experience a **maximal degree of empathy** with neither liked nor disliked, and with non-familiar other in situations where he experiences a similar emotional state to that of the other. The idea is that one is very sensitive to the other's emotion when he is experiencing a similar emotion, and that he can thus also fully empathize with neutrally liked or unfamiliar others.

According to our working definition of empathy, an empathic response to the other's emotion should be more appropriate to the other's situation than to one's own [55] (see

Sections 2.1.2 and 2.1.4). Hoffman [55] emphasizes that an empathic response need not to be a close match to the affect experienced by the other, but can be any emotional reaction compatible with the other's situation. Further, according to the thesis of the dimensional theories (see Section 2.2), emotions are related to one another in a systematic manner and their relationships can be represented in a dimensional model. Accordingly, the modulated empathic emotion $empEmo_{t,mod}$ is facilitated only if it lies in an **immediate neighborhood** to the non-modulated empathic emotion $empEmo_t$. Otherwise, $empEmo_{t,mod}$ is inhibited and the next processing step in our model is not triggered (see Figure 5.12).

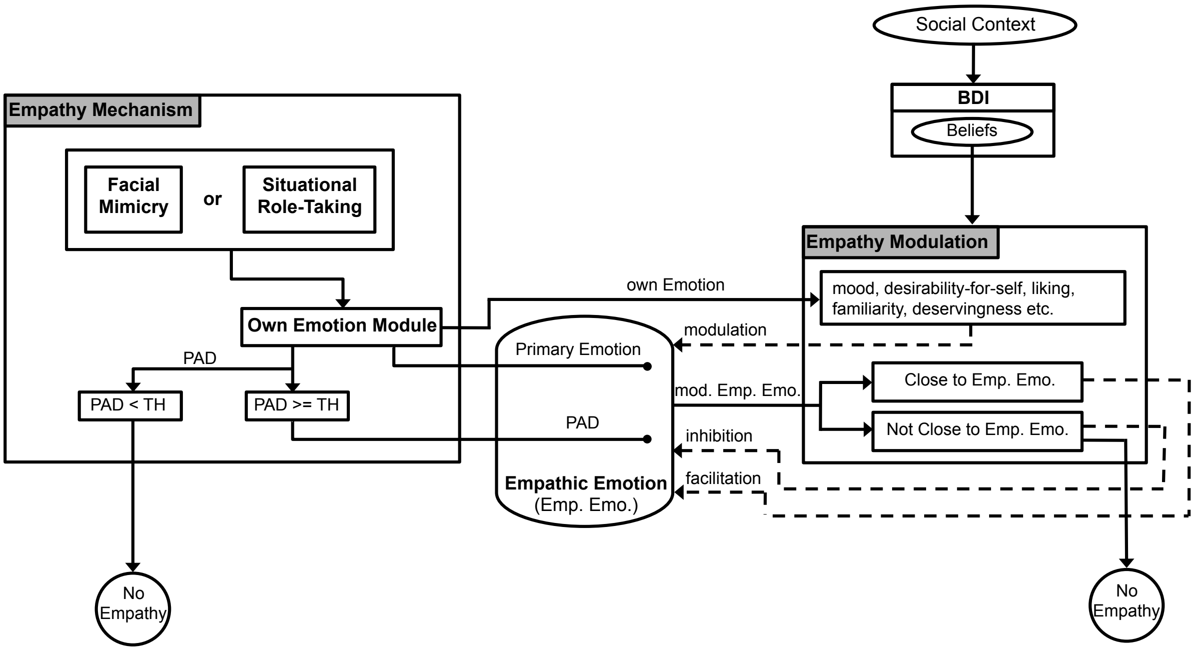


Figure 5.12: **Empathy modulation** as the second processing step in our computational model of empathy.

Hence, for each primary emotion category located within PAD space of the emotion simulation module [8] (see Section 3.1.1), we defined a so called **empathy facilitation region** as a box surrounding the emotion category. For example, Figure 5.11 (right) shows the PA space of positive dominance of the emotion simulation module [8] (see Section 3.1.1) with the defined *empathy facilitation region* for the emotion category *angry*. As depicted in Figure 5.11 (right), the modulated empathic emotion $empEmo_{t,mod}$ has as related emotion category *concentrated* and the non-modulated empathic emotion $empEmo_t$ has as related emotion category *angry*. Accordingly, once the modulated empathic emotion $empEmo_{t,mod}$ enters the *empathy facilitation region* defined for *angry*, it is facilitated or otherwise it is inhibited. Within the *empathy facilitation region* for each emotion category, the modulated empathic emotion $empEmo_{t,mod}$ represents an empathic response that is **compatible** with the other's situation (cf. [55], see above). Thus, the virtual human is allowed to react with an emotion from a **different type** (emotion category) although compatible with the empathic emotion.

Table 5.3 shows the PAD coordinates of the corner points defined for each *empathy facilitation region*. The choice of these values for the *empathy facilitation regions* is a matter of design and evaluation.

Emotion category in PAD space	**Emp. Facil. Reg. PAD coordinates**
Happy: (80, 80, +/-100), (50, 0, +/-100)	(100, 0, +/-100), (100, 100, +/-100), (-25, 0, +/-100), (-25, 100, +/-100)
Surprised: (10, 80, +/-100)	(30, 0, +/-100), (30, 100, +/-100), (-30, 0, +/-100), (-30, 100, +/-100)
Annoyed: (-50, 0, 100)	(25, 0, +/-100), (25, 100, +/-100), (-100, 0, +/-100), (-100, 100, +/-100)
Sad: (-50, 0, -100)	(25, 0, +/-100), (25, 100, +/-100), (-100, 0, +/-100), (-100, 100, +/-100)
Angry: (-80, 80, 100)	(25, 0, +/-100), (25, 100, +/-100), (-100, 0, +/-100), (-100, 100, +/-100)
Fearful: (-80, 80, -100)	(25, 0, +/-100), (25, 100, +/-100), (-100, 0, +/-100), (-100, 100, +/-100)
Concentrated: (0, 0, +/-100)	(30, 0, +/-100), (30, 100, +/-100), (-30, 0, +/-100), (-30, 100, +/-100)

Table 5.3: The PAD coordinates of the defined *empathy facilitation region* (Emp. Facil. Reg) for each emotion category within PAD space.

According to Table 5.3, the *empathy facilitation region* for the emotion category *happy* covers the entire space of positive pleasure with slightly negative pleasure. The *empathy facilitation regions* for *annoyed*, *angry*, *fearful*, and *sad* cover the entire space of negative pleasure with slightly positive pleasure. Further, a modulated empathic emotion with a related emotion category *concentrated* or *surprised* is defined as a compatible empathic response to a non-modulated empathic emotion with any other related emotion category defined in PAD space. However, when the non-modulated empathic emotion has as related emotion category *concentrated* or *surprised*, only a modulated empathic emotion with a related emotion category *concentrated* or *surprised* is defined as a compatible empathic response. In order to enlarge the *empathy facilitation regions* of both *surprised* and *concentrated*, the negative or positive change in pleasure that elicited the empathic emotion (cf. Section 5.2.3) has to be considered. Thus, the space of positive pleasure could be included when the empathic emotion is elicited as a result of a positive change in pleasure, while the negative space of pleasure could be included when the empathic emotion is elicited as a result of a negative change in pleasure.

Once the modulated empathic emotion $empEmo_{t,mod}$ is facilitated, the next processing step in our model, namely, expression of empathy, is triggered (see Section 5.4). Otherwise, the virtual human continues expressing his own emotional state $ownEmo_t$.

Since in our approach, we consider the modulation factor *mood*, one would ask why the empathy modulation is not realized within the **dynamics/mood space** of the virtual human's emotion simulation module (cf. Section 3.1.1) instead of PAD space. In this regard, a modulated empathic emotion could be generated by modulating, with the values $p_{i,t}$, the emotional impulse which allows for a change from the virtual human's own emotional state toward the non-modulated empathic emotion. In this case, a problem is, that some PAD values cannot be reached by means of emotional impulses and thus a change toward an empathic emotion with such PAD values cannot be achieved.

In the following section, we introduce our approach to calculate the degree of empathy once the modulated empathic emotion $empEmo_{t,mod}$ is facilitated.

5.3.2 Degree of empathy

As mentioned in the previous section, the degree of empathy is represented by the **distance** between $empEmo_{t,mod}$ and $empEmo_t$ within **PAD space**. Hence, once the modulated empathic emotion $empEmo_{t,mod}$ enters the *empathy facilitation region*, the degree of empathy is calculated and increases toward the non-modulated empathic emotion $empEmo_t$. Outside the *empathy facilitation region*, the degree of empathy is equal to 0. Within the *empathy facilitation region*, the degree of empathy is calculated by the following equation at each point in time t a modulated empathic emotion $empEmo_{t,mod}$ is facilitated:

$$degEmp_t = (1 - \| \frac{empEmo_{t,mod} - empEmo_t}{maxDistBox} \|)^2 \tag{5.4}$$

The value $degEmp_t$ represents the calculated degree of empathy and ranges within $[0, 1]$. The value $empEmo_{t,mod}$ represents the modulated empathic emotion. The value $empEmo_t$ represents the non-modulated empathic emotion (cf. previous section). The value $maxDistBox$ represents the possible maximum distance between $empEmo_{t,mod}$ and $empEmo_t$ within the *empathy facilitation region* (see Figure 5.13). Note that the distances $\|empEmo_{t,mod} - empEmo_t\|$ and $maxDistBox$ are **weighted distances** in PAD space. That is, we defined values of weights for each dimension within PAD space. A **polynomial function** is chosen in order to get smooth values of the calculated degree of empathy.

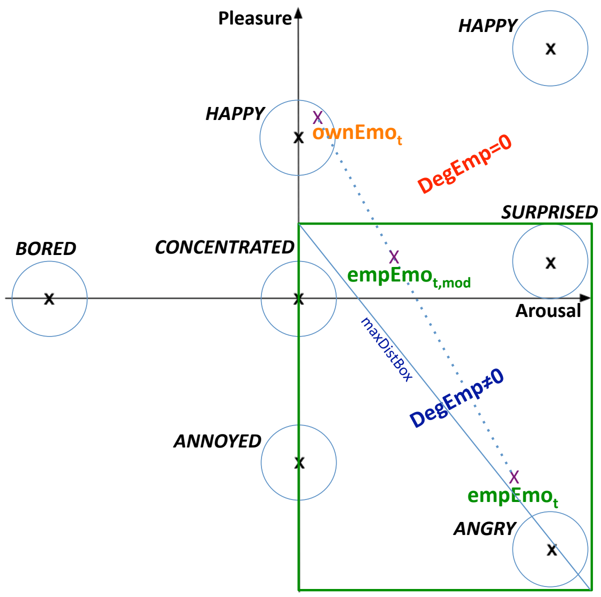


Figure 5.13: The degree of empathy within PA space of positive dominance of the emotion simulation module [8]. Outside the *empathy facilitation region*, the degree of empathy is equal to 0. Inside the *empathy facilitation region*, the degree of empathy increases toward the non-modulated empathic emotion $empEmo_t$. $maxDistBox$ represents the possible maximum distance between $empEmo_{t,mod}$ and $empEmo_t$ within the *empathy facilitation region*.

According to the **dimensional theories** (see Section 2.2), the pleasure dimension is the first, most agreed upon dimension, the arousal dimension is the second agreed upon dimension and the dominance dimension is the third, less agreed upon dimension. Thus, regarding the defined **weight values** for each dimension within PAD space, we assigned a higher weight value to the pleasure dimension, a lower value to the arousal dimension, and a very low value to the dominance dimension. Hence, a modulated empathic emotion that differs only in its arousal value to the non-modulated empathic emotion reflects a higher degree of empathy than one that differs only in its pleasure value. For example, *annoyed* reflects a higher degree of empathy with respect to *angry* than *surprised*. Further, a modulated empathic emotion that differs only in its dominance value to a non-modulated empathic emotion reflects a high degree of empathy. For example, *fearful* reflects a high degree of empathy with respect to *angry* and vice versa.

Now, consider the following example. The virtual human's emotional state $ownEmo_t$ has as its related emotion category *angry*. The modulated empathic emotion $empEmo_{t,mod}$ also has *angry* as its related emotion category, and the non-modulated empathic emotion $empEmo_t$ has as related emotion category *annoyed*. In this case, the calculated degree of empathy has a lower value than when $empEmo_{t,mod}$ has *annoyed* as its related emotion category. Thus, the virtual human has a higher degree of empathy when his empathic response is more appropriate to the other's situation than to his own (cf. [55], see Section 2.1.2).

As long as no further empathic emotion is elicited (cf. Section 5.2.3), the modulated empathic emotion decays over time (cf. [8], see Section 3.1.1). The **decay function** of the modulated empathic emotion is affected by the degree of empathy, that is, the higher the calculated value of the degree of empathy, the slower the decay. The lower the value the faster the modulated empathic emotion decays. This is realized by respectively, increasing and decreasing the values of the **spring mass constants** (cf. [8], see Section 3.1.1) of the modulated empathic emotion. Also the degree of empathy decays with respect to the decaying values of pleasure and arousal of the modulated empathic emotion. Once the modulated empathic emotion's values of pleasure and arousal are equal to 0, the virtual human returns to expressing his emotional state. Accordingly, as long as the PA values of the modulated empathic emotion do not equal 0, the virtual human continues expressing the modulated empathic emotion. However, this may produce inappropriate behavior. For example, consider the virtual human showing a positive empathic emotion that starts to decay after a while. During this time, the virtual human's own emotional state is triggered positively. Thus, the virtual human cannot express his emotional state since he is still expressing the decaying empathic emotion. This issue should be considered in more details in future work.

5.4 Expression of empathy

Hoffman [55] and Davis [34] (see Section 2.1.2) emphasize the role of empathy as motivating **prosocial and cooperative behaviors** such as caring, helping, and justice. They define caring and helping as actions that increase the well-being and the positive emotions of others. Further, Bavelas et al. (cf. Section 2.1.2) state that mimicry conveys involvement and solidarity with the other's situation. Accordingly, in our model, we consider the **expression of empathy** as the process by which actions are triggered to increase the other's well-being and to convey involvement with the other's situation.

The expression of empathy consists of triggering the virtual human's **multiple modalities** by the modulated empathic emotion (see Figure 5.14).

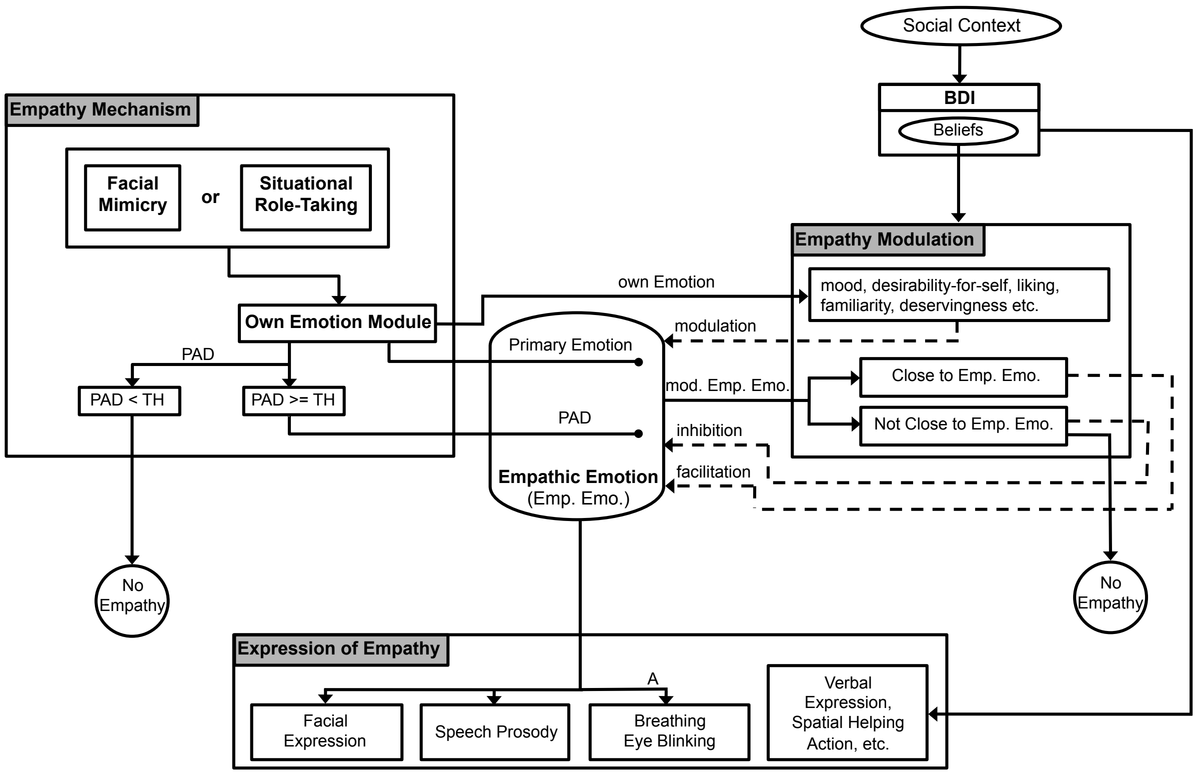


Figure 5.14: The computational model of empathy with the third processing step **expression of empathy**.

As mentioned at the beginning of this chapter, the proposed computational model of empathy is realized for the virtual humans MAX and EMMA. Thus, for EMMA, a facial expression from her facial expression repertoire (cf. Section 4.3) is triggered by the

PAD value of the modulated empathic emotion. Using *EmoSpeak* (see Section 4.2.2), EMMA's speech prosody is modulated by the PAD value of the modulated empathic emotion. For MAX, a facial expression is triggered with respect to the value of awareness likelihood of the primary emotion category related to the modulated empathic emotion. Regarding his speech prosody, MAX's value of emotional pitch is triggered with respect to the primary emotion category related to the modulated empathic emotion.

For both MAX and EMMA, the frequencies of their eye blinking and breathing behaviors are also influenced by the value of arousal of the modulated empathic emotion. Triggering further modalities, such as verbal utterances or spatial helping actions, depends on the scenario's context (see Section 6.1).

5.5 Summary and conclusion

This chapter introduced a computational model of empathy realized for the virtual humans MAX and EMMA. In Section 5.1, we gave an overview on the structure of the model and its integration within the emotion simulation module of an existing cognitive architecture. In light of the requirements for building a computational model of empathy as formulated in Section 1.2 and discussed in Section 3.3.3, the proposed model is based on **three processing steps**, empathy mechanism, empathy modulation, and expression of empathy. Supported by the theoretical background introduced in chapter 2, in Sections 5.2, 5.3, and 5.4 respectively, we discussed **working definitions** for the three processing steps underlying our empathy model, and we introduced the approaches to realize them.

In Section 5.2, we introduced our approach to realize **facial mimicry** based on EMMA's facial expression repertoire where PAD values are linked to AUs. Hence, by combining the meanings of AUs within PAD space as represented by their corresponding intensity functions, PAD values corresponding to perceived facial expressions of emotions are determined. Accordingly, the proposed approach is in line with the emphasis that dimensional emotion models present dense and continuous emotion spaces that are more inclined to characterize the continuity and subtlety of emotion expressions, and that allow for the recognition of a wide range of emotions. A first evaluation of the proposed approach shows that the PAD values determined from MAX's facial expressions of emotion are quite accurate, thus providing **promising results**. However, this approach should be further investigated with human data provided by an automatic AU recognition system, or by a database of facial expressions annotated with AUs and their intensity values. This field is in its pioneering stage and thus currently, such human data are not easy to get. The proposed approach can be **easily adapted** for use by virtual humans other than EMMA that have a similar facial expression repertoire, and can also be easily carried over in different interaction scenarios. Hence, in light of the requirement **Universality** that a computational model of empathy should fulfill (cf. Section 1.2), our realization of facial mimicry contributes to the universality of our empathy model.

Further in Section 5.2, an approach to realize the empathy mechanism **situational role-taking** was introduced. As compared to the empathy mechanism facial mimicry, situational role-taking allows for the consideration of contextual cues which are also another possible source of information based on the appraisal of others' situations. However, our approach to realize situational role-taking requires a context scenario and is thus **domain dependent**. Furthermore, by means of our defined condition of elicitation of an empathic emotion, empathy is not only elicited in response to the saliency of a perceived emotional state but is also elicited with regard to the **dynamic change** in the perceived emotional state. Moreover, different values for the elicitation thresholds can be defined which allow for the modeling of a virtual human's **responsiveness** to the other's perceived emotional state.

In Section 5.3, we introduced an approach to **modulate** an empathic emotion and to calculate a degree of empathy. The proposed approach allows for the modulation of an empathic emotion through the factor *empathizer's mood* as well as through other arbitrary predefined modulation factors that can have values ranging in $[0, 1]$, e.g., *liking* or *familiarity*. Furthermore, it also allows for the assignment of different values of weights to the latter mentioned factors, to define which modulation factors have a more significant impact on the modulation of an empathic emotion and on the degree of empathy than others. In line with Hoffman's emphasis [55] that an empathic response need not to be a close match to the affect experienced by the other, but can be any emotional reaction compatible with the other's situation (cf. Section 2.1.2), not only we modulate the **intensity** of an empathic emotion but also its related **type** (emotion category). Hence, we defined regions of immediate neighborhood for each emotion category within PAD space where a modulated empathic emotion from different type (emotion category) to the non-modulated one represents a compatible empathic response. Further, a **degree of empathy** is calculated as the degree of similarity between the modulated and the non-modulated empathic emotion. In this regard, we rely on the thesis of the dimensional theories (see Section 2.2) that emotions are related to one another in a systematic manner, and that their relationships can be represented in a dimensional model. Hence, in our approach, we exploited the assumed relationships between emotions in PAD space. Once the modulation factors are defined, our proposed approach for the modulation of an empathic emotion and for the calculation of a degree of empathy can be **easily carried over** in different context scenarios as realized in the present thesis (see Section 6.1). Thus, in light of the requirement **Universality**, the proposed approach, as our approach to realize facial mimicry, further contributes to the universality of our empathy model.

In Section 5.4, the **expression of empathy** as triggered by the modulated empathic emotion was introduced. In this regard, facial expressions, speech prosody, and eye blinking and breathing behaviors can be classified as **context independent** modalities while other modalities such as verbal utterances and spatial helping actions can be classified as **context dependent**. Thus, in light of the requirement **Universality**, context independent expressions of empathy further contribute to the universality of our empathy model.

In sum, the proposed computational model of empathy allows a virtual human to empathize with his interaction partner to different degrees, and to express his degree of empathy through different modalities. Thus, our model enhances a virtual human's **social behavior** as one of our main objectives (cf. Chapter 1). Furthermore, within our model, different parameters are defined that allow for the investigation of several theoretical aspects of empathy that are not explicitly defined within theoretical models of empathy. Thus, our model also provides an **experimental tool** for its underlying theories, which is another main objective of this thesis (cf. Chapter 1). While at this point, we might seem to have achieved our thesis objectives, this issue should be further **verified** by the application and evaluation of our model within a context scenario which is subject to the next chapter.

6 Application and evaluation

This chapter introduces the application and evaluation of the computational model of empathy presented in the previous chapter. The application of the proposed model is carried out in the context of two interaction scenarios presented in Section 6.1, a conversational agent scenario and a spatial interaction task scenario. Subsequently, in Section 6.2, an empirical evaluation of our model in the context of the conversational agent scenario is introduced.

6.1 Application scenarios

In this section, the application of our computational model of empathy in the context of two interaction scenarios is introduced. First, in Section 6.1.1, we present a conversational agent scenario involving the virtual humans MAX [62] (see Section 3.1.2) and EMMA (see Chapter 4) and a human interaction partner. Then, in Section 6.1.2, we present a spatial interaction task scenario involving the virtual human MAX and a human interaction partner.

6.1.1 A conversational agent scenario

As emphasized in Section 4.1.1, one important aspect in creating the virtual human EMMA besides the virtual human MAX, is to allow for the consideration of **agent-agent interaction** in addition to human-agent interaction. Hence, the conversational agent scenario presented here involves both virtual humans MAX and EMMA and a human interaction partner. This scenario is an extension of a previous scenario where the virtual human MAX acts as a guide in a public computer museum [62] (see Section 3.1.2). Within this scenario, MAX engages in natural face-to-face conversation with visitors and provides them with information about the museum or the exhibitions. In this regard, MAX conducts multimodal small talk dialogs using speech, gestures, and facial behaviors. During his interaction with human partners, MAX's emotions can be triggered positively or negatively. For example, the perception of persons in MAX's visual field or the interpretation of a human partner's verbal utterance as a compliment trigger MAX's emotions positively. The interpretation of a human partner's verbal utterance as obscene or politically incorrect trigger MAX's emotions negatively. In this scenario, we integrate EMMA as a **third interaction partner**. As with MAX, human interaction partners can also engage in a small talk dialog with EMMA, and her emotions can be triggered in the same way as those of MAX. Figure 6.1 depicts the scenario setting. Both virtual humans reside in a 3D computer graphics simulation of a biosphere (see also Figure 4.1, p. 72). The virtual humans are displayed in their real sizes on two separate panels. Human partners interact with MAX and EMMA using a keyboard as input device to communicate.

Figure 6.1: The virtual humans MAX and EMMA displayed in their real sizes on two separate panels in the conversational agent scenario. The human partner interacts with the virtual humans using a keyboard.

Within this scenario, EMMA empathizes with MAX's emotions based on the computational model of empathy proposed in Chapter 5. During small talk between a human partner and MAX, EMMA follows the conversation by directing her visual attention toward the speaking agent. When attending to MAX, EMMA's empathy process is triggered. As mentioned in Section 4.2, this scenario is the **first interaction scenario** for EMMA. Again, in accordance with EMMA's empathic capabilities within her first interaction scenario, we chose the name 'EMMA' to refer to an **E**mpathic **M**ulti**M**odal **A**gent (cf. Section 4.2). In the following, we illustrate EMMA's empathic behavior in this scenario.

When EMMA's visual attention is directed toward MAX, the empathy mechanism

facial mimicry introduced in Section 5.2.1 is triggered. Accordingly, EMMA internally imitates MAX's emotional facial expression by getting the values of his facial muscles at each point in time and by mapping them to her **A**ction **U**nits (AUs). Subsequently, a **P**leasure-**A**rousal-**D**ominance (PAD) value is determined from the values of AUs provided by internal imitation of MAX's facial expression (cf. Section 5.2.1). As discussed in Section 5.2.1, it is possible that different PAD values result from Equations 5.1 and 5.2. In this case, a mean value of all possible solutions regarding pleasure and arousal is calculated. Regarding dominance, **context related information** is considered as a means to an explicit solution. Since EMMA's value of dominance is always positive within this scenario (cf. [62]), she assumes the same value for MAX. This can be considered as a very simple form of situational role-taking where only the value of dominance is inferred (see Section 5.2.2). The determined PAD value represents a hypothesis about MAX's emotional state, and is input to EMMA's emotion simulation module [8], thus driving the simulated dynamics of the perceived emotional state over time (cf. Section 5.2.3).

The perceived emotional state is represented by a second reference point within EMMA's emotion simulation module, thus allowing her to distinguish between her own and MAX's emotional state (see Figure 6.2, left). The perceived emotional state is asserted as **belief** about MAX's emotional state within EMMA's BDI-module [70]. Once MAX's perceived emotional state exceeds a predefined saliency threshold, or a salient change in MAX's perceived emotional state is detected (cf. Section 5.2.3), an empathic emotion is elicited (see Figure 6.2, middle) and the next processing step, empathy modulation, is triggered.

Per empathy modulation, as introduced in Section 5.3, the elicited empathic emotion is modulated by means of three factors. First, EMMA's changing **mood** over time. Second, **liking** as the degree to which EMMA likes MAX. And third, **familiarity** as the degree to which EMMA is familiar with MAX. OCC [88] distinguish between *momentary* and *dispositional liking*. *Momentary liking* refers to the value of *liking* in a defined moment while *dispositional liking* refers to the basic value of *liking*. For example, one may momentarily feels less liking for an otherwise dispositionally liked person. The values of *momentary* and *dispositional liking* influence each other. With regard to *familiarity*, its value could increase or decrease during the interaction, thus affecting its basic or its dispositional value. In this scenario, we consider the values of *liking* and *familiarity* as dispositional values. Thus, their values are predefined and do not change during the interaction. Changing values of *liking* during the interaction are taken into account in

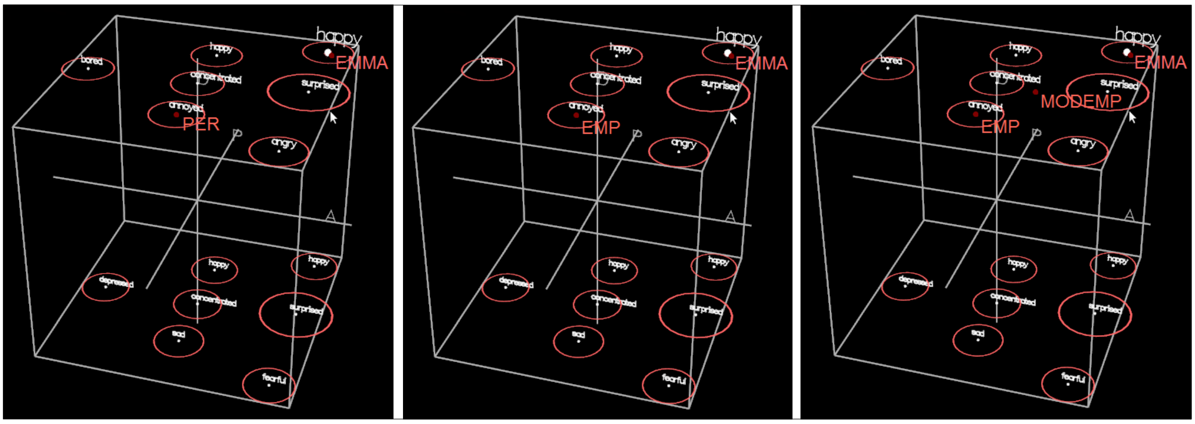


Figure 6.2: The PAD space of EMMA's emotion simulation module [8]. **PER** denotes the perceived emotional state, **EMP** denotes the elicited empathic emotion, and **MODEMP** denotes the modulated empathic emotion.

the scenario presented in Section 6.1.2. Depending on the values and weights of *dispositional liking* and *familiarity*, **different profiles** of EMMA's modulated empathy with MAX can be defined and considered. That is, the higher the values of *liking* and *familiarity*, the higher EMMA's degree of empathy with MAX. Regarding the modulation factor *mood*, the more similar EMMA's emotional state to MAX's perceived emotional state, the higher EMMA's degree of empathy with MAX. For example, when EMMA's emotions are triggered negatively by the human partner, EMMA's degree of empathy with MAX's positive emotions is rather low or equals 0, except when she has a high value of relationship (*liking* and *familiarity*) to MAX. Once the modulated empathic emotion (see Figure 6.2, right) is facilitated, a degree of empathy is calculated and the next processing step, expression of empathy, is triggered.

Based on the processing step expression of empathy, EMMA's context independent, as well as context dependent, modalities are triggered (see Section 5.4). In this scenario, EMMA's **context dependent modality** consists of a verbal utterance triggered when, in turn, the human partner's verbal utterance triggers MAX's emotions. The human partner's input utterance is evaluated with the recognized change in MAX's perceived value of pleasure. The changes in MAX's perceived value of pleasure are detected by EMMA based on calculating the difference of the pleasure values, $P_{t_k} - P_{t_{k-1}}$, at timestamps t_{k-1} and t_k. Accordingly, a positive change in MAX's perceived value of pleasure triggers utterances from EMMA that encourage the human partner to continue being kind to MAX. A positive change that results in a positive value of pleasure triggers utterances such as *That's great! You are so kind to MAX!*. A positive change in the space of negative pleasure triggers utterances such as *This is not enough! try to be kinder to MAX!*. A negative change in pleasure triggers utterances from EMMA that advises the human partner not to be unfriendly to MAX. A negative change in the space of positive pleasure triggers utterances such as *Why are you saying that to MAX! This is nasty!*. A negative change that results in a negative pleasure value triggers utterances such as *Better think about what you are saying! This is really nasty!*.

As long as no further empathic emotion is elicited, EMMA's modulated empathic emotion decays over time (cf. Section 5.3.2). Accordingly, the higher EMMA's degree of empathy with MAX, the slower her modulated empathic emotion decays, and vice versa. Once the PA values of the modulated empathic emotion equal the value 0, EMMA returns to her own emotional state.

In this scenario, we illustrated how by means of our model, EMMA empathizes with MAX's emotions to different degrees depending on her *mood* and on her defined values of relationship (*liking* and *familiarity*) to MAX. An example dialog situation within this scenario is presented in Section 6.2.2. In the following, another application scenario of our model is presented.

6.1.2 A spatial interaction task scenario

A cooperation between the present thesis project and the thesis project of Nhung Nguyen [81] was realized under the supervision of Prof. Ipke Wachsmuth [13]. In this context, our proposed model (see Chapter 5) is applied within the spatial interaction task scenario realized in the work by Nguyen & Wachsmuth [82].

Nguyen & Wachsmuth [82] propose a computational model for structuring and controlling a virtual human's **spatial behavior** in task execution at close distances. In this regard, the space surrounding the virtual human, which they define as *peripersonal space*, and the space in interpersonal interaction, which they define as *interaction space*, are modeled. The peripersonal space is subdivided into *touch space* and *lean-forward space* (see Figure 6.3).

The boundary of touch space corresponds to the lengths of the virtual human's arms. The boundary of lean-forward space corresponds to the maximal reach of the virtual human when bending forward from the waist. The virtual human assumes that another

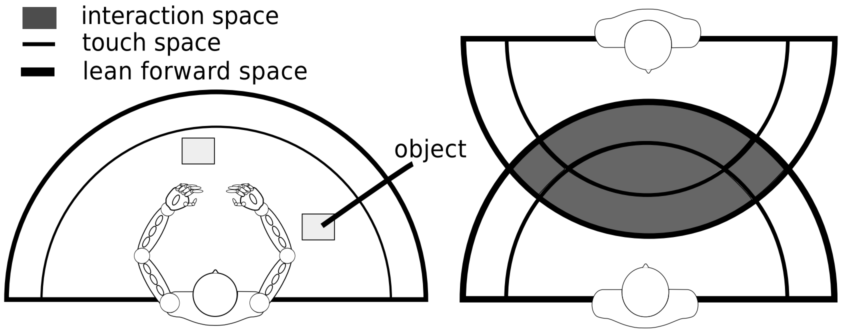


Figure 6.3: Subspaces surrounding a virtual human [82]. Left: the peripersonal space as subdivided into touch space and lean-forward space. The lean-forward space is an extension of touch space. Right: interaction space as the intersection of two partners' peripersonal spaces.

agent entering his proximity is surrounded by the same peripersonal space as his own. Thus, the interaction space is modeled as the intersection of the peripersonal spaces of the virtual human and the other agent. The computational model proposed by Nguyen & Wachsmuth [82] is realized for the virtual human MAX in a spatial interaction task scenario with a human interaction partner.

The spatial interaction task scenario consists of a **cooperative tower building task**. In a virtual reality CAVE-like environment, MAX and the human partner are standing face-to-face at a table. Their overlapping peripersonal spaces form their interaction space. The goal is to cooperatively solve a tower building task with virtual toy blocks by alternatively putting one toy block upon the other. The toy blocks are labeled with different numbers and have different sizes. The numbers descend with lower sizes. At the beginning of the task, the largest toy block is placed by default in the center of interaction space where MAX and the human partner have to place the remaining blocks. In order for a human interaction partner to select and place the virtual toy blocks, a Wiimote controller is used. Figure 6.4 shows an example interaction where the human partner selected block number nine and placed it upon the largest toy block labeled **T**.

The blocks are placed randomly at free locations within MAX's and the human partner's peripersonal spaces. Each partner is assigned a number of blocks with respect to a defined minimum. The following two rules are defined to build a tower. First, *the blocks can be ordered by their numbers*, e.g., block number three can be placed on top of block number four. When all blocks are ordered with respect to their numbers, the highest tower, which we call *ideal tower* is built. Second, *the blocks can be ordered by*

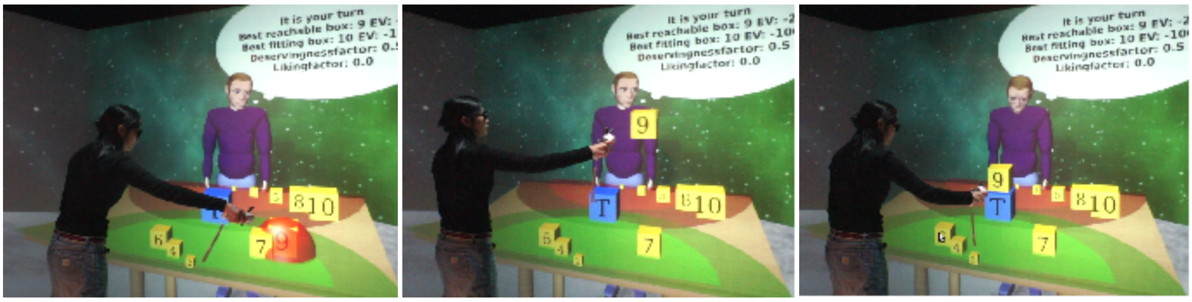


Figure 6.4: The virtual human MAX and a human interaction partner in the spatial interaction task scenario. From left to right, the human partner starting the task by selecting block number nine and by placing it upon the largest block labeled **T**. The red and green semi circles, respectively, represent the peripersonal spaces of MAX and the human partner. The bright arcs represent the lean-forward spaces while the dark semi circles represent the touch spaces. The intersection of both peripersonal spaces represents the interaction space (cf. Figure 6.3). The game situation as perceived by MAX is shown by his thought bubble.

their sizes, e.g., block number three can be placed on top of block number five omitting block number four. In this case, a tower that is smaller than the ideal tower is built. Accordingly, the *most appropriate block* is defined as the block that best fits both of these rules. To successfully achieve the tower building task, each partner should place the most appropriate block located in his peripersonal space. Once the smallest block with number one is placed on the top of the tower, the tower building task is completed.

As mentioned above, the goal of the spatial interaction task is to cooperatively solve the tower building task. For this purpose, a **cooperative spatial action**, called *helping action*, is defined as relocating objects into positions reachable by the partner. For example, it is possible that the most appropriate block needed by one of the partners is not reachable. In this case, the helping action consists of relocating the most appropriate block into a position reachable by the partner. Based on the size and layout of interaction space [82], the virtual human MAX can relocate objects to any free positions reachable by his partner. In this regard, the following question arises: *Which position within interaction space is chosen by MAX to help the human partner?*

According to Hoffman [55] and Davis [34], empathy plays a major role as a motivational basis of cooperative prosocial behavior such as helping and caring (cf. Section 2.1.2). Hence, based on the computational model of empathy proposed in this thesis (see Chapter 5), MAX's **helping action** is triggered and is modulated by his degree of empathy with his interaction partner. In the following, we illustrate MAX's empathic behavior within this scenario.

During the partner's game turn, MAX's attention focus is on the human partner and the empathy mechanism **situational role-taking** introduced in Section 5.2.2 is triggered. As mentioned in Section 5.2.2, the realization of situational role-taking requires a context scenario and is introduced in more details in this section. As defined in Section 5.2.2, situational role-taking refers to the ability to generate a hypothesis about the other's emotional state by appraising the other's situation using one's own appraisal mechanisms. Hence, in the spatial interaction task scenario, MAX generates a hypothesis about his human partner's emotional state by appraising his partner's game situation using his own appraisal mechanisms.

During his game turn, MAX appraises his situation based on a sequence of plans defined in his BDI-module [70]. MAX's **BDI-plan** to achieve the goal of placing a block on top of the tower is depicted by Algorithm 1. Accordingly, once MAX has the turn, this BDI-plan is triggered (line 2). In the Skip-Bo card game scenario [8] (see Section 3.1.2), MAX's value of dominance is set to positive during his turn and to negative during his partner's turn. Similarly, MAX's value of dominance during his game turn in the present scenario is set to 100, while during his partner's game turn it is set to the value -100 (lines 4 and 24). For a successful completion of the tower building task, MAX first searches for the most appropriate block to place on top of the tower (line 6). Depending on the block's position, a corresponding value of emotional valence, also called *emotional impulse* [8] (cf. Section 3.1.1), is triggered according to a defined **cost function** (lines 7 to 11 and 14 to 16). This cost function associates the physical effort of motor actions with emotional valences. The **physical effort** of motor actions is represented by object positions within peripersonal space. That is, with increasing arm-reach distances for objects, more physical effort is needed for humans to reach for them (cf. [72]). This cost function is depicted by Figure 6.5.

According to this cost function, blocks located in MAX's touch space are reachable with less physical effort and are thus associated with lower costs represented by positive values of emotional valences ranging in $[0, 100]$. The value 100 corresponds to the center of MAX's touch space. Blocks located in MAX's lean forward space are reachable with more physical effort and are thus associated with higher costs represented by negative values of emotional valences ranging in $[-100, 0]$. The value 0 corresponds to the boundary of touch space. Blocks located in extrapersonal space are not reachable and are associated

Algorithm 1 MAX places block on top of the tower.

```
Goal: ACHIEVE PLACE-BLOCK-ON-TOP-OF-TOWER
Precondition maxHasTurn
Body
setDominance positive
playedMove = false
while block ← getNextMostAppropriate(freeBlocks) and !playedMove do
   blockPosition ← getPosition(block)
   if blockPosition in reach-space then
      grasp block
      emotionalValence ← costFunction(blockPosition)
      send emotionalValence
      placeOnTopOfTower( block)
      playedMove = true
   else
      emotionalValence ← costFunction(blockPosition)
      send emotionalValence
      numNotReachBlocks += 1
   end if
   if numNotReachBlocks = numAllApproprBlocks then
      fail
   end if
end while
setGameTurn partner
setDominance negative
Effects send emotionalValencePositive
Failure send emotionalValenceNegative
```

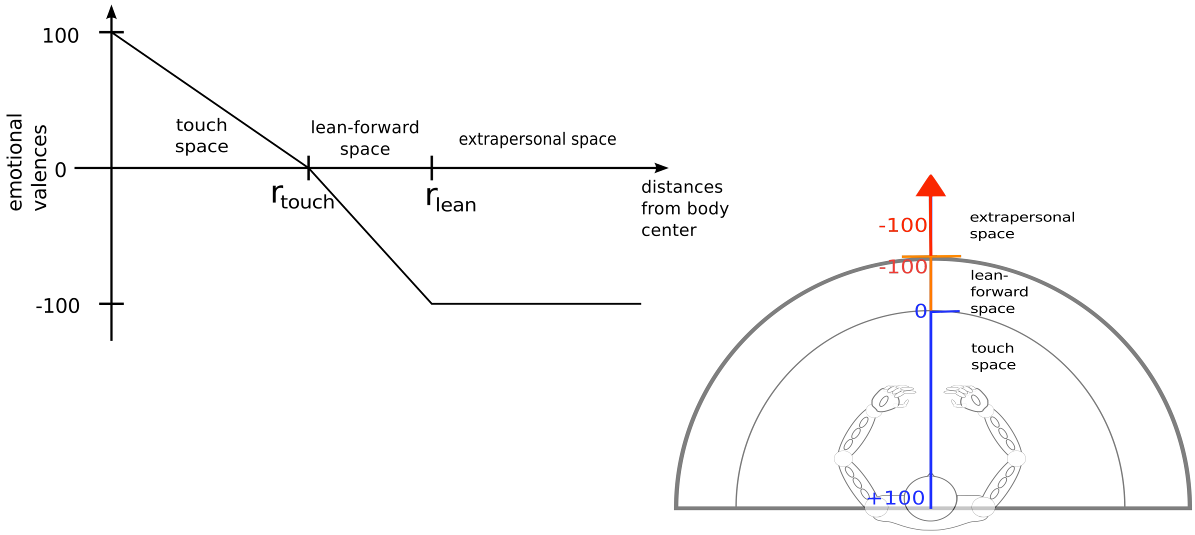


Figure 6.5: The defined cost function associating the physical effort of motor actions with emotional valences.

with a maximal negative value of emotional valence equal to −100. Further, given the case that the most appropriate block needed by MAX is not reachable, MAX searches for the next most appropriate and reachable one. Once such a block is located by MAX, he places it on top of the tower (line 12). This action is interpreted as **goal success** and is rewarded with a positive value of emotional valence equal to 40 (line 25). In case no appropriate and reachable block is available (lines 19 to 21), MAX skipped a turn. This is interpreted as **goal failure** and triggers a negative value of emotional valence equal to −40 (line 26).

As mentioned earlier, during the partner's game turn, MAX's attention focus is on the partner, and the empathy mechanism situational role-taking (cf. Section 5.2.2) is triggered. Thus, MAX appraises the partner's situation using his own appraisal mechanisms based on a sequence of plans defined in his BDI-module. MAX's **BDI-plan** to appraise the partner's achievement of the goal of placing a block on top of the tower is depicted by Algorithm 2. This BDI-plan is triggered once the partner has the turn (line 2). As for himself, MAX assumes the partner's value of dominance is equal to 100 (line 4). Furthermore, MAX searches for the most appropriate block needed by the partner to put on top of the tower (line 5). Once such a block is located, MAX infers the block's position in the partner's peripersonal space (line 6). Note that MAX assumes that his partner is surrounded by a peripersonal space which is the same as his own. Thus, MAX

Algorithm 2 Partner places block on top of the tower.

Goal: ACHIEVE PLACE-BLOCK-ON-TOP-OF-TOWER PARTNER
Precondition $partnerHasTurn$
Body
setDominance $positive$
$block \leftarrow$ getMostAppropriate($freeBlocks$)
$blockPosition \leftarrow$ getPosition($block$)
$emotionalValence \leftarrow$ costFunction($blockPosition$)
send $emotionalValence$
$numNotReachBlocks \leftarrow$ getNumNotReachBlocks($freeBlocks$)
if $numNotReachBlocks = numAllApproprBlocks$ **then**
    fail
end if
$block \leftarrow$ getGraspBlock($freeBlocks$)
$blockPosition \leftarrow$ getPosition($block$)
$emotionalValence \leftarrow$ costFunction($blockPosition$)
send $emotionalValence$
if placeOnTopOfTower($block$) **then**
    success
end if
setGameTurn $self$
setDominance $negative$
Effects send $emotionalValencePositive$
Failure send $emotionalValenceNegative$

creates a model of his partner's peripersonal space by projecting his own peripersonal space to the partner. Using one's own body model to simulate someone else's perspective is known as **embodied simulation** [82] (cf. [43], see Section 2.1.3). A corresponding value of emotional valence is triggered for the partner according to the cost function depicted by Figure 6.5 (lines 7 and 8). Accordingly, embodied simulation, followed by the generation of a value of emotional valence, are triggered as lower-level cognitive processes defined within the reactive layer of the virtual human's cognitive architecture [70] (see Section 5.1.2).

The partner's assumed value of dominance together with the triggered value of emotional valence for the partner are input to MAX's emotion simulation module [8] and drive the simulated dynamics of the partner's perceived emotional state over time (cf. Section 5.2.3). The perceived emotional state is represented by a second reference point within MAX's emotion simulation module, thus allowing him to distinguish between his own and the partner's emotional state (cf. Figure 6.2, left). The perceived emotional state is asserted as **belief** about the partner's emotional state within MAX's BDI-module. Once the partner's perceived emotional state exceeds a predefined saliency threshold, or a salient change in the partner's perceived emotional state is detected (cf. Section 5.2.3), an empathic emotion is elicited (cf. Figure 6.2, middle) and the next processing step, empathy modulation, is triggered.

Per empathy modulation (see Section 5.3.1), the elicited empathic emotion is modulated by means of three factors. First, MAX's **mood** during the game. Second, *liking* as the degree to which MAX likes the human partner during the game. Third, *deservingness* as the degree to which the partner deserves his current game situation. The values of *liking* and *deservingness* range in $[0, 1]$ from neither liked nor disliked to most liked, and from neither deserved nor not deserved to most deserved. Note that positive values of *deservingness* represent deserved positive events and not deserved negative ones. The values of *liking* and *deservingness* change during the interaction and are asserted as **beliefs** in MAX's BDI-module. Thus, we consider the values of *liking* as momentary (cf. OCC [88]); (see Section 6.1.1). In the following, the calculation of the values of *liking* and *deservingness* is introduced.

Liking is calculated as the partner's assumed degree of empathy with MAX. The partner's degree of empathy with MAX is calculated based on the partner's investment in helping actions. That is, the more the partner helps MAX, the higher his assumed degree of empathy, the higher MAX likes his partner. In this regard, we follow Davis' emphasis that one's ability for empathy impacts one's social behavior which is perceived by others and which thus influences one's social relationships with others [34] (cf. Section 2.1.2). The relationship between helping actions and degree of empathy is introduced later in this section.

Deservingness is calculated as the number of reachable appropriate blocks in MAX's touch space divided by the number of all current appropriate blocks. That is, the more reachable appropriate blocks are in MAX's touch space, the higher the value of *deservingness* and vice versa.

The higher the values of *liking* and *deservingness*, the higher MAX's degree of empathy with the human partner. The more similar MAX's emotional state to the partner's perceived emotional state, the higher MAX's degree of empathy with the human partner. Once the modulated empathic emotion (cf. Figure 6.2, right) is facilitated, the next processing step, expression of empathy, is triggered.

Based on the processing step expression of empathy, MAX's context independent as well as context dependent modalities are triggered (see Section 5.4). In this scenario, MAX's **context dependent modality** consists of a helping action, defined earlier in this section, as relocating the most appropriate block toward positions reachable by his partner. MAX's helping action is triggered only if the partner's pleasure value becomes negative, and if the most appropriate block is reachable for MAX. By calculating the difference in the partner's perceived pleasure value as $P_{t_k} - P_{t_{k-1}}$, at time-stamps t_{k-1} and t_k, a helping action is triggered when $P_{t_k} - P_{t_{k-1}} <= 0$ and $P_{t_k} <= 0$. Accordingly, MAX do not always help as soon as the most appropriate block is in the partner's lean forward space, thus leaving some room for the partner to perform the task by himself. Further, MAX's helping action is modulated by his degree of empathy with his partner. MAX's degree of empathy is calculated as introduced in Section 5.3.2.

MAX's **helping action** consists of placing the most appropriate block into a position within interaction space that is reachable by the partner. The boundary of interaction space corresponds to two circular arcs spanned by the lean forward spaces of MAX and his partner (see Figure 6.6, gray area). These circular arcs are denoted respectively as $leanArc_{MAX}$ and $leanArc_{partner}$. The closest position to the partner, where MAX can place a block is denoted as P_m (see Figure 6.6). P_m is defined as the intersection of $leanArc_{MAX}$ and the line segment spanned by MAX's center of peripersonal space and the partner's center of peripersonal space. The corresponding position vector is denoted by $\mathbf{p}_m$. The position vector of the most appropriate block needed by the partner is

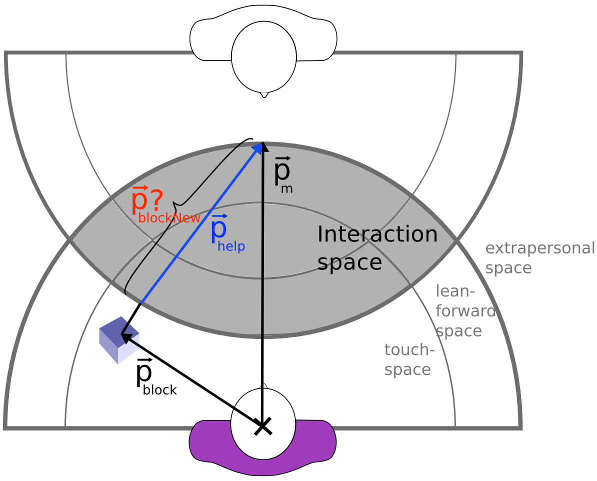


Figure 6.6: MAX's helping action in interaction space.

denoted by $\mathbf{p}_{block}$ (see Figure 6.6). Based on the following equation, the new position of the block is calculated:

$$\mathbf{p}_{blockNew} = \mathbf{p}_b + (degEmp \cdot \mathbf{p}_{help}), \text{ with } \mathbf{p}_{help} = \mathbf{p}_m - \mathbf{p}_{block} \tag{6.1}$$

degEmp denotes the calculated degree of empathy for MAX and determines the distance by which the most appropriate block is moved toward the partner. A maximally distant position, where a most appropriate block can be relocated, is defined as the intersection of $leanArc_{partner}$ and the line segment spanned by $\mathbf{p}_m$ and $\mathbf{p}_{block}$. Hence, *degEmp* modulates only the part of $\mathbf{p}_{help}$ within interaction space. This prevents helping actions by which the new block position is located outside interaction space where the block is still not reachable by the partner.

Furthermore, the **effort quality** of MAX's helping action is also modulated by his degree of empathy (cf. [26], see Section 3.2.1). In this regard, MAX's degree of empathy modulates the velocity of MAX's replacement movement by modulating its *velocity-factor*. The velocity-factor of the replacement-movement takes values between 0 and 1, and modulates the overall duration of the movement. Each replacement-movement has a maximum and a minimum time of execution. The value of 0 denotes the slowest predefined replacement-movement. With increasing value, the duration of the movement decreases linearly. This means that the duration time becomes shorter and the movement is executed faster until it reaches the value of 1, denoting the fastest predefined replacement-movement. The standard predefined value of the maximum overall duration is set to 11 seconds and the minimum is set to 3 seconds. If, for example, the degree of empathy is of value 0.5, the velocity-factor of the replacement-movement is also set to 0.5. This results in an overall duration of 7 seconds for MAX's entire movement. Thus, MAX's degree of empathy also impacts his **behavior expressivity** during his helping action.

MAX's modulated empathic emotion decays over time (cf. Section 5.3.2). Accordingly, the higher MAX's degree of empathy with his partner, the slower his modulated empathic emotion decays, and vice versa. Once the PA values of the modulated empathic emotion equal the value 0, MAX returns to his own emotional state.

Further, the virtual human MAX waits until the partner has performed his game moves. Once the partner grasps a block, MAX infers the block's position in the partner's peripersonal space (lines 13 and 14, Algorithm 2) and triggers a corresponding value of emotional valence for the partner according to the cost function depicted by Figure 6.5 (lines 15 and 16). When the partner puts the block on top of the tower, MAX interprets this action as **goal success** and triggers an emotional impulse equal to 40 for the partner (lines 17 to 19 and line 22). In case no appropriate block is reachable by the partner and MAX's helping action is not triggered, MAX interprets this game situation as **goal failure** and triggers an emotional impulse for the partner equal to -40 (lines 9 to 12 and line 23). These emotional impulses further drive the simulated dynamics of the partner's perceived emotional state over time.

In the following, an example interaction between MAX and the human partner in this scenario is described. Figure 6.7 illustrates an example sequence of **MAX's game moves**. As shown by Algorithm 1, MAX first searches for the most appropriate block, eight in our example situation (see Figure 6.7, top, left). Since the most appropriate block is in MAX's forward lean space, MAX looks at his partner and waits for his partner's help (see Figure 6.7, top, right). By this behavior MAX signalizes to his partner that he would like to be helped. In case the partner does not help, MAX grasps the block, receives a negative emotional impulse of -70, and places it on top of the tower (see Figure 6.7, bottom, left). After positioning the block on top of the tower, MAX receives a positive emotional impulse of 40 as a reward for his goal success (Figure 6.7, bottom, right). However, MAX is expressing a negative emotional state because of the previously triggered negative emotional impulse of -70. The value of *deservingness* is first equal to 0.5 because both MAX and the human partner have the same number of appropriate

Figure 6.7: Example sequence of MAX's game moves.

blocks. This value decreases after MAX places the block number eight on top of the tower thus decreasing the number of his most appropriate blocks. The value of *liking* is equal to 0 because the partner has not helped MAX.

Figure 6.8 illustrates an example situation where the **partner helps MAX**. In this situation, MAX needs block number six which is the most appropriate block. However, the block is in MAX's extrapersonal space. Thus, a negative emotional impulse of -100 is triggered for MAX (see Figure 6.8, top, left). As mentioned earlier, MAX waits for the help of his partner (see Figure 6.8, top, right). In fact, the partner helps MAX by placing the needed block in MAX's reach space (see Figure 6.8, bottom, left). Accordingly, MAX's value of *liking* increases based on the partner's investment in helping him (see Figure 6.8, bottom, right). MAX then places the block on top of the tower resulting first in a negative emotional impulse of -30 since the needed block was placed in his forward-lean space, and then a positive emotional impulse of 40 as a reward for his goal success (see Figure 6.8, bottom, right). Note that during the partner's achievement of the helping action, MAX's negative emotional state triggered at the beginning by the emotional impulse of -100, decays to neutral. As compared to the previously described example situation (see Figure 6.7), MAX is expressing a positive emotional state after

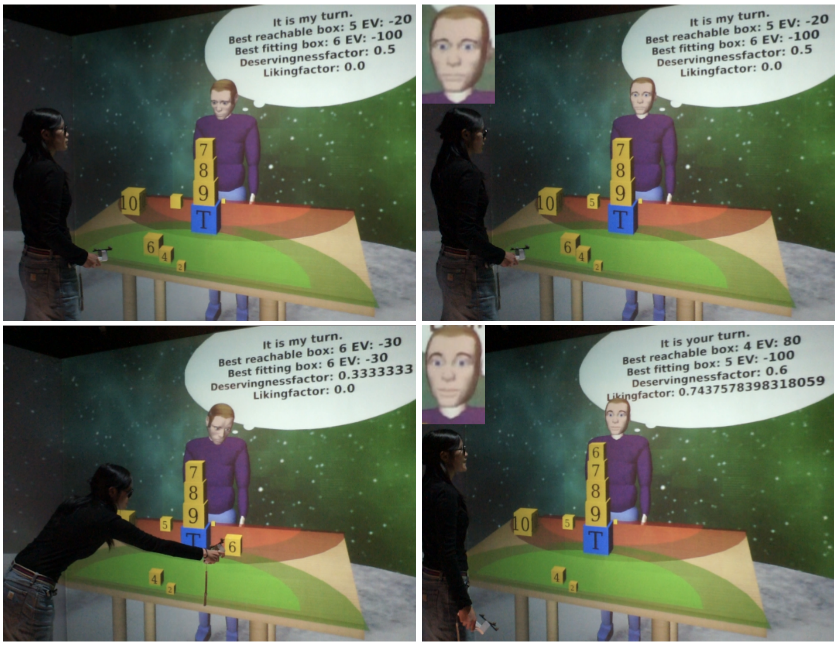


Figure 6.8: Example situation where the human partner helps MAX.

placing the block on top of the tower. This is in line with Davis's definition of helping as an action that at some cost to the self, reduces the negative states or increases the positive states for the other [34] (cf. Section 2.1.2).

Figure 6.9 illustrates an example situation where **MAX helps his partner**. In this situation, MAX searches for the most appropriate block needed by his partner, namely, block number five (Figure 6.9, left). Accordingly, an emotional impulse of -100 is triggered for the partner since the most appropriate block is not reachable by the partner. Subsequently, MAX's empathic emotion is elicited and a degree of empathy is calculated according to the values of *liking*, of *deservingness*, and of MAX's *mood*. Furthermore, MAX's helping action is triggered and is modulated by the value of his degree of empathy (see Figure 6.9, left). Note that this example situation is not a sequel to the one illustrated by Figure 6.8.

In this scenario, we illustrated how MAX's **degree of empathy** with his partner triggers and modulates his helping action. Furthermore, MAX's value of *liking* is based on the partner's investment in helping him. Thus, the more the partner helps MAX, the more MAX likes his partner and the more he empathizes and helps him. In the following section, we introduce the empirical evaluation of our model in the context of the conversational agent scenario presented in Section 6.1.1.

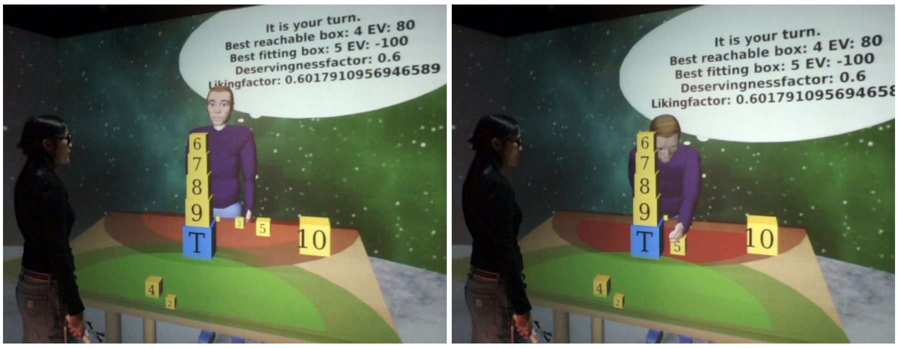


Figure 6.9: Example situation where MAX helps his partner.

6.2 Empirical evaluation

After the successful application of the proposed computational model of empathy within two different interaction scenarios (see Section 6.1), an empirical evaluation of the model was carried out, and is presented in the next sections. This evaluation was performed in **cooperation** with Prof. Pia Knoeferle and Dr. Maria Nella Carminati as members of the psycholinguistics section of project A1 'Modelling Partners' (see Chapter 1).

6.2.1 Hypotheses

The purpose of the empirical evaluation is to investigate how a virtual human's behavior produced by the empathy model is perceived and interpreted by human participants. Accordingly, the empirical evaluation of the empathy model is performed in the context of the conversational agent scenario introduced in Section 6.1.1 to test the following hypotheses:

- **H1**: the virtual human EMMA's expression of empathy is perceivable by the participants.
- **H2**: the virtual human EMMA's expressed degree of empathy is perceivable by the participants.
- **H3**: the human participants acknowledge different values of relationship between EMMA and MAX according to EMMA's expressed degree of empathy.
- **H4**: the virtual human EMMA is perceived as more likable the higher the value of her expressed degree of empathy.

6.2.2 Design and procedure

For the purpose of this empirical evaluation, we designed **24 dialog interactions** to be used in a repeated measures design. The interactions were between EMMA, MAX, and a human partner who we called Lisa. At the beginning of each of the 24 dialog interactions, the virtual humans are in a neutral emotional state. In each of the 24 dialog interactions, Lisa begins by greeting EMMA and then praising her. Consequently, a positive emotional impulse of +100 is input to EMMA's emotion simulation module, activating her primary emotion category *happy*. Simultaneously, EMMA greets Lisa and thanks her for being kind. Then Lisa greets MAX although she then insults him. Thus, a negative emotional impulse of −100 is input to MAX's emotion simulation module, activating his primary emotion category *angry*. Simultaneously, MAX answers with a negative verbal utterance such as *Lisa, you are horrible*. Meanwhile, EMMA empathizes with MAX to different degrees depending on her mood and her defined relationship to MAX. Note that MAX's facial expression for *angry* is perceived by EMMA as showing the emotional state *annoyed* (cf. Table 5.1, p. 111).

Regarding EMMA's **degree of empathy** with MAX, three different conditions are considered for each of the 24 dialog interactions. In the three conditions, EMMA is in the same positive mood after being praised by Lisa. Thus, the modulation factor *mood* is constant across all conditions. Accordingly, the three conditions consist of varying the value of EMMA's relationship to MAX, thus impacting her degree of empathy with MAX. Since the focus is not on the interaction of several relationship modulation factors, only the factor *dispositional liking* is manipulated. The **three conditions** are as follows:

1. EMMA's value of liking toward MAX is equal to 0. Accordingly, EMMA's modulated empathic emotion is inhibited and EMMA's value of degree of empathy is also equal to 0. In this case, EMMA continues expressing her own positive emotional state of happiness triggered by Lisa's praising expression. We call this condition the **neutral liking condition** (see Figure 6.10, left).

2. EMMA's value of liking toward MAX is equal to 0.5. Accordingly, EMMA's modulated empathic emotion is facilitated and has as its related primary emotion category *concentrated*. In this case, EMMA expresses the modulated empathic emotion. EMMA's value of degree of empathy is equal to 0.25. Hence, EMMA's values of liking and degree of empathy are higher than in the first condition. We

call this condition the **medium liking condition** (see Figure 6.10, middle).

3. EMMA's value of liking toward MAX is equal to 1. Accordingly, EMMA's modulated empathic emotion equals the non-modulated empathic emotion which has as its related primary emotion category *annoyed*. In this case, EMMA expresses the non-modulated empathic emotion. EMMA's value of degree of empathy is equal to 1. Hence, EMMA's values of liking and degree of empathy are higher than in the first and second condition. We call this condition the **maximum liking condition** (see Figure 6.10, right).

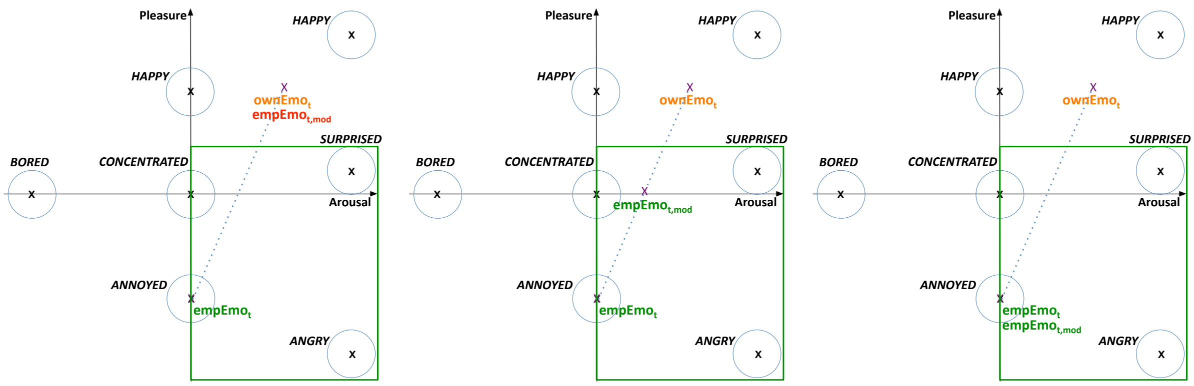


Figure 6.10: EMMA's modulated empathic emotion, $empEmo_{t,mod}$, in PAD space of positive dominance. $ownEmo_t$ refers to EMMA's own emotional state, $empEmo_t$ refers to the non-modulated empathic emotion, and the box represents the *empathy facilitation region* defined for the emotion category *annoyed* (cf. Section 5.3). From left to right, respectively, EMMA's modulated empathic emotion, $empEmo_{t,mod}$, in the neutral liking condition, the medium liking condition, and the maximum liking condition.

In all three conditions, EMMA's degree of empathy is expressed by her **facial expression and speech prosody** while the verbal utterance is the same across the three conditions. Thus, after MAX's response to Lisa, EMMA also responded with a negative verbal utterance such as *You are nasty to MAX*. EMMA's and MAX's secondary behaviors and conversational gestures are deactivated in all three conditions. Figure 6.11 depicts the utterances of EMMA, MAX, and Lisa in two example dialog interactions.

In order to obtain a controlled environmental setting for the evaluation of EMMA's empathic behavior, a total of **72 videos** of the 24 Dialog interactions in the three conditions were recorded. Figure 6.12 depicts the setting for the video recordings.

Example Dialog 1	Example Dialog 2
HP: Hallo EMMA, ich finde dich hübsch (Hello EMMA, you look pretty)	**HP**: Hallo EMMA, ich finde dich klug (Hello EMMA, you are clever)
E: Hallo Lisa, das ist lieb von dir (Hello Lisa, you are nice)	**E**: Hallo Lisa, das ist großartig von dir (Hello Lisa, you are great)
HP: Hei MAX, du bist mir zu hässlich (Hey MAX, you look ugly)	**HP**: Hei MAX, du scheinst mir blöd zu sein (Hey MAX, you seem to be stupid)
M: Nun Lisa, du bist fies (So Lisa, you are nasty)	**M**: Nun Lisa, du bist unhöflich (So Lisa, you are rude)
E: Du bist aber unmöglich zu MAX (You are obnoxious to MAX)	**E**: Du bist aber fies zu MAX (You are nasty to MAX)

Figure 6.11: Two example dialog interactions of EMMA, MAX, and Lisa. **HP** refers to human partner, **E** refers to EMMA, and **M** refers to MAX.

Figure 6.12: Setting for the video recordings.

In order to present a more natural interaction to the participants, the human partner in the videos does not use a keyboard to input his responses. Instead, another person in the background typed them while the human partner spoke. Figure 6.13 presents a snapshot of the last frame of a video showing a dialog interaction and EMMA's facial expression of empathy in the three conditions.

Figure 6.13: Snapshot of the last frame of a video showing a dialog interaction. From left to right: EMMA's facial expression of empathy in the three conditions.

For the purpose of assigning the video recordings to participants, we constructed **three experimental lists** following the Latin Square design. Each list comprises 24 videos from each condition (giving a total of 72 videos) such that each video appeared in each list in only one condition, as shown by Table 6.1. Altogether, **30 participants** (15 female and 15 male) took part in the experiment, with each list assigned to 10 participants, and with an equal number of male and female participants assigned to each list. The 24 videos contained in a list were presented in a random order to each corresponding participant.

To test the hypotheses formulated in Section 6.2.1, each participant was asked to complete a **questionnaire after each shown video**. The questionnaire comprises five questionnaire items that were rated using a 7-point Likert scale ranging from -3 to $+3$. The items are listed in Table 6.2. The first two items were used to test hypothesis **H1**. The third, fourth, and fifth items were respectively used to test hypotheses **H2**, **H3**,

Dialog Number	List 1	List 2	List 3
D1	video(cond1)	video(cond2)	video(cond3)
D2	video(cond2)	video(cond3)	video(cond1)
D3	video(cond3)	video(cond1)	video(cond2)
D4	video(cond1)	video(cond2)	video(cond3)
D5	video(cond2)	video(cond3)	video(cond1)
…	…	…	…
D24	video(cond3)	video(cond1)	video(cond2)

Table 6.1: The recorded videos as arranged in three lists according to the Latin Square design. 'cond1' refers to the neutral liking condition, 'cond2' refers to the medium liking condition, and 'cond3' refers to the maximum liking condition.

and **H4**.

Measure	Questionnaire item	Scale
Expression of empathy (**H1**)	"In the last frame of the video, EMMA's face shows: " "In the last frame of the video, EMMA's speech prosody is: "	-3 = very negative mood +3 = very positive mood -3 = very negative +3 = very positive
Degree of empathy (**H2**)	"In this video, EMMA is: "	-3 = very cold to MAX +3 = feeling with MAX
Values of relationship (**H3**)	"In this video, EMMA has: "	-3 = very negative relationship to MAX +3 = very strong relationship to MAX
Likability (**H4**)	"In this video, EMMA is overall: "	-3 = very unlikable +3 = very likable

Table 6.2: A schematic overview of the questionnaire presented after each shown video to test hypotheses **H1**, **H2**, **H3**, and **H4**.

After having watched all 24 videos, the participants were further asked to complete **two post-experimental questionnaires**. The first questionnaire comprises 21 questionnaire items. The first item was to evaluate the participants' understanding of the content of the presented videos. As such, the participants were asked to describe the general story presented in the videos, and if they can discern any perceived differences in their content. The remaining questionnaire items are grouped into four categories and were rated using a 7-point Likert scale ranging from -3 to $+3$. Accordingly, nine questionnaire items were used to evaluate EMMA's general behavior in the videos (see Table 6.3). Three questionnaire items were used to evaluate the naturalness of EMMA's behavior in the videos (see Table 6.4). Four questionnaire items were used to evaluate EMMA's social behavior over all presented videos (see Table 6.5). Finally, four questionnaire items were used to evaluate the general likability of EMMA (see Table 6.6).

Measure		Questionnaire item	Scale
EMMA's general behavior in the videos	Facial behavior	"EMMA's face was expressive: "	-3 = strongly disagree +3 = strongly agree
		"EMMA's facial expressions were exaggerated: "	-3 = strongly disagree +3 = strongly agree
	Speech behavior	"EMMA's pronunciation was clear: "	-3 = strongly disagree +3 = strongly agree
		"EMMA's speech was difficult to understand: "	-3 = strongly disagree +3 = strongly agree
	Attention behavior	"EMMA had eye contact with Lisa: "	-3 = strongly disagree +3 = strongly agree
		"EMMA never had eye contact with MAX: "	-3 = strongly disagree +3 = strongly agree
	Empathic response trigger	"In the last frame of all presented videos, EMMA's response toward Lisa is triggered by MAX's facial expression: "	-3 = strongly disagree +3 = strongly agree
		"In the last frame of all presented videos, EMMA's response toward Lisa is triggered by Lisa's utterance toward MAX: "	-3 = strongly disagree +3 = strongly agree
	Behavior in general	"The behavior of EMMA over all presented videos was exaggerated: "	-3 = strongly disagree +3 = strongly agree

Table 6.3: Questionnaire items to measure EMMA's general behavior in the videos.

The second post-experimental questionnaire comprises 14 questionnaire items to measure the **I**nterpersonal **R**eactivity **I**ndex (IRI) [33] of the participants thus measuring their empathy. This measure was collected so as to allow us to test any **correlation** that might exist between the participants' IRI and the ratings of the five questionnaire

Measure		Questionnaire item	Scale
Naturalness of EMMA's behavior in the videos	Facial behavior	"EMMA's facial expressions appeared: "	-3 = artificial +3 = natural
	Speech behavior	"EMMA's pronunciation appeared: "	-3 = artificial +3 = natural
	Attention behavior	"EMMA's attention behavior appeared: "	-3 = artificial +3 = natural

Table 6.4: Questionnaire items to measure the naturalness of EMMA's behavior in the videos.

Measure		Questionnaire item	Scale
EMMA's social behavior over all presented videos	Cooperative	"EMMA is a cooperative agent: "	-3 = strongly disagree +3 = strongly agree
	Caring	"EMMA is an uncaring agent: "	-3 = strongly disagree +3 = strongly agree
	Protective	"EMMA is a protective agent: "	-3 = strongly disagree +3 = strongly agree
	Social	"EMMA is an asocial agent: "	-3 = strongly disagree +3 = strongly agree

Table 6.5: Questionnaire items to measure EMMA's social behavior over all presented videos.

items after each shown video. The hypothesis here is: *The higher the value of IRI, the more participants can differentiate between the three conditions of EMMA's empathy with MAX in their ratings for the five questionnaire items*. This hypothesis is not central to the present thesis, and is thus not listed with the four hypotheses formulated in Section 6.2.1. The results regarding this hypothesis can be considered as a by-product of the present work. A German version of the questionnaire was used [92]. The questionnaire items were rated using a 5-point Likert scale. Note that the participants were advised about the scale change before they began the questionnaire.

At the end of the experiment, an **interview** was conducted with each participant to ask about the criteria on the basis of which they rated EMMA's likability (cf. Table 6.2)

Measure		Questionnaire item	Scale
General likability of EMMA	Trust	"You would trust EMMA if she were your friend: "	-3 = strongly disagree +3 = strongly agree
	Friendship	"You could not imagine EMMA to be your friend: "	-3 = strongly disagree +3 = strongly agree
	Interaction	"You would like to interact with EMMA: "	-3 = strongly disagree +3 = strongly agree
	Interest	"You are not interested in EMMA: "	-3 = strongly disagree +3 = strongly agree

Table 6.6: Questionnaire items to measure the general likability of EMMA.

after each presented video.

The questionnaire to test the hypotheses and the first post-experimental questionnaire were designed in the context of the present study according to the general criteria for a **questionnaire design**. Thus, the questions are written in everyday language using short and simple sentences. In order to avoid the acquiescence bias, half of the questions in the post-experimental questionnaire are worded in reverse. With regard to the questionnaire after each video, the questions were presented with an inverted scale for half of the participants. All questions were randomized for each participant to prevent order effects.

The videos and the questionnaires were presented to participants on a computer screen. A headset was used during the presentation of the videos (see Figure 6.14).

Figure 6.14: Setting for running the experiment.

Further, to ensure that the experimental design is clear and appropriate for the human participants, a **pretest** was performed with three participants who were unfamiliar with the research subject. The three participants, respectively, evaluated the videos contained in the three lists provided by the Latin Square design (cf. Table 6.1). The clarity of the experiment was confirmed by these participants such that no further changes were required.

As mentioned earlier, the study was conducted with a total of 30 participants (15 female, 15 male). The participants were unfamiliar with the subject and purpose of the experiment. The **participants' age** ranged between 21 and 38 years. Before the beginning of the experiment, the participants were asked to read and sign a consent form. Further, the participants received written instructions as to how to proceed during the experiment. They also watched an example video and got some example questions to become familiar with the experiment. The example video was not taken from the 72 recorded videos. Before the participants started with the experiment, they were asked if they had any clarifying questions. The experiment duration was between 30 and 45 minutes. At the end of the experiment, each participant received five euros and a button depicting EMMA's facial expression of *happy* as a thank-you gift for their willingness to participate.

6.2.3 Results

After successful collection of the participants' ratings, a statistical analysis of the collected data was conducted. In the following, the results of the statistical analysis of the data are reported for the questionnaire presented after each shown video and for the post experimental questionnaires.

Questionnaire after each video

We began the analysis of the rating data by calculating the **mean rating** by condition for each of the five questionnaire items (see Table 6.2) for participants and items (i.e. videos) separately. Next, we performed an **omnibus mixed design two-way ANOVA** using participants and items as random effects. Condition (3 levels: neutral liking, medium liking, maximum liking) was a within-participant factor, while gender (2 levels: male, female) was a within-participants in the ANOVA by participants and a within-items in the ANOVA by items. The results of the omnibus ANOVA show a significant effect of condition for all five questionnaire items. However, no significant effect of gender was found nor did the gender factor interact with condition in any of the questionnaire items. Accordingly, we did not carry out any further tests involving the gender factor.

To assess how the conditions differ from each other, we next performed a series of planned **pairwise comparisons** by participants and items. As the participant and item comparisons yielded similar results, in the following sections for simplicity's sake, we present the results for the participants' analyses. To address the concern that some of our data may not be normally distributed, we also performed **non-parametric tests** (Wilcoxon signed-rank test and Mann-Whitney test). These gave the same results as the parametric pairwise comparisons which are reported in the following sections. In the following, the names of the three conditions are abbreviated to *neutLike* for the neutral liking condition, to *medLike* for the medium liking condition, and to *maxLike* for the maximum liking condition.

Expression of empathy Participants' perception of EMMA's expression of empathy (facial expression and speech prosody) was measured using the first two questionnaire items listed in Table 6.2. The mean values of the ratings and their standard errors for each condition are listed in Table 6.7 and are visualized in Figure 6.15 respectively for the first and the second questionnaire item. Further, the results of the pairwise comparisons of the ratings in the three conditions are listed in Table 6.8.

Measure		Condition		
		neutLike	***medLike***	***maxLike***
Expression of empathy	Facial expression	M = 0.883 (SE = 0.157)	M = -0.483 (SE = 0.103)	M = -1.554 (SE = 0.135)
	Speech prosody	M = 0.521 (SE = 0.173)	M = -0.550 (SE = 0.110)	M = -1.592 (SE = 0.123)

Table 6.7: Mean values (M) and their standard errors (SE) showing participants' perception of EMMA's expression of empathy.

Regarding the ratings of EMMA's **facial expression** in the last frames of the presented videos, the mean values show that EMMA's facial expression was rated as showing

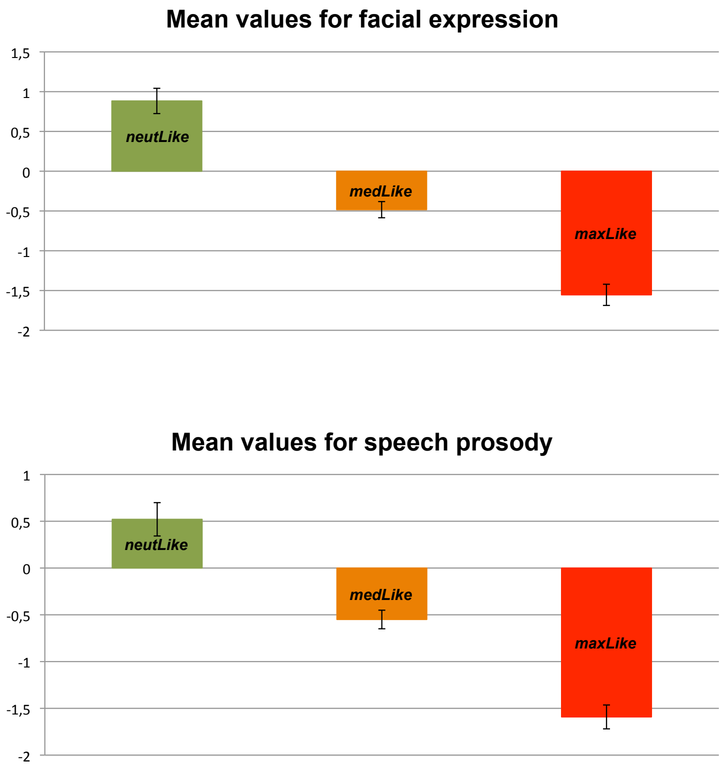


Figure 6.15: Visualization of the mean values and their standard errors showing participants' perception of EMMA's expression of empathy.

Measure		**Pairwise Comparisons**		
		neutLike* vs. *medLike	***medLike* vs. *maxLike***	***maxLike* vs. *neutLike***
Expression of empathy	Facial expression	*MD* = 1.367*** (*SE* = 0.156) (*CI* = [1.047:1.687])	*MD* = 1.071*** (*SE* = 0.131) (*CI* = [0.802:1.339])	*MD* = 2.438*** (*SE* = 0.195) (*CI* = [2.038:2.837])
	Speech prosody	*MD* = 1.071*** (*SE* = 0.173) (*CI* = [0.717:1.425])	*MD* = 1.042*** (*SE* = 0.150) (*CI* = [0.735:1.348])	*MD* = 2.113*** (*SE* = 0.241) (*CI* = [1.619:2.606])

Table 6.8: Pairwise comparisons of the ratings in the three conditions regarding participants' perception of EMMA's expression of empathy. *MD* refers to mean difference, *SE* refers to standard error of the difference, *CI* refers to the 95% confidence interval for the difference; $*** = p < .001$.

a positive mood in the neutral liking condition ($M = 0.883$), as showing a slightly negative mood in the medium liking condition ($M = -0.438$), and as showing a more negative mood in the maximum liking condition ($M = -1.554$). The results of the pairwise comparisons show that the three conditions were rated as significantly different from each other ($p < .001$).

Regarding the ratings of EMMA's **speech prosody** in the last frames of the presented videos, the mean values show that EMMA's speech prosody was rated as slightly positive in the neutral liking condition ($M = 0.521$), as slightly negative in the medium liking condition ($M = -0.550$), and as more negative in the maximum liking condition ($M = -1.592$). As for the ratings of EMMA's facial expression, the results of the pairwise comparisons show that the three conditions were rated as significantly different from each other ($p < .001$).

Degree of empathy Participants' perception of EMMA's expressed degree of empathy was measured using the third questionnaire item listed in Table 6.2. The mean ratings for this item and their standard errors for each condition are listed in Table 6.9 and are visualized in Figure 6.16. Further, the results of the pairwise comparisons of the ratings in the three conditions are reported in Table 6.10.

Measure	Condition		
	neutLike	***medLike***	***maxLike***
Degree of empathy	*M* = 0.458 (*SE* = 0.189)	*M* = 0.992 (*SE* = 0.132)	*M* = 1.608 (*SE* = 0.144)

Table 6.9: Mean values (M) and their standard errors (SE) showing participants' perception of EMMA's expressed degree of empathy.

The mean values show that EMMA was rated as slightly feeling with MAX in the neutral liking condition ($M = 0.458$) and as progressively more feeling with MAX in the medium liking condition ($M = 0.992$) and the maximum liking condition ($M = 1.608$) respectively. As for EMMA's expression of empathy, the results of the pairwise comparisons show that the three conditions were rated as significantly different from each other ($p < .001$).

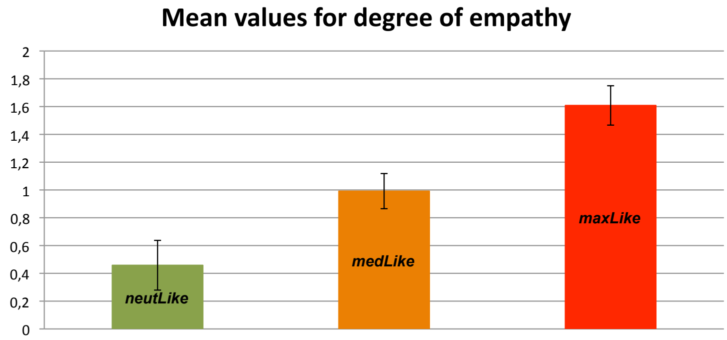


Figure 6.16: Visualization of the mean values and their standard errors showing participants' perception of EMMA's expressed degree of empathy.

Measure	Pairwise Comparisons		
	neutLike *vs.* *medLike*	*medLike* *vs.* *maxLike*	*maxLike* *vs.* *neutLike*
Degree of empathy	*MD* = 0.533*** (*SE* = 0.135) (*CI* = [0.256:0.810])	*MD* = 0.617*** (*SE* = 0.124) (*CI* = [0.363:0.871])	*MD* = 1.150*** (*SE* = 0.219) (*CI* = [0.702:1.598])

Table 6.10: Pairwise comparisons of the ratings in the three conditions regarding participants' perception of EMMA's expressed degree of empathy. *MD* refers to mean difference, *SE* refers to standard error of the difference, *CI* refers to the 95% confidence interval for the difference; $*** = p < .001$.

Values of relationship Participants' acknowledgment of different values of relationship between EMMA and MAX was measured using the fourth questionnaire item listed in Table 6.2. The values of the mean ratings and their standard errors for each condition are reported in Table 6.11 and are visualized in Figure 6.17. The results of the pairwise comparisons of the ratings in the three conditions are reported in Table 6.12.

The mean values show that EMMA's value of relationship to MAX was rated as slightly positive in the neutral liking condition ($M = 0.325$), and as progressively more positive in the medium liking condition ($M = 0.888$) and the maximum liking condition ($M = 1.442$) respectively. As for EMMA's expression of empathy and her expressed degree of empathy, the results of the pairwise comparisons show that the three conditions were rated as significantly different from each other ($p < .001$).

Measure	Condition		
	neutLike	***medLike***	***maxLike***
Values of relationship	M = 0.325 (SE = 0.177)	M = 0.888 (SE = 0.127)	M = 1.442 (SE = 0.120)

Table 6.11: Mean values (M) and their standard errors (SE) showing participants' acknowledgment of different values of relationship between EMMA and MAX.

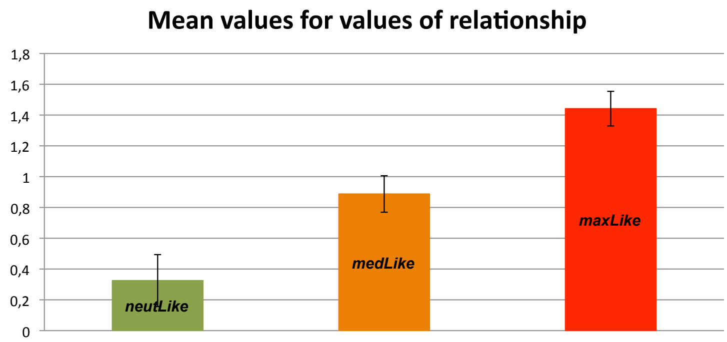


Figure 6.17: Visualization of the mean values and their standard errors showing participants' acknowledgment of different values of relationship between EMMA and MAX.

Measure	Pairwise Comparisons		
	neutLike* vs. *medLike	***medLike* vs. *maxLike***	***maxLike* vs. *neutLike***
Values of relationship	MD = 0.563*** (SE = 0.132) (CI = [0.292:0.833])	MD = 0.554*** (SE = 0.105) (CI = [0.338:0.770])	MD = 1.117*** (SE = 0.196) (CI = [0.716:1.518])

Table 6.12: Pairwise comparisons of the ratings in the three conditions regarding participants' acknowledgment of different values of relationship between EMMA and MAX. MD refers to mean difference, SE refers to standard error of the difference, CI refers to the 95% confidence interval for the difference; $*** = p < .001$.

Likability Participants' likability of EMMA was measured using the fifth questionnaire item listed in Table 6.2. The values of the mean ratings and their standard errors for each condition are reported in Table 6.13 and are visualized in Figure 6.18. The results of the pairwise comparisons of the ratings in the three conditions are reported in Table

6.14.

Measure	Condition		
	neutLike	***medLike***	***maxLike***
Likability	*M* = 0.250 (*SE* = 0.125)	*M* = 0.500 (*SE* = 0.118)	*M* = 0.746 (*SE* = 0.161)

Table 6.13: Mean values (M) and their standard errors (SE) showing participants' likability of EMMA.

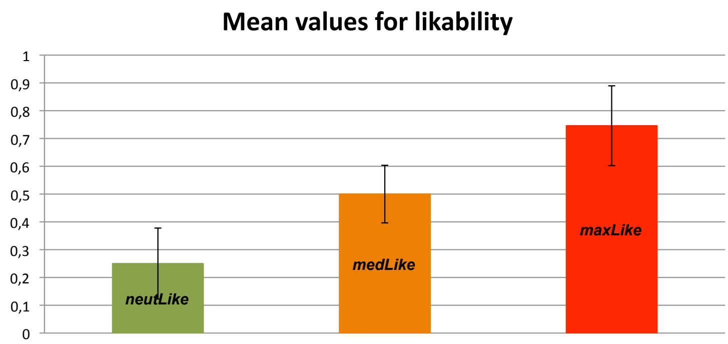


Figure 6.18: Visualization of the mean values and their standard errors showing participants' likability of EMMA.

Measure	Pairwise Comparisons		
	neutLike* vs. *medLike	***medLike* vs. *maxLike***	***maxLike* vs. *neutLike***
Likability	*MD* = 0.250^{ns} (*SE* = 0.134) (*CI* = [-0.025:0.525])	*MD* = 0.246^{ns} (*SE* = 0.136) (*CI* = [-0.033:0.525])	*MD* = 0.496** (*SE* = 0.195) (*CI* = [0.096:0.896])

Table 6.14: Pairwise comparisons of the ratings in the three conditions regarding participants' likability of EMMA MD refers to mean difference, SE refers to standard error of the difference, CI refers to the 95% confidence interval for the difference; $** = p < .05, ns = p > .07$.

The mean values show that EMMA was rated as slightly likable in the neutral liking condition ($M = 0.250$) and as progressively more likable in the medium liking condition

($M = 0.500$) and the maximum liking condition ($M = 0.746$) respectively. The results of the pairwise comparisons show a significant difference only between the first and third conditions ($p < .05$).

First post-experimental questionnaire

In the following, the results of the evaluation of participants' ratings for the questionnaire items in the first post-experimental questionnaire (see Tables 6.3, 6.4, 6.5, and 6.6) are reported. After adjusting for **reverse scoring**, the **mean values** of the ratings were calculated for each questionnaire item except for the first where participants were asked to report their understanding of the content presented in the videos.

Understanding of the content presented in the videos Participants' understanding of the content presented in the videos was evaluated by asking them to describe the general story shown in the videos, and if they could discern any perceived differences in their content.

In their answers, most of the participants reported that Lisa first praised EMMA and then insulted MAX; that MAX defended himself through his answer to Lisa and that EMMA reacted to Lisa's insult toward MAX differently. Some of the participants described EMMA's reaction as defending MAX to different degrees. Others described the story as showing different values of relationship between EMMA and MAX. Five participants reported having seen that EMMA's facial expression changed during MAX's response to Lisa, and not immediately after Lisa's utterance toward MAX. One of the participants described this behavior of EMMA as unnatural and irritating since EMMA did not immediately react to Lisa's utterance toward MAX. Another participant described it as natural and likable. A further participant pointed out that EMMA's reaction coincided with MAX's facial expression.

The reported differences perceived in the content of the videos were mainly regarding EMMA's facial expression and speech prosody. Seven of the participants reported having perceived differences only in the facial expressions. Two participants reported having perceived differences only in the speech prosody. One participant reported differences in the utterances while three reported no perceived differences at all. With regard to the perceived differences in EMMA's facial expressions, some participants described the differences as showing a neutral, positive, or negative mood while others just described the position of the lip corners (upward, downward) and the eye brows (constricted or not). With regard to EMMA's speech prosody, the differences were described in terms of speed, volume, and pitch.

EMMA's general behavior in the videos EMMA's general behavior in the videos was measured using the questionnaire items listed in Table 6.3. The mean values of the ratings and their standard errors are listed in Table 6.15.

Measure		Questionnaire item	Mean	Standard Error
EMMA's general behavior in the videos	Facial behavior	"EMMA's face was expressive: "	-0.500	0.238
		"EMMA's facial expressions were exaggerated: " (*reverse scored*)	1.500	0.213
	Speech behavior	"EMMA's pronunciation was clear: "	-0.666	0.260
		"EMMA's speech was difficult to understand: " (*reverse scored*)	-0.200	0.289
	Attention behavior	"EMMA had eye contact with Lisa: "	1.666	0.210
		"EMMA never had eye contact with MAX: " (*reverse scored*)	1.933	0.266
	Empathic response trigger	"In the last frame of all presented videos, EMMA's response toward Lisa is triggered by MAX's facial expression: "	-0.366	0.242
		"In the last frame of all presented videos, EMMA's response toward Lisa is triggered by Lisa's utterance toward MAX: "	1.566	0.320
	Behavior in general	"The behavior of EMMA over all presented videos was exaggerated: " (*reverse scored*)	1.400	0.256

Table 6.15: Mean ratings of EMMA's general behavior in the videos. Note that for the questionnaire items that are *reverse scored*, the mean values correspond to these questionnaire items worded in the reverse.

The mean values of the ratings of the first and second questionnaire items show that EMMA' facial expressions were rated as rather inexpressive and as not exaggerated. The mean values concerning the ratings of the third and fourth questionnaire items show that EMMA's speech was rated as rather difficult to understand and as unclear. The mean values of the ratings of the next two questionnaire items show that the participants agreed that EMMA's attention was toward Lisa and also toward MAX in the presented videos. Regarding the last two questionnaire items, the mean values show that participants rather disagreed that MAX's facial expression triggered EMMA's response toward Lisa and agreed that Lisa's utterance triggered EMMA's response.

Naturalness of EMMA's behavior in the videos The naturalness of EMMA's behavior in the videos was measured using the questionnaire items listed in Table 6.4. The mean values of the ratings and their standard errors are listed in Table 6.16.

Measure		Questionnaire item	Mean	Standard Error
Naturalness of EMMA's behavior in the videos	Facial behavior	"EMMA's facial expressions appeared: "	-1.266	0.248
	Speech behavior	"EMMA's pronunciation appeared: "	-2.100	0.173
	Attention behavior	"EMMA's attention behavior appeared: "	-0.466	0.316

Table 6.16: Mean ratings of the naturalness of EMMA's behavior in the videos.

The mean values corresponding to the ratings of the first and second questionnaire items show that EMMA's facial expressions and pronunciation were rated as artificial. The mean values corresponding to the ratings of the third questionnaire item show that EMMA's attention behavior was rated as rather artificial. The mean value of the ratings of all questionnaire items shows that EMMA's overall expressive behavior was rated as artificial.

EMMA's social behavior over all presented videos EMMA's social behavior over all presented videos was measured using the questionnaire items listed in Table 6.5. The mean values of the ratings and their standard errors are listed in Table 6.17. The mean values of the ratings of all questionnaire items show that EMMA was rated as a cooperative, a caring, a protective, and a social agent.

Measure		Questionnaire item	Mean	Standard Error
EMMA's social behavior over all presented videos	Cooperative	"EMMA is a cooperative agent: "	1.033	0.188
	Caring	"EMMA is an uncaring agent: " (*reverse scored*)	1.200	0.285
	Protective	"EMMA is a protective agent: "	1.200	0.241
	Social	"EMMA is an asocial agent: " (*reverse scored*)	1.333	0.300

Table 6.17: Mean ratings of EMMA's social behavior over all presented videos. Note that for the questionnaire items that are *reverse scored*, the mean values correspond to these questionnaire items worded in the reverse.

General likability of EMMA The general likability of EMMA was measured using the questionnaire items listed in Table 6.6. The mean values of the ratings and their standard errors are listed in Table 6.18.

Measure		Questionnaire item	Mean	Standard Error
General likability of EMMA	Trust	"You would trust EMMA if she were your friend: "	0.100	0.312
	Friendship	"You could not imagine EMMA to be your friend: " (*reverse scored*)	-0.133	0.361
	Interaction	"You would like to interact with EMMA: "	-0.200	0.360
	Interest	"You are not interested in EMMA: " (*reverse scored*)	0.166	0.314

Table 6.18: Mean ratings of the general likability of EMMA. Note that for the questionnaire items that are *reverse scored*, the mean values correspond to these questionnaire items worded in the reverse.

The mean values of the ratings of all questionnaire items show that participants would neither trust nor distrust EMMA as a potential friend, neither imagine nor not imagine EMMA as a potential friend, neither like nor not like to individually interact with EMMA, and that participants were neither interested nor not interested in EMMA.

Second post-experimental questionnaire

A second post-experimental questionnaire measured participants' **I**nterpersonal **R**eactivity **I**ndex (IRI) [33] as a means to assess empathy. This measure was collected to test for any correlation that might exist between the participants' IRI and their ratings of the five questionnaire items after each shown video. A German version [92] of the questionnaire proposed in [33] was used. The questionnaire proposed in [33] comprises four subscales with corresponding questionnaire items to measure different aspects of participants' empathy. With the German version of the IRI questionnaire, two aspects of empathy can be measured using the perspective taking scale and the empathic concern scale. Accordingly, the mean values for each empathy scale were calculated for each participant. Further, for each of the five questionnaire items listed in Table 6.2, the mean values of the differences between the ratings of the neutral liking condition and the maximum liking condition were calculated for each participant. A **Pearson Correlation** between the mean values of the perspective taking and empathic concern scales and the mean values indicating participants' differentiation between the neutral and maximum liking conditions was conducted. No significant correlations were found between participants' IRI and their differentiation between the neutral and maximum liking conditions. Accordingly, our hypothesis that *the higher the value of IRI, the more participants can differentiate between the three conditions of EMMA's empathy with MAX in their ratings for the five questionnaire items* (cf. Section 6.2.2) was not confirmed. Finally, we report the results of the analysis of participants' answers during the post-experimental interview.

Interview: Criteria for rating EMMA's likability after each presented video

In order to determine the criteria by which participants rated the likability of EMMA in the three conditions, we conducted a post-experimental interview. The analysis of the participants' answers shows that 20 participants rated the likability of EMMA on the basis of her different reactions toward MAX. That is, the more EMMA was feeling with MAX, the more she was rated as likable. Two participants reported rating EMMA's likability on the basis of her reaction toward Lisa. That is, the more EMMA was friendly to Lisa, the more she was rated as likable. One of these participants also emphasized that *one's rude behavior cannot be justified through others' rude behavior and thus that EMMA should not be rude because Lisa is behaving that way.* Three participants rated

EMMA's likability on the basis of her feeling with MAX and also of her positive reaction toward Lisa at the beginning of the video. Two participants reported that EMMA was rated as neither likable nor unlikable because they were not involved in the observed interaction. Two other participants also rated EMMA as neither likable nor unlikable because her observed behavior in the videos was deemed unconvincing and was unnatural. One participant was not able to describe any criteria by means of which he rated EMMA's likability after each presented video. In the following, the results of the present study are discussed and interpreted.

6.2.4 Discussion and conclusion

The results of evaluating participants' perception of EMMA's expression of empathy show that EMMA's **facial expression** of *happy* in the neutral liking condition was rated as showing a positive mood, that EMMA's facial expression of *concentrated*, as the neutral emotional state, in the medium liking condition was rated as showing a slightly negative mood, and that EMMA's facial expression of *annoyed* in the maximum liking condition was rated as showing a negative mood. The results also show that EMMA's **speech prosody** for *happy* in the neutral liking condition was rated as slightly positive, that EMMA's speech prosody for *concentrated* in the medium liking condition was rated as slightly negative, and that EMMA's speech prosody for *annoyed* in the maximum liking condition was rated as negative. The perceived differences in EMMA's expression of empathy are also significant in all three conditions. In sum, the results show the appropriate recognition of EMMA's expression of empathy as positive in the first condition, and as progressively more negative in the second and third conditions respectively. Hence, the results confirm our **first hypothesis**, **H1**, that EMMA's expression of empathy is perceivable by the participants, and also suggest the appropriate modeling of EMMA's facial expressions and speech prosody.

The results of evaluating participants' perception of EMMA's expressed **degree of empathy** show that EMMA was rated as slightly feeling with MAX in the neutral liking condition and as progressively more feeling with MAX in the medium liking and the maximum liking conditions respectively. The perceived differences in EMMA's expressed degree of empathy are also significant in all three conditions. Thus, EMMA's degree of empathy with MAX was rated as significantly higher in the maximum liking condition than in the other two conditions, and as significantly higher in the medium liking condition than in the neutral liking condition. Hence, the results confirm our **second hypothesis**, **H2**, that EMMA's expressed degree of empathy is perceivable by the participants. Descriptively, Figures 6.15 and 6.16 show that the more similar participants' ratings of EMMA's expression of empathy to MAX's expressed emotional state, the higher the rated value of EMMA's degree of empathy with MAX. This is in line with our definition of the degree of empathy as the degree of similarity between one's empathic emotion and the other's perceived emotion (cf. Section 5.3). That is, the more similar one's empathic emotion to the other's perceived emotion, the higher the degree of empathy. Accordingly, these findings further substantiate the theoretical assumption underlying our model that empathy occurs to different degrees (cf. Section 2.1). Furthermore, the results show that EMMA's facial expression and speech prosody are reliable indicators of her different degrees of empathy, thus providing further support for the appropriate modeling of EMMA's facial expressions and speech prosody.

The results of evaluating participants' acknowledgment of different **values of relationship** between EMMA and MAX show that EMMA's value of relationship to MAX was rated as slightly positive in the neutral liking condition, and as progressively more positive in the medium liking and maximum liking conditions respectively. The acknowledged differences in EMMA's value of relationship to MAX are also significant in all three conditions. Thus, EMMA's acknowledged value of relationship was rated as significantly higher in the maximum liking condition than in the other two conditions, and as significantly higher in the medium liking condition than in the neutral liking condition. Hence, the results confirm our **third hypothesis**, **H3**, that human participants acknowledge different values of relationship between EMMA and MAX according to EMMA's expressed degree of empathy. Descriptively, Figures 6.16 and 6.17 show that the higher the rated value of EMMA's expressed degree of empathy, the higher the rated value of her relationship to MAX. This is in line with our definition of the impact of relationship modulation factors such as *liking* or *familiarity* in our model. That is, the higher the values of such modulation factors, the higher the similarity between the empathic emotion and the other's perceived emotion, the higher the degree of empathy (cf. Section 5.3). These findings further substantiate the theoretical assumption underlying our proposed model that empathy is modulated by several modulation factors such as the relationship between the empathizer and the observed other (cf. Section 2.1). Moreover, the results also show that EMMA's facial expression and speech prosody are reliable indicators of her different values of relationship to MAX thus providing further support for the appropriate modeling of EMMA's facial expressions and speech prosody.

The results of evaluating participants' **likability** of EMMA show that EMMA was rated as slightly likable in the neutral liking condition, and as progressively more likable in the medium liking and maximum liking conditions respectively. However, a significant difference was found only between the first and third conditions. These results only partially support our **fourth hypothesis**, **H4**, that EMMA is perceived as more likable the higher the value of her expressed degree of empathy. In this regard, the analysis of the reported criteria for rating EMMA's likability in the three conditions shows that while most of the participants reported having rated EMMA's likability on the basis of her degree of empathy with MAX, others rated EMMA's likability on the basis of, e.g., her positive emotions in the presented video. That is, the more EMMA was friendly toward Lisa the more she was rated as likable. Accordingly, these participants may have interpreted EMMA's expression of a positive emotion, together with a verbal expression of empathy as a more convenient way to communicate the inappropriateness of Lisa's behavior toward MAX. Thus, a more conservative empathic behavior in such conflict situations seems to be more appreciated by these participants. This issue could be addressed by taking **social norms and rules** into account in the computational modeling of empathy, since they may allow for masking the 'true feeling' of strong empathy in such situations.

Evaluation of participants' understanding of the content presented in the videos shows that the participants clearly understood the scenario shown in the videos. Most of the participants reported differences with regard to EMMA's facial expression and speech prosody, while in turn, most of these participants reported differences only in terms of EMMA's facial expression. Some of the participants reported the perceived differences in terms of expressive features such as lip corners (upward, downward), eye brows (constricted or not) and speech volume, speed, and pitch which characterize EMMA's facial expressions and speech prosody. These findings provide further support for the appropriate modeling of EMMA's facial expressions and speech prosody and further confirm our **first hypothesis H1**. However, the results of evaluating EMMA's general behavior in the videos show that EMMA's facial expressions were rated as rather inexpressive and artificial and that her speech was rated as rather unclear and also artificial. In this regard, further evaluations of EMMA's facial expressions and speech prosody are needed to improve their expressiveness and naturalness and to thus further improve their appropriate modeling.

Further, most of the participants agreed that Lisa's response to MAX triggered EMMA's response to Lisa, and disagreed that MAX's facial expression triggered EMMA's response to Lisa. However, some of the participants reported having seen that EMMA's facial expressions changed during MAX's utterance and not immediately after Lisa's utterance, and only one participant recognized that EMMA's reaction coincided with MAX's facial expression. Thus, it was difficult to recognize that EMMA's reaction toward Lisa was in response to her perception of MAX's facial expressions during facial mimicry (cf. Sections 6.1.1 and 5.2.1). While we expected to obtain this result, we were interested to know if some of the participants recognized this behavior in EMMA, and in fact it was interesting to see that one of the participants did.

In general, EMMA was rated as neither likable nor unlikable. This could be explained by the fact that the participants were not directly involved in the interaction with EMMA. This aspect was reported in the criteria of rating EMMA's likability as neutral in the three conditions. However, overall EMMA was rated as a cooperative, a caring, a protective, and a social agent.

According to the results of the study reported in Section 4.3, EMMA's **neutral facial expression** was rated by participants as displeased, as aroused, and as submissive (see Table 4.1, p. 83). As compared to this, in the present study, EMMA's facial expression in the medium liking condition, which is very similar to EMMA's neutral facial expression, was rated as showing a slightly negative mood, and thus as more neutral than EMMA's facial expression in the previous study. A crucial difference between the studies is that in the above mentioned study, EMMA's facial expression was presented in a static context, while in the present study EMMA's facial expression was presented in a **dynamic context** alongside other facial expressions of emotion. This further supports Russell's [100] emphasis that the meaning of facial expressions is not absolute but **relative**, and that facial expressions do not come with single values. However, EMMA's appearance in the above mentioned study was a little different from her current appearance, which may have influenced participants' ratings. Thus, more evaluations should be done in order to further investigate this issue.

Altogether, the results of the present study show that EMMA's expression of empathy, as generated by our computational model of empathy was appropriately recognized by the participants and is a reliable indicator of EMMA's degree of empathy, and her value of relationship to MAX. Accordingly, the results further corroborate the theories and findings underlying our model. A **highlight** is that the present study is one of the first studies in this field to consider three conditions showing three different degrees of empathy thus allowing for a more refined and fine grained evaluation of the output of the model and of its underlying parameters.

6.3 Summary and conclusion

This chapter introduced the application and evaluation of our computational model of empathy. In Section 6.1, we presented the application of our model in a conversational agent scenario and in a spatial interaction task scenario. In the **conversational agent scenario**, the virtual humans MAX [62] and EMMA (see Chapter 4) and a human interaction partner were involved. In this scenario, EMMA empathizes with MAX to different degrees depending on her *mood* and defined values of relationship (*liking* and *familiarity*) to MAX. In this scenario, we applied the empathy mechanism **facial mimicry** (see Section 5.2.1) and showed how by means of a very simple form of the empathy mechanism situational role-taking (see Section 5.2.2), EMMA infers an explicit value of dominance for MAX. This can be considered a first step toward combining the hypotheses about the other's emotional state as generated by both mechanisms to get a more adequate hypothesis. EMMA's different degrees of empathy are expressed by her facial expression, speech prosody, and breathing and eye blinking behaviors. EMMA also expresses a verbal utterance advising the human partner to continue being kind to MAX or to stop being unfriendly to MAX. In this scenario, we illustrated the different processing steps underlying our empathy model. In light of the requirement **Adequacy** (cf. Section 1.2), different values of the modulation factors, *liking*, *familiarity*, and EMMA's *mood* were tried and combined to check the adequacy of the empathic behavior provided by our model. Accordingly, EMMA's empathic behavior reflected the empathic behavior generated by our model, and was as expected in line with the underlying theories.

Further in Section 6.1, we presented the application of our model in a **spatial interaction task scenario**. As compared to the conversational agent scenario where agent-agent interaction was considered, this scenario involved the virtual human MAX and a human interaction partner. In this scenario, MAX empathizes with his human interaction partner to different degrees depending on the values of the modulation factors *liking*, *deservingness*, and MAX's *mood*. In this scenario, we applied the empathy mechanism **situational role-taking**. A highlight in this scenario is that MAX's degree of empathy triggers and modulates his spatial helping action. That is, the quality and distance of the helping movement are modulated by MAX's calculated value of degree of empathy. Accordingly, the more MAX empathizes with his partner, the more he helps the partner by rapidly putting needed objects in front of him. This is in line with Hoffman's and Davis's emphasis that empathy is a motivational basis for prosocial cooperative behavior such as helping and caring [55] [34]. On the other hand, MAX's value of *liking* is based on the partner's investment in helping actions. That is, the more the partner helps MAX, the more MAX likes his partner, empathizes, and in turn, helps him. This is in line with Davis's emphasis that the ability for empathy impacts one's social behavior and thus one's relationship with others [34]. In light of the requirement **Adequacy**, MAX's empathic behavior according to different values of the modulation factors reflects the empathic behavior generated by our model, and was as expected in line with the underlying theories.

In light of the requirement **Universality**, our approach to empathy modulation was easily carried over to both considered scenarios. In this regard, extensions were made only for defining *liking* and *familiarity* as modulation factors in the first application scenario, and *liking* and *deservingness* in the second.

In order to further investigate the adequacy of the empathic behavior generated by our model, an **empirical evaluation** was conducted with the purpose of evaluating human participants' perception and interpretation of the generated behavior. This was subject to Section 6.2. The empirical evaluation was conducted in the conversational agent scenario to evaluate EMMA's empathic behavior toward MAX. In this regard, we defined three conditions where different degrees of empathy were considered, a neutral liking condition, a medium liking condition, and a maximum liking condition. The results show that EMMA's different degrees of empathy were perceived by human participants, and that the human participants also acknowledged different values of relationship between EMMA and MAX according to EMMA's degree of empathy. Thus, EMMA's expression of empathy, as a reflection of the empathic behavior generated by our model was a reliable indicator of her different degrees of empathy and of her different values of relationship to MAX. Thus, in light of the requirement **Adequacy**, our findings further substantiate the adequacy of the empathic behavior generated by our model.

In conclusion, the application and evaluation of our model shows that it provides adequate empathic behavior, and that the virtual human EMMA is perceived as capable of exhibiting different degrees of empathy and values of relationship with MAX. Furthermore, the findings of our empirical evaluation provided further support for the theories and assumptions underlying the model. Thus, our model enhances a virtual human's **social behavior** and provides an **experimental tool** for underlying theories.

7 Resume

As motivated in the introduction (see Chapter 1), the goal of the present thesis is to provide a computational model of empathy to enhance an artificial agent's social behavior, and to provide an experimental tool for the psychological theories shaping the model. In light of the requirements to achieve these objectives (cf. Chapter 1), the results and contribution of this thesis are summarized and discussed in Section 7.1 of this chapter. Subsequently, in Section 7.2, future desirable extensions to further improve the presented work are described.

7.1 Results and contribution

In our attempt to achieve our thesis objectives, a careful consideration of the theoretical background on empathy and of previous works on empathic artificial agents, as presented respectively in Chapters 2 and 3, was carried out. Accordingly, we defined the requirements to achieve our objectives (see Sections 1.2 and 3.3.3). In this regard, we defined three requirements for building a computational model of empathy that addresses three central processes to empathy, **Empathy Mechanism**, **Empathy Modulation**, and **Expression of Empathy**. As such, the present thesis provided a computational model of empathy (see Chapter 5) based on three processing steps: first, the empathy mechanism as the process by which an empathic emotion is produced; second, the empathy modulation as the process by which the empathic emotion produced in the first step is modulated; and third, the expression of empathy as the process by which the empathic emotion modulated in the second step is expressed. Our empathy model is realized for the virtual humans MAX [62] and EMMA (see Chapter 4). Furthermore, we defined two requirements that a computational model of empathy should fulfill, **Adequacy** and **Universality**. In light of the above mentioned requirements, the results and contribution of the present thesis are summarized and discussed in the following.

Empathy mechanism

Following the theoretical background on empathy presented in Section 2.1, we realized two empathy mechanisms for our model, facial mimicry (see Section 5.2.1) and situational role-taking (see Section 5.2.2).

Facial mimicry as defined in our model has found little attention in previous works on empathic artificial agents (cf. Section 3.3), except in the work by Breazeal et al. [21] where only the imitation step was considered. Accordingly, by means of our realization of facial mimicry, the new virtual human EMMA internally imitates a perceived facial expression by first mapping perceived facial features to **A**ction **U**nits (AUs) (internal imitation), and by subsequently determining its related emotional state (emotional feedback). Our realization of facial mimicry is based on EMMA's facial expression repertoire as a **shared representational system** where AUs are linked to **P**leasure-**A**rousal-**D**ominance (PAD) values. That is, by combining the meanings of AUs within PAD space, a PAD value is determined from a perceived emotional facial expression. A first evaluation of our realization of facial mimicry shows **promising results** in that it determines quite accurate PAD values from the virtual human MAX's facial expressions of emotion, and in that the determined PA values mimic the course of intensities of his facial muscles. Thus, our approach allows for the recognition of a **wide range of emotional states** and is not limited to the recognition of a predefined number of emotion categories. This is in line with the emphasis that dimensional emotion models allow for the recognition of a wide range of emotional states, and that they are more convenient for characterizing the continuity and subtlety of emotion expression. However, in previous works on automatic emotion recognition, little attention was devoted to emotion recognition using a dimensional rather than a categorical approach, in particular, regarding emotion recognition from facial expressions (cf. Section 3.2.2).

As compared to the computational model of empathy provided by Rodrigues et al. [98], both facial mimicry and situational role-taking are considered separately in our model. However, a first step toward combining these mechanisms is that we used the value of dominance inferred by **situational role-taking**, as **context related information** to improve emotion recognition from facial expressions in our realization of facial mimicry (see Section 6.1.1).

The perceived emotional state produced by facial mimicry or situational role-taking is simulated within the virtual human's emotion simulation module [8] and is thus represented in his PAD space (see Section 5.2.3). Accordingly, by means of our defined condition of elicitation of an empathic emotion, empathy is not only elicited in response to the saliency of a perceived emotional state but also in response to the **dynamic change** in the perceived emotional state. Thus, empathy can be elicited, e.g., in response to a neutral emotional state as resulting from a rapid and salient change in emotions. This further substantiates the convenience of dimensional emotion models in characterizing the continuity and dynamics of emotional states. Furthermore, the simulation of an empathic emotion within the virtual human's emotion simulation module allowed for the simulation of the **time course** of an empathic emotion, and of its interaction with the virtual human's emotional state. As in the theoretical models of empathy, this issue was not explicitly and clearly addressed in most of the previous works.

Empathy modulation

By means of our realization of empathy modulation (see Section 5.3), a virtual human is allowed to empathize to different degrees depending on several **modulation factors**, e.g., his mood and his relationship to the other. Accordingly, our approach allows for the modulation of an empathic emotion through the factor *empathizer's mood* and through further arbitrary predefined factors that can have values ranging in $[0, 1]$, e.g., *liking*, *familiarity*, and *deservingness*. Furthermore, it also allows for the assignment of different weight values for these factors to define which have a more significant impact on modulating the empathic emotion than others. Further, regions of immediate neighborhood for each emotion category located in PAD space were defined where a modulated empathic emotion from different type (emotion category) but compatible with the non-modulated one is facilitated. Accordingly, a **degree of empathy** is calculated as the degree of similarity between a modulated empathic emotion and a non-modulated one within the defined regions of immediate neighborhood. In this regard, we rely on the thesis of the dimensional emotion theories that emotions are related to one another in a systematic manner and that their relationships can be represented in a dimensional model (cf. Section 2.2). Hence, we exploited the assumed relationships between emotions in PAD space.

In previous works on empathic artificial agents, little attention was devoted to the consideration of different degrees of empathy which is a crucial aspect in further enhancing an artificial agent's **social behavior** (cf. Section 3.3). Furthermore, in previous works, where this issue is addressed, only the intensity of an empathic emotion is modulated. In contrast, our realization of empathy modulation also allows for the modulation of the type (emotion category) of the empathic emotion. This is in agreement with Hoffman's underscoring that an empathic response need not to be a close match to the affect experienced by the other, but can be any emotional reaction compatible with the other's situation [55] (cf. Section 2.1.2).

Expression of Empathy

In our computational model of empathy, a virtual human is allowed to have a **multimodal expression of empathy** (see Section 5.4). That is, the virtual human is allowed to express empathy based on his repertoire of multiple modalities such as facial expressions, speech prosody, eye blinking and breathing behaviors, verbal utterances, and spatial helping actions. A **highlight** is the virtual human EMMA's elaborate model of facial expressions (cf. Section 4.3) which also underlies our realization of facial mimicry, and hence EMMA's ability for facial mimicry. In this regard, we investigated the **meaning** of single AUs within PAD space and how it contributes to the meaning of a whole facial expression of emotion. As a result, we provided three-dimensional intensity functions for each AU within PAD space, which we combined to reconstruct a facial expression repertoire. The provided repertoire comprises **quite accurate facial expressions** of EMMA that reflect the trajectories of the time course of her emotions within PAD space. Thus, EMMA is allowed to express a wide range of emotional states rather than a limited set of emotion categories. This is in line with the emphasis that dimensional emotion models allow for the expression of a wide range of emotional states, and that they are more inclined for characterizing the continuity and subtlety of emotion expression (cf. Section 3.2.1).

Adequacy

The adequacy of the empathic behavior provided by our empathy model was verified in the context of its application and evaluation in a context scenario (see Chapter 6). Thus, we applied our model in **two different interaction scenarios**, a conversational agent scenario and a spatial interaction task scenario. Furthermore, an empirical evaluation of our model in the context of the conversational agent scenario was performed.

The application of our model in the **conversational agent scenario** (see Section

6.1.1) shows that the virtual human EMMA is able of empathizing with the virtual human MAX to different degrees, depending on her mood and her defined values of relationship to MAX. In this regard, different values and weights of the modulation factors *liking* and *familiarity* as well as different values of EMMA's *mood* were considered to check the adequacy of EMMA's empathic behavior in consequence to the changing values of these factors. EMMA's empathic behavior in this scenario reflected the empathic behavior generated by our model, and was as expected in line with the theories underlying the model.

The application of our model in the **spatial interaction task scenario** (see Section 6.1.2) shows that the virtual human MAX is able of empathizing with a human interaction partner to different degrees depending on his *mood* and on the values of *liking* and *deservingness* as further modulation factors. As compared to the conversational agent scenario where the values of the modulation factors *liking* and *familiarity* are predefined, in the spatial interaction task scenario, the values of *liking* and *deservingness* are inferred according to the interaction context. A **highlight** is that MAX's spatial helping action is triggered and modulated by his degree of empathy with his interaction partner. In this regard, the quality of MAX's spatial helping movement, as well as its distance, are modulated by the degree of empathy. That is, the more MAX empathizes with his partner, the closer and faster he places needed objects in front of the partner and thus the more he invested in helping him. This is in line with Hoffman's and Davis' emphasis that empathy is a motivational basis of prosocial and cooperative behaviors such as helping and caring [55] [34] (cf. Section 2.1.2). Furthermore, MAX's value of *liking* is inferred based on his partner's investment in helping actions and thus on his partner's hypothesized degree of empathy. This is in line with Davis' emphasis [34] that one's ability for empathy impacts one's **social behavior** which is perceived by others, and which thus influences one's social relationships with others (cf. Section 2.1.2). MAX's empathic behavior during interaction to cooperatively build a tower of virtual toy blocks reflected the empathic behavior generated by our model, and concurred with the theories underlying the model. This further substantiates the adequacy of the empathic behavior generated by our model.

In order to investigate how the empathic behavior generated by our model is perceived and interpreted by human participants and thus to further verify its adequacy, we conducted an **empirical evaluation** of the model (see Section 6.2). The evaluation was performed in the context of the conversational agent scenario where **three different**

conditions regarding EMMA's degree of empathy were considered. The results show that EMMA's expression of empathy (facial expression and speech prosody) was appropriately recognized in all three conditions and that it was a reliable indicator of EMMA's degree of empathy and her value of relationship to MAX. These findings further support our claim that the empathic behavior generated by our model concurs with underlying theories, and further substantiates them. A **highlight** is that this study is one of the first studies in this field to consider three conditions showing different degrees of empathy, thus allowing for a more fine-grained evaluation of the output of the model as well as of its underlying parameters. Taken together, the application and evaluation of our model in a context scenario suggest that our model provides **adequate empathic behavior** which is in line with underlying theories and hypotheses.

Universality

Regarding the **empathy mechanism**, our realization of facial mimicry can be easily carried over in different interaction scenarios of the virtual human EMMA. In contrast, our realization of situational role-taking requires a context scenario and is thus domain dependent. Regarding our realization of **empathy modulation**, once the modulation factors are defined, the modulation of the empathic emotion and the calculation of different degrees of empathy can be easily carried over to different context scenarios, e.g., the conversational agent scenario and the spatial interaction task scenario considered in the present work. In this regard, adaptations were only carried out for the definition of the modulation factors such as *liking* and *familiarity* in the first scenario, and *liking* and *deservingness* in the second. Regarding the **expression of empathy**, we defined context independent expressions such as facial expressions, speech prosody, and eye blinking and breathing behaviors. All in all, our model is not entirely universal but is so in several of its components making it easier to apply and adapt to different scenarios as demonstrated in the present thesis.

Conclusion

In light of the **objectives** of this thesis, a computational model of empathy is provided that allows a virtual human to empathize with his interaction partner to different degrees, and to express his degree of empathy by means of different modalities. The adequacy of the empathic behavior provided by our model was verified and confirmed in the context of its application in two interaction scenarios, and of its empirical evaluation. In this regard, the findings show that the virtual human was perceived by human participants as capable of exhibiting different degrees of empathy, and of having different values of relationship with his interaction partner. Thus, our model satisfies our objective that it enhances a virtual human's **social behavior**.

Furthermore, different parameters are defined in our model that allow for the concretization and investigation of several theoretical aspects of empathy that are not explicitly and clearly defined in the theoretical models of empathy. Hence, our model satisfies our objective to provide an **experimental tool** for underlying theories and assumptions. This issue was further supported by the findings of the empirical evaluation of the model where a fine-grained evaluation of the values of our model parameters was successfully performed, thus substantiating the model's underlying theories and hypotheses.

7.2 Future work

While much effort was invested in our attempt to achieve the thesis objectives, more effort is needed to further improve our work on empathy. Accordingly, our approach to determine PAD values from AUs displaying emotions should be further investigated on the basis of **human data** provided by an automatic AU recognition system, or by a **database** of emotional facial expressions annotated with AUs and their intensity values. Thus, the aim is to allow for the application of facial mimicry, as defined in our model, in the context of human-agent interaction. As presented in Section 3.2.2, there are several works on automatic AU recognition. However, it was not possible to get such a system in the context of the present thesis since this field is still in its beginning stages. Further, while expressive cues such as facial expressions may present a valid source of information about others' emotional states, contextual cues remain another possible source of information. Thus, the integration of information about the other's emotional state as provided by expressive and by contextual cues may allow for a more adequate recognition of the other's emotional state. This issue was addressed in previous works on empathic artificial agents (e.g., [98] and [69], see Section 3.3) and should be taken into account in our model. In this regard, a future work is to **combine the hypotheses** about the other's emotional state provided by facial mimicry and situational role-taking to show how they can complement each other to obtain an explicit and a more adequate hypothesis. A first step was carried out by using the value of dominance inferred by situational role-taking as context related information to improve emotion recognition from facial expressions in our realization of facial mimicry. Another interesting aspect is the distinction between primary and secondary emotions as defined in the virtual human's emotion simulation module [8]. In this regard, our model generates only primary emotions as empathic emotions. Accordingly, by means of facial mimicry, a facial expression showing the secondary emotion *relief* is likely to be falsely interpreted as showing the primary emotion *happy*. Thus, the aim is to enhance our realization of situational role-taking to the generation of **secondary emotions** as empathic emotions such as *relief*. Furthermore, also **individual role-taking** could be addressed to allow for the consideration of the other's beliefs, desires, and goals, and to thus allow for non-egocentric empathy. In the context of the project part A1 'Modelling Partners' (see Chapter 1), a joint-attention framework was developed by Pfeiffer-Lessmann et al. [93] that allows for cognitive partner modeling, e.g., the partner's beliefs, desires, and goals. Thus, this information could be used for the realization of individual role-taking.

Regarding our realization of empathy modulation, the proposed approach can be extended to retrieving the values of modulation factors such as *liking* and *familiarity* from a person memory as proposed by Mattar & Wachsmuth [74]. Further, personality factors are defined in the virtual human's emotion simulation module [8], for example, factors impacting how emotional (e.g., temperamental vs. lethargic) the virtual human is. Accordingly, the impact of these **personality factors** on modulating the virtual human's empathic emotion could be investigated in our future work. We also aim to consider further modulation factors (cf. Table 5.2) as, for example, *similarity*. An interesting aspect might be to extend our approach to the consideration of **negative empathic emotions** such as *gloating* or *resentment* by extending the values of modulation factors such as *liking* or *deservingness* to negative ones (cf. [88], see Section 2.1). This issue would, for example, allow for the modeling of a virtual human's competitive behavior such as placing needed objects out of reach from his interaction partner in the spatial interaction task scenario presented in Section 6.1.2.

In his theoretical model of empathy, Davis [34] distinguishes between affective and cognitive intrapersonal outcomes of empathic processes (cf. Section 2.1.2). While in our empathy model only affective outcomes are considered, **cognitive outcomes** could be subject to future work. For example, consider the situation where one is feeling sadness in response to the other's emotional state of sadness and one knows the other prefers that one shows a more positive emotional state to alter the other's negative feeling. In such situations, the affective outcome of sadness could be masked by showing a positive emotional state according to the cognitive outcome of estimating the other's preferences, thus resulting in a more convenient behavior toward the other. Further, our evaluation of EMMA's empathic behavior in the conversational agent scenario shows that some participants preferred the situations where EMMA was showing a more conservative empathic behavior. Also in such situations, the affective outcome could be masked by showing a behavior dictated by the cognitive outcome, resulting from **social norms and rules** for a more appropriate social behavior. Such a behavior could be displayed by so called 'complex facial expressions' as expressions that are modified with respect to some socio-cultural rules following the approach proposed by Niewiadomski et al. [84].

In our empirical evaluation of the empathic behavior provided by our model, we considered three different conditions where the virtual human respectively expresses a positive emotional state in the neutral liking condition, a neutral emotional state in the medium liking condition, and a negative emotional state in the maximum liking condition. According to our realization of empathy modulation, the virtual human would express the emotional state of *surprised* in the medium liking condition when the empathic emotional state has *anger* as its related emotion category (cf. Section 6.2). Accordingly, a future work could involve evaluating human participants' perception of the virtual human's expressed emotional state *surprised* in the medium liking condition and its impact on the perception of his degree of empathy. Furthermore, the evaluation of human participant's perception of the modulation factor *empathizer's mood* and how it impacts the degree of empathy could also be considered in future work. Empirical evaluation of the model in **other context scenarios**, such as the spatial interaction task scenario presented in Section 6.1.2, could be performed to further investigate the adequacy of the proposed model. In the spatial interaction task scenario, the more the partner helps MAX the more MAX likes his partner, empathizes, and helps him. Thus, by his modulated helping actions, MAX signalizes his human partner's investment in cooperating with him. Accordingly, in future work, we aim at empirically evaluating how MAX's modulated helping action impacts the partner's engagement in achieving a successful cooperation.

While significant advances have been made in allowing artificial agents to empathize, much work remains before they can be considered holders of the human 'helping genes'! [55].

Bibliography

[1] I. Albrecht, J. Haber, M. Schröder, and H. P. Seidel. Mixed feelings: Expression of non-basic emotions in a muscle-based talking head. *Virtual Reality*, 8(4):201–212, 2005.

[2] B. Aschenberner and C. Weiss. Phoneme-viseme mapping for German video-realistic audio-visual-speech-synthesis. *IKP – Working Paper NF 11, Universität Bonn*, pages 1–11, 2005.

[3] M. S. Bartlett, G. C. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1(6):22–35, 2006.

[4] C. D. Batson and N. Ahmad. Empathy-induced altruism in a prisoner's dilemma II: what if the target of empathy has defected? *European Journal of Social Psychology*, 31(1):25–36, 2001.

[5] J. B. Bavelas, A. Black, N. Chovil, C. R. Lemery, and J. Mullett. Form and function in motor mimicry topographic evidence that the primary function is communicative. *Human Communication Research*, 14(3):275–299, 1988.

[6] J. B. Bavelas, A. Black, C. R. Lemery, and J. Mullett. Motor mimicry as primitive empathy. In N. Eisenberg and J. Strayer, editors, *Empathy and its Development*, pages 317–338. Cambridge University Press, 1987.

[7] C. Becker, H. Prendinger, M. Ishizuka, and I. Wachsmuth. Evaluating affective feedback of the 3D agent Max in a competitive cards game. In *Affective Computing and Intelligent Interaction (ACII)*, volume 3784 of *LNCS*, pages 466–473. Springer Berlin Heidelberg, 2005.

[8] C. Becker-Asano. *WASABI: Affect Simulation for Agents with Believable Interactivity*. PhD thesis, Bielefeld University, 2008.

[9] T. Bickmore and R. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions On Computer-Human Interaction (TOCHI)*, 12(2):293–327, 2005.

[10] M. Botvinick, A. P. Jha, L. M. Bylsma, S. A. Fabian, P. E. Solomon, and K. M. Prkachin. Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *NeuroImage*, 25(1):312–319, 2005.

[11] H. Boukricha. A first approach for simulating affective theory of mind through mimicry and role-taking. In *The Third International Conference on Cognitive Science, Symposium: Emotional Computer Systems and Interfaces*, pages 529–530, 2008.

[12] H. Boukricha, C. Becker, and I. Wachsmuth. Simulating empathy for the virtual human Max. In *Proceedings of the Second International Workshop on Emotion and Computing, in conj. with the German Conference on Artificial Intelligence (KI2007)*, pages 22–27, 2007.

[13] H. Boukricha, N. Nguyen, and I. Wachsmuth. Sharing emotions and space – empathy as a basis for cooperative spatial interactions. In S. Kopp, S. Marsella, K. R. Thorisson, and H. H. Vilhjalmsson, editors, *Intelligent Virtual Agents (IVA)*, volume 6895 of *LNAI*, pages 350–362. Springer Berlin Heidelberg, 2011.

[14] H. Boukricha and I. Wachsmuth. Towards facial mimicry for a virtual human. In *Proceedings of the Fourth International Workshop on Emotion and Computing – Current Research and Future Impact, in conj. with the German Conference on Artificial Intelligence (KI2009)*, pages 32–39, 2009.

[15] H. Boukricha and I. Wachsmuth. Empathy-based emotional alignment for a virtual human: A three-step approach. *Künstliche Intelligenz*, 25(3):195–204, 2011.

[16] H. Boukricha and I. Wachsmuth. Mechanism, modulation, and expression of empathy in a virtual human. In *Proceedings of the IEEE SSCI2011 - Symposium Series on Computational Intelligence, Workshop on Affective Computational Intelligence (WACI)*, pages 30–37. IEEE, 2011.

[17] H. Boukricha and I. Wachsmuth. Modeling empathy for a virtual human: How, when and to what extent? In K. Tumer, P. Yolum, L. Sonenberg, and P. Stone,

editors, *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, volume 3, pages 1035–1036. IFAAMAS, 2011.

[18] H. Boukricha, I. Wachsmuth, A. Hofstätter, and K. Grammer. Pleasure-arousal-dominance driven facial expression simulation. In *Third International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII)*, pages 119–125. IEEE, 2009.

[19] S. Brave, C. Nass, and K. Hutchinson. Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agents. *International Journal of Human-Computer Studies*, 62(2):161–178, 2005.

[20] C. Breazeal. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1):119–155, 2003.

[21] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*, 11(1-2):31–62, 2005.

[22] R. A. Brooks. Intelligence without reason. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 569–595, 1991.

[23] J. Cao, H. Wang, P. Hu, and J. Miao. PAD model based facial expression analysis. In *Advances in Visual Computing*, volume 5359 of *LNCS*, pages 450–459. Springer Berlin Heidelberg, 2008.

[24] L. Carr, M. Iacoboni, M. C. Dubeau, J. C. Mazziotta, and G. L. Lenzi. Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the U.S.A.*, 100(9):5497–5502, 2003.

[25] Y. Cheng, C. P. Lin, H. L. Liu, Y. Y. Hsu, K. E. Lim, D. Hung, and J. Decety. Expertise modulates the perception of pain in others. *Current Biology*, 17(19):1708–1713, 2007.

[26] D. Chi, M. Costa, L. Zhao, and N. Badler. The EMOTE model for effort and shape. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '00*, pages 173–182. ACM Press/Addison-Wesley Publishing, 2000.

[27] W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979.

[28] W. S. Cleveland and S. J. Develin. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403):596–610, 1988.

[29] S. E. Costin and D. C. Jones. Friendship as a facilitator of emotional responsiveness and prosocial interventions among young children. *Developmental Psychology*, 28(5):941–947, 1992.

[30] M. Courgeon, J. C. Martin, and C. Jacquemin. User's gestural exploration of different virtual agents' expressive profiles. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, volume 3, pages 1237–1240. IFAAMAS, 2008.

[31] A. Damasio. *Decartes' Error: Emotion, Reason, and the Human Brain.* Gosset/Putnam Press, 1994.

[32] C. Darwin. *The Expression of the Emotions in Man and Animals*, volume 526. The University of Chicago Press, 1965.

[33] M. H. Davis. A multidimensional approach to individual differences. *JSAS Catalog of Selected Documents in Psychology*, 10:85, 1980.

[34] M. H. Davis. *Empathy: A Social Psychological Approach.* Westview Press, 1994.

[35] F. De Vignemont and T. Singer. The empathic brain: How, when and why? *Trends in Cognitive Sciences*, 10(10):435–441, 2006.

[36] J. Dias and A. Paiva. Feeling and reasoning: A computational model for emotional characters. In C. Bento, A. Cardoso, and G. Dias, editors, *EPIA'05 Proceedings of the 12th Portuguese conference on Progress in Artificial Intelligence*, volume 3808 of *LNAI*, pages 127–140. Springer Berlin Heidelberg, 2005.

[37] N. Eisenberg, C. L. Shea, G. Carlo, and G. P. Knight. Empathy-related responding and cognition: A "chicken and the egg" dilemma. In W. M. Kurtines and J. L. Gewirtz, editors, *Handbook of Moral Behavior and Development*, volume 2, pages 63–88. Lawrence Erlbaum Associates, 1991.

[38] P. Ekman. Facial expressions. In *Handbook of Cognition and Emotion*, 16, pages 301–320, 1999.

[39] P. Ekman and W. V. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, 1978.

[40] P. Ekman, W. V. Friesen, and J. C. Hager. *Facial Action Coding System: Investigator's Guide.* Research Nexus, a subsidiary of Network Information Research Corporation, Salt Lake City UT, USA, 2002.

[41] R. El Kaliouby and P. Robinson. Mind reading machines: Automated inference of cognitive mental states from video. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 1, pages 682–688. IEEE, 2004.

[42] C. Elliott. Using the affective reasoner to support social simulations. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 194–200. Morgan Kaufmann, 1993.

[43] V. Gallese. Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, 4(1):23–48, 2005.

[44] V. Gallese, C. Keysers, and G. Rizzolatti. A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8(9):396–403, 2004.

[45] A. P. Goldstein and G. Y. Michaels. *Empathy Development, Training and Consequences.* Lawrence Erlbaum Associates, 1985.

[46] K. Grammer and E. Oberzaucher. The reconstruction of facial expressions in embodied systems: New approaches to an old problem. In *ZiF:Mitteilungen*, volume 2/2006, pages 14–31, 2006.

[47] J. Gratch and S. Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, 2004.

[48] J. Gratch and S. Marsella. Evaluating the modeling and use of emotion in virtual humans. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, volume 1, pages 320–327. IEEE, 2004.

[49] X. Gu and S. Han. Attention and reality constraints of the neural processes of empathy for pain. *Neuroimage*, 36(1):256–267, 2007.

[50] H. Gunes and M. Pantic. Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions*, 1(1):68–99, 2010.

[51] H. Gunes, M. Piccardi, and M. Pantic. From the lab to the real world: Affect recognition using multiple cues and modalities. In O. Jimmy, editor, *Affective Computing: Focus on Emotion Expression, Synthesis, and Recognition*, pages 185–218. InTech Education and Publishing, 2008.

[52] G. Hein and T. Singer. I feel how you feel but not always: The empathic brain and its modulation. *Current Opinion in Neurobiology*, 18(2):153–158, 2008.

[53] E. T. Higgins. Role-taking and social judgment: Alternative developmental perspectives and processes. In J. H. Flavell and L. Ross, editors, *Social Cognitive Development: Frontiers and Possible Futures*, pages 119–153. Cambridge University Press, 1981.

[54] M. L. Hoffman. Interaction of affect and cognition in empathy. In C. E. Izard, J. Kagan, and R. B. Zajonc, editors, *Emotions, Cognition, and Behavior*, pages 103–131. Cambridge University Press, 1984.

[55] M. L. Hoffman. *Empathy and Moral Development.* Cambridge University Press, 2000.

[56] A. Hofstätter. Implementierung des Facial Action Coding Systems in einem Avatar. Master's thesis, Universität Wien, 2008.

[57] S. Hyniewska, R. Niewiadomski, M. Mancini, and C. Pelachaud. Expression of affects in embodied conversational agents. In K. R. Scherer, T. Bänziger, and E. B. Roesch, editors, *Blueprint for Affective Computing*, pages 213–221. Oxford University Press, 2010.

[58] W. James. What is an emotion? *Mind*, 9(34):188–205, 1884.

[59] R. W. Jr Hill, J. Gratch, S. Marsella, J. Rickel, W. R. Swartout, and D. R. Traum. Virtual humans in the mission rehearsal exercise system. *Künstliche Intelligenz*, 17(4):32–38, 2003.

[60] K. Karpouzis, G. Caridakis, L. Kessous, N. Amir, A. Raouzaiou, L. Malatesta, and S. Kollias. Modeling naturalistic affective states via facial, vocal, and bodily expressions recognition. In *Artificial Intelligence for Human Computing*, volume 4451 of *LNCS*, pages 91–112. Springer Berlin Heidelberg, 2007.

[61] R. Koenen, F. Pereira, and L. Chiariglione. MPEG-4: Context and objectives. *Signal Processing: Image Communication*, 9(4):295–304, 1997.

[62] S. Kopp, L. Gesellensetter, N. Krämer, and I. Wachsmuth. A conversational agent as museum guide – design and evaluation of a real-world application. In *Intelligent Virtual Agents (IVA)*, volume 3661 of *LNCS*, pages 329–343. Springer Berlin Heidelberg, 2005.

[63] S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth. Max – a multimodal assistant in virtual reality construction. *Künstliche Intelligenz*, 17(4):11–17, 2003.

[64] S. Kopp and I. Wachsmuth. Synthesizing multimodal utterances for conversational agents. *Computer Animation and Virtual Worlds*, 15(1):39–52, 2004.

[65] D. Krebs. Empathy and altruism. *Journal of Personality and Social Psychology*, 32(6):1134–1146, 1975.

[66] C. Lamm, C. D. Batson, and J. Decety. The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, 19(1):42–58, 2007.

[67] E. Lee, J. I. Kang, I. H. Park, J. J. Kim, and S. K. An. Is a neutral face really evaluated as being emotionally neutral? *Psychiatry Research*, 157(1):77–85, 2008.

[68] I. Leite, C. Martinho, A. Pereira, and A. Paiva. iCat: An affective game buddy based on anticipatory mechanisms. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, volume 3, pages 1229–1232. IFAAMAS, 2008.

[69] I. Leite, A. Pereira, G. Castellano, S. Mascarenhas, C. Martinho, and A. Paiva. Modelling empathy in social robotic companions. In *Advances in User Modeling*, volume 7138 of *LNCS*, pages 135–147. Springer Berlin Heidelberg, 2012.

[70] N. Lessmann, S. Kopp, and I. Wachsmuth. Situated interaction with a virtual human – perception, action, and cognition. In G. Rickheit and I. Wachsmuth, editors, *Situated Communication*, pages 287–323. Mouton de Gruyter, 2006.

[71] R. W. Levenson, P. Ekman, and W. V. Friesen. Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, 27(4):363–384, 1990.

[72] L. S. Mark, K. Nemeth, D. Gardner, M. J. Dainoff, J. Paasche, M. Duffy, and K. Grandt. Postural dynamics and the preferred critical boundary for visually guided reaching. *Journal of Experimental Psychology: Human Perception and Performance*, 23(5):1365–1379, 1997.

[73] S. Marsella, J. Gratch, and P. Petta. Computational models of emotion. In K. R. Scherer, T. Bänziger, and E. B. Roesch, editors, *Blueprint for Affective Computing*, pages 21–45. Oxford University Press, 2010.

[74] N. Mattar and I. Wachsmuth. Who are you? On the acquisition of information about people for an agent that remembers. In *ICAART 2012 - Proceedings of the 4th International Conference on Agents and Artificial Intelligence*, volume 2, pages 98–105. SciTePress, 2012.

[75] D. N. McIntosh. Facial feedback hypotheses: Evidence, implications, and directions. *Motivation and Emotion*, 20(2):121–147, 1996.

[76] S. W. McQuiggan, J. L. Robison, R. Phillips, and J. C. Lester. Modeling parallel and reactive empathy in virtual agents: An inductive approach. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, volume 1, pages 167–174. IFAAMAS, 2008.

[77] A. Mehrabian and N. Epstein. A measure of emotional empathy. *Journal of Personality*, 40(4):525–543, 1972.

[78] A. Mehrabian and J. A. Russell. *An Approach to Environmental Psychology*. The MIT Press, 1974.

[79] A. N. Meltzoff and M. K. Moore. Imitation of facial and manual gestures by human neonates. *Science*, 198(4312):75–78, 1977.

[80] A. N. Meltzoff and M. K. Moore. Explaining facial imitation: A theoretical model. *Early Development and Parenting*, 6(34):179–192, 1997.

[81] N. Nguyen. *Modellierung von peripersonalem Raum und Interaktionsraum durch ein Körperschema für humanoide Agenten.* PhD thesis, Bielefeld University, 2012.

[82] N. Nguyen and I. Wachsmuth. From body space to interaction space – modeling spatial cooperation for virtual humans. In K. Tumer, P. Yolum, L. Sonenberg, and P. Stone, editors, *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1047–1054. IFAAMAS, 2011.

[83] M. A. Nicolaou, H. Gunes, and M. Pantic. Output-associative rvm regression for dimensional and continuous emotion prediction. *Image and Vision Computing*, 30(3):186–196, 2012.

[84] R. Niewiadomski, M. Mancini, S. Hyniewska, and C. Pelachaud. Communicating emotional states with the Greta agent. In K. R. Scherer, T. Bänziger, and E. B. Roesch, editors, *Blueprint for Affective Computing*, pages 256–268. Oxford University Press, 2010.

[85] M. Ochs, R. Niewiadomski, C. Pelachaud, and D. Sadek. Intelligent expressions of emotions. In *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction (ACII)*, volume 3784 of *LNCS*, pages 707–714. Springer Berlin Heidelberg, 2005.

[86] M. Ochs, D. Sadek, and C. Pelachaud. A formal model of emotions for an empathic rational dialog agent. *Autonomous Agents and Multi-Agent Systems*, 24(3):410–440, 2012.

[87] B. L. Omdahl. *Cognitive Appraisal, Emotion, and Empathy.* Lawrence Erlbaum Associates, 1995.

[88] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions.* Cambridge University Press, 1988.

[89] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum. *The Measurement of Meaning.* University of Illinois Press, 1957.

[90] J. Ostermann. Animation of synthetic faces in MPEG-4. In *Proceedings of Computer Animation 98*, pages 49–55. IEEE, 1998.

[91] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall, and C. Zoll. Learning by feeling: Evoking empathy with synthetic characters. *Applied Artificial Intelligence*, 19(3–4):235–266, 2005.

[92] C. Paulus. Der Saarbrücker Persönlichkeitsfragebogen SPF(IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index, 2009.

[93] N. Pfeiffer-Lessmann, T. Pfeiffer, and I. Wachsmuth. An operational model of joint attention – timing of gaze patterns in interactions between humans and a virtual human. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society (CogSci)*, pages 851–856, 2012.

[94] R. Plutchik. *Emotion: A Psychoevolutionary Synthesis*. Harper & Row, 1980.

[95] H. Prendinger and M. Ishizuka. The empathic companion: A character-based interface that addresses users' affective states. *Applied Artificial Intelligence*, 19(3–4):267–285, 2005.

[96] G. Rizzolatti. The mirror neuron system and its function in humans. *Anatomy and Embryology*, 210(5):419–421, 2005.

[97] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2):131–141, 1996.

[98] S. H. Rodrigues, S. F. Mascarenhas, J. Dias, and A. Paiva. "I can feel it too!": Emergent empathic reactions between synthetic characters. In *Third International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII)*, Amsterdam, Netherland, 2009. IEEE.

[99] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[100] J. A. Russell. Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective. In J. A. Russell and J. M. Fernández-Dols, editors, *The Psychology of Facial Expression*, pages 295–320. Cambridge University Press, 1997.

[101] J. A. Russell and A. Mehrabian. Distinguishing anger and anxiety in terms of emotional response factors. *Journal of Consulting and Clinical Psychology*, 42(1):79–83, 1974.

[102] J. A. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294, 1977.

[103] M. V. Saarela, Y. Hlushchuk, A. C. D. C. Williams, M. Schürmann, E. Kalso, and R. Hari. The compassionate brain: Humans detect intensity of pain from another's face. *Cerebral Cortex*, 17(1):230–237, 2007.

[104] K. R. Scherer. The component process model: Architecture for a comprehensive computational model of emergent emotion. In K. R. Scherer, T. Bänziger, and E. B. Roesch, editors, *Blueprint for Affective Computing*, pages 47–70. Oxford University Press, 2010.

[105] H. Scholsberg. Three dimensions of emotion. *Psychological Review*, 61(2):81–88, 1954.

[106] M. Schröder. Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions. In *Affective Dialogue Systems*, volume 3068 of *LNCS*, pages 209–220. Springer Berlin Heidelberg, 2004.

[107] M. Schröder and J. Trouvain. The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology*, 6(4):365–377, 2003.

[108] Y. Shin. Facial expression recognition based on emotion dimensions on manifold learning. In *Computational Science – ICCS 2007*, volume 4488 of *LNCS*, pages 81–88. Springer Berlin Heidelberg, 2007.

[109] M. Shugrina, M. Betke, and J. Collomosse. Empathic painting: Interactive stylization through observed emotional state. In *Proceedings of the Fourth International Symposium on Non-Photorealistic Animation and Rendering*, pages 87–96. ACM, 2006.

[110] T. Singer, B. Seymour, J. P. O'Doherty, K. E. Stephan, R. J. Dolan, and C. D. Frith. Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439:466–469, 2006.

[111] J. Snodgrass. *Judgment of feeling states from facial behavior: A bottom-up approach.* PhD thesis, University of British Columbia, 1992.

[112] N. Stoiber, R. Seguier, and G. Breton. Automatic design of a control interface for a synthetic face. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, pages 207–216. ACM, 2009.

[113] E. Stotland. Exploratory investigations of empathy. In L. Berkowitz, editor, *Advances in Experimental Social Psychology*, volume 4, pages 271–314. Academic Press, 1969.

[114] F. Strack, L. L. Martin, and S. Stepper. Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54(5):768–777, 1988.

[115] N. T. Termine and C. E. Izard. Infants' responses to their mothers' expressions of joy and sadness. *Developmental Psychology*, 24(2):223–229, 1988.

[116] M. Valster and M. Pantic. Fully automatic facial action unit detection and temporal analysis. In *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, pages 149–149. IEEE, 2006.

[117] J. D. Velásquez. Modeling emotions and other motivations in synthetic agents. In *Proceedings of the 14th National Conference on Artificial Intelligence and Ninth Conference on Innovative Applications of Artificial Intelligence*, AAAI'97/IAAI'97, pages 10–15. AAAI Press, 1997.

[118] B. Wicker, C. Keysers, J. Plailly, J. P. Royet, V. Gallese, and G. Rizzolatti. Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, 40(3):655–664, 2003.

[119] A. Wojdel, L. J. M. Rothkranz, and J. C. Wojdel. Fuzzy-logical implementation of co-occurence rules for combining AUs. In *Computer Graphics and Imaging (CGIM)*, pages 1–6, 2003.

[120] M. Wooldridge. *An Introduction to Multiagent Systems.* John Wiley & Sons, 2002.

[121] S. Zhang, Z. Wu, H. M. Meng, and L. Cai. Facial expression synthesis using PAD emotional parameters for a chinese expressive avatar. In A. Paiva, R. Prada, and R. Picard, editors, *Proceedings of the Second International Conference on Affective Computiong and Intelligent Interaction (ACII)*, volume 4738 of *LNCS*, pages 24–35. Springer Berlin Heidelberg, 2007.

Appendices

A EMMA's facial expression repertoire

A.1 The Facial Action Coding System (FACS)

In this section, two tables showing the list of Action Units (AUs) as defined in FACS and the rules to link AUs to the six basic emotion categories [40] are presented.

AU Number	AU Name	AU Number	AU Name
Upper Face AUs		**Lower Face AUs**	
1	Inner Brow Raise	9	Nose Wrinkle
2	Outer Brow Raise	10	Upper Lip Raise
4	Brow Lowerer	11	Nasolabial Furrow Deepener
5	Upper Lid Raise	12	Lip Corner Puller
6	Cheek Raise	13	Sharp Lip Puller
7	Lids Tight	14	Dimpler
43	Eyes Close	15	Lip Corner Depressor
45	Blink	16	Lower Lip Depressor
46	Wink	17	Chin Raiser
70	Brows Not Visible	18	Lip Pucker
71	Eyes Not Visible	20	Lip Stretch
Head Positions		22	Lip Funneler
51	Turn Left	23	Lip Tightener
52	Turn Right	24	Lip Presser
53	Head Up	28	Lip Suck
54	Head Down	72	Lower Face Not Visible
55	Tilt Left	**Miscellaneous AUs**	
56	Tilt Right	8	Lips Toward Each Other
57	Forward	19	Tongue Show
58	Back	21	Neck Tightener
Eye Positions		29	Jaw Thrust
61	Eyes Left	30	Jaw Sideways
62	Eye Right	31	Jaw Clencher
63	Eyes Up	32	Bite
64	Eyes Down	33	Blow
65	Walleye	34	Puff
66	Crosseye	35	Cheek Suck
Lip Parting and Jaw Opening		36	Tongue Bulge
25	Lips Part	37	Lip Wipe
26	Jaw Drop	38	Nostril Dilate
27	Mouth Stretch	39	Nostril Compress

Table A.1: The list of AUs from FACS [40].

Emotion	Prototypes	Major Variants
Surprise	1+2+5B+26 1+2+5B+27	1+2+5B 1+2+26 1+2+27 5B+26 5B+27
Fear	1+2+4+5+20+25, 26, or 27 1+2+4+5+25, 26, or 27	1+2+4+5+L or R20+25, 26, or 27 1+2+4+5 1+2+5Z, with or without 25, 26, 27 5+20, with or without 25, 26, 27
Happy	6+12 12C/D	
Sadness	1+4+11+15B, with or without 54+64 1+4+15, with or without 54+64 6+15, with or without 54+64	1+4+11, with or without 54+64 1+4+15B, with or without 54+64 1+4+15B+17, with or without 54+64 11+15B, with or without 54+64 11+17
	25 or 26 may occur with all prototypes or major variants	
Disgust	9 9+16+15, 26 9+17 10 10+16+25, 26 10+17	
Anger	4+5+7+10+22+23+25, 26 4+5+7+10+23+25, 26 4+5+7+23+25, 26 4+5+7+17+23 4+5+7+17+24 4+5+7+23 4+5+7+24	Any of the prototypes without any one of the following AUs: 4, 5, 7, or 10

Table A.2: Rules to link AUs to the six basic emotion categories [40].

A.2 EMMA's Action Units

In this section, a list of figures showing EMMA's AUs is provided. EMMA's AUs are modeled with the help of experienced FACS coders in cooperation with the department of anthropology at the University of Vienna [56].

Neutral Face

Action Unit 1:
Inner Brow Raiser

Action Unit 2:
Outer Brow Raiser

Action Unit 4:
Brow Lowerer

Action Unit 5:
Uppder Lid Raiser

Action Unit 6:
Cheek Raiser

Action Unit 7:
Lower Lid Raiser

Action Unit 9:
Nose Wrinkler

Action Unit 10:
Upper Lid Raiser

Action Unit 11:
Nasolabial Furrow Deepener

Action Unit 12:
Lip Corner Puller

Action Unit 13:
Sharp Lip Puller

Action Unit 14:
Dimpler

Action Unit 15:
Lip Corner Depressor

Action Unit 16:
Lower Lip Depressor

Action Unit 17:
Chin Raiser

Action Unit 18:
Lip Pucker

Action Unit 20:
Lip Strechter

Action Unit 22:
Lip Funnel

Action Unit 23:
Lip Tightener

Action Unit 24:
Lip Presser

Action Unit 25:
Lips Part

Action Unit 26:
Jaw Drop

Action Unit 27:
Mouth Stretch

Action Unit 28:
Lip Suck

Action Unit 29:
Jaw Thrust

Action Unit 30:
Jaw Sideways

Action Unit 31:
Jaw Clencher

Action Unit 32:
Bite Lip

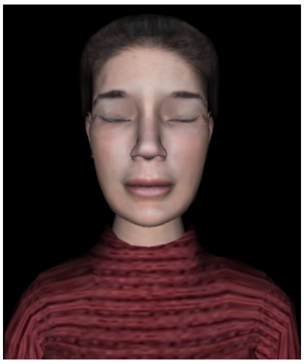
Action Unit 33:
Blow

Action Unit 34:
Puff

Action Unit 35:
Suck Cheeks

Action Unit 38:
Nostril Dilatator

Action Unit 39:
Nostril Compressor



Action Unit 43:
Eyes Close

Action Unit 51:
Head Turn Left

Action Unit 52:
Head Turn Right

Action Unit 53:
Head Turn Up

Action Unit 54:
Head Turn Down

Action Unit 55:
Head Tilt Left

Action Unit 56:
Head Tilt Right

Action Unit 61:
Eyes Left

Action Unit 62:
Eyes Right

Action Unit 63:
Eyes Up

Action Unit 64:
Eyes Down

Action Unit 90:
Body Turn Left

Action Unit 91:
Body Turn Right

A.3 EMMA's Visemes

In this section, a list of figures showing EMMA's visemes is provided. EMMA's visemes are modeled following the work of Aschenberner & Weiss [2] (see Table A.3).

Phoneme (BOSS)	Viseme	Example
p, b	**P**	Pause, Bitte
t, d, k, g	**T**	Tonne, Dach, König, Gier
n, @n, l, @l	**N**	Nadel, raten, Liebe, Igel
m	**M**	Mutter
f, v	**F**	Finder, Vase
s, z	**S**	Fass, Sein
S, Z, tS, dZ	**Z**	Schein, Garage, Tscheche, Dschungel
h, r, x, N	**R**	Hase, Reden, Dach, Wange
j, C	**C**	Junge, Wicht
i:, I, e:, E:, E	**E**	Bier, Tisch, Weg, Räte, Menge
a:, a	**A**	Wange, Watte
o:, O	**O**	Wolle, Wogen
u:, U	**U**	Buch, Runde
@, 6	**Q**	Bitte, Weiher
y:, Y, 2:, 9	**Y**	Tür, Mütter, Goethe, Götter

Table A.3: Phoneme-viseme mapping as defined by Aschenberner & Weiss [2].

Viseme P
AU17(0.6)+AU24(0.2)+AU34(0.2)

Viseme T
AU20(0.1)+AU25(0.4)+AU26(0.4)+AU27(0.3)

Viseme N
AU16(0.5)+AU25(0.1)+AU26(0.2)

Viseme M
AU17(0.3)+AU23(0.3)+AU24(0.4)+AU26(0.1)

Viseme F
AU10(0.2)+AU11(0.4)+AU17(0.1)+AU18(0.4)+AU22(0.2)+AU26(0.2)+AU27(0.3)+AU33(0.4)

Viseme S
AU10(0.2)+AU12(0.3)+AU14(0.3)+AU18(0.1)+AU22(0.3)+AU24(0.1)+AU26(0.2)+AU27(0.5)+AU33(0.4)

Viseme Z
AU18(0.2)+AU22(0.7)+AU26(0.1)+AU27(0.4)+AU33(0.4)

Viseme R
AU10(0.1)+AU11(0.3)+AU16(0.2)+AU25(0.6)+AU26(0.5)+AU27(0.2)

Viseme C
AU16(0.2)+AU25(0.3)+AU26(0.3)+AU27(0.2)+AU33(0.2)

Viseme E
AU12(0.4)+AU16(0.3)+AU20(0.2)+AU25(0.7)+AU26(0.7)

Viseme A
AU25(0.7)+AU26(0.8)+AU27(0.3)

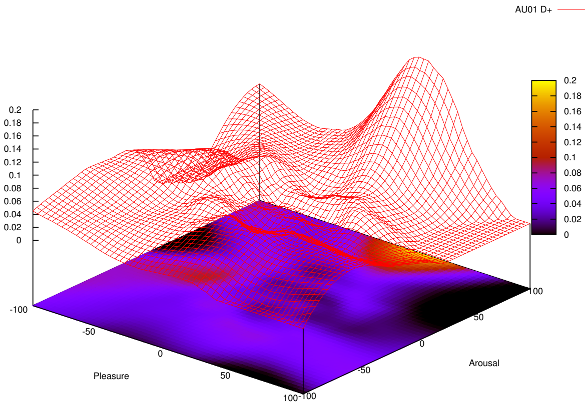
Viseme O
AU18(0.8)+AU22(0.3)+AU25(0.1)+AU26(0.5)+AU27(0.3)

Viseme U
AU18(0.8)+AU22(0.4)+AU26(0.1)+AU27(0.1)

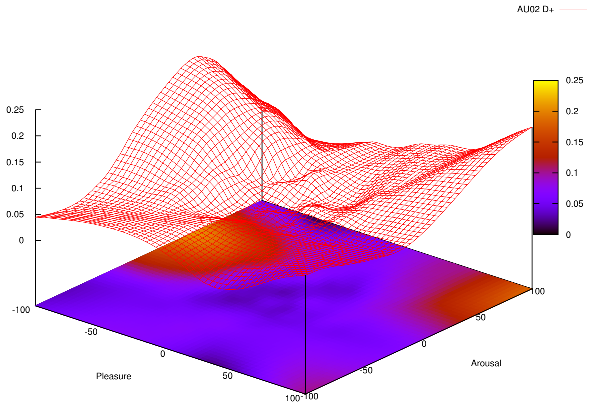
Viseme Q
AU25(0.5)+AU26(0.5)

Viseme Y
AU18(0.8)+AU22(0.4)+AU26(0.2)+AU27(0.1)

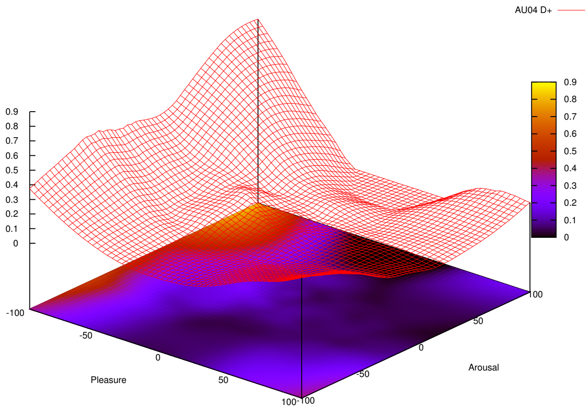
A.4 Non-linear Regression Planes - PA Space of Positive Dominance



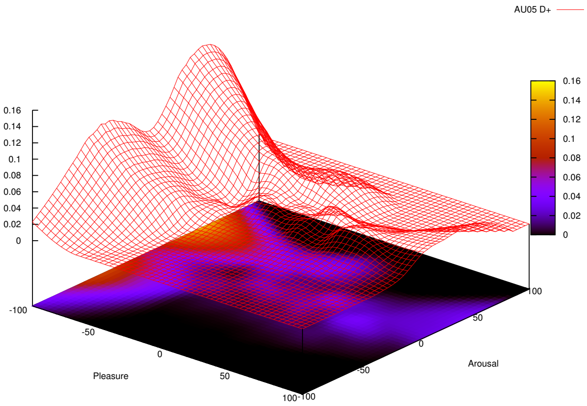
AU1



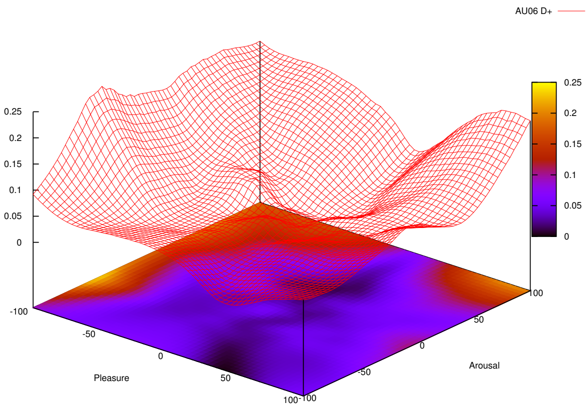
AU2



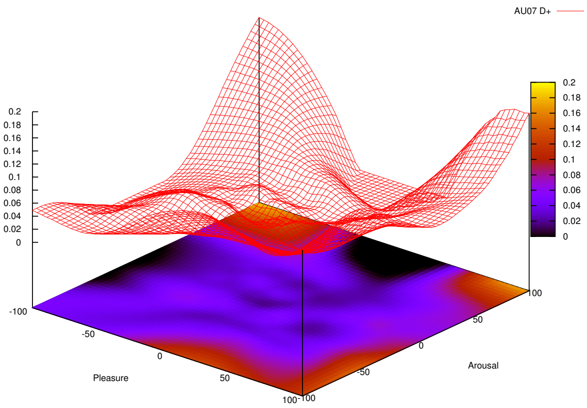
AU4



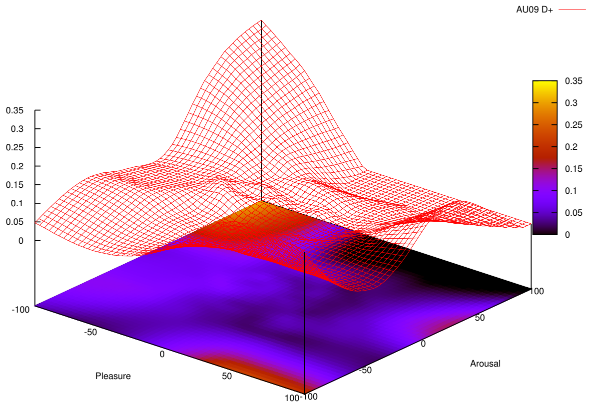
AU5



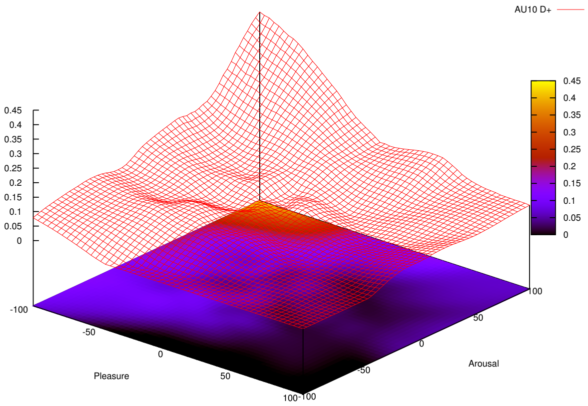
AU6



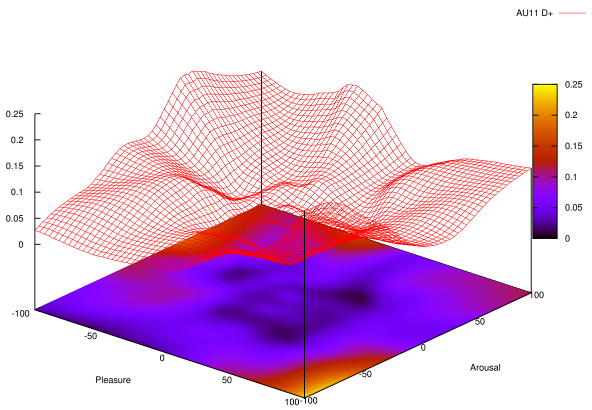
AU7



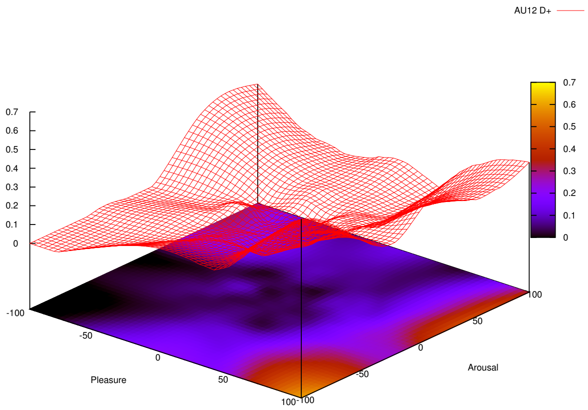
AU9



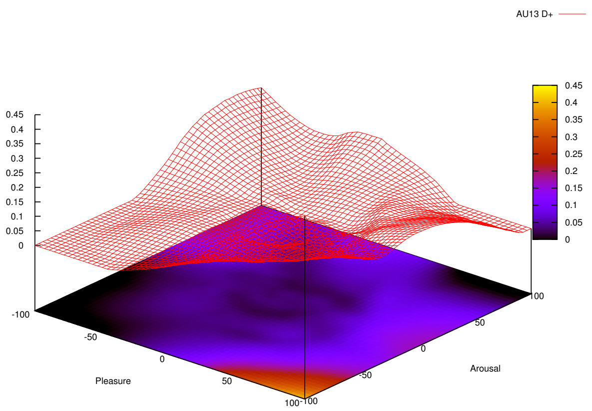
AU10



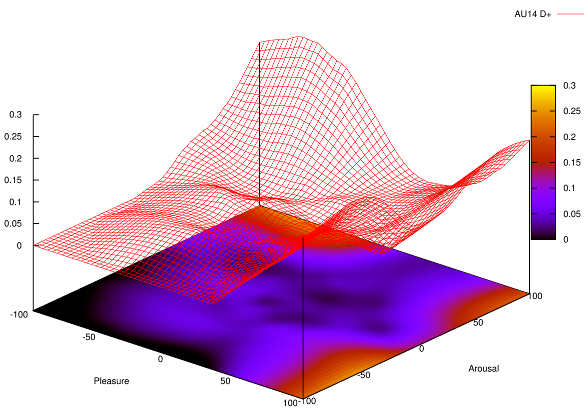
AU11



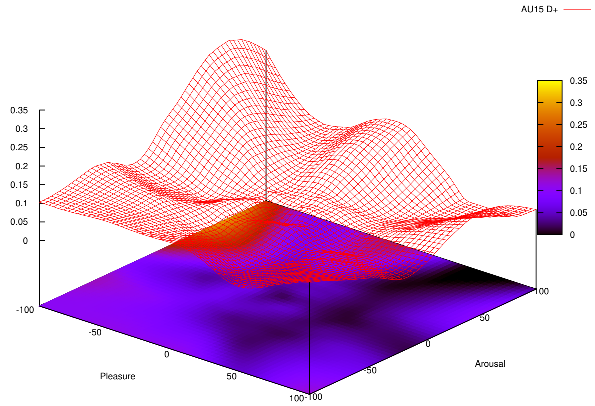
AU12



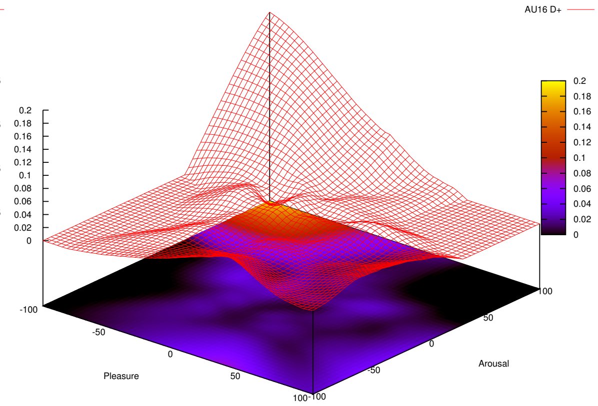
AU13



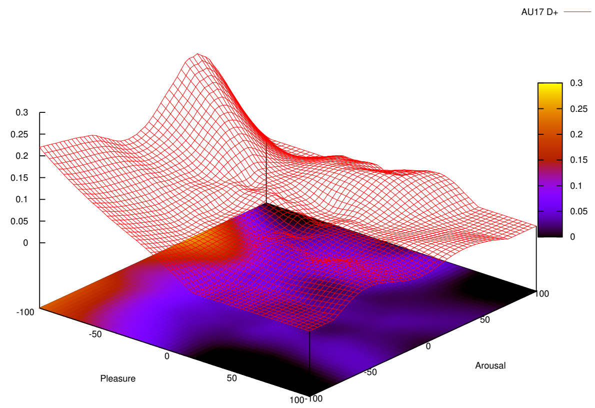
AU14



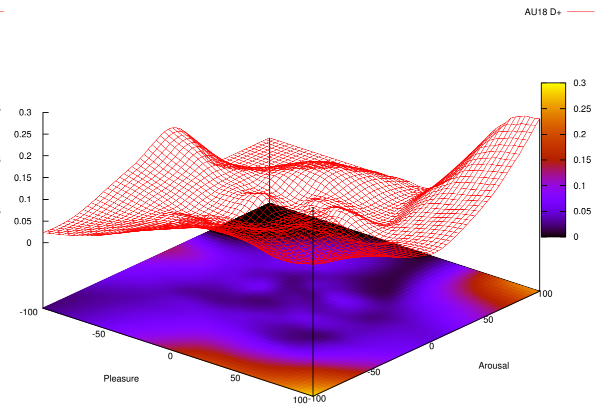
AU15



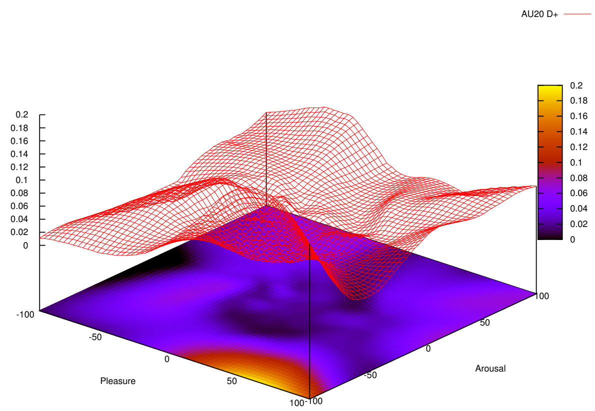
AU16



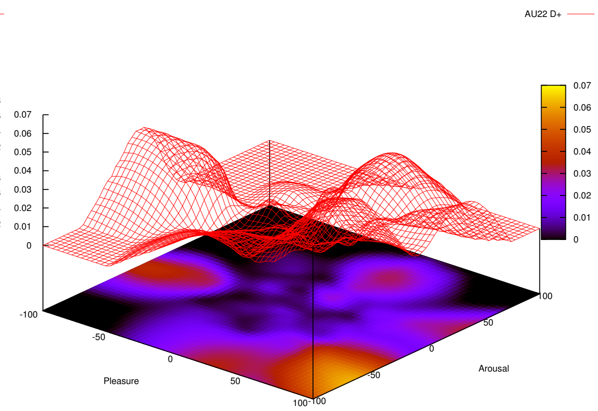
AU17



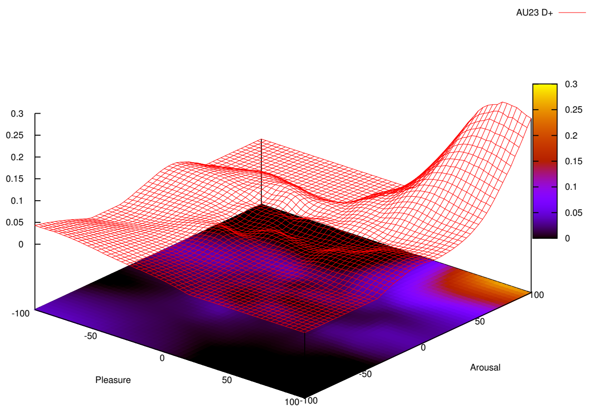
AU18



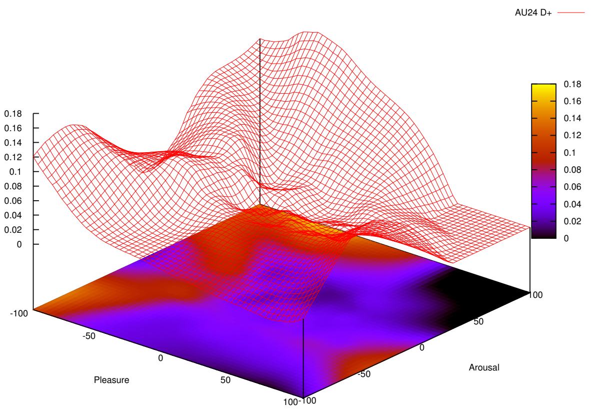
AU20



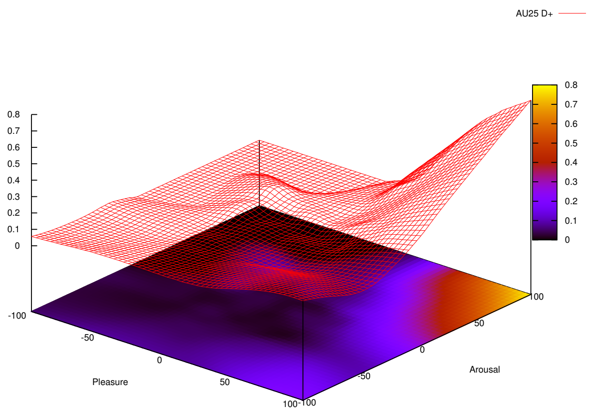
AU22



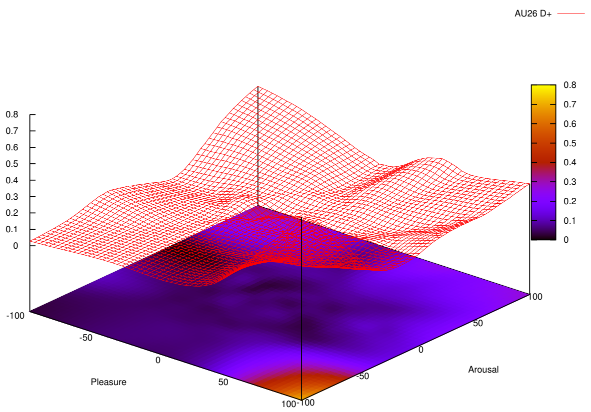
AU23



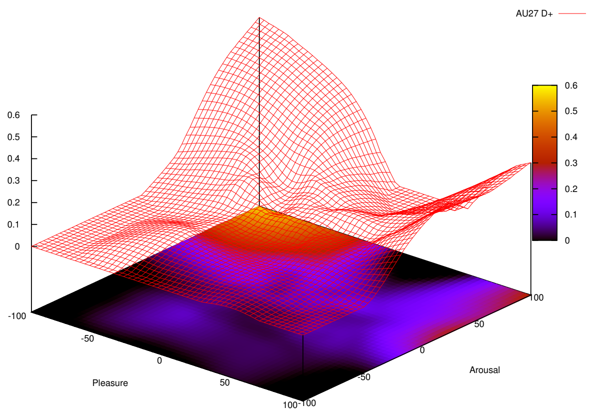
AU24



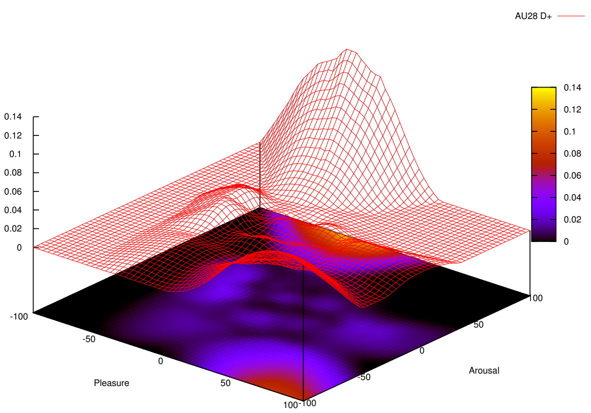
AU25



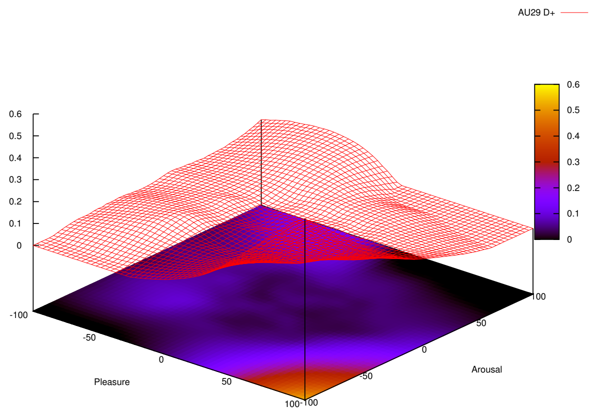
AU26



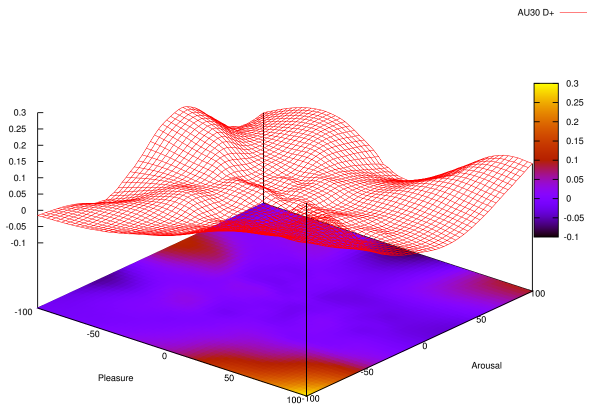
AU27



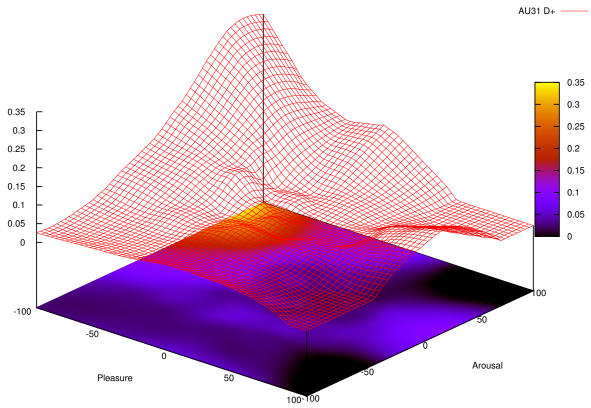
AU28



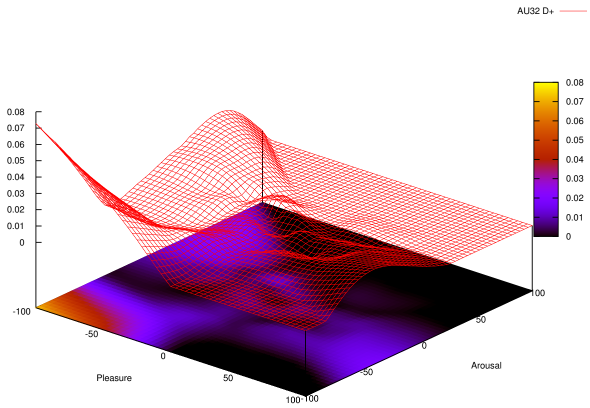
AU29



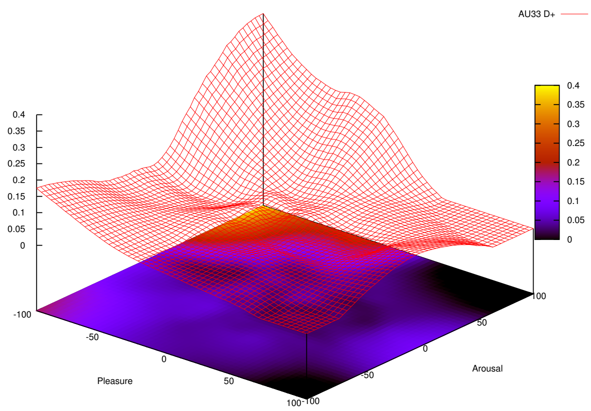
AU30



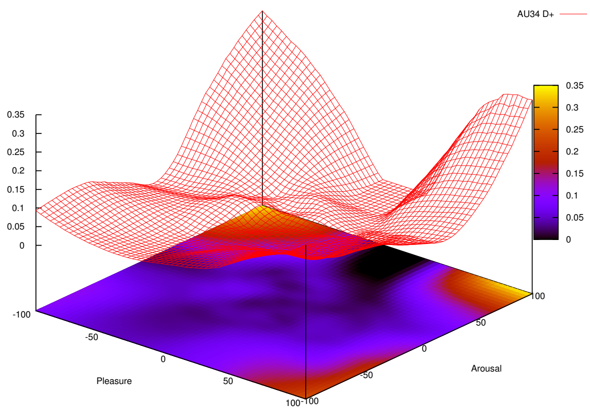
AU31



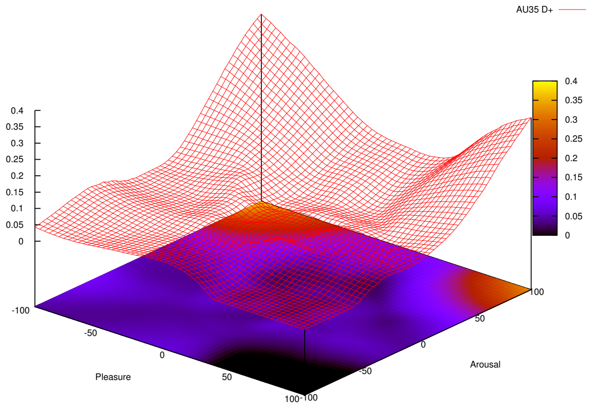
AU32



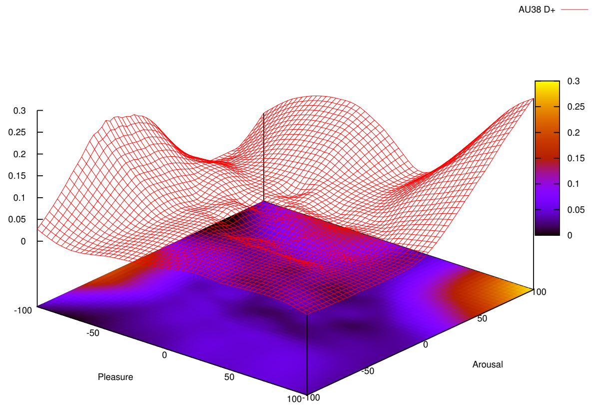
AU33



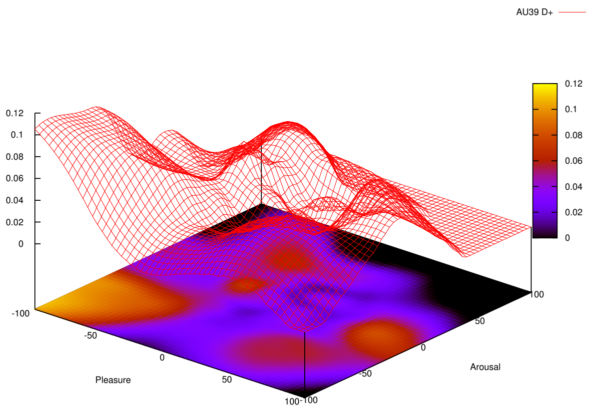
AU34



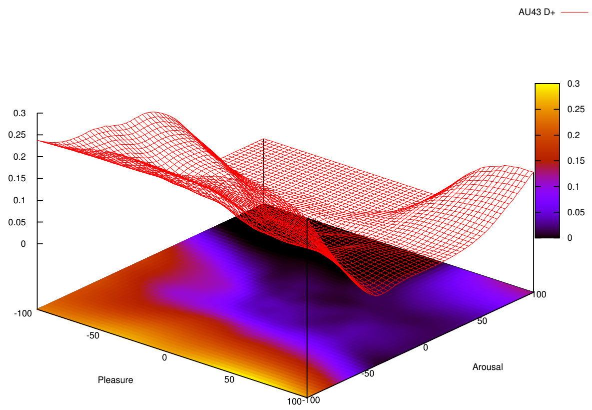
AU35



AU38

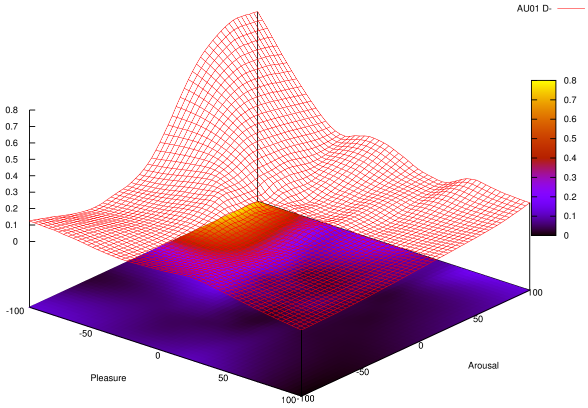


AU39

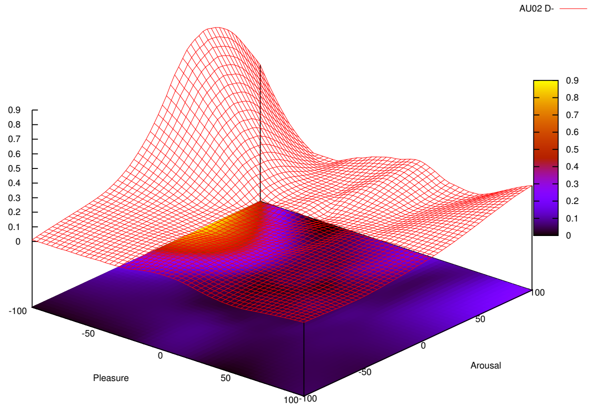


AU43

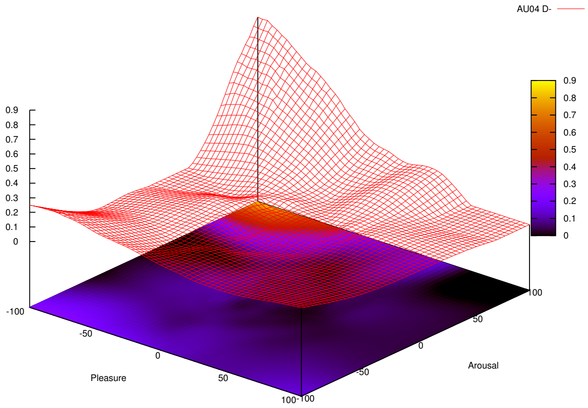
A.5 Non-linear Regression Planes - PA Space of Negative Dominance



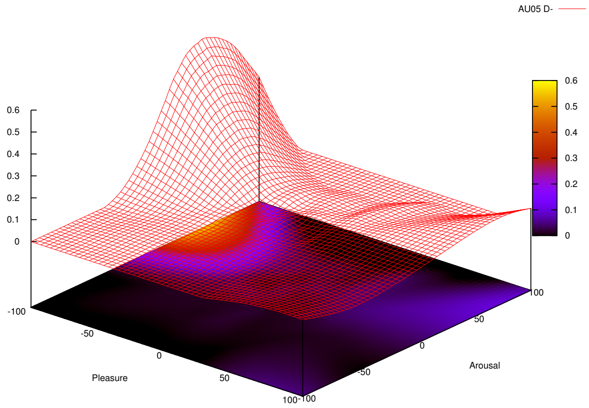
AU 01



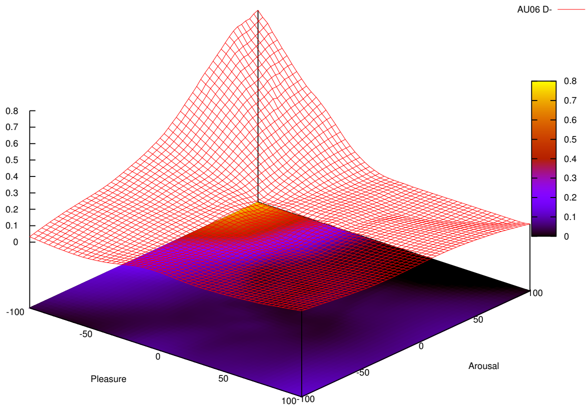
AU 02



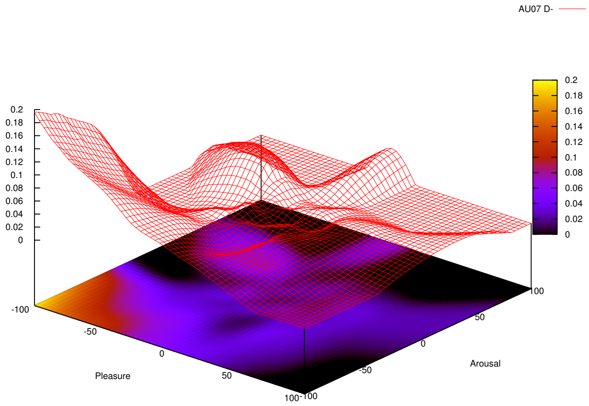
AU 04



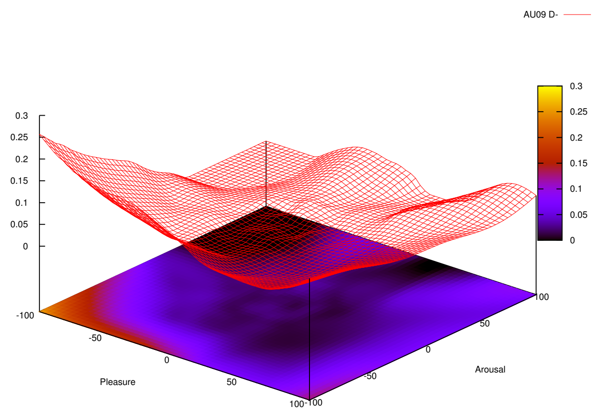
AU 05



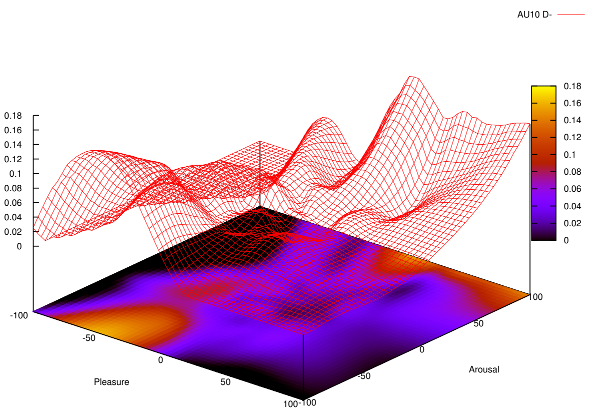
AU 06



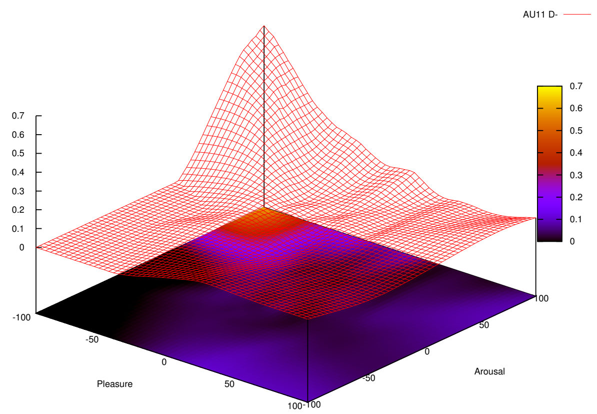
AU 07



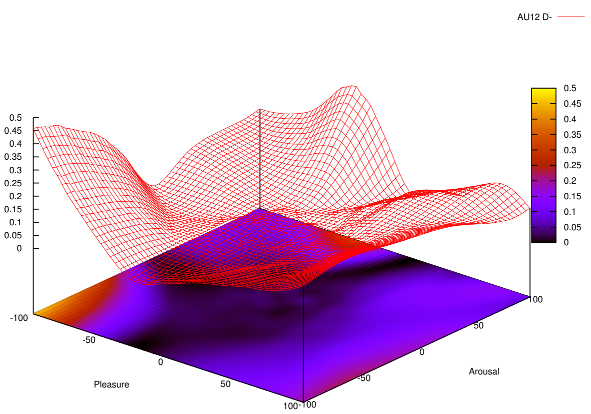
AU9



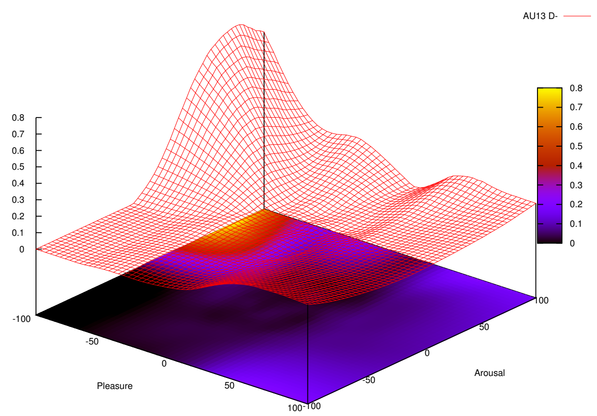
AU10



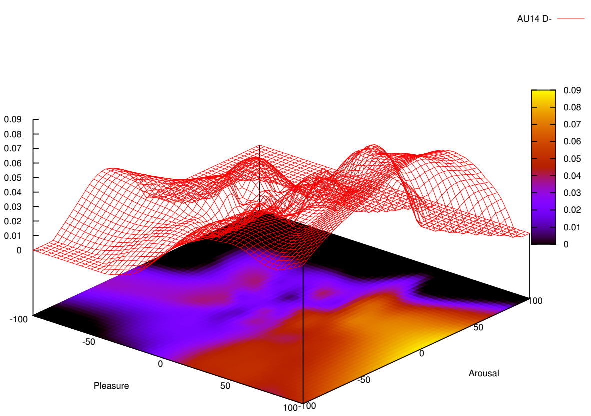
AU11



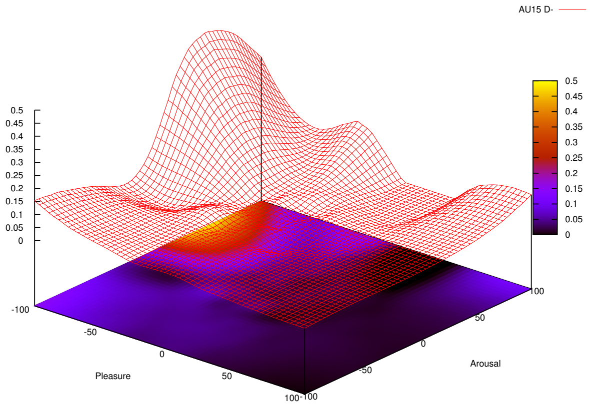
AU12



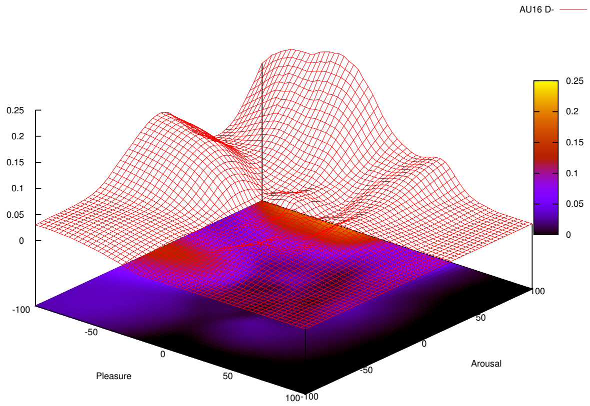
AU13



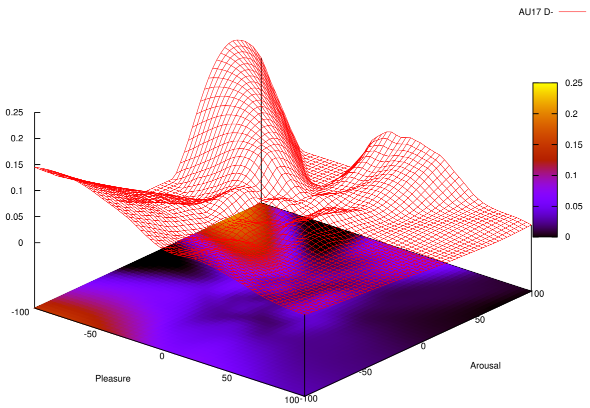
AU14



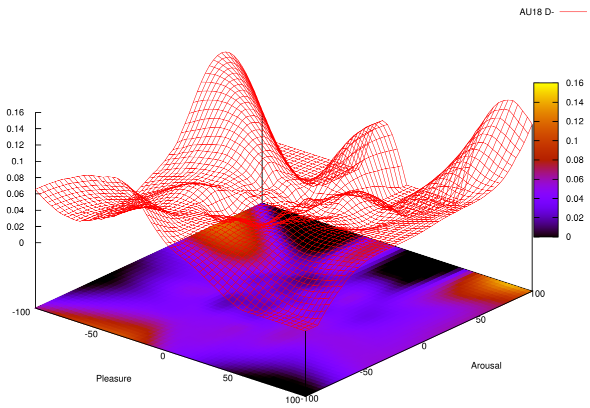
AU15



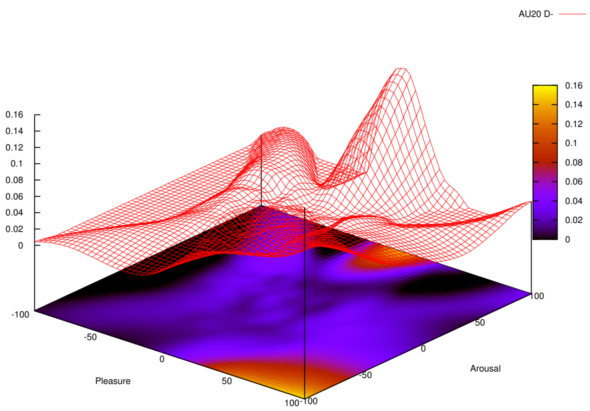
AU16



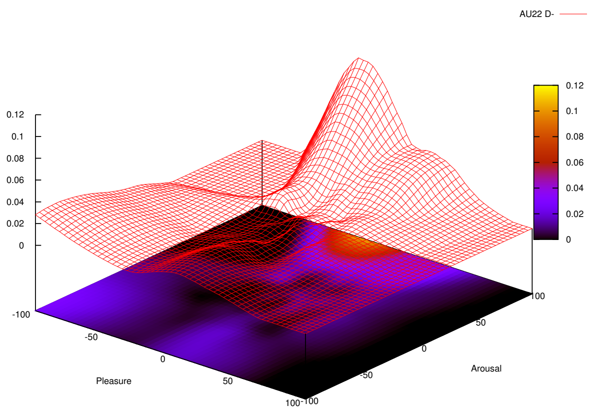
AU17



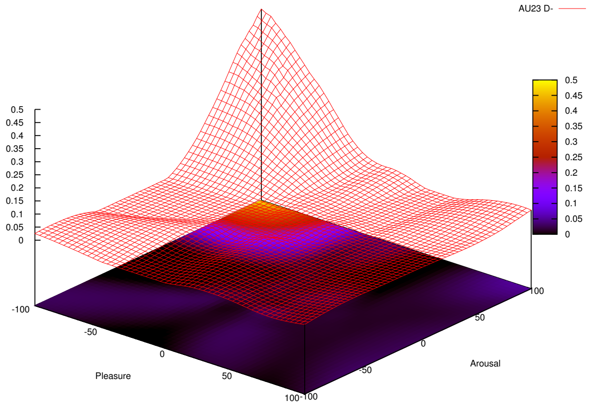
AU18



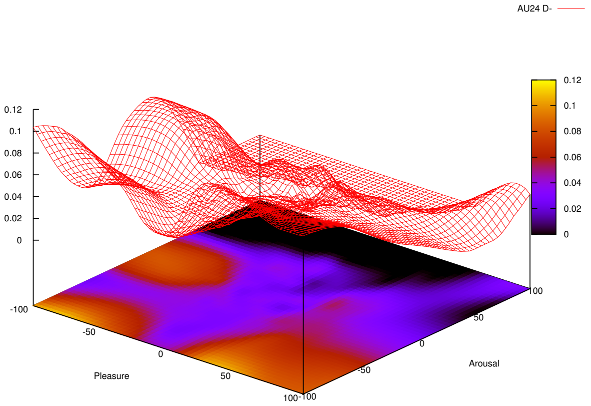
AU20



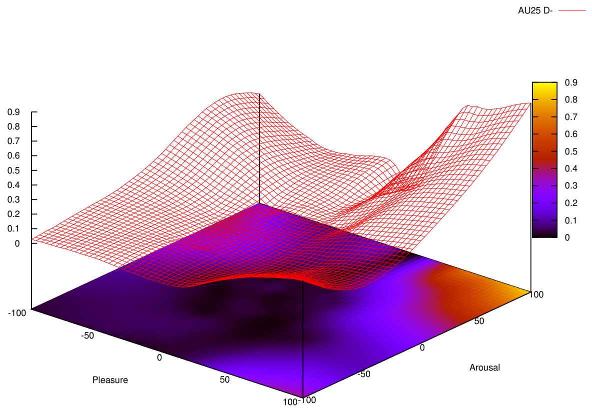
AU22



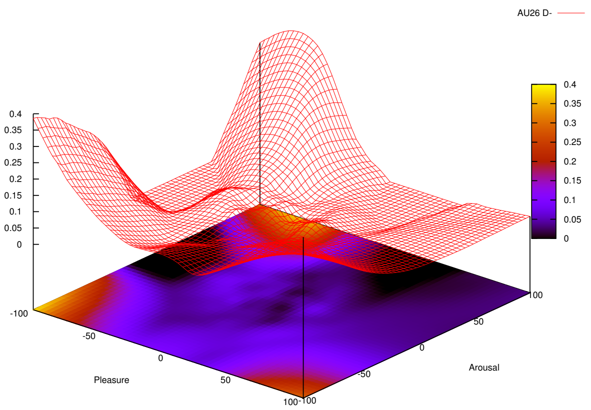
AU23



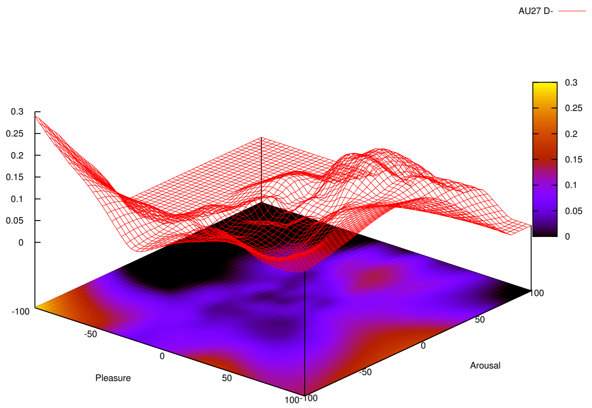
AU24



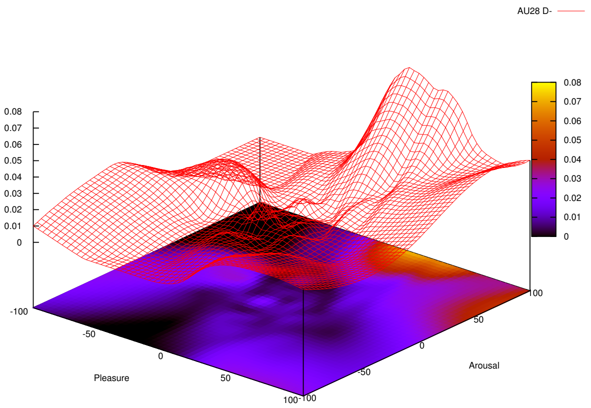
AU25



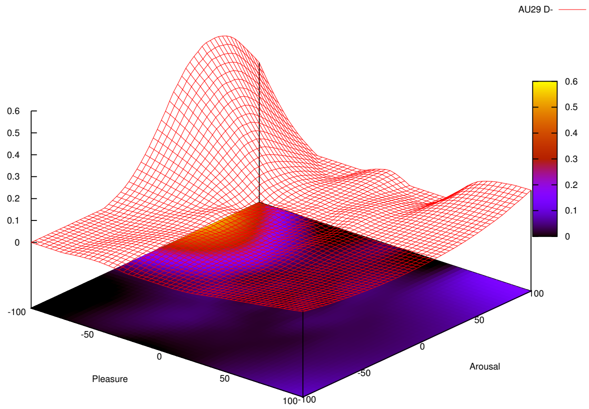
AU26



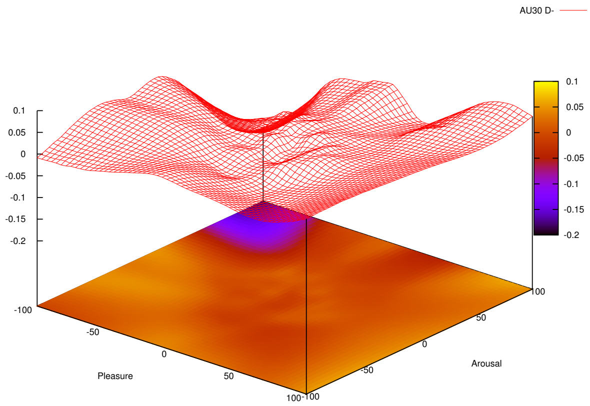
AU27



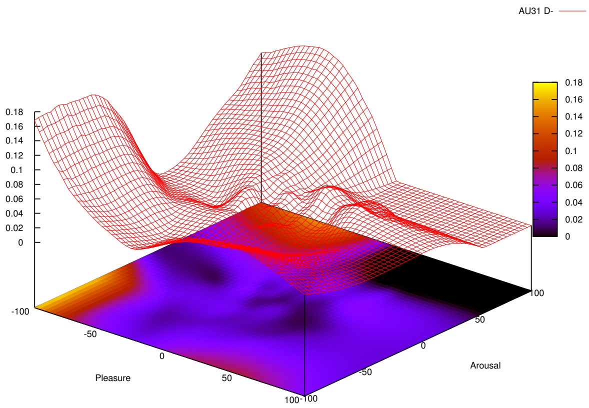
AU28



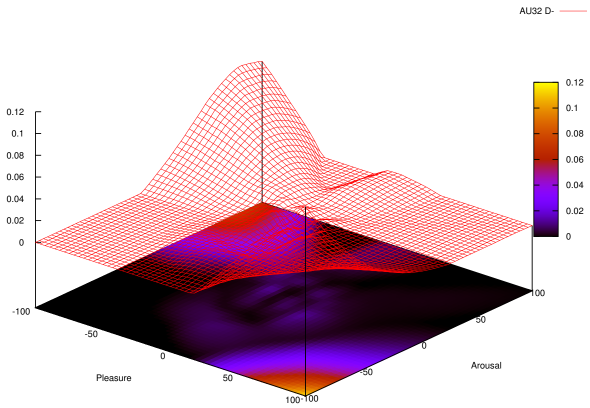
AU29



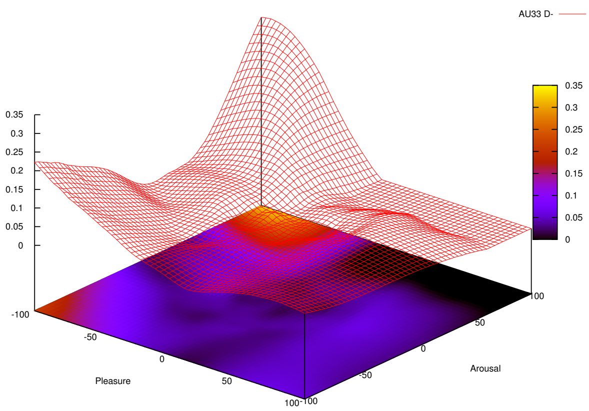
AU30



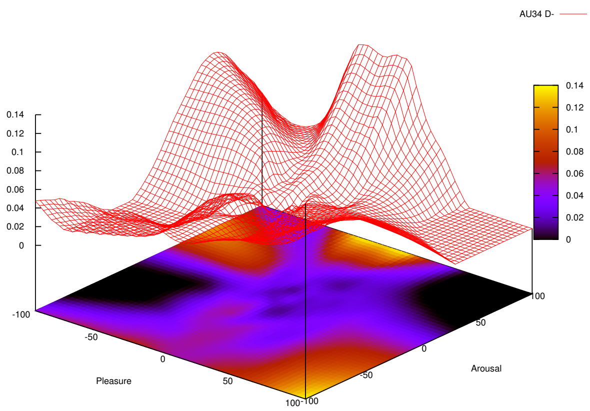
AU31



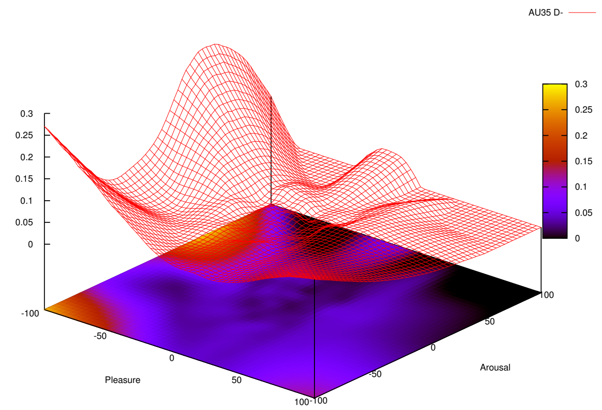
AU32



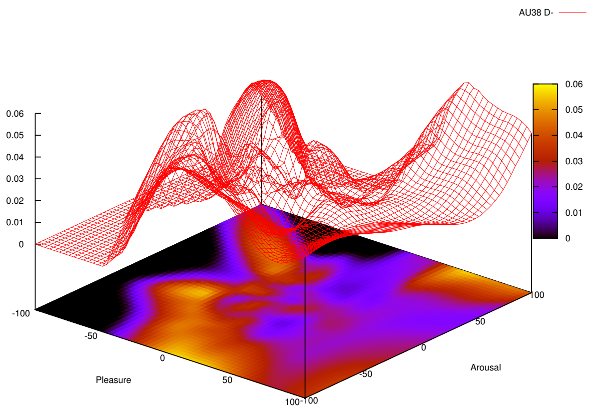
AU33



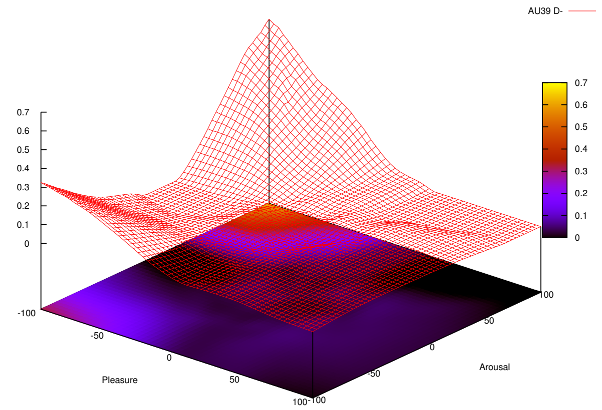
AU34



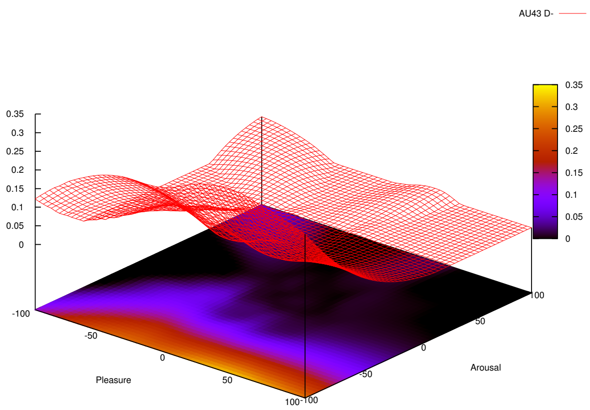
AU35



AU38

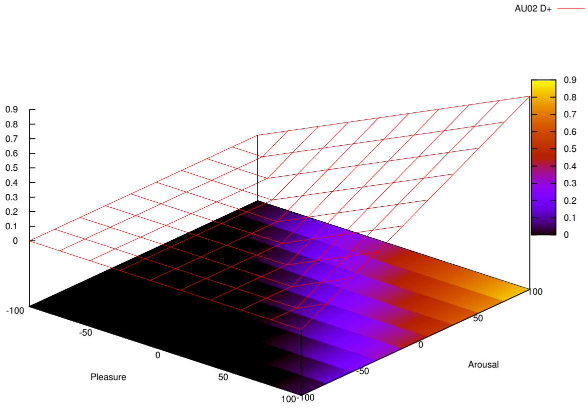


AU39

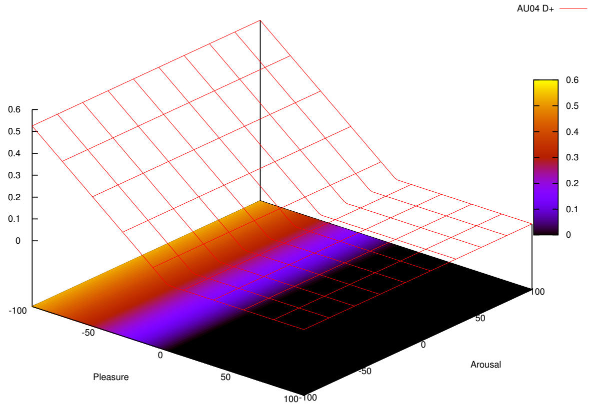


AU43

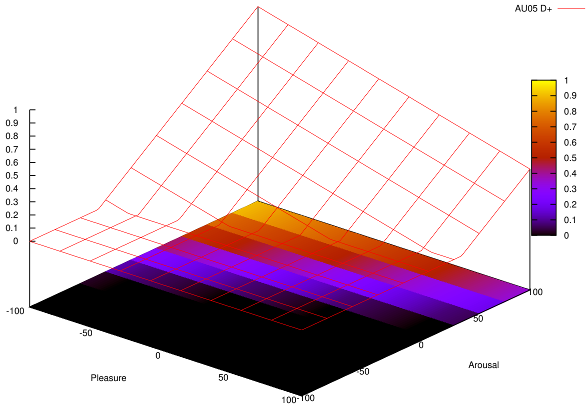
A.6 Linear Regression Planes - PA Space of Positive Dominance



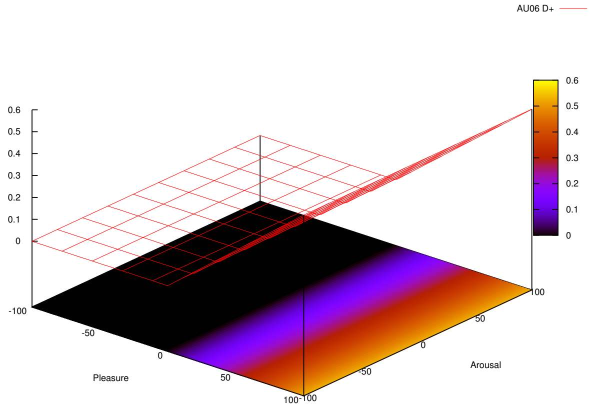
AU2



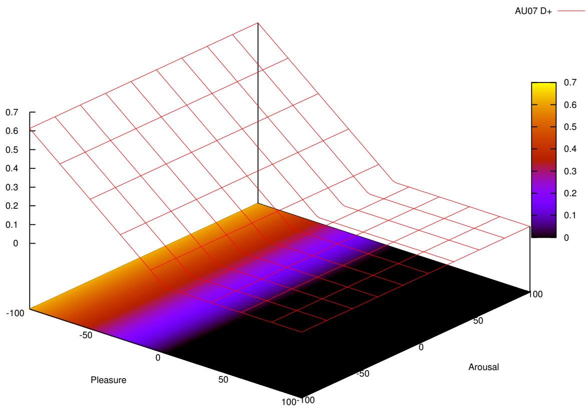
AU4



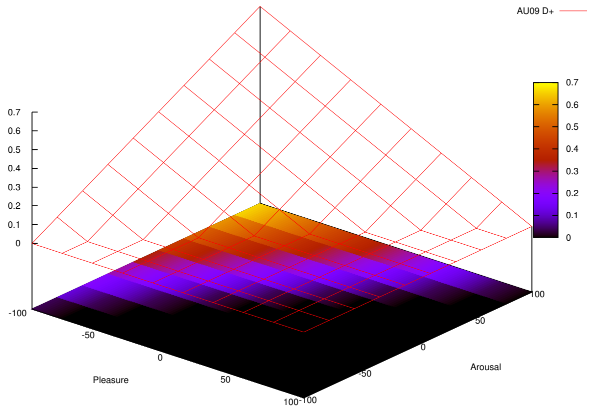
AU5



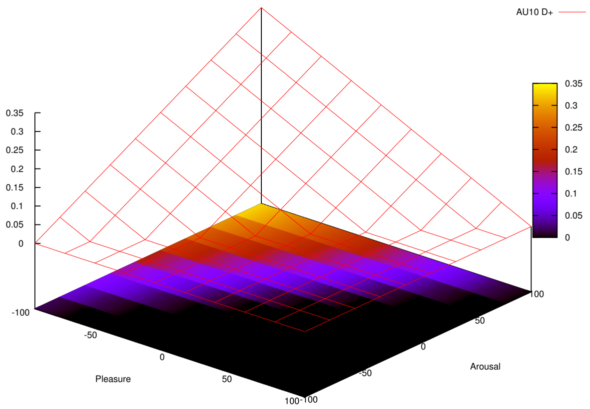
AU6



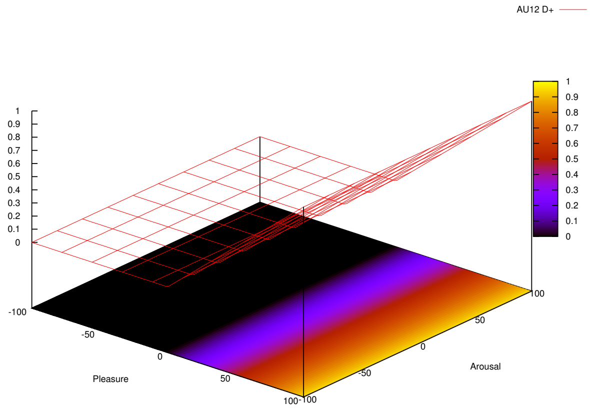
AU7



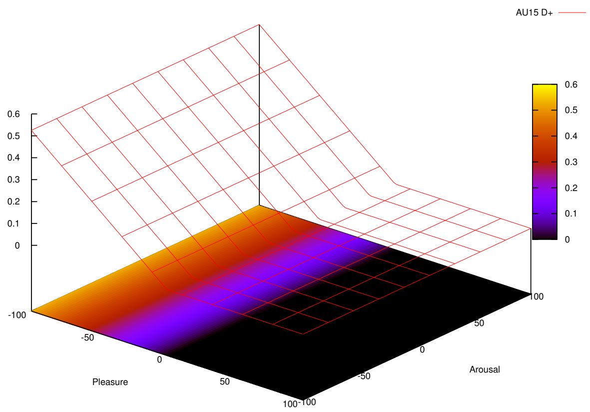
AU9



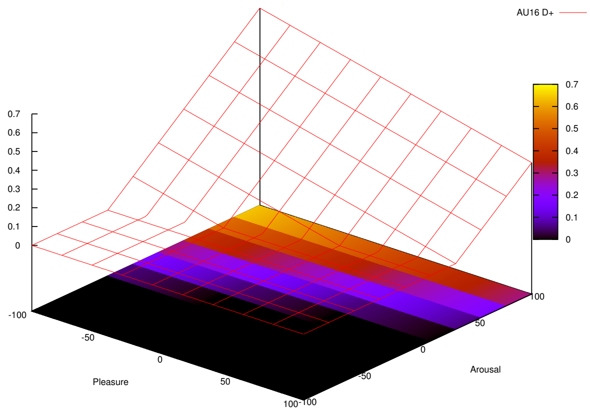
AU10



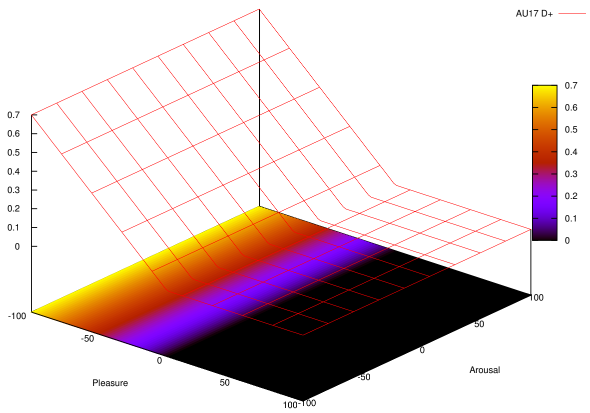
AU12



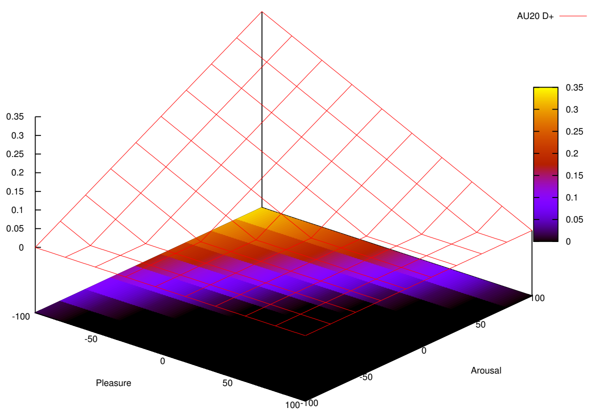
AU15



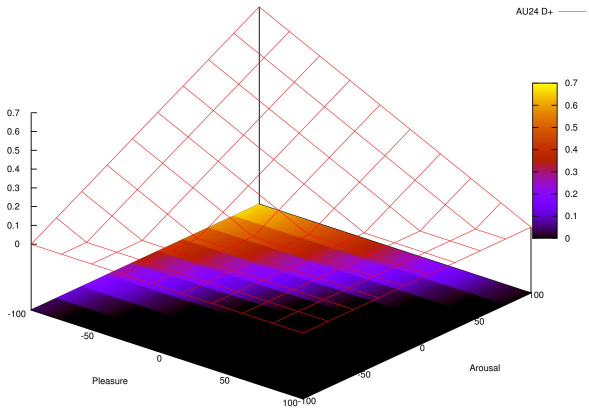
AU16



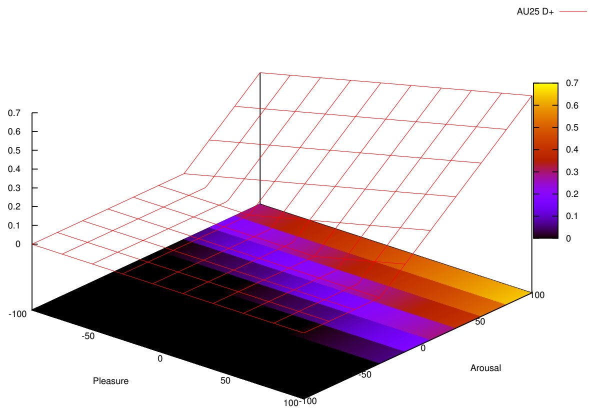
AU17



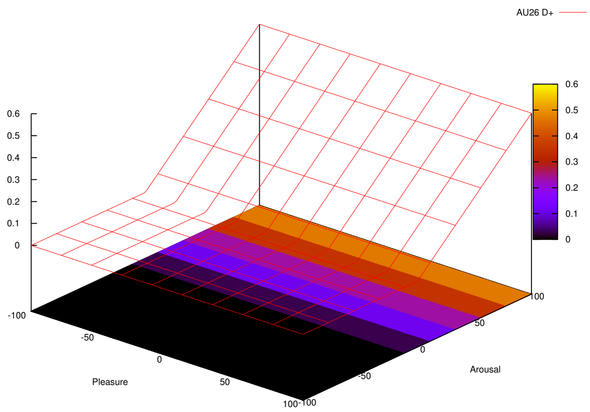
AU20



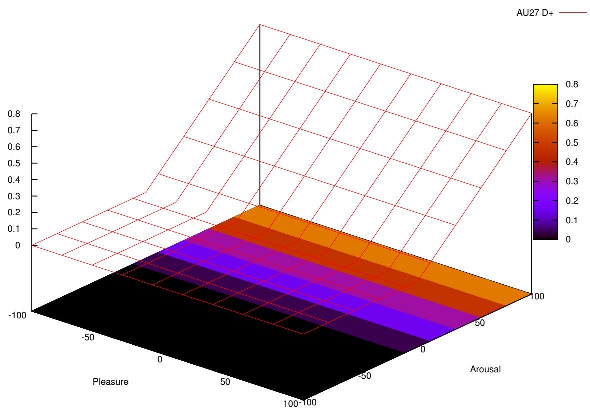
AU24



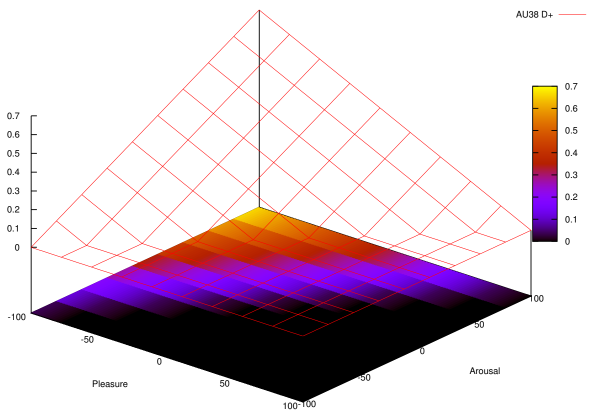
AU25



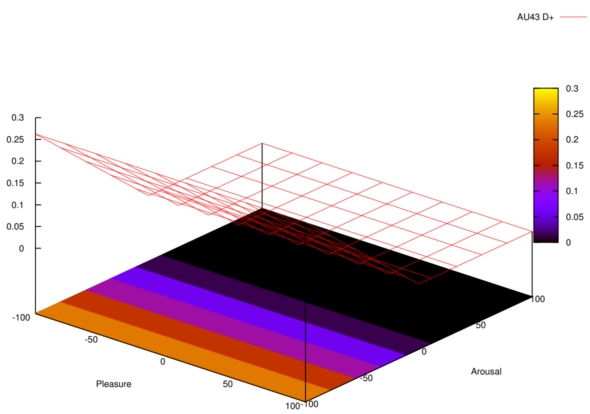
AU26



AU27

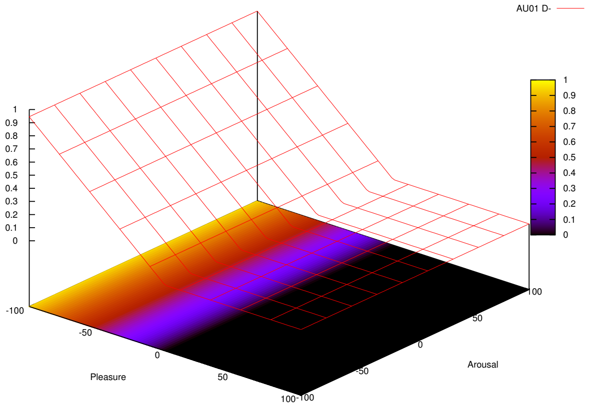


AU38

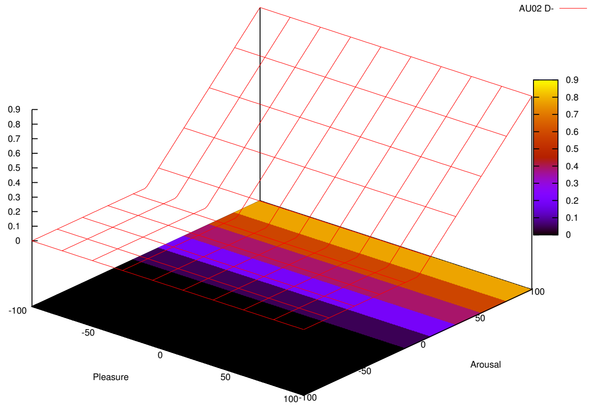


AU43

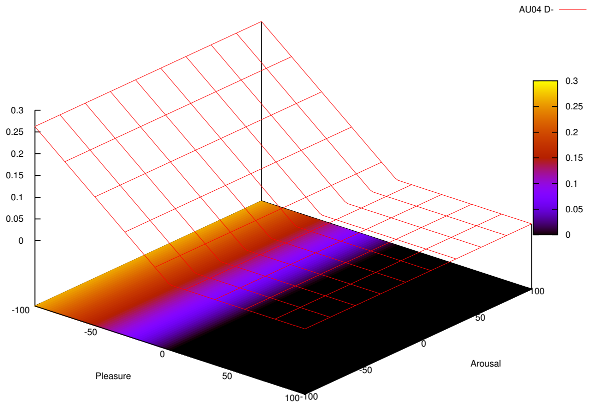
A.7 Linear Regression Planes - PA Space of Negative Dominance



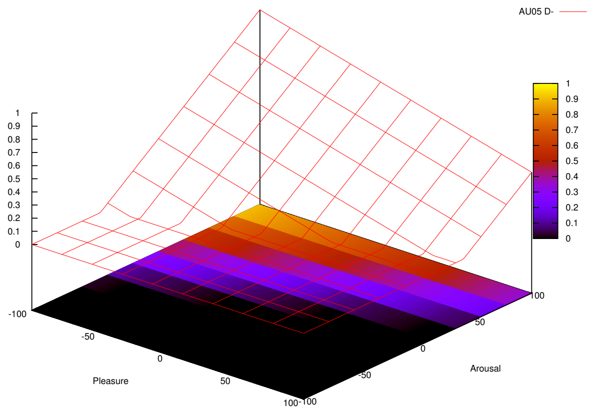
AU1



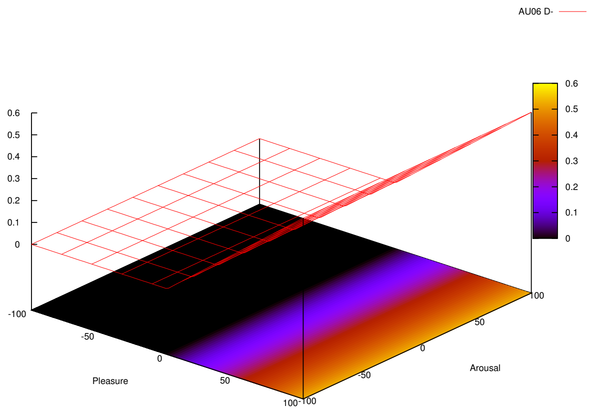
AU2



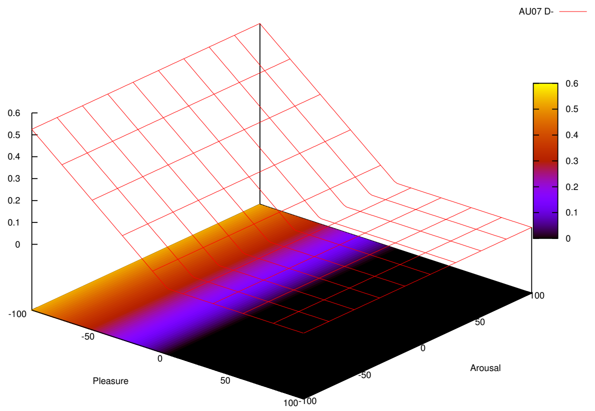
AU4



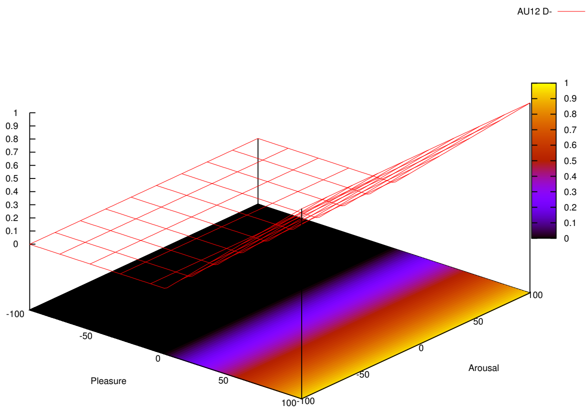
AU5



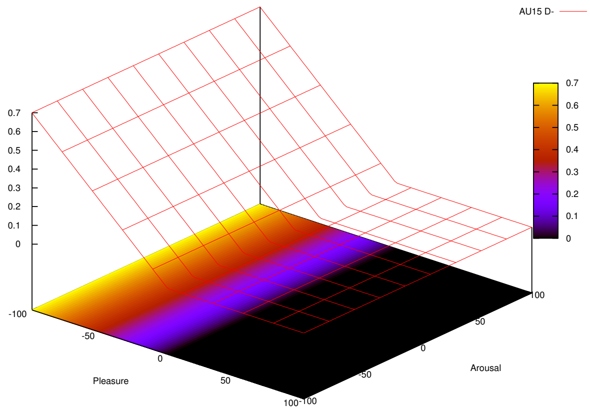
AU6



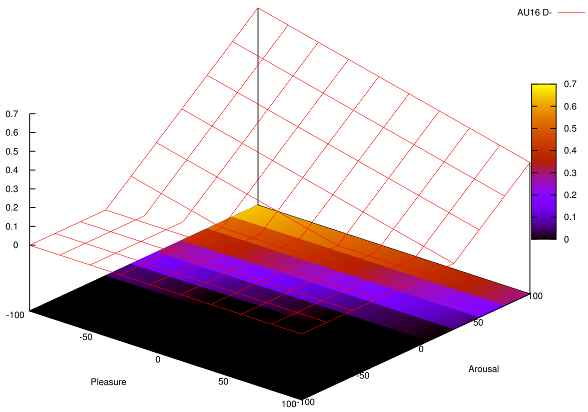
AU7



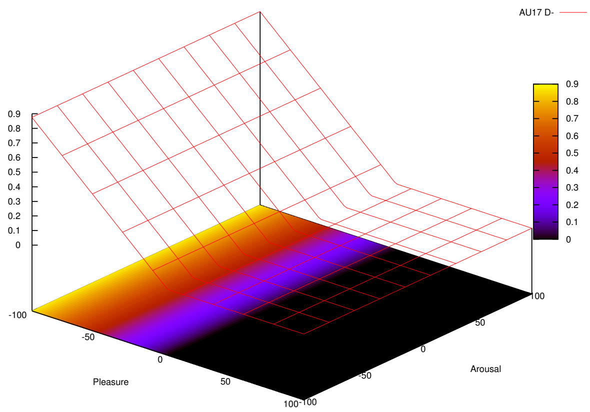
AU12



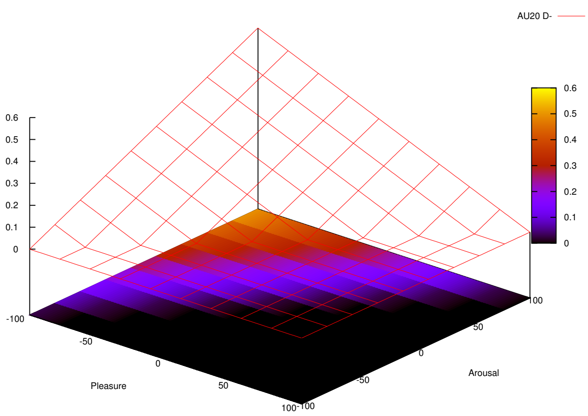
AU15



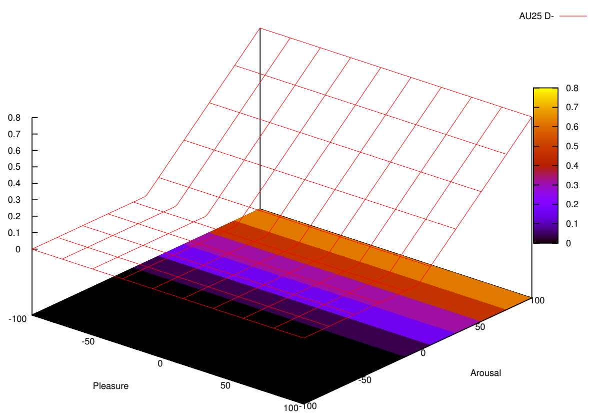
AU16



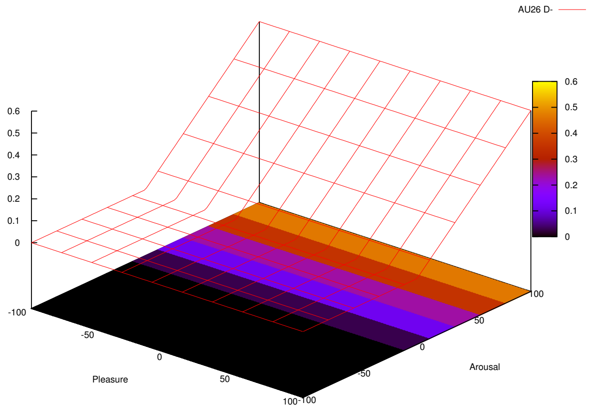
AU17



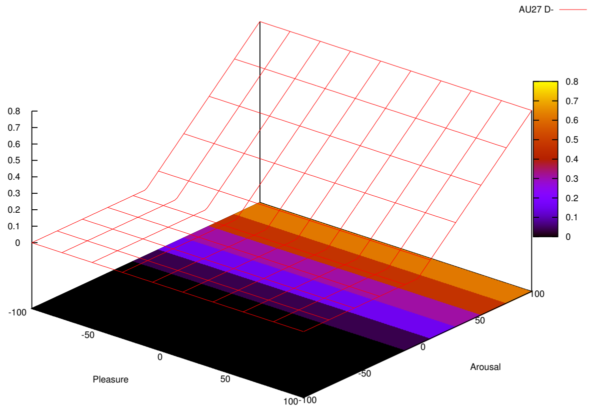
AU20



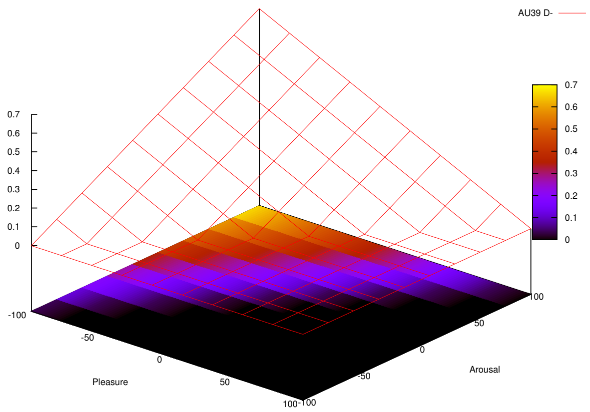
AU25



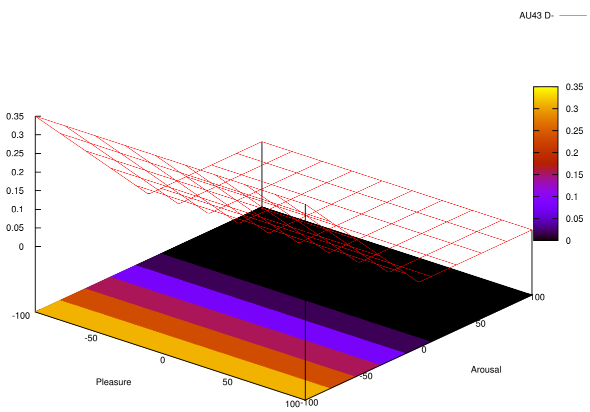
AU26



AU27



AU39



AU43