eMuu – An Embodied Emotional Character for the Ambient Intelligent Home

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof.dr. R.A. van Santen, voor een commissie aangewezen door het College voor Promoties in het openbaar te verdedigen op maandag 28 oktober 2002 om 16.00 uur.

door

Christoph Bartneck

Geboren te Bünde, Duitsland

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr. D.G. Bouwhuis

prof.dr. E.H.L. Aarts

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

Bartneck, Christoph

eMuu : an embodied emotional character for the ambient intelligent home / by Christoph Bartneck. - Eindhoven : Technische Universiteit Eindhoven, 2002. -

ISBN 90-386-1887-5

NUR 984

Subject headings: Emotional embodied characters / Ambient intelligent homes

© Christoph Bartneck

All rights reserved. Reproduction in whole or in part is prohibited without the written consent of the copyright owner.

Acknowledgements

This study was supported by:

- Stan Ackerman Institute (SAI), Eindhoven
- J.F. Schouten School for User-System Interaction Research (JFS), Eindhoven
- IPO, Center for User-System Interaction (USI), Eindhoven
- Advanced Telecommunication Research Institute International (ATR), Kyoto
- Philips Research, Eindhoven

The author would like to thank the following persons, in alphabetical order, for their contributions:

Emile Aarts, Don Bouwhuis, Magdalena Bukowska, Elmo Diederiks, Yassine Faihe, Jettie Hoonhout, Hu Jun, Yasuhiro Katagiri, Armin Kohlrausch, Michio Okada, Boris de Ruyter and Toru Takahaschi.

Contents

1

Introduction

Many companies, universities and research institutes are working on the home of the future. Besides Microsoft and IBM, Philips Research is one of the key players and made recently a considerable step forward by opening a first prototype, called “Home Lab” (Aarts, 2002). One of the central concepts in the idea for the HomeLab is Ambient Intelligence (see box 1) (Aarts, Harwig, & Schuurmans, 2001). One of the key components of ambient intelligence, as described above, is the natural interaction between the home and the user. The most natural human interaction is speech and therefore the ambient intelligent home should use speech technology. Not only the TV and video recorder might be voice controlled, but also most other electronic devices such as ovens, the lights and heating. The house itself might become voice controlled.

One of the difficulties of communication between two partners, such as the ambient intelligent home and its user, through speech is that none of them can ever be sure that the information transmitted will be perceived successfully and understood correctly. This difficulty can be termed the “uncertainty principle” of communication (Bouwhuis, 1991). Moreover, the receiver of the original message cannot be certain that his or her answer is received successfully and understood correctly by the original sender either. The uncertainty is mutual and can be described as the "recursive uncertainty principle" (Bouwhuis, 1991).

> In the near future our homes will have a distributed network of intelligent devices that provides us with information, communication, and entertainment. Furthermore, these systems will adapt themselves to the user and even anticipate on user needs. These consumer systems will differ substantially from contemporary equipment through their appearance in peoples environments, and through the way users interact with them. Ambient Intelligence is the term that Philips uses to denote this new paradigm for in-home computing and entertainment. Salient features of this new concept are ubiquitous computing, natural interaction, and intelligence. Recent developments in technology, the Internet, the consumer electronics market, and social developments indicate that this dream might become reality soon. First prototypes of ambient intelligent home systems have been developed, but the realization of true ambient intelligence calls for much additional research of multidisciplinary teams consisting of technologists, designers, and human behaviour scientists.

Box 1: Definition of Ambient Intelligence (Aarts et al., 2001).

Nevertheless, in many situations the uncertainty of human communication is reduced to a minimum by employing various means to ensure intelligibility and interpretability of the messages. The loudness of the speech and its rate, for example, may be adjusted. Furthermore, the speaker may repeat utterances, paraphrase the utterance or provide additional context information. The meaning of a certain message might also be amplified by employing gestures, body posture and emotional facial expressions. The ambient intelligent home itself, however, does not have an anthropomorphic body to do so.

If the means to ensure intelligibility and interpretability of the messages, as described above, do not result in a successful communication, the speakers have to resort to feedback. Dialogue control acts form the majority of this feedback (Bunt, 1989). Dialogue control acts do not transmit information itself but are concerned with the status of the information transmission. They may consist of additional utterances, such as "What did you say?", or body language, such as facial expression, glance and gestures. These acts may account for up to 50% of all utterances in spoken language and therefore it appears reasonable to distinguish in a dialogue the content and the protocol. The content is the information to be transferred and the protocol the ways in which the transmission is executed (Bouwhuis, 1991).

To be able to employ the full range of dialogue control acts and to amplify the meaning of a message with emotional expression the ambient intelligent home needs an anthropomorphic entity to execute facial expressions and gestures. Instead of talking to an empty room and receiving voice messages through a central speaker system the user would talk to one central anthropomorphic entity. This entity could employ the dialogue control acts simultaneously to the speech acts, so that the user has natural and constant feedback about the status of the information transmission. The visual dialogue control acts create the awareness, or at least the impression, that the entity is listening and fully involved in the communication process. Kellner et al. (2000) suggested that a home character using speech technology would indeed be the most appropriate interaction style for the ambient intelligent home.

The visual feedback might be particularly useful for hearing impaired users, since they have problems or are not able to receive acoustic feedback of the status of the communication. If the character displayed correct lip movements while talking, the hearing impaired user might even be able to lip-read the message.

Such an anthropomorphic entity can have many forms and names. The most familiar is obviously another human being, but artificial anthropomorphic entities also exist, such as avatars, robots or the well-known NEC PaPeRo (see Figure 15).

Since the terminology of the artificial anthropomorphic entities research is diverse it is important to first define the terminology of this study. The most important term is "character". Webster defines a character as "one of the persons of a drama, novel or film". Considering that a computer can be viewed as a form of theatre (Laurel, 1991) this term seems suitable to describe the overall category of computer controlled artificial social entities, such as on-screen animated figures, avatars and robots. A "home character" is therefore a character that resides at the home of the user.

The home character will display intentional behaviour and therefore the users will assume it plays a certain role. A father, for example, shows intentional behaviour of raising his kids. Not all roles are socially acceptable for the home environment. Most people, for example, would prefer not to live together with their boss from work. The most suitable role for a home character in the ambient intelligent home appears to be the one of a butler, because butlers have played an accepted role in the home for centuries. In the same way that a certain behaviour is expected of the role of the father, the butler character will also raise certain expectations for its behaviour, such as loyalty, discretion and general helpfulness. The fact that the home character is not a human will not change these expectations, because humans tend to treat computers in the same way as they treat other humans (Nass & Reeves, 1996).

A character playing its specific role is often referred to as a persona. To avoid possible confusions with the area of personality research, this study continues to us the term “character” instead of “persona” to describe an artificial anthropomorphic entity for the home, even though this term might not completely describe all aspects of this entity.

On a higher level the messages of human communication consists of four features: facts, relationship, appeal and self-revelation (Schulz, 1981). The home character has to consider all four features to successfully play its role as a butler.

The facts feature contains the content of the message and the relationship feature contains the sender’s opinion of the receiver and in what relationship they are. The appeal feature contains the information what the sender wants the receiver to do and the self-revelation feature contains information about the state of the sender in particular his or her emotional state. The relationship, appeal and self-revelation features are not communicated through *what* is said but through *how* it is said.

The sender encodes all four features into his or her message and the receiver interprets the four features of the perceived message. Only in a successful communication are the sent features of a message similar to the interpreted feature. A mismatch between the sent and the interpreted features of the message can explain many failures of communication.

Consider this classic example: A husband and wife drive together in a car and the wife steers the car. They approach a traffic light. The man says: ”The traffic light is green.” and the woman replies: “Am I driving or you?”. The wife interpreted the message of her husband as: The traffic light is green (fact), I am a better driver than you (self-revelation), I want you to drive faster (appeal) and I am your teacher (relationship). This might have not been the message that the husband sent. An alternative interpretation could be: The traffic light is green (fact), I am paying attention to the traffic to help you (self-revelation), and we are on an equal level (relationship). This simple example demonstrates the complexity of human conversations that the home character should in principle be able to manage.

Another key component of the ambient intelligent home as described in Box 1 is that the devices and computers themselves will move into the background, becoming invisible to the user. The advantage of a home character would be that the user could address all these devices and computers through one central and visible entity that would execute the given commands.

A considerable number of these devices in the home are for entertainment purposes, such as TVs, stereos and video recorders. Since several studies (Doyle, 1999; Koda, 1996; Rizzo, 1999) suggested that the entertainment area is a suitable application domain for characters, it is likely that these devices will be the first to take advantage of a home character.

1.1. Scenarios

The following scenarios describe how the user might interact with a home character. They are archetypical examples of how the character might behave and help to derive requirements for the design of the character in Chapter 4. The text in the brackets [] denotes the emotional facial expressions of the character James.

1.1.1. "Internet download" scenario

Helena: 8.30 in the morning. That late already? I am going to be late for work again. And I still need to download all the music, videos and programs for tonight's karaoke party.

Helena walks to the nearest screen. Upon her approach James, her home character, appears on the screen.

Helena: James, I need some files from the internet. Can you please download the following songs: Queen - We will rock you, ABBA - Dancing Queen, Whitney Houston - One moment in time.

The list goes on and on.

James: Helena, [smile] I see that you want to download a lot of greatest hits songs, shall I also download Michael Jackson - Thriller?

Helena: Yes, that would be great. Henry is always so funny when he tries to do the moonwalk.

James: When do you need the files?

Helena: For tonight's karaoke party, lets say 18.00?

James: 18.00? That will be tight [light sad]. The network traffic is usually very high during the day. Considering the available bandwidth I think I will not be able to download Queen - We will rock you and Wham - Wake me up before you go.

Helena: Yeah, they are promising sufficient bandwidth for 5 years already, but whenever you really need some files it is never enough. And I really like George Michael!

James: Okay Helena. If I download George Michael then I cannot download Whitney Houston. Is that okay? [light smile]

Helena: Hmm, alright then. But hey, my favourite movie Titanic is not on the download list.

James: That movie takes really long to download [very light anger]. I would not be able to download even half of your selected songs. Do you also need it for tonight?

Helena: No, I wanted that movie for tomorrow. I guess it can wait a little bit. But keep it on the list!

James: Okay, I will download it as soon as this session is over [light smile]. Do you want anything else?

Helena: No, that would be it. I have to run, otherwise I will be late for work. See you tonight

James: Bye.

Helena rushes off to work and James starts to search and download the songs from the most affordable online music vendor. Towards 17.30 he send an email to Helena.

Helena: Oh, it's from JAMES. Download is ready. Great! Now I only need to pick up the drinks and some snacks and the party can start.

1.1.2. "Dinner" scenario

Helena approaches the door to her flat and James recognizes her. He unlocks the door, switches on the light in the corridor and welcomes her with a big smile.

James: Hello Helena, how was your day? [strong smile]

Helena: Don't ask! I had no time to get lunch and I am really hungry.[sad] So tell me, what's on my diet plan tonight?

James: You did not eat anything for lunch? [light sad] You know that you should have eaten at least some fruit.

Helena: But I had no time! Please, tell me what's on for dinner? I'm starving.

James: Well, considering that you had no lunch you could eat the following combination and still stay within your daily calorie limit. [light smile]

A list with several starters, main dishes and deserts shows up.

Helena: I had a terrible day. I need some chocolate for desert.

James: Okay, then you can have a mixed salad as a starter and some vegetables with rice for the main dish. [light smile]

Helena: Doesn't sound very exciting. Can I have a soup for the starter?

James: Only if you accept some light yoghurt for dessert. [light sad]

Helena: Alright, I know I shouldn't eat chocolate. What a day… *sigh*

James: Shall I playback the Titanic movie I downloaded for you today?

Helena: Yeah, that would be nice. I need to relax. Can you display the recipes?

James: Of course. [light smile]

Helena starts preparing the food with the recipe that JAMES displays on a screen. Titanic starts playing as soon as Helena sits on the couch in front of the TV.

1.2. Focus of this thesis

These two scenarios demonstrate that the design of a home character has to consider many aspects. All of the technical aspects of the ambient intelligent home such as the sensing of the user, the control of all the home's function and the adaptation of the home's behaviour to the user are not in the focus of this study. Instead this study focuses on the design and evaluation of a home character that functions as the interface between the user and the home. In particular it focuses on the effect of the character's emotional expressiveness and its embodiment on its usability.

As mentioned above, body language is an important component for human-character interaction. This study focuses on emotional facial expressions because they provide natural and continuous feedback to the user about the status of the communication and therefore play an important role in the design of characters (Picard, 1997b; Koda, 1996).

Clearly, a home character is not alive but based on machines, such as the computer that controls it or the screens that display it. Therefore the question arises whether machines can express emotions and if they can do so convincingly.

Machines are of course able to convey emotions. Almost all recent experiments that tested emotional expressions presented their stimuli to the participants using machines, such as speakers, tape recorders and computers. Only few experiments used actors performing live in front of the participants, because the possible inconsistent presentations of emotions by the actor introduces an additional source of variance in the experiment. By using the emotional expressions of humans, machines can easily convey emotions. However, all emotional expressions of machines are by definition abstractions of human expressions. Even actors expressing emotions to each other on the TV are not real people but mediated characters (Nass et al., 1996) as described in the introduction of this chapter and therefore their expressions are also only abstractions. The more abstract an expression is, the more likely it is that the user will perceive it as an expression of a machine. However, machines do not have their own non-human emotions or the ability to express them. Humans would also not be able to understand non-human emotions and their expressions without additional learning. This additional learning is not necessary for human emotions and their expressions, because human-human interaction already trained the user. Emotional expressions are learned from a very young age and probably belong to the earliest communicative devices of infants (Bornstein, 1982). Therefore, machines should mimic human expressions and their abstractions to communicate emotions to the user.

However, it is not obvious that the user will perceive the emotional expressions of machines as being convincing. On the one hand one could argue that humans might think that machines do not have emotions and therefore all expressions would be perceived as fake. On the other hand, they might think that machines, in particular computers, do not have bad intentions, knowledge about the user or reasons to lie and therefore the expressions would be perceived as convincing. Computers have a very high creditability and even able to fool lawyers (Kluger & Adler, 1993; Dijkstra, 1998).

The embodiment is another important aspect of the home character since it provides the physical basis for its emotional expression. The two types frequently found in literature and products are screen characters and robotic characters. The perhaps best known and possibly most disliked screen character is Microsoft's Paperclip (see Figure 7), a little help agent for the Microsoft Office software. The other type are robots, such as Kismet (Breazeal, 1999) or Papero (NEC, 2001) (see Figure 10 and Figure 15), which are not yet commercially available. Giving a character a physical body moves it closer to a real life form. It takes the character out of the virtual screen world and introduces it into the real physical world. Little is known about how the different embodiments influence the human-character interaction. Several characters have been proposed (see Chapter 3), but no comparative study is available yet.

The largest effect that the emotional expressiveness and the embodiment of the character might have on the interaction between a user and the home may be found in the enjoyability of the interaction. Enjoyability is one of the subcategories of user satisfaction, which is again a subcategory of usability, as defined in ISO 9241. The other categories of usability, efficiency and effectiveness, might be affected as well, but the enjoyability aspect of user satisfaction appears to be the most promising one since even the most critical studies (Hollnagel, 1999) about the benefits of emotional characters for human-machine interaction consider the increase of enjoyment as a possible contribution.

This leads to the three main research questions of this study:

1. How convincing are the emotional expressions of machines?
2. Will the user perceive the interaction with a character that uses emotional expression more enjoyable than with a character that does not use emotional expressions?
3. Will the user perceive the interaction with a robotic character more enjoyable than with a screen character?

1.3. Outline of the thesis

The ambient intelligent home requires new and socially acceptable interfaces as described in the introduction of this chapter. An embodied interface character, referred to as the "Home Character", might suit the new challenges, such as providing a social entity for speech dialogues and offering natural feedback to the user. This thesis investigates the influence of the embodiment and the emotional expressiveness of a home character on the enjoyability of the interaction.

In Chapter 2 the convincingness of emotional expressions of machines is investigated experimentally. The results, in combination with a literature review on interface characters in Chapter 3, are integrated into the requirements of a home character described in Chapter 4. Chapter 5 describes the implementation of the home character based on the requirements in the previous Chapter. In Chapter 6 the influence of the embodiment and emotional expressiveness of the home character on the enjoyability of the interaction is investigated experimentally. The result of the experiments, their consequences for the design of characters and suggestions for further research directions are discussed in the final Chapter 7.

2

Emotional expressions of machines

This chapter presents a literature review on emotional expressions of humans and machines, as well as an experiment, which evaluated a model for convincingness and the influence of several factors on the convincingness of emotional expressions.

The literature review investigated how the emotional expressions of humans and machines through speech, music and body language are perceived by humans. It summarizes the quality and quantity of their parameters and successful examples of synthesis. Next, a model for the convincingness of emotional expressions, based on Fogg and Hsiang Tseng (Fogg & Hsiang, 1999), was developed and tested. Furthermore, it was experimentally investigated if the type of emotion (happiness, sadness, anger, surprise, fear and disgust), knowledge about the source (human or machine), the level of abstraction (natural face, computer rendered face and matrix face) and medium of presentation (visual, audio/visual, audio) of an emotional expression influences its convincingness and distinctness.

2.1. Introduction

Most of the previous studies and implementations described below concentrated their evaluations on the distinctness of the emotional expression (Bartneck, 2000), which they measured through the recognition accuracy of the participants. The recognition accuracy measured the correct identifications of emotional expressions. This study expanded the evaluation with another important attribute: the convincingness of the emotional expression. Moreover, this study investigates possible differences in the perception of emotional expressions of humans and machines. Are emotional expressions of machines as convincing as emotional expressions of humans? What factors influence convincingness? These questions are particular important for the design of emotional home character.

Two main viewpoints to describe emotions can frequently be found in the literature. One considers emotions as discrete categories (Ekman, 1973; Izard, 1977; Plutchik, 1980), such as happiness, sadness or anger. The other characterizes emotions as points in a multidimensional space (Schlossberg, 1954; Osgood, Suci, & Tannenbaum, 1957; Russel, 1979). Arousal and valence could, for example, define such a two dimensional space. The two viewpoints are not as different as they might seem. The discrete categories, for example, can be described as clusters of points in the dimensional approach. Frijda (Frijda, 1986) argued that on the one hand the number of dimensions may prove to be large (Nowlis, 1966; Frijda, 1969; Smith & Ellsworth, 1985; Schiano, Ehrlich, Rahardja, & Sheridan, 2000), which moves the dimensional viewpoint toward the categorical. On the other hand, the discrete emotions vary along common dimensions (Izard, 1977) and can be ordered in terms of similarities and as pairs of opposites (Plutchik, 1980). This pushes the categorical viewpoint towards the dimensional.

For reasons of experimental convenience, this study takes the categorical viewpoint on emotions. Many studies (Ekman, Friesen, & Ellsworth, 1972) used the categories happiness, surprise, fear, anger, sadness and disgust. This study takes the same categories as a point of departure to take advantage of this well established theoretical framework.

2.2. Literature review of emotional expressions

2.2.1. The importance of emotions

The importance of emotions has been analysed in numerous studies (Frijda, 1986; Lazarus, 1991), including several ones on the role of emotions in cognitive processes (Norman, 1981). Prior studies found that emotions play an important role in problem solving (Feist, 1994) and decision making (Barnes & Thagard, 1996) by providing information on the emotional desirability of the options available, therefore reducing and limiting reasoning to only those that induce positive feelings. Emotions also guide actions and control resources (Oatley & Jenkins, 1996). Emotions play an important role in the design of interfaces (Picard, 1997b; Nielsen, 1994) because people interact with machines as if they were social actors (Nass et al., 1996). It is not unusual, for example, to hear people yelling at their computer just as if it might feel sorry and change its behaviour (Picard, 1997a).

Emotions in characters might not only improve the interaction with them but could also be beneficial for the characters themselves. Emotions are crucial for humans to manage goals and actions in a complex and rapidly changing world. Therefore one may assume that characters that face the same world as humans and have similar goals might benefit from an emotional system as well (Sloman & Croucher, 1981; Gadanho & Hallam, 1998; Botelho & Coelho, 1998). That would hold true, for instance, for autonomous robotic characters managing multiple goals in the real world. However, there are still several qualities of humans, such as reproduction, that machines have not yet imitated. Today's toaster does not even have a self preservation instinct. It does not care if it functions properly or if it gets crushed with a hammer.

2.2.2. Emotional expressions of humans

Humans express their emotions through actions, which can be perceived through the visual, auditory and tactile modality. Body language, such as facial expressions and gestures, are the main elements perceived by the visual modality. Speech and music are the main elements perceived by the auditory modality. Actions perceived by the tactile modality, for example petting and punching, are, due to their low relevance for human-computer interaction, beyond the scope of this study.

Expressing emotions is a natural act for humans. The ingenuous ease of it contrasts with the difficulty of describing it scientifically. Furthermore, the capability to express emotions can be refined through the performing arts, such as acting and singing. All music students spend hours with their teachers learning to play music not just as it is written in the score, but also in the appropriate emotion. Even students who have learned to do it are usually still unable to explain how they do it.

Many studies have been performed to find out how humans express emotions. In the following paragraphs some of their results are summarized to provide the reader with a starting point for further literature review. They focus on speech, music and body language. These results were applied as the base for the design of emotional expressions for the characters described in Chapter 5.

Speech

Speech is a powerful method to communicate emotions. If your friend, for example, does not show up for a meeting with you, you can express your anger through a telephone call. You are restricted to speech, but your friend will most likely understand the emotional state you are in. You may convey your emotional state by the content of your message, such as "I am angry" but also through the sound of your voice.

The most influential parameters for non-content emotional expressions in speech are pitch (level, range and variability), tempo and loudness. Many other studies used these parameters and Scherer summarized their results (Scherer, 1979). Murray and Arnott (Murray & Arnott, 1992) conclude in their literature review that in general, the vocal effects caused by particular emotions are consistent between authors and between the different studies carried out, with only minor differences being apparent. However, Murray's and Arnott's quantification of these speech parameters are rather vague. A more concrete approach is the Affect Generator (Cahn, 1990), a software tool to synthesize emotional speech. It allows settings on a scale from -10 to +10 for each of its 19 parameters, such as pitch range and speech rate. Zero represents the parameter influences to obtain a neutral effect, while -10 and +10, the minimum and maximum influence, respectively. Unfortunately, this scale does not translate to results of other studies as it is only meaningful within this software tool.

A more general approach is to quantify parameters in percentage of the neutral setting. Mozziconacci quantified this way optimal pitch and tempo settings for certain emotions (Mozziconacci, 1998). However, calibrating the neutral setting remains difficult.

Music

Music is a difficult method to communicate emotions because culture (Davis, 1978; Crowder, 1984), skills of the performer (Bresin & Friberg, 1999; Juslin, 1997a) and age of the listener influence the perception. The widely accepted association between mode (major and minor) and emotion (happy and sad) develops, for example, during childhood (Cunningham & Sterling, 1988; Gerardi & Gerken, 1995; Kastner & Crowder, 1990).

Scherer and Oshinsky (Scherer & Oshinsky, 1977) demonstrated that 66% to 75% of the variance in the emotional attributes of music can be explained by manipulation of amplitude, pitch (level, variation and contour), tempo, envelope and filtration. Furthermore, they argue that their results overlap with the findings in emotional expressions in speech. Juslin (Juslin, 1997b) summarized expressive principles which he obtained by a series of studies using several different instruments, performers and melodies.

Body language

Pantomimes use only facial expressions and bodily movements to express emotions. Their success is amazing considering the abstract vocabulary of movements available to them. The main components of body language are facial expressions, gestures and body movement. There is no difference in the relative importance of the components of body language (Ekman, Friesen, O'Sullivan, & Scherer, 1980) for their distinctness. Emotions can be communicated through all of them equally well.

• *Facial expression*

Expressing emotions through the face is so natural for humans that it takes a considerable amount of effort to mask them (Ekman et al., 1972). Keeping a "poker face" in a critical situation is difficult. The main components used to express emotions are mouth, cheeks, eyes, eyebrows and forehead. Ekman and Frieser (Ekman & Frieser, 1976) compiled archetypes of emotional expressions in the human face. Humans do not need high quality photos or photo-realistic computer renderings to perceive emotions in facial expressions. The study of Etcoff and Magee (Etcoff & Magee, 1992) used drawings of the human face, generated by the caricature generator (Brennan, 1985). The drawing consisted of only 37 lines, but the participants were still able to identify the emotional expressions accurately.

• *Gesture*

90% of the gestures only occur during speech (McNeill, 1992). They convey some information, but they are not richly informative and the information conveyed is largely redundant in the presence of speech (Krauss, Morrel-Samuels, & Colasante, 1991). Still, people pay attention to them (Nobe, Hayamizu, Hasegawa, & Takahashi, 1997) and gestures certainly make speech more lively. An easy and precise vocabulary, such as notes for music, is, due to its variance and inconsistency, not available for gestures. However, McNeill (McNeill, 1992) grouped gestures into categories, such as Iconic, Metaphoric, Deictics and Beats.

• *Body movement*

Most of the descriptive studies on emotional body movement are informal (Frijda, 1986). Table 1 summarizes Frijda's analyses:

Table 1: Description of body movements for several emotions. From (Frijda, 1986)

Emotion	Body movement
Fear	Forceful eye closure or staring at source, frowning by drawing the eyebrows together, bending the head, hunching the shoulders, bending the trunk and knees
Surprise	Widening of the eyes, brief suspension of breating, general loss of muscle tone, mouth falls open
Anger	Teeth bared, fierce glance (fixed stare, eyes slightly widened, eyebrows contracted) , clenching fists (optional), lips compressed,Movements are vigorous and brisk, body tense
Sadness	Depressed corners of the mouth, lowered muscle tone, turning inward, weeping (optional)
Happiness	High frequency of unfounded and goalless changes in direction and the preponderance of movements orthogonal to the direction of locomotion, smiling, laughing (optional)

2.2.3. Emotional expressions of machines

Many companies and researchers are working on the synthesis of emotional expressions. The quality of synthesized facial expression is high, which is demonstrated by the computer animated movies of Pixar (Pixar, 1998). They are not only able to create convincing expressions, but also to animate them. A ready to use tool for synthesis and animation of facial expression is the CSLU Toolkit (CSLU, 1999). It is software for speech recognition and synthesis, which includes an animated character called Baldi. Massaro (Massaro, 1998) showed that humans perceive Baldi's emotional facial expressions accurately. The results of this study support his findings.

The quality of synthesized speech is far behind compared to the developments in synthesized facial expression and body language, because speech is a far more complex and difficult to control medium of communication. Toy Story and all other computer-animated movies up to this point are good examples for this. They all successfully used computer-generated characters, but they all fall back to real actors for the voices. The most promising synthesis of emotions in speech is the Affect Generator by Cahn (Cahn, 1990) mentioned above. She successfully applied 17 parameters, which resulted in a recognition accuracy of 78.7%.

The Director Musices (Bresin et al., 1999) is a promising synthesis program for emotional expression in music performance. It is a rule-based software tool for automatic music performances. By altering about 20 rules, such as phrasing and intonation, they have been able to reach a recognition accuracy of 64% at a 14% chance level.

A promising synthesis of body language and speech is the work of several researchers at the Department of Computer & Information Science at the University of Pennsylvania (Cassell et al., 1998). They implemented a system which automatically generates and animates conversations between multiple human-like agents with appropriate and synchronized speech, intonation, facial expression and hand gestures.

Figure 1: An emotional expressive car.

Already today products which express emotions are available. Sony's entertainment robot "Aibo" (Sony, 1999) is able to express six emotions and their blends (see Figure 11). Even a car, named Pod, that expresses emotions (Toyota, 2002) has been developed (see Figure 1). The horse shoe shaped lights that dominate the front view of the car radiate specific colours according to the mood of the vehicle. They are orange when the car is happy, for example when the user approaches Pod. They are dark blue when it is sad such as when it has run out of fuel. When it registers sharp braking or changing of direction they shine red with anger.

Therefore, it is uninteresting for this study to ask if machines can express emotion. More important is the question if there is a difference in the user's perception of emotions expressed by either a machine or a human and what attributes of the emotional expression are most important for the user's perception.

2.3. The experiment

This study considers *convincingness* as a synonym for *believability*. This term is used to prevent confusion with Fogg's and Hsiang Tseng's (Fogg et al., 1999) model of computer believability. They described a model of *believability* in which the perceived *trustworthiness* and *expertise* of a system predicts its *believability*. They defined *trustworthiness* as the perceived goodness or morality of the system and *expertise* as the perceived knowledge and skill of the system.

Their model was used as the basis for this study's model of *convincingness* of emotional expression. However, applying their definition of *expertise* to emotional expressions is not easy. The perceived appropriateness of emotional expressions was considered as a measure for emotional *expertise* of the system. A system that displays the right emotion at the right time, which is appropriate, has *expertise* in the field of emotional expressions. Furthermore, this study assumes that the *intensity* and *distinctness* of an emotional expression has influence on its *convincingness*. A big smile, for example, is likely to be perceived as more convincing than an indifferent rise of the eyebrow. This leads us to a first model of *convincingness*, in which the perceived *expertise*, *trustworthiness*, *intensity* and *distinctness* of an emotional expression predicts its *convincingness* (see Figure 2).

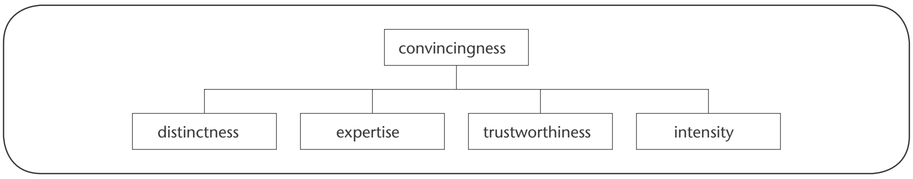

Figure 2: Model of convincingness.

To judge *trustworthiness* and emotional *expertise*, it is necessary to know the context in which the specific emotion is expressed. In real life, context information is always available. The jamming of an emotional CD player, for example, would be associated with its sad emotional expression. For this experiment the context of a simple dice game was chosen, because it required only a small amount of learning from the participants and it was easy for the participants to evaluate (see Appendix 9.3). The participants, however, did not participate in the dice game because their own emotional state might influence their perception. Therefore, it was decided that they only had to observe the game.

The participants judged the emotional expression of one player, which was shown on a computer screen (see Figure 4). The participants were told that the opponent of this player was sitting behind a wall, which was shown on the same computer screen. Thus the opponent was invisible to the participants (see Figure 4). This setup ensured that the participants could not sympathize with one player, due to the gender, attractiveness or type (human or machine). Moreover, by focusing on one player the participants did not need to constantly revalue the situation from opposing points of view. A certain result in the game would be an advantage for one player and naturally a disadvantage for the other). It was stated that none of the players bluffed or cheated since such a behaviour would be useless and against the rules of the game. None of the players would gain an advantage in the game by doing so. Keeping a poker face did not help the players to win.

The source of the expression was included as a factor, because humans might consider emotional expressions from machines less or more convincing than expressions from humans. On the one hand one could argue that humans might think that machines do not have emotions and therefore all expressions would be perceived as fake. On the other hand, they might think that machines, in particular computers, do not have intentions, knowledge about the user, or reasons to lie and therefore the expressions would be perceived as convincing.

A program was used to present the stimuli, because, unlike for humans, it is easy for a program to repeatedly produce exactly the same emotional expressions. To distinguish the two conditions for the source of the expression the players were labelled either “Human” or “Machine”. Furthermore, different background pictures were used. In the human condition a person was sitting at the table and in the machine condition a computer was placed on the table (see Figure 4). It was expected that the situation in which each expression occurred would have an influence on its perception. Therefore, a script, that was based on a pre-test (Bartneck, 2000) controlled the program and paired each stimulus with its specific situation in the game. This ensured that the appropriate emotional expression would occur in each specific situation. This script set the context.

Another factor is the type of the emotional expression. Machines need a clearly distinguishable vocabulary of those. Six emotional expressions, plus a neutral expression, provide enough complexity to act appropriately in most situations. A higher number of expressions might exceed the human capacity to process information as described in the 7 ± 2 rule (Miller, 1956). The neutral expression was shown by default and changed into an emotional expression for evaluation by the participants.

However, the expressive abilities of the machine might be limited. A mobile phone, for example, has only a small LCD display. It is impossible to present a human face in all its details on it. Therefore, it is important to test if the abstractions of an expression are convincing as well. Three levels of abstraction were tested, which were based on typical products in the area of home consumer electronics (see Table 2).

Table 2: Levels of abstraction of the emotional expressions.

Product category	Product examples	Level of abstraction
Screens	TV, Monitor, Projector	Natural human face
Screen characters	Games, Virtual newsreader	Baldi (Real time 3D face)
Small devices	Mobile phone, PDA	Matrix face (10x10 pixel)

Humans would use their own face and not an abstraction of it to express an emotion. It is highly unlikely and therefore unconvincing by default for a human to express happiness by displaying a smiling matrix face on a piece of paper instead of smiling with his or her own face. Therefore, only the *natural* abstraction level was tested in the human condition of the source. This will set the benchmark to which the machine's expressions will be compared.

Even though no single modality predominates the perception of emotions, a combination of modalities might be more convincing than each modality alone (Ekman et al., 1980). Machines, such as mobile phones or TVs, are capable of presenting multimedia expressions. To reduce the complexity of the experiment the media factor was tested only in combination with the matrix face (see Table 3). Either only the matrix face or the audio or the combination of both were presented. The participants were asked to evaluate the visual and audio stimuli as one expression if they appeared simultaneously. For practical reasons, this study focused on content free media, such as facial expressions and auditory signals.

2.3.1. Manipulation

A 2 (source) x 3 (abstraction) x 3 (media) x 6 (emotion) within participants experiment was conducted. Certain factors were limited to certain conditions (see table Table 3 and section 2.3 for an explanation). Altogether 36 conditions were tested.

Table 3: The 36 conditions of the experiment.

Source	Abstraction	Media	Emotion
Human	Natural face	Visual	happiness, sadness, anger, surprise, fear and disgust
Machine	Natural face	Visual	happiness, sadness, anger, surprise, fear and disgust
Machine	Baldi face	Visual	happiness, sadness, anger, surprise, fear and disgust
Machine	Matrix face	Visual	happiness, sadness, anger, surprise, fear and disgust
Machine	Matrix face	Audio	happiness, sadness, anger, surprise, fear and disgust
Machine	Matrix face	Audio/Visual	happiness, sadness, anger, surprise, fear and disgust

2.3.2. Measures

Convincingness, *expertise*, *trustworthiness* and *intensity* were measured by answering the following questions:

- How convincing is this expression?
- How appropriate is this expression in this situation?
- How trustworthy is this expression?
- How intense is this expression?

on a 1-7 scale (e.g. 1=very unconvincing, 7=very convincing). The *distinctness* of an expression was measured by the recognition accuracy of the participants, which consisted of a forced choice between the 7 categories.

2.3.3. The Participants

33 employees (20 male 13 female) of the Technische Universiteit Eindhoven, at the age between 21 to 61, participated in the experiment.

2.3.4. The stimuli

Three actors produced facial expressions that were photographed with a digital camera. They were asked to imagine an event in which each emotion was felt strongly. With the help of a pre-test the distinctness of their expressions were analysed and for the final experiment the expressions of the most successful actor were used. Baldi (CSLU, 1999) was used as an example for a typical Real-Time-3D Character. The quality of his expressions has been tested earlier (Etcoff et al., 1992). Professional designers created the matrix faces and the audio stimuli based on the information in Chapter 2.2. They were optimised through several iterative circles of design and evaluation prior to this study. The audio stimuli consisted of short pieces of abstract music, similar to the beeps of R2D2 in the movie Star Wars.

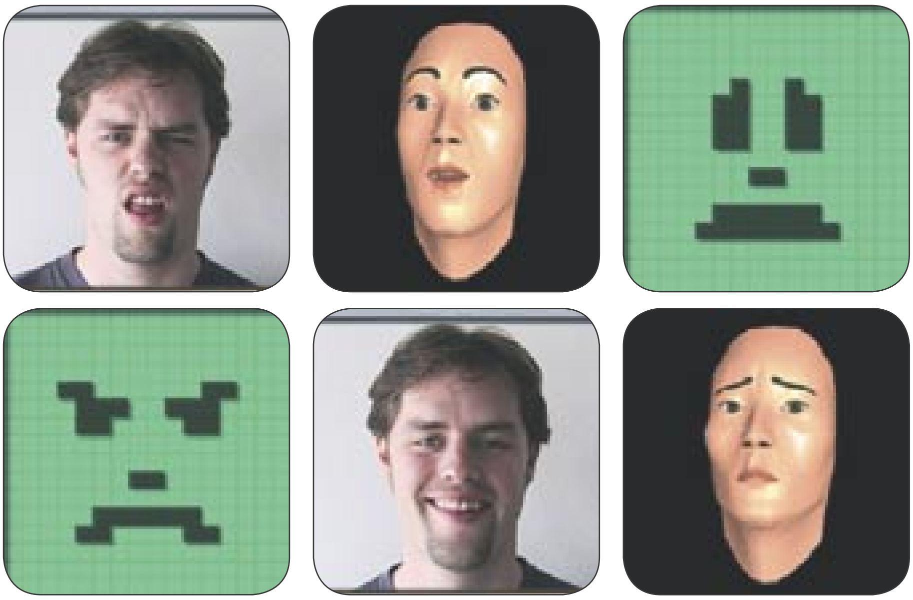

Figure 3: Examples of the stimuli. From top left: disgust, surprise, sadness, anger, happiness and fear.

2.3.5. Procedure

Before the experiment, the participants read an introduction about the experiment and the dice game. They were explicitly instructed to distinguish between the trustworthiness and convincingness of the emotional expression (original text: "A car salesperson might be convincing, but not necessarily trustworthy.") and between the type of player (human or computer). It was clarified that none of the players bluffed or cheated since such a behaviour would be useless and against the rules of the game. None of the players would gain an advantage in the game by doing so. After reading the instructions the participants played the game against the experimenter to become familiar with the rules.

Then, the participants observed 4 training games with the program to get used to the interface. The program showed the questions and recorded the answers. In these training games they were confronted with all stimuli and all questions. It can be assumed that this training was sufficient for all participants because the complexity of the task was low.

In a short pause before the start of the experiment, the experimenter answered questions the participants might have had about the process and the program. Afterwards the experimenter left the room. The participants observed 6 games, each consisting of 30 rounds. Each round consisted of one turn for each player. One emotional expression occurred per round, to which the participants had to answer one question by clicking with the mouse on a response button, such as a 1-7 scale or the list of emotions. The expressions of the players appeared either before or after throwing the dice. Fear, for example appeared before and happiness afterwards. The core experiment took 45 minutes to complete with a pause of 5 minutes in the middle. The participants received a little gift for their participation after the experiment.

Figure 4: Screenshot of the program. Left: machine condition. Right: human condition. In the human condition on the right: The opponent that was hidden behind the wall on the right rolled 11. The human player on the left rolled a 6 and therefore lost the round. Thus he expresses sadness.

2.3.6. Equipment

A program showed the game, the facial stimuli and the questions to the participants on a screen. A lap-top with a 14” screen and 800x600 pixels resolution was used to run the program. The stimuli were presented in a screen area at the top-left, the questions and possible answers were presented in a screen area at the right (see Figure 4). A set of stereo-speakers was connected to the lap-top to play the audio stimuli.

2.3.7. Results

This chapter focuses on the results for the model of *convincingness* and the scores for *distinctness* and *convincingness*. A more detailed description of this study is available elsewhere (Bartneck, 2000). Analyses of variance (ANOVA) were conducted on all dependent measures. Furthermore, a multiple regression analysis and several t-tests were performed on certain measures. The level was set to 0.05 for all tests. *Convincingness* was measured on a 7-point scale ranging from 1=very unconvincing and 7=very convincing and *distinctness* was measured by the recognition accuracy (Table 5 – Table 8).

Table 4: Pearson correlation coefficients for variables predicting convincingness across all conditions. * not significant =.05

	Convincingness	Distinctness	Intensity	Expertise
Distinctness	*0.380	-	-	-
Intensity	0.677	*0.280	-	-
Expertise	0.787	0.377	0.418	-
Trustworthiness	0.874	*0.180	0.736	0.666

Table 4 presents the results of a regression analysis. It shows the correlation matrix for variables predicting *convincingness* across all conditions. 84.1% of the variance in *convincingness* can be predicted from *distinctness*, *intensity*, *trustworthiness* and *expertise*. *Distinctness* is only weakly correlated (r=.380) with *convincingness* and is not a significant (sig.=.107) predictor. Both, *convincingness* (r=.874) and *intensity* (r=.736) are strongly correlated with *trustworthiness*. *Intensity* is not a significant (sig.=.462) predictor for *convincingness* when *trustworthiness* is already considered in the analysis due to a collinearity. *Trustworthiness* alone predicts 75.6 % of the variance in *convincingness*.

The type of emotion, such as happiness or sadness, has significant influence on *convincingness* ($F[5,160]=29.696$, $p<.001$). Surprise and happiness were more convincing ($t[32]=3.974$, $p<.001$) than sadness, disgust and anger which were more convincing ($t[32]=3.562$, $p=.001$) than fear.

Table 5: Average convincingness and distinctness scores for each emotion across all conditions.

Emotion	Convincingness	Distinctness
Surprise	5.68	93%
Happiness	5.71	95%
Sadness	5.25	90%
Disgust	5.05	68%
Anger	4.67	71%
Fear	4.02	70%

The *distinctness* of the emotional expressions, which was tested by the recognition accuracy of the users, is significantly influenced by the type of emotion expressed ($F[5,160]=17.011$, $p<.001$). The data suggest two groups of emotions which differ on their distinctness. The lowest score in the "higher" group was sadness which was significantly above the highest score in the "lower" group, which was anger (t[32]=4.478, p<.001). There was no significant difference within the "higher" group, consisting of sadness, happiness and surprise and within the "lower" group consisting of disgust anger and fear.

Knowledge about the *source* of the emotional expression, which was either human or machine, had no significant influence on its convincingness (F[1,32]=.379, p=.542). Only the scores for *distinctness* were slightly influenced (see Table 6) (F[1,32]=4.238, p=.048). However, this statistical small difference might be negligible.

Table 6: Average convincingness and distinctness scores for each source across all the natural face conditions (see Table 3).

Source	Convincingness	Distinctness
Human	5.14	84%
Machine	5.08	89%

The *abstraction* of an emotional expression, which was tested with a natural, baldi and matrix face, has no significant influence on its *convincingness* (F[2,64]=.008, p=.992). Only the scores for *distinctness* were influenced significantly (F[2,64]=20.873, p<.001). The scores for the Baldi faces were higher (t[32]=2.262, p=.031) than for the natural faces, which were above (t[32]=4.455, p<.001) the ones for the matrix faces (see Table 7).

Table 7: Average convincingness and distinctness scores for each visual abstraction level of the machine condition.

Abstraction	Convincingness	Distinctness
Natural	5.08	89%
Baldi	5.10	94%
Matrix	5.10	77%

The *medium* used to express an emotion, which could be either audio, visual or the combination of both, has significant influence on its *convincingness* (F[2,64]=4.332, p=.017). Visual and audio/visual expressions were slightly more convincing than audio expressions (t[32]=2.089, p=.045). *Distinctness* was not significantly influenced (see Table 8).

Table 8: Average convincingness and distinctness scores for three media in the matrix face condition.

Media	Convincingness	Distinctness
Visual	5.10	77%
Audio/Visual	5.19	75%
Audio	4.77	68%

2.3.8. Discussion

Distinctness is, against the expectations of this study, not a significant predictor for *convincingness*. The reason for this outcome can probably be found in the methodology of the experiment. It was impossible for the participants to evaluate their choice in the recognition task, because no feedback about the correctness of their interpretation was provided. If, for example, the participant would identify a certain expression and fill in the forced choice question, no feedback whether the identification was correct was given. Therefore, they rated the convincingness of the expressions independently of whether they interpreted the emotion correctly or not. They could make up their own interpretation of why this expression makes sense in this context. To confirm this finding a control experiment in which both, matching and mismatched information about the type of the emotion is provided would need to be performed. Even though *distinctness* is not a good predictor for *convincingness*, communication would fail between the machine and the user if the expression was frequently misinterpreted. The expression would convince the user of the wrong circumstances.

The participants were explicitly instructed to distinguish between *trustworthiness* and *convincingness* (text from instruction: "A car sales person might be convincing but not necessarily trustworthy"). The strong correlation between *trustworthiness* and *convincingness* and the finding that *trustworthiness* alone predicts 75.6 % of the variance in *convincingness* suggests that the difference between these two concepts is very small. The participants might have even treated the two words as synonyms.

In summary, there are three problems with the suggested model for *convincingness*. First, *distinctness* is not a good predictor for *convincingness;* second, the concepts of *trustworthiness* and *convincingness* are too closely related to each other to be evaluated separately; last, there is a collinearity problem for *intensity* and *trustworthiness*. Therefore the results of this study suggest a new model for convincingness that resolves these problems. It merges *convincingness* and *trustworthiness* into a new variable, denoted *convincingness'* , and leaves out *distinctness* (see Figure 5). Through these changes also the collinearity problem for *intensity* and *trustworthiness* is solved.

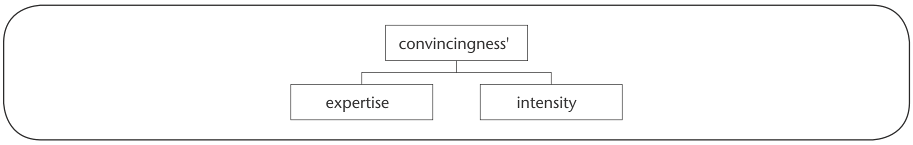

Figure 5: New model of convincingness.

The *convincingness'* score for each subject was calculated by taking the average of the *convincingness* and *trustworthiness* scores. A second regression analysis for this model was performed (see Table 9). 75.6% of the variance in *convincingness'* can be predicted from *intensity* and *expertise* and both are significant predictors (sig.<.001). The new model seems to fit the data of this study much better than the originally suggested.

Table 9: Pearson correlation coefficients for variables predicting convincingness'.

	Convincingness'	Intensity
Intensity	0.732	-
Expertise	0.747	0.418

2.4. Conclusions

In this study, the type of emotion had the strongest influence on convincingness. The participants rated the two "positive" emotions happiness and surprise highest on almost all measurements. Anger and especially fear were rated lowest. The results also suggest that certain emotional categories, such as fear, are less distinct than others and therefore endanger the success of the communication. The user might have frequently misinterpreted the expression and would therefore be convinced of the wrong circumstances. A possible solution for this problem is to avoid using the emotional category fear for a home character.

Highly abstracted faces were as convincing as natural faces. Only the distinctness of an expression was influenced by its abstraction. Interestingly, the Baldi face scored higher (94%) than the natural face (89%). The distinctness of synthetic facial expression has reached the level of natural faces. Both scores are rather high compared to results of other studies (Bartneck, 2000). However, most of those studies did not provide context information with their stimuli, which would always be available for a home character. The results showed that even abstracted home characters are able to express convincing emotional expressions. They do not need to have realistic human faces.

Emotional expressions of machines, including screen characters, are perceived at least as convincing as emotional expressions of humans. This result is in line with the media equation (Nass et al., 1996). A home character might therefore use emotional expressions, because they are not perceived as fake.

Fogg's and Hsiang Tseng's (Fogg et al., 1999) model defined believability by two components trustworthiness and expertise. The data of this study suggest that the concepts of convincingness, which is similar to their concept of believability, and trustworthiness are not distinct enough to be evaluated separately. The improved model of convincingness (see Figure 5) suggests that an emotional embodied character should express the right emotion at the right time with the right intensity.

The literature review of emotional expressions provided knowledge to create a vocabulary of facial and audio expressions that was improved through several iterative cycles prior to this study. The experiment proved this vocabulary to be more distinct than most previous designs (Bartneck, 2000). This vocabulary was the base for the design of the emotional expression of the home character as described in Chapter 5.7.

3

Overview of characters in literature and products

This chapter provides an overview of the studies and products that use characters, such as computer games, interface agents and robots. Commercially available products are included as well as a review of Philips Research's efforts in this area. This overview helped to define requirements in Chapter 4 and understand possible problems and advantages of certain characters.

3.1. Computer games

Computer games were among the first applications of characters that interact with the user. Frequently the character played the role of the user's enemy, but increasingly often that of their companion. The designers of the game characters wanted to make them more and more lifelike and engaging. Good examples of integrating emotions into game characters are The Sims (Electronic Arts, 2000) (see Figure 6), Creatures (Grand, Cliff, & Malhotra, 1997) (see Figure 6) and Petz (PF Magic, 2000).

In all of them the user has to interact with characters that are able to express their emotional states through body language or acoustic signals. The success of these products shows that the interaction with character can be very enjoyable.

Figure 6: The Sims, Creatures and Tamagotchi.

The computer game Petz is also available for the Gameboy, a handheld gaming computer. Other examples for games on handheld devices are Portapets (Eruptor, 2000) for the Palm Pilot and the world famous Tamagotchi (Bandai, 2000). The Tamagotchis swept the markets in the late 90s. The owner of a Tamagotchi had to nurse its Tamagotchi creature so that it evolved into a nice dragon (see Figure 6). Neglecting the creature resulted in miss-development and eventually death of the character. The Tamagotchis were eventually banned from many public schools because their beeping distracted the class and the students paid more attention to their Tamagotchi Creature than to their teacher. The history of the Tamagotchi shows how strong an emotional binding between a character and a user can become.

The computer games Petz, Portapets and Tamagotchi are all based on mobile devices and show, that it might be possible to integrate a home character not only in a TV, but also on remote controls, Personal Digital Assistants (PDA) and mobile phones.

The examples mentioned in this section show that the interaction with character can be very engaging and enjoyable, but by definition this is still within the context of a game. The interaction with a character when a real world task is at hand might be different.

3.2. Interface Agents

This section focuses on interface agents that use emotional expressions. Several other so-called "talking heads" are not included, because their main focus is on the visual representation of dialogue systems. Their main features are lip synchronization and conversational gestures, such as glance and nodding. Dehn (Dehn & van Mulken, 2000) reviewed empirical research on the impact of screen based animated interface agents on the user and provided a valuable starting point. In the following paragraphs Dehn's results will be used and expanded.

Microsoft created life like characters in their Persona Project (Bell et al., 1997) (see Figure 7). Their software characters including the necessary development software is available for free at Microsoft's website (Microsoft, 2000). The persona project is essentially an agent development platform and has no built-in complete set of emotional expressions or an emotion management system. However, it has an extended set of conversational and entertainment expressions, such as idle behaviours and confirmation gestures.

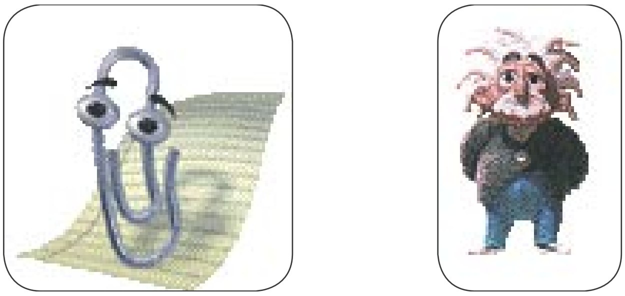

Figure 7: Microsoft Agents.

Fin Fin, developed by Fujitsu Interactive (Fujitsu, 1997), is a half-dolphin half-bird character that interacts with its world and the user (see Figure 8). It appears to be based on Microsoft's agent software mentioned above and can be integrated into Fujitsu's Internet relay chat (IRC) system CHOCOA (Fujitsu, 2002a). An internet relay chat system is a different application domain than a home character and there is also no evaluation of FinFin available. This is also the case for the Imigos agents (see Figure 8) from Plan_B (plan_b media, 2000)

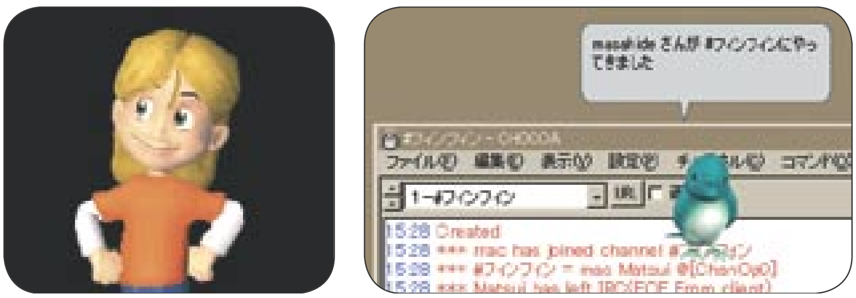

Figure 8: Imigos agent "Cedy" from Plan_B and Fin Fin from Fujitsu.

Tomoko Koda developed and evaluated poker-playing agents (see Figure 9) (Koda, 1996). She demonstrated that emotional agents helped users to engage in a poker playing task and that they are well suited for the entertainment domain. The facial expressions of a poker player are the key components to win a game of poker. This will not be the case for a home character. The facial expressions might be valuable feedback and therefore improve the communication, but they are not the key component to a successful task completion. Asking the home character to record a certain movie will not fail if the user is not able to correctly recognize the facial expression of the character.

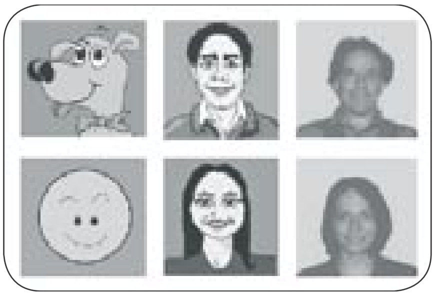

Figure 9: Koda's poker playing agents expressing happiness.

Elliott (Elliott, 1992), O'Reilly (O'Reilly, 1996) and Cañamero (Cañamero, 1997) created detailed multi-agents worlds in which the agents interacted with each other and expressed their emotions to the user and each other. However, the agents did not interact with the user, which a home character will have to do.

Several other studies created emotional agents, but offer only a very limited evaluation (El Nasr, Ioerger, & Yen, 1999) or no evaluation at all (Studdard, 1995; Martinho & Paiva, 1999). Therefore, it is difficult to compare their results with other studies, or assess their success at all.

3.3. Robots

Many agents are created as physical embodiments, such as toy puppets and robots. A more detailed review on robots can be found elsewhere (Bartneck & Okada, 2001b).

Fumio Hara has been developing realistic animated talking heads at the Hara Labs (Menzel & D'Aluisio, 2000). His robotic faces become more and more indistinguishable from humans. Their biggest strength is at the same time their biggest weakness. The high degree of realism of the heads raises high expectations in the conversational abilities of the heads that the current speech and dialogue technology cannot live up to.

The MIT, Siemens and IBM each designed less realistic but not less expressive robotic faces. Kismet (Breazeal, 1999) (see Figure 10), Mexi (C-Lab (University of Paderborn and Siemens), 2002) (see Figure 10) and Pong (IBM, 2000) (see Figure 10) have considerable expressive capabilities. All of these robots are research platforms. However, their metallic and mechanical appearance might not blend very well in a home environment.

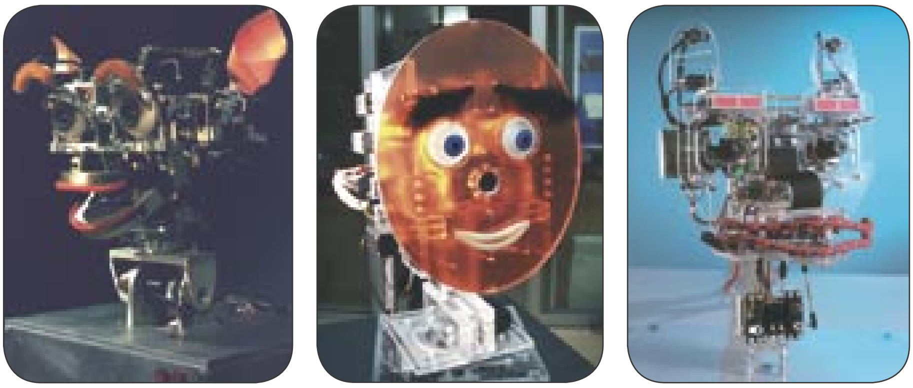

Figure 10: Kismet , Pong and Mexi.

Felix, a robot built by Cañamero and Fredslund (Cañamero & Fredslund, 2000) deserves special attention because it is built from LEGO bricks. It is a cheap and easy method to build robotic faces. However, the robot might be perceived only as a toy due its construction toy appearance and the prior experience of playing with LEGO that many people have. Felix has been used as a research platform for human robotic interactions study.

Other companies produced emotional pets. Sony created a robotic dog called Aibo (Sony, 1999) (see Figure 11) and Microsoft (Microsoft, 1999) developed a product line called Actimates with its particularly well-known Barney character. Omron created a robotic cat named NeCoRo (Omron, 2002), which is designed to interact with humans (see Figure 11). RoboScience (RoboScience, 2002) created a robotic dog named RS-01 RoboDog (see Figure 12), which is currently the largest and strongest. Small children are able to actually ride the dog.

The facial emotional expressiveness of these pets is usually limited to flashing LEDs. Mechanical animated eyebrows and lips might be much more expressive than flashing lights.

Figure 11: Aibo and NeCoRo.

Another application domain for robots is home security. Sanyo Electric, in co-operation with Tmsuk, developed a robotic watch dog called T7S Type 1 (Sanyo, 2002), which has a built in camera and mobile phone, so that it can take a picture of a burglar and send it wirelessly to the home owner (see figure Figure 12). Home security has a considerably different emphasis on certain requirements for a robot, such as its toughness and reliability.

Figure 12: RS-01 RoboDog and Sanyo T7S Type 1.

Several companies investigate humanoid robots. Sony released with its Sony Dream Robot (SDR-4X) (Sony, 2002) the first commercially available product (see figure Figure 13) designed for the consumer market. Honda has a long history in building humanoid walking robots and their robot Asimo (Honda, 2002) (see Figure 13) is regularly showcased on robotic festivals, such as the RoboFesta (RoboFesta, 2002) or the Robodex (Toshitada Doi, 2002). Robovie (Michita Imai, 2002), a humanoid robot developed by ATR (see Figure 13), is designed to interact with humans and exist as a partner in human society. Fujitsu developed a miniature humanoid robot called HOAP-1 (Fujitsu, 2002b), which is designed for wide application in research and development of robotic technologies (see Figure 14). This robot is commercially available. Another humanoid robot is Pino (see Figure 14). This robot was developed in the framework of the Kitano Symbiotic Systems Project (Kitano Symbiotic Systems

Project, 2002) and is now owned by ZMP (ZMP, 2002). ZMP seemed to have licensed PINO to the Tsukuda Original (Tsukuda Original, 2002) toy company, which sells the robot through their internet site.

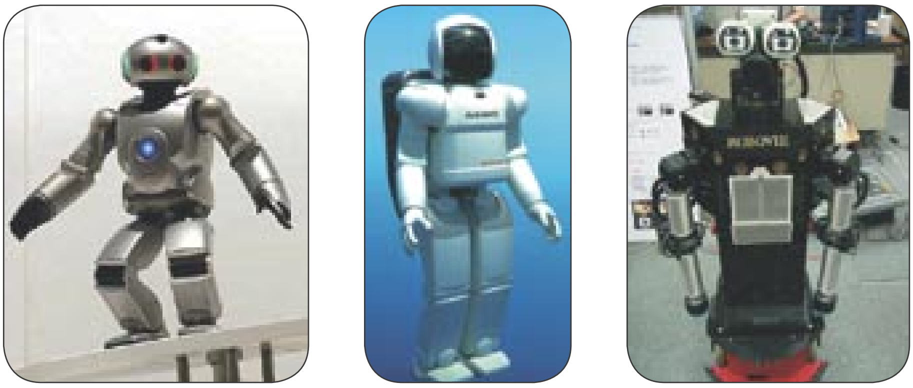

Figure 13: Sony SDR-4X, Asimo and Robovie.

Even though these humanoid robots are milestones in the development of robotics, they still have a long way to go to meet the visions of the 60s, such as a dish washing humanoid robot. A home character might not need to be able to walk on two legs, since it either does not need locomotion at all or could use wheels to drive around the home. The considerable effort required for walking on two legs does not yet seem to meet its benefits, such as the increased manoeuvrability.

Figure 14: Pino and HOAP-1.

PaPeRo (NEC, 2001), a little robot from NEC (see Figure 15), is not only able to communicate with humans, but also to control domestic appliances, such as TVs and video recorders. It is also able to communicate emotions through its face. This robot meets almost all requirements mentioned below, but was unfortunately not available for study.

Figure 15: PaPeRo and Kuma.

Kuma (Matsushita, 1999) is a robotic companion developed by Panasonic, which has been designed to keep company to elderly people (see Figure 15). However, it might have been a better idea to develop appropriate communication tools for elderly people than creating a robotic companion, which solves the problem of loneliness only superficially.

3.4. Philips Research

This study was carried out at Philips Research in Eindhoven, The Netherlands. It is relevant to take a look at their previous projects in the area of embodied interface characters. Philips developed several anthropomorphic interfaces and animated characters, starting in 1997 with the Mediators project (Ginn, Heister, & Kohar, 1997), which integrated characters into a TV recommendation system. The characters were presented by movie clips of real actors. Again, the results of a simple qualitative user evaluation showed that the high realism of the characters evoked expectations to their conversational abilities that the system could not live up to. Therefore, the next version of anthropomorphic interfaces, called Blobs (Heister, 1999), used far more abstract characters. No results of user evaluations are available. The same holds true for the Degos project (Kabala, 2000), which visualised user profiles with an abstract anthropomorphic character. The next version, called Licons (Diederiks, 2000), used animated characters, such as a jellyfish, penguin and sea star, to visualise stereotypical user profiles (see Figure 16). The profiles, based on a marketing study, were developed to ease the cold start problem of an adaptive electronic program guide. A user evaluation showed that most users liked the Licons, but were not convinced of their functional benefit. Another result was that the technology behind the Licons was not sophisticated enough to engage the user permanently. The latest version of an animated anthropomorphic interface is Bello (Diederiks & Sluis, 2000), a character for a voice controlled TV (see Figure 16). It provides the user with feedback through conversational gestures, such as nodding, and different types of barking.

Figure 16: Bello and Licons.

The analysis of trends and current status in the field of non-industrial robots suggests that a robotic home character is a good business opportunity (Crucq, 2000). This has led to an ongoing project on a robotic user interface.

It is not possible to compare the results of the various user tests performed on some of the Philips projects mentioned above, because they differ significantly in their methodology. The same hold true for the majority of evaluation of interface character. A standardised user test would be necessary to monitor its progress in creating user interfaces. It might be impossible to create such a test for all possible interfaces, but it should be possible to create one for related interfaces, such as animated interface characters.

4

Requirements for the characters

This chapter describes the requirements of the home character based on the scenarios of Chapter 1, the experimental results of Chapter 2 and the overview of other studies and products in Chapter 3. Each of these requirements could be the object of further research, since they might influence the enjoyability of the home character. Due to limitations in time and budget careful design decisions were made for these requirements resulting in the described solutions. The decisions were based as much as possible on results of previous studies. The following paragraphs list all the requirements under consideration.

- **Application domain**

The home character is intended for the ambient intelligent home. However, it was not possible to directly test a character in an ambient intelligent home context, because such a home was not yet available. Philips Research's Home Lab (Aarts, 2002) became only available at the very end of this study. Thus, it was, for example, not possible to recognize the user when her/she approaches the home and unlock the door as described in the scenarios in Chapter 1.1.

Therefore an interaction context had to be created that allowed to study the resulting behaviour of the home character and the user. This interaction context should be similar to the interaction context of an ambient intelligent home. Such an interaction is described in the scenarios in Chapter 1.1 and can be characterized as a cooperative negotiation. It is also necessary for this context to be sufficiently complex. A very simple command of the user, such as “lights out”, does not offer enough content for the character to be emotional about (Rizzo, 1999). A negotiation context provides enough emotionally complexity. Previous studies (O’Reilly, 1996) successfully used a negotiation task for a user evaluation as well.

• Match of appearance and abilities

An important problem with the implementation of a home character is not the character itself, but the technology behind it. No interface character has passed the Turing test yet. Their conversational skills are still very limited. The interaction with a character, which is based on speech technology, will not lead to a higher enjoyability if the speech recognition does not work properly. Therefore it is important to match the appearance and interaction of the character with its technical abilities.

Besides the limitations of the speech recognition accuracy it is also not yet possible to systematically express emotions convincingly through synthesized speech. This has two important consequences which both lead to the conclusion that the character should not speak at all.

First, it would be unconvincing if the character showed a certain facial emotional expression and talked with a neutral voice. This does not exclude non-speech audio, such as beeps. It is very important for the convincingness of a character to present the character consistently across all media available to it (Nass & Gong, 2000).

Second, the user would evaluate the facts, relationship, appeal and self-revelation feature of each utterance of the character as described in Chapter 1. The inability of the character to reveal its emotional state through speech would possibly be interpreted by the user as missing sympathy. It would sound strange if the character, for example, opened the front door of the house for the user to enter and spoke with an absolute neutral voice: ”Welcome home”.

An animal character is a good representation for the home character (Diederiks et al., 2000) since the communication abilities of animals match the technical abilities of the current system. A dog for example, is able to understand simple human speech, but still makes mistakes. This mirrors the quality of current speech recognition software. Moreover, animals are not able to talk, which matches the requirement mentioned above. An animal embodiment simply lowers the expectations of the user.

• **Relationship between character and user**
If the home character is going to play the role of a butler in the ambient intelligent home, it should be the friend of the user and not its foe, even though a large number of crime stories suggest otherwise. It should co-operate with the user as much as possible. The negotiation context used in this study was set up to allow for an integrative solution (Beach, 1997) hence, promote co-operation.

The absence of speech simplifies the character's communication with the user. As mentioned in Chapter 1, the user would interpret four features of each voice message of the character. Since the character does not speak it does also not need to actively control the four features.

• **Multiple rooms**
The home character should be accessible in every room in the home. In case of a screen character one screen would need to be installed in every room to display the character and in case of a robotic character the robot would either need to be able to move around or one robot per room must be provided. For direct control voice commands, such as "lights on", the presence of a character might be less important. This requirement has strong practical consequences. To decrease the complexity of this study the user interacted with only one character in one room.

• **Number of users and characters**
Multiple users such as a complete family may live in one home. Therefore it is necessary that the character is able to interact with multiple users, possibly simultaneously in multiple rooms. Each member of the home could also have his or her own character, which would result in a multiple user/multiple character situation. To decrease the complexity of this study only one user interacted with only one character.

• **Communication channel of the character**
The communication channel of the character consists of sending and receiving information. Both can be verbal or non-verbal. In the verbal interaction one can discriminate between speech and visually displayed text. The character could talk or display text to the user and vice versa. The non-verbal interaction channel consists of facial expressions, body language and gesture. The character could, for example, show facial expressions and analyse the gestures of the user.

The most natural communication channel with a character is most likely, due to its anthropomorphic appearance, speech. The more anthropomorphic it is the more humans will expect it to act like another human. The character should therefore be able to listen and speak. Due to the limitations of the available speech synthesis technology to express emotions in speech this study focuses on listening (see details above). In addition the character will be able to display facial expression without being able to analyse the user's expressions. The appearance of the character is simplified to lower the expectations of the user (see details above).

• Initiative

The interaction between a character and a user can be classified by who takes the initiative. Either only the user or only the character may initiate the interaction. Telephony applications are typical examples. The initiative can also be mixed among the participating parties.

The mixed initiative is the more suitable interaction for a home character. The character should be able to inform the user about certain events without a user requesting it and the user should be able to start interacting with the character at any time.

• Presence of additional interfaces

More complex tasks of the user might require additional user interfaces in parallel to the character. Complex information, such as a recipe (see dinner scenario in Chapter 1.1), is much easier to communicate on a graphical screen than through speech. Therefore this study uses an additional graphical user interface (GUI) in parallel to the character to present information.

• Integration of character in task

Many characters, in particular robotic characters, have their focus on the entertainment value. Sony's Aibo (Sony, 1999) is a toy in the first place. The home character may also have some entertainment value, but its main focus is to be a functional interface to the home.

The application domain described above allows the character to be a functional part in the interaction. The character is the co-operative competitor to the user in the negotiation and therefore an integral part of the task.

• Emotional expressions

The home character should use convincing emotional expressions to give natural feedback to the user. The results of the experiment described in Chapter 2 show that certain aspects need to be considered to enable the home character to express emotions convincingly.

First, the emotional expressions need to be distinct enough, so that the user does not misinterpret them. Following the results of Chapter 2 only the most distinct emotional categories (happiness, sadness and anger) are used for the home character.

Second, the experiment showed that the abstraction level of emotional expressions had no influence on their convincingness. However, several studies (Brennan, 1985; Thomas & Johnson, 1981; O'Reilly, 1996) show that abstracted characters can exaggerate their expression better and hence be more convincing. It is not more convincing to use anthropomorphic expressions. However, the abstraction level is not measurable on an absolute scale and therefore it is difficult to precisely specify the required abstraction level.

The cartoon like face of Muu2 (Okada, 2001) appears to be a reasonable abstraction level (see Figure 24). An eyebrow and lip was added to Muu2 to enable the robot to express emotions (see Figure 25). This new robot is called emotional Muu, or in short eMuu.

Third, the source of the emotional expression, which could be either human or machine, had no influence on the perceived convincingness. The emotional expressions of the home character will therefore not be perceived as a fake.

Fourth, the new model of convincingness of emotional expressions (see Figure 5) is based on expertise and intensity. The latter means that the character needs to have several grades of intensity for each category of emotional expressions. In Chapter 2, the perceived appropriateness of emotional expressions was considered as a measure for emotional *expertise* of the system. A system that displays the right emotion at the right time, which is appropriate, has *expertise* in the field of emotional expressions. To be able to do so the home character requires an emotional model. The requirements of the emotional model are described in the following paragraphs.

Last, the emotional state should have influence on the actions of the character in the same way as the human emotions influence our behaviour (Barnes et al., 1996; Oatley et al., 1996). It would be unconvincing, for example, if the character expressed strong anger on its face, but still acted very cooperative and helpful.

• Emotion model

To be able to show emotional expressions convincingly the home character needs to show the right emotion with the right intensity at the right time (see Chapter 2). To do so an emotion model is necessary. This model should enable the character to argue about emotions the way people do. An event that would make a human sad, for example the loss of a resource such as money, should also make the character sad.

However, it is not necessary to model a precise human emotion system. A "Black Box" approach (Wehrle, 1998) appears to be sufficient do develop a believable character. The purpose of this approach is to produce outcomes or decisions that are similar to those resulting from humans, disregarding both the processes whereby these outcomes are attained as well as the structures involved.

Moreover, the homes of people are very diverse and might need different types of characters. It might not be possible to define an emotion model for all possible interface characters because they considerably differ in complexity, application context, functionality, sensory and motoric abilities.

The best approach is therefore to start with the human's emotional model and adapt it to every specific character. The existing models, described below, enable us to argue how a toaster might feel if it had emotions.

The emotion model must be able to evaluate all situations that might occur during the interaction with the user and must also provide a structure for variables influencing the intensity of an emotion. The right intensity of an emotional expression is very import for the convincingness of the character (see Chapter 2).

Ortony, Clore and Collins (Ortony, Clore, & Collins, 1988) developed a computational model, called OCC, in which they specify 22 emotion types based on valenced reactions to situations constructed either as being goal relevant events, as acts of an accountable agent (including itself), or as attractive or unattractive objects. Moreover, it offers a structure for the variables, such as likelihood of an event or the familiarity of an object, which determines the intensity of the emotion types. However, it does not offer a translation from the 22 emotion types to the often used basic emotional expressions (Ekman et al., 1976).

Roseman, Antoniou and Jose (Roseman, Antoniou, & Jose, 1996) created a model that is closely related to the OCC model. Many of the classifications and variables in the OCC model can also be found in their model. However, it does not offer a structure for variables influencing the intensity of an emotion. Roseman, Antoniou and Jose empirically evaluated their model and changed it according to the results.

Sloman (Sloman, 1999) developed a model by applying an evolutionary view on emotions. He distinguishes between 3 architectural layers: the reactive layer, the deliberative layer and the meta-management layer. His model remains far more abstract compared to the OCC model and does not offer a structure for variables influencing the intensity of an emotion. No empirical evaluation of the model is available.

The OCC model is most suitable for this project because it offers sufficient complexity and detail to cover most situations an emotional interface character might have to deal with. It is also the only model that offers a structure for variables influencing the intensity of an emotion. These might be the reasons why most other projects in this area applied the OCC model (Elliott, 1992; Studdard, 1995; Koda, 1996; O'Reilly, 1996). However, all the rules, standards, goals and attitudes of the character that the OCC model requires need to be specified by the designer of the character up front.

An alternative method of implementation would be an adaptive algorithm. The character would learn what expression to use in a particular situation. To do so it would need feedback from a trainer. The role of the trainer could be filled in by a large number of people to avoid personal preferences of a single trainer.

If one assumes that the sample size of human trainers is sufficiently large and that the technical implementation of the adaptive algorithm is perfect, then one could assume that the output of the algorithm resembles the average emotional behaviour of humans. The adaptive algorithm would set the benchmark to which the OCC model would be compared. However, the average human behaviour might not be optimal. O'Reilly (O'Reilly, 1996) argued that unrealistic characters, such as in Disney animation movies, may be much more believable, because they are more abstracted and use exaggerated expressions.

Another advantage of the adaptive algorithm would be that the implementation of complex behaviour is easier than with models. Our experiment would not be complex enough to take advantage of this characteristic, but the real home might very well reach that level of complexity.

Moreover, adaptive algorithms have the ability to generalize, i.e. the capability of processing successfully situations that have not previously been encountered. The static rule system of the OCC model cannot do that. The adaptive algorithm would also able to continue its learning process to adapt to single users. From this perspective, the training session could be considered the boot strapping of the algorithm.

There are also some disadvantages of using an adaptive algorithm. Designing the personality and behaviour of the character is much more difficult, because it is not possible to systematically manipulate parameters in the algorithm. To create a dominant personality, for instance, the algorithm would need to be trained with dominant human behaviour. Only certain variables would need to be changed in the OCC model. In addition the development process of a stable and functional adaptive algorithm is much more difficult than implementing an OCC model, particularly since several OCC implementations are already available (O'Reilly, 1996; Elliott, 1992).

The questions whether the OCC model or the adaptive algorithm is the better implementation will become of interest as soon as it can be shown that the addition of emotions to the interaction with an embodied character is beneficial for the user satisfaction in the first place. Therefore this study first focuses on the OCC model since it is much easier to implement. Afterwards the implementation with an adaptive algorithm might be considered.

• Summary

In the framework of a negotiation task one cartoon like animal character interacted with one user. The character was capable of speech recognition, distinctive facial expressions and its emotion model was based on the OCC model. The complex interaction with the user was set up to promote cooperative behaviour with mixed initiative. A graphical user interface presented additional information.

5

Implementation of the characters

This chapter describes the implementation of the screen character and the robotic character that was developed according to the requirements mentioned in Chapter 4. This chapter explains the overall system architecture, the software architecture, the physical implementation and the control software of the characters.

5.1. Functionality of the characters

The character was able to negotiate with the user about stamps based on the rules described in Appendix 9.1. Stamps were chosen for the object of negotiation because they do not carry individual values for the participants of the experiment (see Chapter 6.1 for details). The character used exactly the same actions, such as making a partial offer or refusing a deal, as the user. The character evaluated the negotiation situation from a rational and emotional point of view. Depending on the outcome of these evaluations it made its next game move and expressed its emotional state through its face. The character was able to express happiness, sadness and anger by changing the form of its eyebrow and lip.

5.2. The System architecture

The requirements listed in Chapter 4 resulted in the system architecture shown in Figure 17 and Figure 18. Both architectures were part of the same control program. In principal all components of both architectures were operational all the time, but depending on the type of character, either robotic or screen, certain components were deactivated. In case of the screen character, for example, the robot would be switched off and covered with a box. The architectures described below show only the active components. Before describing each component in more detail, the relationship between the requirements and the architecture is given.

The specified negotiation between one character and one user resulted in a component for the user player, one for the computer player and one for the negotiation board. The latter contained all the objects of the negotiation and the facilities to trade them. They were displayed on an additional screen interface, as specified in Chapter 4. The user player component did not contain any intelligence, since the user was exclusively responsible for the planning and decision. It only converted the user's input, which was received through speech and clicking on the board, to negotiation events.

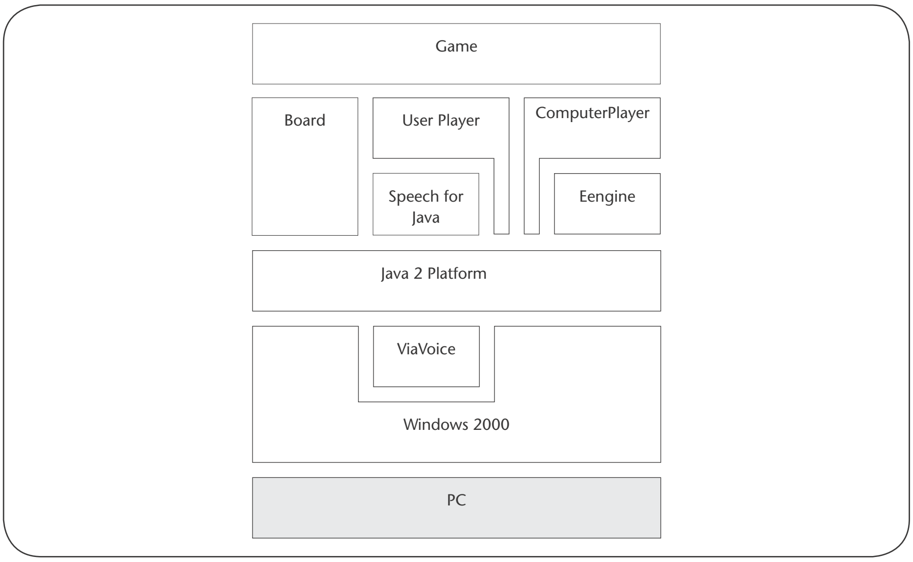

Figure 17: The system architecture for the screen character. The grey colored boxes denote the hardware level.

The computer player component could either be a robotic character or a screen character. The system architecture for the robotic computer character (see Figure 18) differs from the system architecture necessary for the screen computer character (see Figure 17). The Emotion Engine (Eengine) of a screen character is integrated into the software running on the PC and therefore the infrared protocol, JavaCom, LeJos and the robot hardware were deactivated. These components are described in the following paragraphs.

The robotic computer player component is based on two columns (see Figure 18). The left column describes the components implemented on the PC and the right column the components implemented on the Robotic Command Explorer (RCX; see Chapter 5.5.1). This split was necessary to keep the action control loop of the robot as short as possible, which again was necessary to ensure precise facial expressions of the robot. The infrared communication with the PC would be far too slow and unreliable to constantly control the expressions. Therefore the control of the emotional state and its expression had to be a dedicated application running on the RCX. This included the emotion decay function, which constantly reduces the intensity of the emotional state, because the reduction occurred in too short intervals to be controlled through the slow infrared connection. On a more general level the software on the PC can be seen as the "Brain" of the character and the software on the RCX as its "Heart".

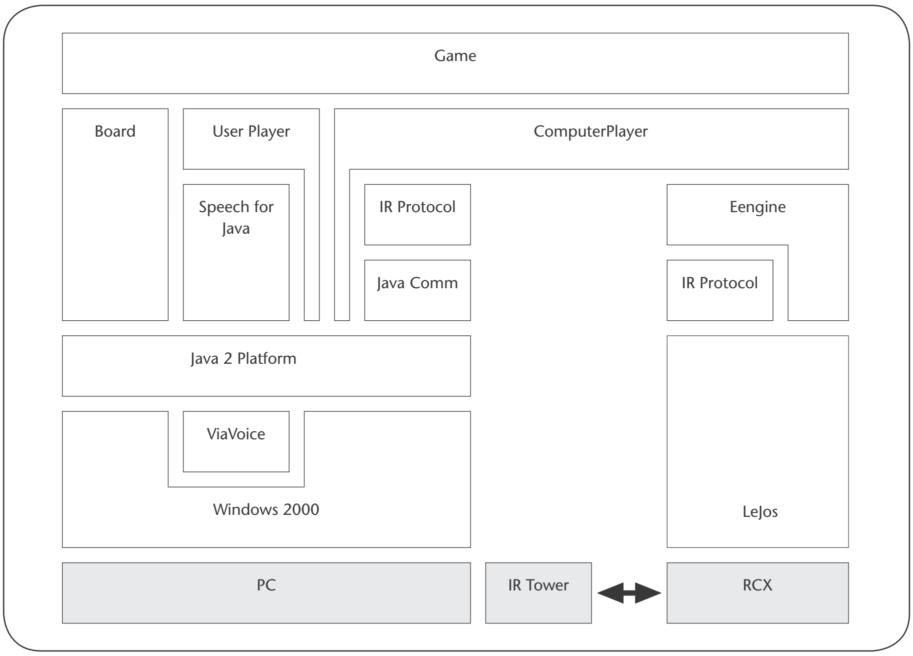

Figure 18: The system architecture for the robotic character. The gray colored boxes denote the hardware level.

Both architectures have several horizontal layers. The bottom is the hardware level, which includes the PC, the infrared tower and the RCX. Next, is the operating system layer, in this case Windows 2000 and LeJos. On top of this layer is the Java 2 Platform. LeJos is the operating system and a Java platform at the same time. On the top is the application layer, which includes all components necessary for the negotiation. In the following sections each of these components is described in more detail.

5.2.1. Game

The game component manages the progress of the experiment, such as the order of the negotiation sessions and the type of players. It displayed the negotiation session for the given duration and instructed the user to fill in the questionnaire after each session. See Chapter 6.7 for a detailed description of the process of the experiment.

5.2.2. Board

The board component manages the negotiation board that is displayed on an additional screen (see Figure 32). It uses Java Swing to display all of its visual components, such as the pads and stamps, which were used for the negotiation (see Chapter 6.4 for details).

5.2.3. User player

The user character component implements the interface between the participants and the game. It uses IBM's "Speech for Java" API to communicate with IBM's "Via Voice", a speech recognition and synthesis software (see below). It did not require any intelligence, since all the planning and decisions were carried out by the user.

5.2.4. Computer player

The computer player component manages the perception, reasoning and acting of the character. It contained all the necessary intelligence and knowledge of the negotiation and its rules. In case of a robot player it uses a protocol to communicate the emotional state through the Java Communication API (Sun, 2002b) and an infrared tower with the Eengine running on the RCX. In case of the screen player the Eengine was integrated in the computer character module and no infrared connection was necessary.

5.2.5. Eengine

The Eengine manages the emotional status of the character and expresses it through its face. In case of the robot it would move the robot's eyebrow and lip and in case of the screen character it would animate the screen character eyebrow and lip. The robot's Eengine runs on the RCX and uses a protocol to communicate the emotional state to the PC.

5.2.6. Infrared protocol

The LeJos operating system uses LEGO's Opcode (Proudfoot, 2002) to send and receive messages from the PC. Each message was limited to a size of 5 bytes and therefore a protocol was developed to encode the emotional states and direct commands into 5 byte messages. See Appendix 9.2 for the detailed protocol.

5.2.7. Java 2 Platform

The Java 2 Platform (Sun, 2002a) provides a cross-platform architecture for building applications. Besides the programming language and the object oriented model the Java 2 Platform offers the easy to use Swing user interface toolkit (Sun, 2002d).

5.2.8. LeJos

The original LEGO RCX firmware was created to teach children programming and combined with LEGO's visual development environment for the PC it certainly offers them an easy and engaging entry. However, it is not flexible enough for this project and therefore LeJos (Solorzano, 2002) was used instead. LeJos is a small Java Virtual Machine that completely replaces the original firmware. Due to the small size of available memory only a small subset of the Java Virtual Machine is implemented by LeJos. It occupies only 17Kb and already contains several standard libraries, such as java.lang, java.io and java.util. It has distinct advantages over LEGO's software, such as support for threads, floating point, string constants and multi program download.

5.2.9. Speech for Java and Via Voice

Speech for Java (IBM, 2002a) is a Java programming interface for incorporating IBM's ViaVoice (IBM, 2002b) speech technology into Java programs. The combination of Speech for Java and ViaVoice is currently the only implementation of the Sun's Java Speech standard (Sun, 2002c) and was therefore used.

5.3. The Class Model

The following figure illustrates the class model for the software of eMuu.

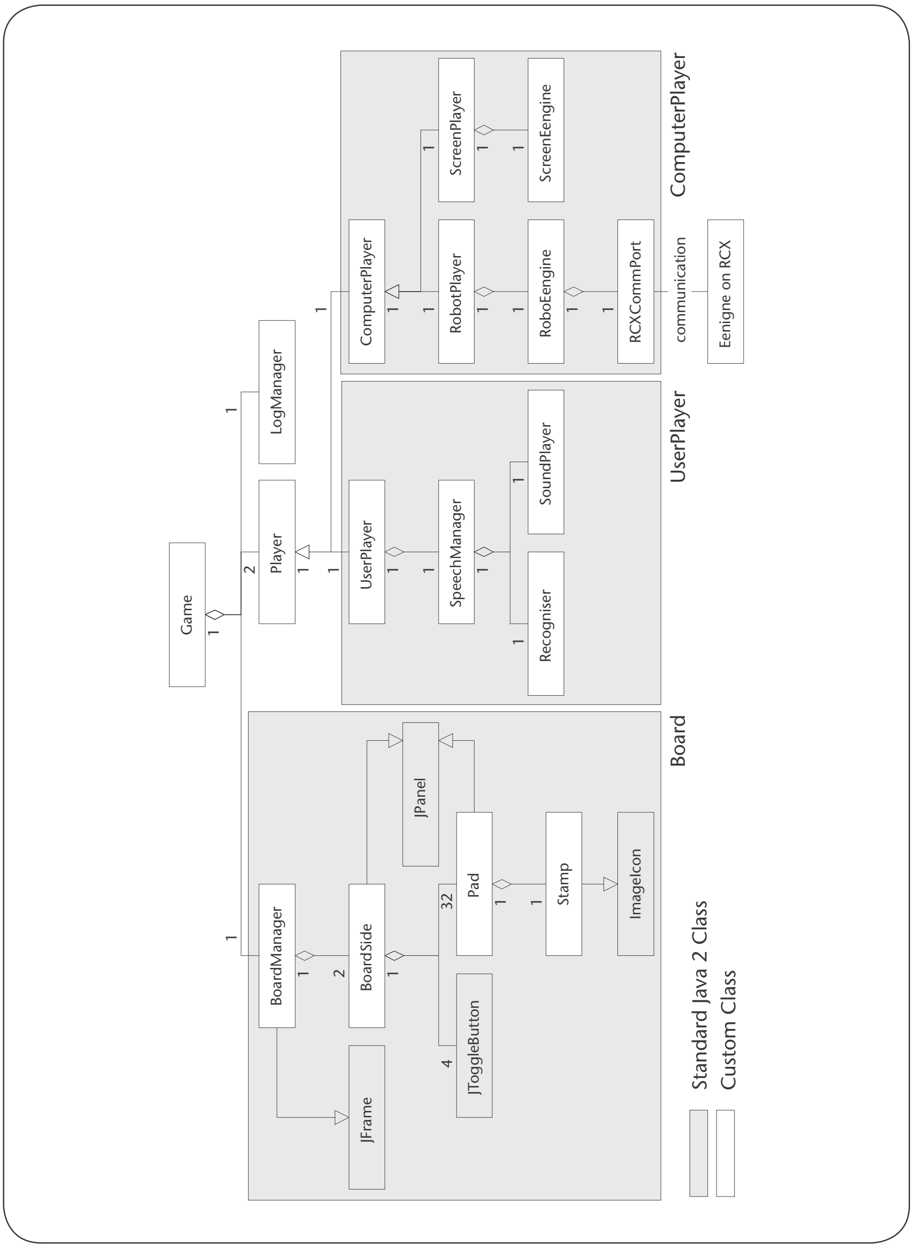

Figure 19: The class model.

The BoardManager, BoardSide, Pad and Stamp class are all part of the Board component mentioned above in the system architecture model (see Figure 17and Figure 18). Likewise, the UserPlayer, SpeechManager, Recogniser and SoundPlayer class were part of the UserPlayer component and the ComputerPlayer, RobotPlayer, ScreenPlayer, RoboEengine, ScreenEengine, RCXCommPort were part of the ComputerPlayer component. In the following paragraphs each of the classes will be described in more detail.

5.3.1. Game

The game class is the root class and manages the BoardManager, the LogManager and the two Players. It also controls the complete experiment session, by starting the negotiation sessions, creating the different negotiators and instructing the participants to fill in the questionnaire. Within a negotiation session it also manages the turn taking of the two negotiators. The game objects allowed the creation of any combination of characters. Besides the combination of user character versus screen/robotic character used for the experiment, two screen characters or two users could negotiate or one screen character and one robotic character could negotiate with each other.

5.3.2. BoardManager

The BoardManager controlled the display by showing either the two board sides or instructions to the participant.

5.3.3. BoardSide

The bottom BoardSide was assigned to the user player and the top BoardSide to the computer player. They displayed the deposit pads, trading pads and action indicators. In addition it controlled the turn indicator and the score display. If both negotiators agreed on a deal it would also exchange the stamps on the trading pads.

5.3.4. Pad

Each pad contained either an actual stamp or an invisible dummy stamp. Each pad was able to exchange its stamp with other pads if it was clicked on. If, for example, the user clicked on a stamp in his deposit pads, it searched for a trading pad that contained a dummy stamped and exchanged its stamp with it.

5.3.5. Stamp

Each stamp contained an ImageIcon, a standard Java component to display images and a value for the each negotiator. The ImageIcon contained the picture for each stamp.

5.3.6. LogManager

The LogManager logged all actions and scores of both negotiators. In addition it logged the speech recognition accuracy by counting the recognized and unrecognised utterances of the user. All the information was stored in a text file on the hard disk. This information was used in the analyses of the experiment to investigate the user's behaviour.

5.3.7. Player

The Player class controlled all the functionality all players share: turn taking. All other functionalities are different for the various player types.

5.3.8. UserPlayer

The UserPlayer would receive the user's input from the SpeechManager and forward the user's action to the BoardSides. The user's input through clicking was directly received by the BoardSides.

5.3.9. SpeechManager

The SpeechManager evaluated each utterance of the user and sent the event to the UserPlayer. The valid utterances were stored in a grammar text file outside the software. This grammar file defined the utterance for each valid action. The utterances were exactly the same as on the action indicators on the negotiation board. To evoke a certain action the user had to speak out its name, such as "make deal".

5.3.10. Recogniser

The Recogniser analysed the sound it received through a microphone and converted it to words. These were then send to the SpeechManager.

5.3.11. SoundPlayer

The SoundPlayer played the "ehhh?" sound if the Recogniser picked up an utterance, but the SpeechManager could not map the utterance to a valid user action. This typically happened if the Recogniser interpreted the sound incorrectly or if the user said something unrelated that could not be mapped to a valid action.

5.3.12. ComputerPlayer

The ComputerPlayer class contained the intelligence of the character, by parsing an exhaustive table, stored in an external text file, into an 6 dimensional array. This array contained all possible game states and the valid moves for each state. Furthermore the class perceived the board and the actions of the opponent and performed a rational and an emotional reasoning about the game state. The emotional reasoning resulted in an emotional stimulus, which was sent to the Eengine. The ComputerPlayer also managed the rational and emotional action potentials, which were used to decide upon the next game move. This class then executed this game move. The detailed process of decision-making is described in Chapter 5.4.

5.3.13. RobotPlayer/RoboEengine

These two classes send the result of the emotional reasoning, an emotional stimulus, to the RCX using an infrared protocol. This protocol is described below.

5.3.14. Eenginge on RCX

This class, running on the RCX, received the messages of the PC using the infrared protocol. It updated the emotional state and expressed it on the face of the robot. It would then send back the updated emotional state to the PC. It also contained the decay function for the emotional state, which would linearly decrease the intensity of the emotional state. Every four seconds it decreased the intensity of the emotional state by 10 percent points.

5.3.15. ScreenPlayer

This class managed the ScreenEengine.

5.3.16. ScreenEengine

The ScreenEengine class is similar to the Eengine running on the RCX. It updated the emotional state and expressed it on the face of the screen character. It also contained the decay function for the emotional state.

5.4. The negotiation process

The negotiation process for the robot and screen character is complex by nature. Before describing each component in detail a concise overview is given.

First the character perceives the negotiation game state by analysing the stamps on the trading pads and the last action of the opponent. Second, it evaluates the state from a rational and emotional point of view. The rational reasoning is based on standards that were defined up front, and results in action potentials. The emotional reasoning applies a sub section of the OCC model (Ortony et al., 1988) to evaluate the negotiation game state. The resulting emotional state is mapped to action potentials. The two action potentials are summed and define the probabilities for each action. The character's action is randomly chosen based on the action potential probabilities. Once the action is chosen, profitable stamps for the trading pads are selected. Should no stamps be available to execute the desired action, another action is selected until a combination of action and stamps is found. Last, the desired action is executed.

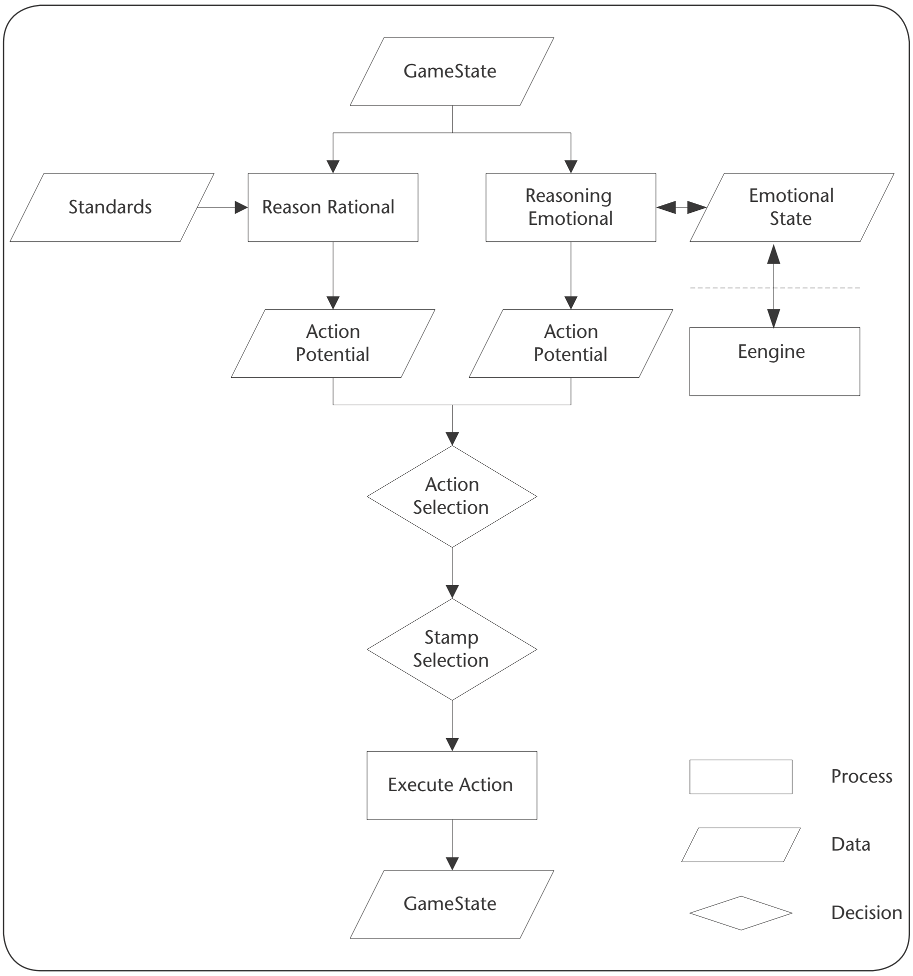

Figure 20: Overview of the negotiation process.

All of these processes take place within the ComputerPlayer class, mentioned above in the class model. Only The Eengine process takes place externally in another class or in the RCX. The following sections describe each negotiation step in more detail.

5.4.1. The game state

The game state is defined by the last user action (Partial Request, Partial Offer, Full Offer, Refuse, Accept) and the values of the stamps on the trading pads. The profit of the character is calculated by subtracting the values of the stamp in its own trading pads from the values of the stamps in the opponents trading pads.

5.4.2. The rational reasoning

The rational reasoning maps a certain user action to appropriate action potentials based on typical negotiation behaviour described in literature, such as Bazerman (Bazerman, Curhan, Moore, & Valley, 2000). A partial request, for example, would usually be answered by a full offer. An action potential consist of a probability for each of the 5 possible actions (partial request, partial offer, full offer, refuse and accept)

5.4.3. The emotional reasoning

The emotional reasoning is based on the OCC model (Ortony et al., 1988), which provides a cognitive structure for human emotions. The requirements mentioned in Chapter 4 favour a “Black Box” approach (Wehrle, 1998) for the emotional model. It is not necessary for the character to internally process events in the same way humans do. It is sufficient if the character’s emotional model produces outcomes similar to those resulting from humans.

This is particularly important, since the requirements in Chapter 4 state that the character should be a cartoon animal and that the appearance of the character should match its abilities. Animals are likely not able to process complex situations in the same way humans do, even though some cat owners might disagree. The emotion model should therefore be simplified to better match the ones of animals by not considering the consequences of events for others and aspects of objects.

The latter would not result in any valuable outcome, because in the negotiation context objects only occur with the action of an agent. If, for example, the user offered a particular good stamp to the character, then already the evaluation of the event and the action of the user will result in a positive emotional state. The evaluation of the stamp as an object would not add any useful information.

The subsection chosen from the OCC model (see Figure 21) focuses on the well-being type, creating a character that is able to communicate its internal emotional state to the world. It is a reactive character that is not able to mask its emotional expressions to, for example, gain an advantage in the negotiation. The remaining model offers enough options to deal with all situations that may occur in the negotiation game (see Figure 22).

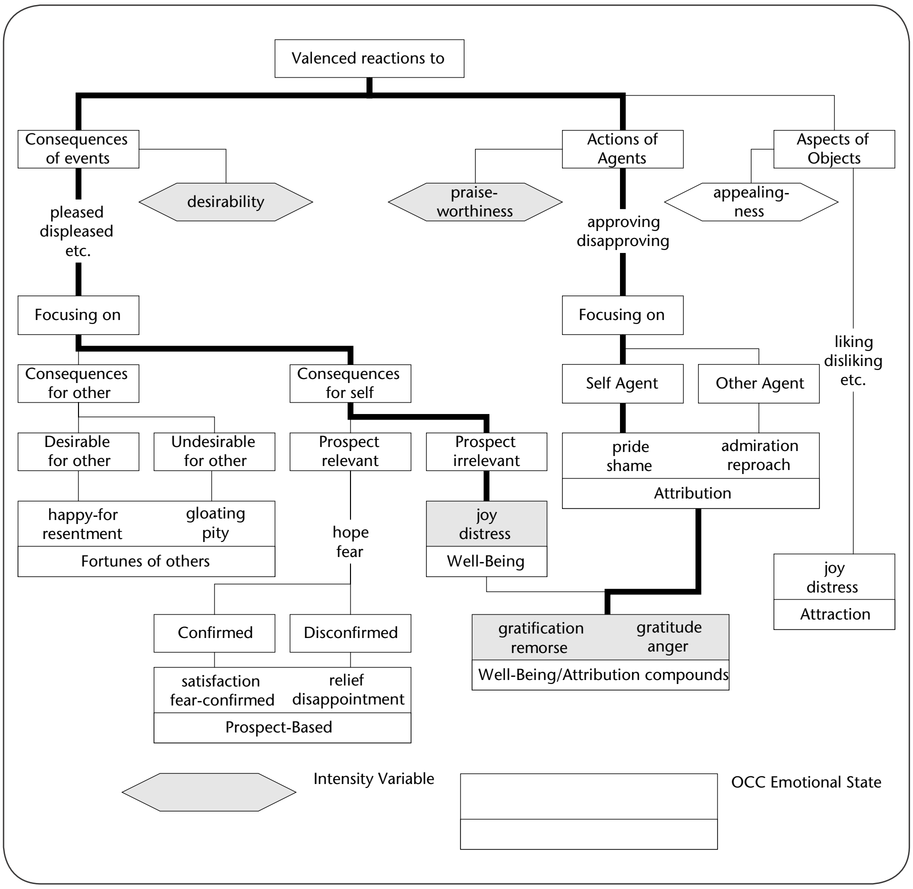

Figure 21: The OCC model and highlighted the chosen subsection.

The complete OCC model distinguishes 22 emotional states; however, it is not possible to distinguish this large number of categories of emotional facial expressions. This complexity can only be attained by including context information with the expression. The smile of a mother next to her son receiving an academic degree could be interpreted as pride, but exactly the same smile towards her husband could be interpreted as love. Moreover, there is only one facial expression available to communicate the positive emotions, which is a smile, whereas there are several facial expressions available to express different negative emotions. In the framework of this study it would have been impossible, for example, to create distinctive expression for joy, gratification and gratitude. Therefore the OCC model was further simplified by mapping the 6 emotional states (joy, distress, gratification, remorse, gratitude and anger) into three basic states (happiness, sadness and anger) that relate directly to the emotional expressions of the character (see Figure 22).

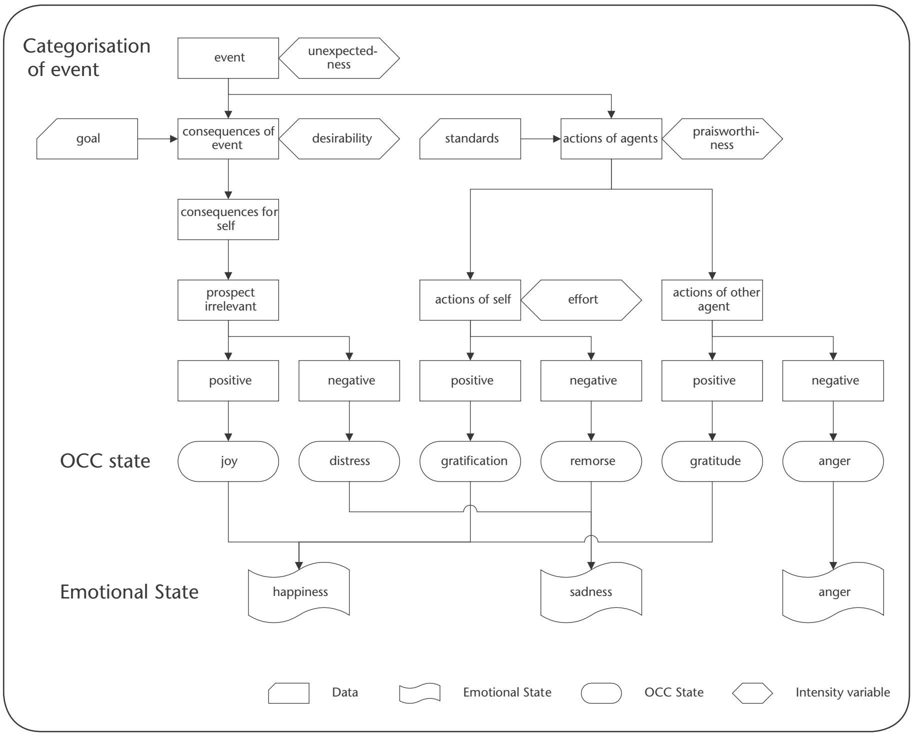

Figure 22: The emotional evaluation of an event.

Happiness, sadness and anger define the emotional state of the character. Each can have an intensity from 0-100.

The emotional reasoning is based on the perception of the game state as described above and has two components: the evaluation of the event and the evaluation of the action of the opponent. After the character has executed its action it will also perform an emotional evaluation of its own action.

The evaluation of events calculated the desirability of the event with the following formula that offered a sensitive adjustment of the game state to an emotional stimulus.

$$d(p)=\text{sign}(p)\times 100\times(1-(1-|p|/950)^{ps})\times ie$$

where

- **d(p)** is desirability of an event given the profit p
- **p** is the profit calculated by summing the values of the stamps in the opponents trading pads minus the sum of values of the stamps in the characters trading pads. The profit has a range from +950 to –950

- **ps** is the ProfitSensitivity, a constant that controls the strength of influence that the profit has on the desirability. This constant determines the form of profit/desirability curve shown in Figure 23.
- **ie** is the ImportanceOfEvent constant which is defined by the goals of the character. A full offer is, for example, more important than a partial request, because it will more likely result in a profitable deal.

If the resulting desirability has a positive value then it is assigned to happiness, otherwise to sadness. The following graph illustrates the formula:

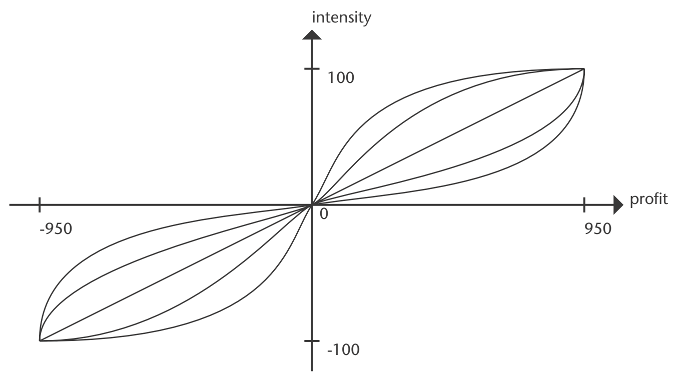

Figure 23: Adjustment of game state to emotional stimulus.

The evaluation of the action of the opponent is calculated in a similar formula for desirability as mentioned above:

$$pw(p)=sign(p)\times 100\times(1-(1-|p|/950)^{ps})\times ia$$

where

- **pw(p)** is the praiseworthiness of an action given the profit p
- **ia** is the ImportanceOfAction constant which is defined by the standards of the character, which are based on typical negotiation behaviour derived from literature, such as Bazerman (Bazerman et al., 2000).

If the resulting praiseworthiness has a positive value then it is assigned to happiness, otherwise to anger. This formula offers a sensitive adjustment of the game state to an emotional stimulus.

Since happiness is affected by two sources, desirability and praiseworthiness, and sadness and anger only by one, a compensation factor c was introduced preventing an overrepresentation of the happy expression.

The emotional stimulus is then derived from:

EmotionalStimulus = (happiness, sadness, anger)

Which is calculated by:

EmotionalStimulus = (d+pw)/c, -d, - pw

This emotional stimulus is then sent to the Eengine and the emotional reasoner receives the updated emotional state back. This state is then mapped to action potentials. Happiness will result in cooperative negotiation behaviour, sadness in more passive behaviour and anger in uncooperative behaviour.

After the character executes its action it evaluates its own action similar to the evaluation of the opponent mentioned above. In this case a negative value for praiseworthiness is assigned to sadness and not to anger.

The emotional state was then mapped to the emotional action potentials. Positive emotional states increase the negotiator's tendencies to select a cooperative strategy (Forgas, 1998). Therefore the happiness state increased the action potential for a partial offer/request slightly and the action potential for full offer and accept strongly. At the same time it decreased the action potential for the refuse action. Anger makes negotiators more

self-centred and less cooperative (Allred, Mallozzi, Fusako, & Raia, 1997; Loewenstein, Thompson, & Bazerman, 1989). Therefore the anger state strongly decreased the action potentials for partial request/offer, full offer and accept. Simultaneously, it increased the action potential for the refuse action. Sadness tends to decrease the general motivation to perform any action and thus the sadness state decreased all action potentials slightly.

5.4.4. The Eengine (Emotion Engine)

The Eengine is located in the RCX in case of the robot character and in the PC in case of the screen character. It receives an emotional stimulus and adds this stimulus to the current emotional state. This most intense emotion in the estate is then expressed on the face. No blends of emotions are expressed, only categories. The updated emotional state is sent back to emotional reasoner. The emotional state, and therefore also its expression, decays over time.

5.4.5. The action selection

The rational and emotional action potentials are summed up to a common action potential. The strength of the influence of the emotional action potential on the common action potential is controlled by a factor. This factor controls therefore how much the actions of the character are influenced by emotions. This factor has some analogies to self-control in humans.

Next a random action based on the probabilities of the action potentials is chosen. If for example the common action potential should result in {10,10,20,40,20} then there is a 10% chance that the character will make a partial request, a 10% chance that it will make a partial offer, a 20% chance for a full offer, a 40% chance for a refuse and a 20% chance for an accept.

5.4.6. The stamp selection

Humans tend to follow negotiation trails. If one party makes an offer it is more likely that the other party modifies the offer than create a complete new counteroffer. Thus, the character modified only a maximum of one trading pad per turn. A modification can either be an exchange of a stamp on a trading pad with another on a deposit pad, removal of a stamp from a trading pad to a deposit pad or adding a stamp from a deposit pad to a trading pad. The character can also choose not to modify any stamps, for example, if it wants to accept a full offer.

The goal of the character is to make profit and therefore it will only make full offers or accept deals that are profitable for it. However, if the character used a simple utility function, then it would always ask for exactly the same combination of stamps, which would make the negotiation process boring for the user. Therefore, it chooses a random stamp from the set of profitable stamps.

Certain rules apply to each character action. If, for example, the character wants to make a full offer then there has to be at least one stamp on the trading pads of each side of the board. See Appendix 9.1 for the list of the rules. The complete knowledge about what the character can do, given it has chosen an action, in each and every game state is stored in an exhaustive 6 dimensional array. This array is stored in a separate text file and parsed into the program.

5.5. The robotic character

5.5.1. Hardware

Michio Okada designed the hull of the robot for his robot Muu2. It is a redesign of his previous robot Muu (Okada, 2001). The main design objectives for the form of the hull was to maximise its cute appearance by applying form factors of infants, such as a big head, big eyes and round shapes. The hull is made from polyurethane and therefore very soft. The design of the hull proved to be very successful during a presentation at the RoboFesta (Bartneck & Okada, 2001a). Especially children touched, hugged and interacted with the robots with great enthusiasm.

Figure 24: eMuu and Muu 2 at the RoboFesta festival, Muu2 and Marvin.

Polyurethane is a soft foam and it was necessary to use a thickness of 2cm to gain enough stability for the hull. The extension on top is of solid foam and therefore adds considerably to the weight of the hull and at the same time lifts the centre of gravity.

The inside of Muu2 is out of metal and can therefore easily support the hull. For eMuu (see Figure 25), however, this study used Lego Mindstorms (LEGO, 2002) to move and control the robot, because it was much faster to build and also more affordable. The disadvantages of Lego are that it is far less stable and has greater tolerances. Lego motors are also rather weak. The design of eMuu's mechanical inside was based on experiences gained in the Marvin projects (Bartneck, 2002), which included the creation of several robotic characters (see Figure 24), such as an actor performing a play by Oscar Wilde.

Figure 25: eMuu.

eMuu used tilting and panning of its head to express its aliveness especially in situations where it would not express any emotions. To be able to tilt and pan the head efficiently, the pivot point had to be as close as possible to the centre of gravity of the hull, which was just below the extension in the inside of the hull. The hinge was placed on the top of a tower that had to be sufficiently thin to give enough space for tilting the head (see Figure 26). The motors to control the head were placed as low as possible to increase the steadiness of the robot. The two RCX were placed at the very bottom for the same reasons. Each of them contains 6 AA batteries and in combination they turned out to be a solid foot for the robot. Moreover, it gave access to the infrared interfaces of the RCXs, which would be covered if they would be placed inside of the robot.

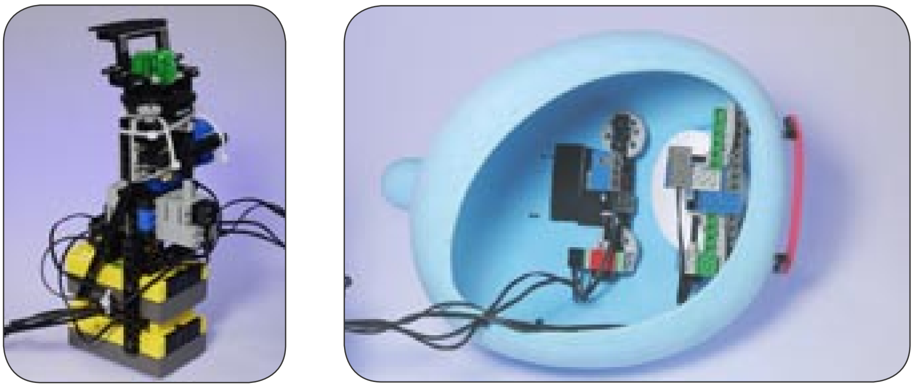

Figure 26: The mechanical interior of eMuu and the Motor unit of the eyebrow.

The motors for the control of the eyebrow and lip were directly attached to the hull (see Figure 26). The thick wall of the hull offered just enough stability to attach a micro motor to it. However, it was necessary to attach them as close as possible to the hull to prevent the motor unit to bend down. These micro motors are very weak and a gear ratio of at least 3:1 was necessary to enable them to move the lip and eyebrow.

Lego does not offer step motors and therefore normal motors were used, which had to be controlled by switching them on or off. Due to the impreciseness and slip of the motor it was necessary to attach a rotation sensor to each motor. Only through this close action and control loop was it possible to control the movements of the robot with the desired degree of precision. However, the rotation sensors itself had a limited resolution of 16 steps per full turn and tended to fail at higher speeds. To increase the resolution of the movement the sensors were placed close to the motors, without exceeding their rotation speed limitations.

The material for the eyebrow/lip had to be soft enough to be easily bent by the weak micro motors, but also stiff enough to form arches. Several materials including, metal springs, a plastic Lego tube and other plastics were considered but only foam fulfilled both requirements satisfactorily.

A basic problem with designing the eyebrow and lip was that the motors only provide movement in form of rotations (see Figure 27). A bended lip, however, is horizontally shorter than a straight one (A≠B). Stretching the lip and eyebrow itself, for example by using a spring for the eyebrow, would not result in the necessary arches to display a friendly smile. Therefore, the pivot point of the rotation had to move to the inside as well (see Figure 27). A small lever attached between the axles and the end of the eyebrow and lip solved the problem because the distance C remains constant.

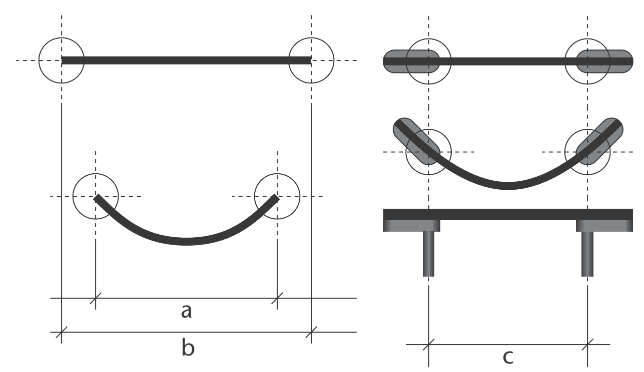

Figure 27: Animation of eyebrow and lip.

Even though only axles and toothed wheels were used to transmit the forces in the robot, a considerable amount or slip occurred. After interacting with the robot for 15 minutes the neutral position of the facial expressions would have moved. It was not possible to identify the actual source for the slip; however, it was suspected that the rotation sensor did not operate correctly. To minimize the slip the gear ratio for the eyebrow and lip was increased from 3:1 to 8:1 (see Figure 28). The motor had to turn more often to rotate the lip a certain degree, which increased the resolution of the rotation sensor. If the sensor would miss to sense one step it would result in a proportional smaller misplacement of the lip. By this change it also became easier for the motors to move the eyebrow and lip. It would not make sense to increase the gear ratio any further, because it would increase the response time of the robot too much. The relationship of a certain event and its expression would be weakened due to the delay. If, for example, the character receives a particularly valuable stamp, but it takes him 10 seconds to express a smile then the user might no longer consider the stamp as the source for the character's happiness.

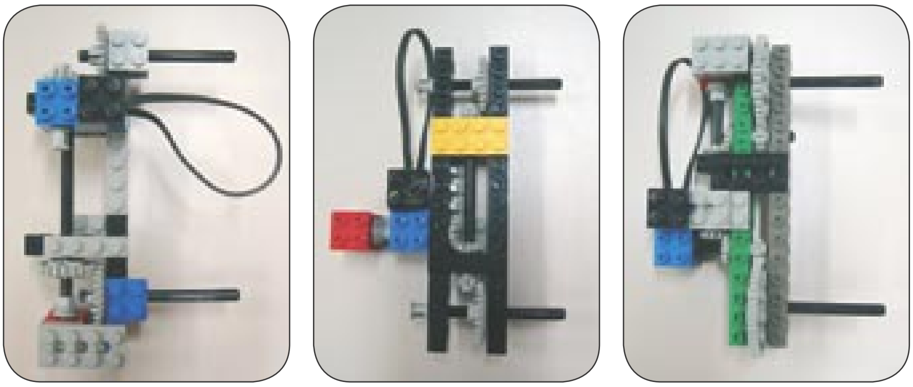

Figure 28: Motor Units with different gear ratio (1:1, 3:1 and 8:1).

The complete mechanics of the robot consisted of:

- 2 RCX controllers (LEGO 9709)
- 1 Infrared Tower (LEGO 9713)
- 2 Motors for turning pan and tilting the head (LEGO 5225)
- 2 Micro motors to animate lip/eye (LEGO 5119)
- 4 Rotation sensors (LEGO 2977)
- ~100 Lego pieces

The Lego Mindstorms Robotic Command Explorer (RCX) is a programmable brick, which has three input ports for sensors such as light, touch, rotation and temperature, and three output ports for actuators such as several different motors, lamps and buzzers. It contains a Hitachi H8/3292 micro controller and a Hitachi H8/300 processor, which has 16KB of ROM and 32Kb of RAM. The RCX also contains a little speaker and a LCD display which were only used for debugging support. The RCX uses a built-in infrared port to communicate with a PC. These specifications put strong constraints on the hardware and software development.

5.5.2. Software

The development process for RCX software has its own difficulties. First, the software is written and compiled on a PC. Next, it is downloaded into the RCX via the infrared connection. This was implemented by using RCX Tools (Rinkens, 2002) which is a very useful tool to compile and download the software. Afterwards the program can be run on the RCX. The main difficulty is the debugging since only a five-digit LCD display and some sounds are available to give feedback to the programmer while the program is running. If an error occurs, the program needs to be edited, compiled and downloaded before it can be tested again. Even an RCX emulation on the PC does not help the debugging process much, because the software has to react to the sensor readings and motors which are only available in the real RCX.

In the robot software running on the RCX two methods of acquiring the sensor values are available. The variables can be read manually in an infinity loop, which is called “pulling” or can be sent automatically through a SensorListener of the LeJos operating system, which is called “pushing”. In the latter case the SensorListener calls a method every time the sensor value changes. The pushing method proved, after various tests, to be more reliable and precise than the pulling method and was therefore implemented in the software.

5.6. The screen character

Compared to the difficulties in construction the robot it was rather easy to design the screen character. A frontal photograph was the base for the design. The two characters should look similar, while stressing their difference in embodiment. The appearance of the screen character was designed in plain 2D without any spatial clues, such as bevels or shadows. Besides a random movement of the eye, which is comparable to the random movement of the robot’s head, only the eyebrow and lip were animated. Since slip only occurs in the physical world the control of robot’s expressions is precise and reliable.

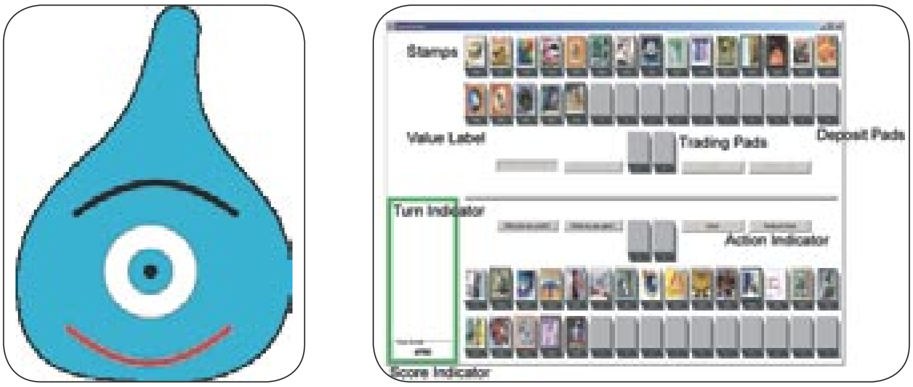

Figure 29: Screen character and the Negotiation Board.

5.7. The emotional expressions of the characters

The eMuu character is based on the Muu2 robot (see Figure 24), which is not able to express emotions in its face. However, eMuu has to be able to express the required emotions happiness, sadness and anger (see Chapter 2). Therefore one lip and one eyebrow were added to eMuu. Since eMuu has only one eyebrow it became necessary to adapt the design of the emotional expression from Chapter 2 to the physiology of eMuu. The shape that the two eyebrows form on the faces of the characters in Chapter 2 was merged to one form. Figure 30 shows the adaptation of the expressions of the Matrix Character of Chapter 2 to eMuu as a screen character and as a robotic character. Following the requirements in Chapter 4 eMuu is able to express an emotion in a grade of intensities where a straight horizontal eyebrow and lip mark the neutral expression.

Due to the physiological constraints eMuu would also not be able to express other emotions. The pre-tests of the first experiment described in Chapter 2, for example, showed that it is necessary for a character to have a nose to be able to express disgust (Bartneck, 2000). Since eMuu does not have a nose he would therefore not be able to express disgust.

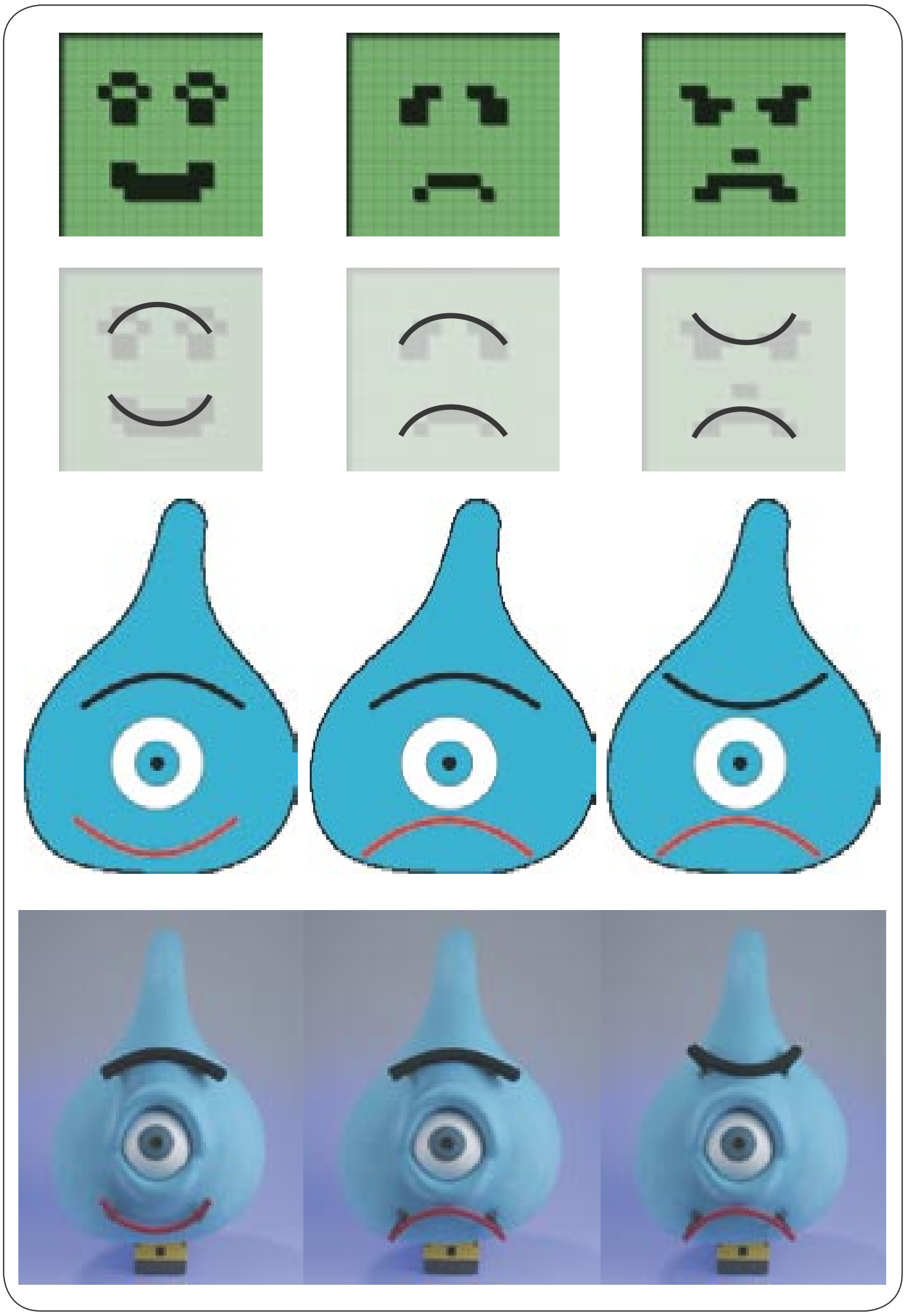

Figure 30: The adaptation of the emotional expressions to eMuu. From left to right: happiness, sadness and anger.

5.8. The negotiation board

The graphical user interface for the negotiation task consisted of a negotiation board on a Wacom LCD Touch Screen. The board was divided into two sides. The top side was assigned to the character and the bottom side to the user. Each side had 30 deposit pads on which the 20 stamps were placed. Since the negotiators were able to trade one stamp against two others it was necessary to have more deposit pads than initial stamps. None of the negotiators acquired more than 30 stamps in a negotiation session. Furthermore each side had two trading pads on which the negotiators could place the stamps they would like to trade. Each deposit and trading pad had a value label at its bottom, showing the value of each stamp to the participant. The value that each stamp had for the character is described in Chapter 6.1.

The chosen action of each negotiator was displayed on action indicators. This also helped the participants to remember what actions they could possibly take and what they have to say to choose them. A turn indicator on the left side showed whose turn it was. The participant's side at the bottom of the screen had an additional score field to display the current score, which is the sum of values of all stamps in his or her possession.

The participants had to click with a pen on a stamp to move it and pronounce their desired action. An "ehmm?" sound (Diederiks et al., 2000) was played if the speech recogniser did not understand the participant's utterance.

The robot and the screen character had the same behaviours available to them. None of them used specific advantages of their embodiment. The screen character, for example, could have easily followed the cursor movement of the participants with its eye and therefore appeared more attentive. Since it was not possible to implement the same behaviour in the robot the screen character did not show this behaviour either. Furthermore, it would have been considerably more difficult to build a robot that drives around and this behaviour would have not added useful functionality for the user. Therefore also the screen character did not move around on the screen.

6

The Evaluation

Most studies and products described in Chapter 3 did not evaluate the interaction between their characters and the user. Some studies performed simple user evaluations, but the value of their results is limited due to methodological deficiencies. O'Reily, for example, used only a very small number of participants in his study (O'Reilly, 1996). Koda tested her characters through a web page (Koda, 1996). Only 25 out of approximately 1000 people who visited the webpage filled in the relevant questionnaire. It cannot be excluded that the people who disliked her characters simply left the webpage instead of filling in the questionnaire. This bias cannot be controlled and is a typical problem of online evaluations. Only recently, efforts started to create standardized evaluation methods for characters (AAMAS2002, 2002). Such methods are necessary to compare the results of the various studies and thus monitor the progress of this research area (Jun Xiao, Stasko, & Catrambone, 2002). At this point in time no standardized test is available and therefore this study used an evaluation method that was optimised to answer the research questions described in Chapter 1.2. These questions are what influence the emotional expressiveness and embodiment of a home character has on the enjoyability of the interaction. A home character was implemented as described in Chapter 5 and this chapter describes an experiment in which this home character was used to test the following hypotheses:

H1: Users will perceive the interaction with a character that uses emotional expression as more enjoyable than with a character that does not use emotional expressions.

H2: User will perceive the interaction with a robotic character as more enjoyable than with a screen character.

6.1. Negotiation task setup

In the experiment the interaction between the user and the character was based on a negotiation task as mentioned in the application domain requirement of Chapter 4. It was important for the experiment that the participants did not have a personal preference for any of the items in the negotiation, as this would introduce an additional uncontrollable factor. If, for example, the negotiation was about songs, as mentioned in the scenarios, then it might be possible that certain participants particularly like or dislike Michael Jackson's Thriller. In the worst case, a participant could already own all the songs he or she likes and dislike all the songs of the character.

O'Reilly's agents (O'Reilly, 1996) negotiated about baseball cards. These cards are far less known in the Netherlands than in the USA and would therefore not be a good negotiation items. Stamps are very ordinary items and the concept of stamp trading is commonly known in the Netherlands. Moreover, trading stamps is not too exciting and will therefore not mask the effects of the character or the emotions. Trading the currently popular Pokemon cards, for example, could have been too enjoyable in itself.

Therefore, a random selection of stamps was used as items of the negotiation. Since none of the participants in the experiment had in depth knowledge about the stamp market, they were also not able to judge the real values of the stamps. The value for each stamp for the user was displayed below each stamp and these values were the same for all the participants in the experiment.

Both negotiators started with 20 stamps, which had a different value for each negotiator. A certain stamp might have a value of 100 Euro for one negotiator and 200 for the other.

A value structure defined the values that each stamp in the possession of one negotiator had for each of the negotiators. The value structures were identical for both negotiators to ensure a fair starting position of the negotiation. See Table 10 for a detailed description of the value structures. However, the pictures that were assigned to each stamp and the order of the stamps on the board were randomised. Both parties had exactly the same starting conditions without noticing it.

Table 10: The values of the stamps in euro

The User's Stamps		The Character's Stamp	
Value for user	Value for character	Value for user	Value for character
50	500	50	500
100	450	100	450
150	400	150	400
200	350	200	350
250	300	250	300
300	250	300	250
350	200	350	200
400	150	400	150
450	100	450	100
500	50	500	50
50	50	50	50
150	150	150	150
250	250	250	250
350	350	350	350
450	450	450	450
50	150	150	50
100	200	200	100
150	250	250	150
200	300	300	200
250	350	350	250
300	400	400	300
350	450	450	350

Moreover, the value structure of the stamps was set up to allow for joint gains and therefore give the negotiation a cooperative character. A certain stamp would be of low value to its initial owner but very valuable for the opponent and vice versa. However, to keep a certain level of difficulty not all stamps were of this type. Some of them were more valuable for their initial owner than his or her opponent and some stamps had an equal value for both negotiators.

The character's value of each stamp had to be hidden from the user.

Otherwise the negotiation task would be reduced to a simple combination task. The user would have been able to quickly assemble the most profitable deals because all the necessary information to optimally perform the task would have been available on the negotiation board. The presence of the character would have been superfluous. Also the joint gain value had to be hidden, because the user could have easily deducted the character's value of a certain stamp by subtracting his or her gain from the joint gain. The task would have been again reduced to a simple combination task.

The participants had to negotiate in three sessions for reasons mentioned in Chapter 6.5 and therefore new pictures were shown for all stamps in each session. The pictures were a random selection of international stamps and none of them was particulary famous or valuable. The pictures were too small to actually identify the origin of a particular stamp but large enough to easily recognize them on the board.

6.2. Pre-test

A qualitative pre-test was performed for the negotiating task to investigate if the participants were able to quickly learn the rules of the negotiation (see Appendix 9.1 for the full description of the rules) and if the task would be interesting enough throughout the experiment. This test took place before the implementation of the program and the characters using a rapid prototype made of Lego bricks (see Figure 31). To implement a value structure as described in Chapter 6.1 each Lego brick was labelled from opposite sides with a number from 1 to 10. This ensured that the participants could only see the value that each brick had for them.

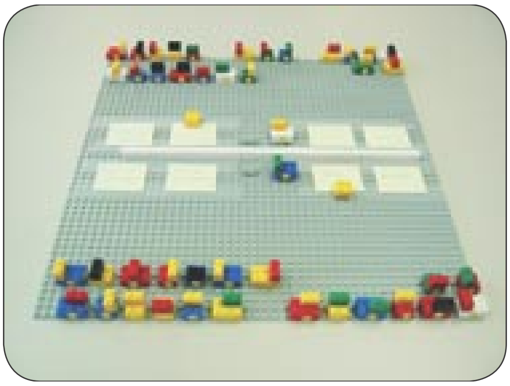

Figure 31: Rapid Prototype of the Negotiation Game.

Six members of Philips Research participated in the experiment and were observed by the experimenter. After reading the instructions the participants negotiated for 10 minutes, while the experimenter observed and logged their behaviour. The participants negotiated with each other in couples and were not allowed to talk to each other. However, they were able to see each other. Afterwards, the experimenter interviewed both participants about the negotiation task.

The rules of the game were quickly understood and exploited in full detail in the negotiation task. The task also offered enough complexity to keep the negotiation interesting. The 20 Lego bricks on each side offered enough trading options. However, the participants were not always motivated to negotiate seriously. They would, for example, accept deals that did not offer them any profit.

The negotiation task setup was adjusted to provide a stronger motivation for the participants to negotiate more seriously. The undefined values from 1 to 10 were replaced by Euro values from 50 to 500. Furthermore the participants of the final experiment were told that the 3 best performing negotiators would receive an extra gift.

6.3. Pilot test

A pilot test was performed at Philips Research Home Lab to test the experimental setup. The Home Lab provided a relaxed home atmosphere and successfully separated the participants from their busy work life. Besides various bug fixes in the software the colour of the trading pads was changed to better distinguish them from the deposit pads. Due to the low accuracy of the speech recogniser the phrases the participants needed to speak to choose an action were changed to make them more distinct from each other. The phrase “deal” was changed to “make deal” and the phrase “no deal” was changed to “refuse deal”.

6.4. Measurements

6.4.1. Enjoyability

None of the studies described in Chapter 3 focused in their evaluation on the enjoyability of the interaction with the character. Sometime a single question, such as “How entertaining was the interaction” is included in the evaluation (O’Reilly, 1996; Koda, 1996), but no serious treatment of this dimension is performed. Only recently, more attention has been paid to the enjoyability of the interaction with computers and products and first workshops took place (Monk, 2002). This research area is only at its beginning and no validated instrument for measuring the enjoyability of interacting with characters is available at this point in time.

Possible physiological measurements of enjoyability, such as heart rate and skin conductance were considered, but it was not clear if enjoyability is a physiological phenomenon in the first place. The influence of stress is physiologically measurable, but the absence of stress does not automatically mean that the participant experiences joy. Moreover, physiological measurements may include noise that mask the physiological effect of enjoyability. In addition, certain physiological measurement, such as heart rate and skin conductance, requires the participant to wear sensors on his or her body. This might be perceived as obtrusive and mask the enjoyability effect. The participant might simply feel uncomfortable with the sensors and cables attached. Therefore, a questionnaire developed by Hoonhout (Hoonhout, 2002; Stienstra & Hoonhout, 2002) was used to measure enjoyability. She based the instrument on a literature review and distinguished several concepts that are important for an enjoyable experience. She collected items, using, for example, existing satisfaction questionnaires, and other related instruments (Malone & Lepper, 1987; Jordan & Servaes, 1995; Jordan & Servaes, 1998; Hassenzahl, 2002). For concepts for which no items could be found, new items were constructed. With this set of items, an enjoyability profile can be obtained - giving more than just one value of how enjoyable using a particular device has been. The questionnaire was in Dutch, which ensured that the participants, who were all native Dutch, had an optimal chance to understand the concepts in the questionnaire. The complete questionnaire will be published in an upcoming paper (Hoonhout, 2002). The 7-point scale questionnaire consisted of 59 items in 8 concepts. An example item in Dutch and the English translation is given for each concept in the following list. Please notice that positive and negative items were used for each concept.

1. Enjoyable, <u>entertaining</u>
   e.g. Ik vond het gebruik van deze opstelling saai.
   I think that using this set-up was boring.
2. Intrinsically <u>motivating</u>, engaging
   e.g. Als je eenmaal bezig bent met deze opstelling, wil je er vanzelf mee doorgaan.
   When you are busy with this set-up, you want to keep on working with it.
3. <u>Challenging</u>, stimulating, exciting
   e.g. Ik vond het een uitdaging om met deze opstelling te werken.
   I thought that it was a challenge to work with this set-up.
4. Exploration, curiosity, fantasy, <u>novelty</u>, surprise
   e.g. Deze opstelling nodigt uit om uit te zoeken wat je ermee kan doen.
   This invites you to find out what you can do with it.

5. Concentration, attention focus, immersiveness
   e.g. Toen ik eenmaal bezig was met deze opstelling, vond ik het moelijk om er weer mee te stoppen.
   When I started using this set-up, I found it difficult to stop working with it.
6. User control, assurance, efficacy
   e.g. Ik voelde me op mijn gemak toen ik met deze opstelling werkte.
   I felt comfortable when I worked with this set-up.
7. Satisfaction, usability, utility
   e.g. Ik ben tevreden over deze opstelling.
   I am satisfied with this set-up.
8. Pride
   e.g. Ik zou in de toekomst deze opstelling niet gaan gebruiken.
   I would not use this set-up in the future.

In the following text these eight concepts are referred to by their underlined word and written in the *italic* style. The complete questionnaire is referred to as the "enjoyability profile".

6.4.2. Negotiation performance of the user and character

Positive moods tend to increase negotiator's tendencies to select a co-operative strategy (Forgas, 1998) and enhance their ability to find joint gains (Carnevalve & Isen, 1986). Angry negotiators are less accurate in judging the interests of opponent negotiators and achieve lower joint gains (Allred et al., 1997). Anger also makes negotiators more self-centred in their preferences (Loewenstein et al., 1989). Bazerman et al. (2000) provide an excellent overview of psychological studies of negotiation.

Therefore it was expected that the participants who enjoy the interaction with the character would achieve higher joint gains than participants who do not. Thus, the joint gain is an indicator for the enjoyability of the interaction. The joint gain was calculated by adding the user's and character's score. The joint gain, the user's and the character's score were automatically logged in the experiment and were analysed at a later state.

In the following text the negotiation performance is referred to by "joint gain", "user's score" and "character's score".

6.4.3. Negotiation behaviour

The negotiation behaviour of the participants may vary and therefore introduce a bias. To control this factor the negotiation behaviour of each participant was automatically logged and analysed at a later state. The behaviour is characterised by the frequencies of the actions chosen by each participant during the negotiation. The available actions for the participant were:

- Partial Request (what do you want?)
- Partial Offer (what do you offer?)
- Full Offer (make deal)
- Refuse (refuse deal)
- Accept (make deal)

In the following text the name of the actions is set in the italic style to better distinguish them from the text. Notice that the logging software distinguished two “make deal” actions. The *full offer* action is recorded if the participant chose the “make deal” action; however, if the character made a full offer before and the user agreed to it by choosing the “make deal” action, the deal was completed and the stamps were exchanged. This action was logged as *accept*.

6.4.4. Speech recognition accuracy

The participant interacted with the home character through speech. However, the speech recognition accuracy may vary between participants and introduce a bias. To control this factor the recognition accuracy was automatically logged and analysed at a later stage. The results of the analysis are presented in Chapter 6.10.

6.4.5. Summary

The following items were measured in the experiment:

- enjoyability profile consisting of eight concepts on the basis of a questionnaire
- negotiation performance consisting of the user’s and character’s scores in Euro on the basis of an automatically generated log file
- negotiation behaviour consisting of 5 different actions on the basis of an automatically generated log file
- speech recognition accuracy on the basis of an automatically generated log file

6.5. Manipulation

A mixed 2x2 between/within participants experiment was performed. The between participant factor was the embodiment, which was either a screen character or a robotic character. The within participant factor was the emotional expressions which were either present or not. Table 11 presents the coding of the conditions, where N denotes the neutral condition, S the screen character condition, R the robotic character condition, SE the emotional screen character condition and RE the emotional robotic character condition.

There might have been individual differences between the participants in the negotiation task itself. Some participants might have liked the negotiation tasks more than others. Therefore a baseline (N) was used in which the participants negotiated without the presence of a character.

The data from this baseline condition is used as a Covariant in the analyses of the influence of the embodiment on the measurements (see Chapter 6.10.6). It eliminates the influence that the negotiation task itself might have had on the measurements. The procedure is not necessary for the analyses of the influence of the emotional expressiveness on the measurement, since this is a within participant condition. Only the *difference* between the emotional and non-emotional conditions is analysed. Unfortunately, it is not possible to have an emotion baseline, since emotional expression requires the presence of a character.

Table 11: Coding of the conditions

	Emotional Expressions	
	Present	Not Present
Screen character	SE	S
Robotic character	RE	R

The participants were assigned randomly to one of the four sequences of the conditions (see Table 12).

Table 12: Presentation sequence order

Sequence	First	Second	Third
1	N	S	SE
2	N	SE	S
3	N	R	RE
4	N	RE	R

The speech recognition accuracy and the negotiation performance are two factors that could not be controlled and might have an influence on the other measurements. Therefore, they were considered independent variables and covariants for certain statistical tests.

6.6. Participants

53 participants, 34 male and 19 female at the age from 18 to 53 took part in the study. All of them were native Dutch. None of the participants had knowledge about collecting stamps. None of them were professional business/sales persons or received negotiation training. The participants received a little gift after the experiment.

6.7. Procedure

The experiment was conducted in a usability lab at the Eindhoven University of Technology. First, the experimenter welcomed the participants. Then they read an introduction explaining the purpose and process of the experiment and the negotiation task (see Appendix 9.1).

The experimenter trained the participants to correctly pronounce the phrases that they could use in the negotiation game. Furthermore, they were given tips, such as not to increase their loudness or separate the words in case the speech recogniser was not able to understand them.

The participants performed a training session of the negotiation tasks. Having finished the training they could ask questions about the process and the rules of the negotiation before the experimenter left the room. In each of the three parts of the experiment, the participants interacted with the system for ten minutes and afterwards filled in a questionnaire. After the three parts the experimenter entered the room, thanked the participant for his or her effort and handed out a little gift.

6.8. Equipment

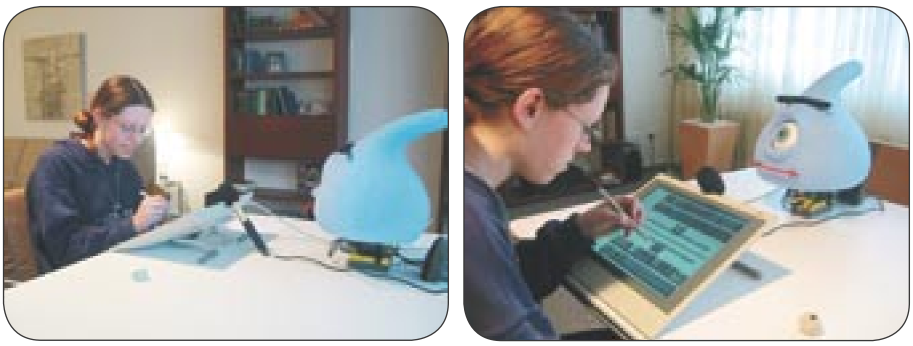

Figure 32: Experiment setup.

The negotiation game was presented on a Wacom PL-400 flat panel touch screen. The participants used a pen to click on the stamps and their voice to select an action (see Figure 32). A Philips SBC-ME400 uni directional microphone and IBM's ViaVoice for the speech recognition was used. A detailed description of the robot and the software can be found in Chapter 5. The participants filled in the questionnaire on a separate laptop.

6.9. Data preparation

The score for every participant for each of the 59 items in the questionnaire were transformed to z-scores. This normalisation created comparable, individual scores that were re-scaled to have a mean of 0 and a standard deviation of 1.

Given the relatively recent development of the questionnaire, a reliability analysis was performed to test its internal consistency. This analysis was performed on the data of the neutral condition, because this condition was presented to all participants first. This condition could have not been influenced by the other conditions and therefore all 53 participants were included in the analysis.

Table 13: Cronbach's Alpha, Mean correlation coefficients and correlation variance for the eight concepts of the enjoyability profile before and after the exclusion of certain items.

	before			after		
	alpha	mean corr.	corr. variance	alpha	mean corr.	corr. variance
Entertaining	0.86	0.52	0.02	0.86	0.52	0.02
Motivating	0.74	0.32	0.02	0.76	0.40	0.01
Challenging	0.64	0.20	0.03	0.69	0.27	0.01
Novelty	0.75	0.25	0.05	0.83	0.47	0.02
Concentration	0.83	0.39	0.01	0.83	0.39	0.01
User Control	0.88	0.47	0.02	0.88	0.47	0.02
Usability	0.82	0.32	0.03	0.82	0.36	0.04
Pride	0.83	0.45	0.01	0.83	0.45	0.01
total	0.96			0.96		

Items that considerably reduced Cronbach's Alpha, a coefficient of reliability, were excluded from further analyses. For example, the item "Only if I really had to I would use this set-up again" from the *motivating* concept was excluded, which resulted in an increase of Cronbach's Alpha from .64 to .69. The statement might have been too extreme. Table 13 presents Cronbach's Alpha before and after the exclusion of certain items in the questionnaire. An Alpha value of .70 is a sufficient level of reliability for single concepts and .90 for the whole questionnaire (Nunally, 1967). Given the high Alpha values for the eight concepts and the high overall Alpha, the questionnaire can be considered sufficiently reliable. In addition, a regression analysis was performed to test the correlation between the questions within each concept. Table 13 presents the mean correlation and correlation variance before and after the exclusion of certain items in the questionnaire. The regression analysis confirmed the reliability analysis. The exclusion of certain items resulted in an increase of the mean correlation and a decrease of the correlation variance.

The reliability analysis was repeated for all of the four conditions to verify the questionnaire. The values for Cronbach's Alpha were in most cases higher in the four experimental conditions than in the initial neutral condition (see Table 14). The questionnaire proved to be reliable under different experimental conditions.

Table 14: Cronbachs Alpha in all five conditions.

	Neutral	Screen		Robot	
	-	Character	E.Character	Character	E.Character
Entertaining	0.86	0.93	0.95	0.92	0.88
Motivating	0.76	0.85	0.90	0.89	0.90
Challenging	0.69	0.94	0.89	0.89	0.91
Novelty	0.83	0.88	0.84	0.89	0.89
Concentration	0.83	0.89	0.87	0.84	0.76
User Control	0.88	0.80	0.91	0.90	0.92
Usability	0.82	0.87	0.84	0.86	0.91
Pride	0.83	0.91	0.90	0.89	0.90
total	0.96	0.98	0.98	0.97	0.98
N	53	25	25	28	28

The mean of all remaining items belonging to one concept was calculated for each of the eight concepts in the enjoyability profile. Since the relative importance of each concept for enjoyability is not known, the best approach is to assume that they are all of equal importance.

The frequencies of the different actions were transformed to proportions by dividing the frequency of each action by the total number of actions performed. These proportions were transformed to obtain a continuous variable that can be assumed to be normally distributed with the following formula: $\ln((p+1)/(1-p+2))$ where p is the proportion which could range from zero to one. The addition of the constants was necessary to avoid problems with very low proportions. The graphs still display the original proportions, because these values give a better overview.

The speech recognition accuracy was calculated by dividing the number of correct recognitions by the total number of utterances.

Missing data for certain participants causes the different degrees of freedom in the various statistical tests. The alpha level was set to 0.05 for all tests.

6.10. Results

The results of the various statistical tests are presented in the following paragraphs. To make clear on what conditions the tests were performed their abbreviation based on the condition coding Table 11 is used, where N denotes the neutral condition, S the screen character condition, R the robotic character condition, SE the emotional screen character condition and RE the emotional robotic character condition.

6.10.1. Order effects

Order effects in the enjoyability profile

Two Analyses of Variance (ANOVA) were performed to test if the order of the presentation of the characters has an effect on the enjoyability profile (see Table 15).

Table 15: Mean z-Scores for the eight concepts in all sequences.

		Screen		Robot	
		1	2	3	4
Character	Entertaining	-0.04	0.05	-0.03	0.22
	Motivating	-0.13	-0.27	-0.21	-0.19
	Challenging	0.12	0.15	-0.01	0.12
	Novelty	-0.07	-0.21	-0.28	0.09
	Concentration	0.20	0.18	0.28	0.29
	User Control	-0.10	0.02	0.19	-0.18
	Usability	0.00	0.13	0.08	-0.25
	Pride	-0.01	-0.11	-0.08	-0.07
E. Character	Entertaining	0.17	-0.03	-0.05	0.26
	Motivating	-0.04	-0.17	-0.40	-0.24
	Challenging	0.09	0.10	-0.01	0.27
	Novelty	0.04	-0.08	0.00	0.27
	Concentration	0.00	0.28	0.14	0.15
	User Control	0.04	-0.08	0.11	-0.40
	Usability	-0.21	-0.01	0.22	-0.22
	Pride	-0.04	-0.04	-0.18	-0.06

Only the scores for the emotional robot on *user control* (F[1,26]=7.073, p=.013) and *usability* (F[1,26]=9.604, p=.005) were affected significantly. The score for *user control* was significantly higher for the emotional robot (RE) if it was presented after the non-emotional robot (R) (t[26]=2.875, p=.0.13). The score for *usability* was significantly higher for the emotional robot (RE) if it was presented after the non-emotional robot (R) (t[26]=3.104, p=.0.05).

Since the presentation order of the characters was cross-balanced, sequence 1-2 and 3-4 were merged (see Figure 33 and Figure 34).

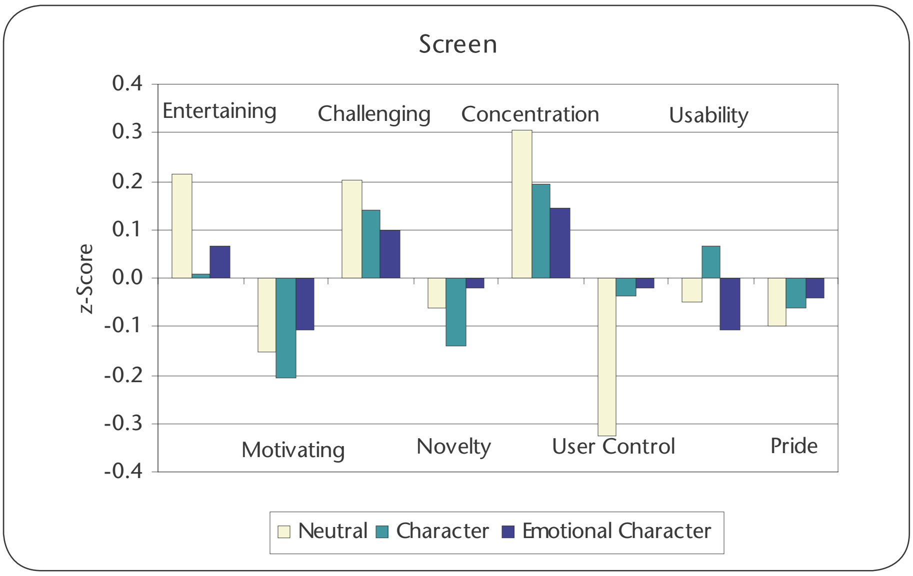

Figure 33: Mean z-Scores of the eight concepts in the in the screen condition.

It cannot be excluded that the differences between the neutral condition (N) and the character conditions (S,R,SE,RE) are due to an order effect. The neutral condition was presented to the participants always first.

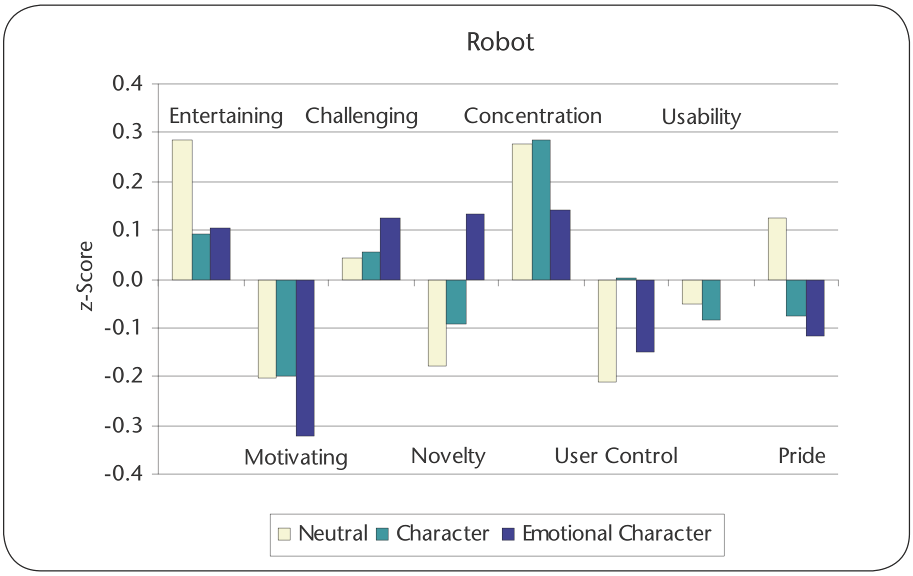

Figure 34: Mean z-Scores of the eight concepts in the robot condition.

Order effects in the negotiation behaviour

An ANOVA was performed to test the influence of the order of the presentation of the characters on the negotiation behaviour. No order effect could be found in the robot conditions (R,RE). However, in the non-emotional screen condition (S) a just significant difference was found (F[1,23]=4.418, p=.047). The proportion of the *refuse* action was significantly higher (t[23]= 2.10, p=.047) if the non-emotional character (S) was presented before the emotional (SE) one (0.04/0.01). The same significant order effect was also found in the emotional screen character condition (SE) (F[1,23]=6.254, p=.020). The proportion of the *refuse* action was significantly higher (t[23]= 2.50, p= .02) if the non-emotional character (S) was presented before the emotional (SE) one (0.04/0.01). Another significant difference was found in the emotional screen character condition (SE) (F[1,23]=6.709, p=.016). The proportion of *request* actions was significantly higher (t[23]= -1.29, p= 0.21) if the emotional character (SE) was presented before the non-emotional (S) one (0.13/0.19). Since the presentation order of the characters was cross-balanced, sequence 1-2 and 3-4 were merged (see Figure 34 and Figure 35).

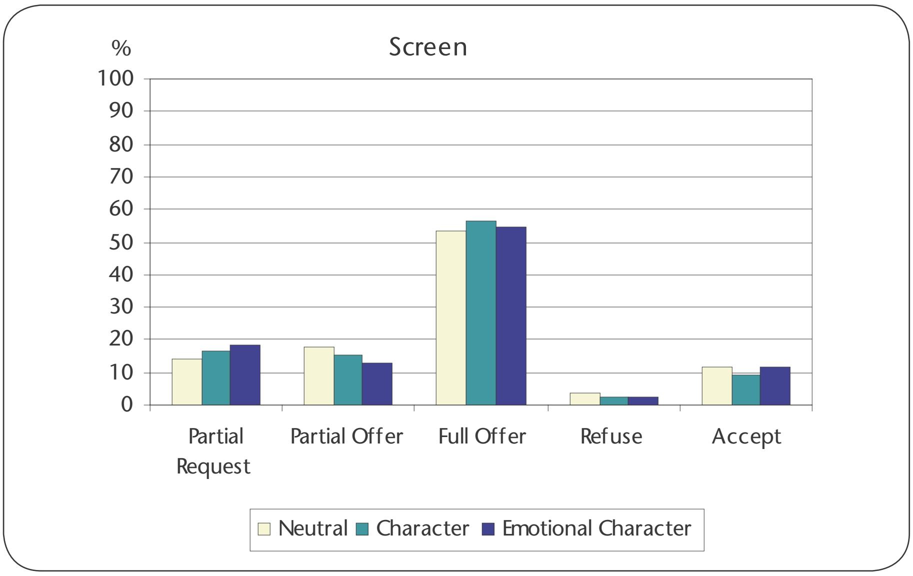

Figure 35: Negotiation behaviour in the screen condition.

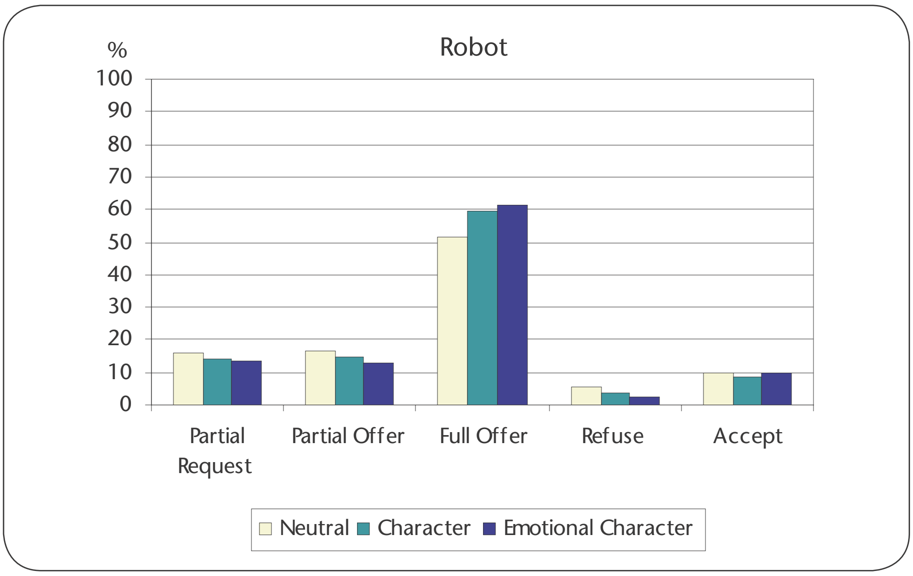

Figure 36: Negotiation behaviour in the robot condition.

Order effects in the negotiation performance

No order effect for the negotiation performance, which consisted of the user's score, character's score and joint gain, was found.

6.10.2. The effect of gender

The gender of the participants had no significant influence on any of the dependent variables.

6.10.3. The recognition accuracy effect

Figure 37 presents the mean speech recognition accuracies for both embodiments. An ANOVA was performed to test if there was a difference in the speech recognition accuracy. There was a significant difference within the robot condition (R,RE) ($F[2,21]=56.136$, $p<.001$). The recognition accuracy in the neutral (N) condition (70%) was significantly higher ($t[23]=6.512$, $p<.001$) than the accuracy in the character (R) condition (56%), which was again significantly higher ($t[24]=3.654$, $p=.001$) than in the emotional character (RE) condition (43%). There was no significant difference within the screen conditions (S,SE).

An ANOVA was performed to test the difference in speech recognition accuracy between the screen and robot conditions. There was a significant difference in the emotional character condition (SE,RE) ($F[1,50]=5.661$, $p=.021$). The recognition accuracy was significantly higher in the screen (SE) condition (58%) than in the robot (RE) condition (43%) ($t[50]=3.903$, $p=.021$).

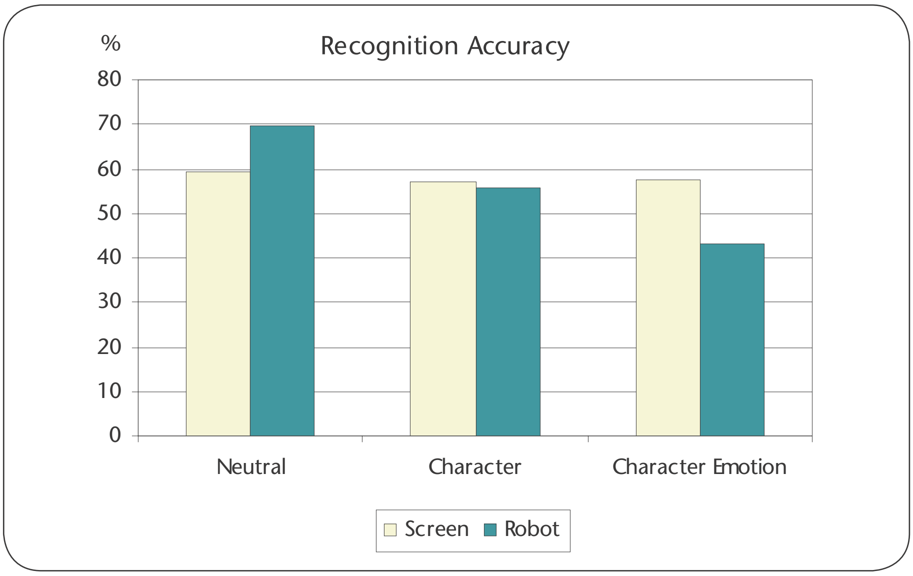

Figure 37: Mean recognition accuracies (proportion) for different embodiments.

There was a large variance between participants. The recognition accuracy varied between 11% and 97% between participants. This resulted in large standard deviations (see Table 16).

Table 16: Mean recognition accuracy (proportion) and standard deviation for the screen (S, SE) and robot conditions (R, RE).

Character	Condition	Mean	Std Deviation
Screen	N	0.59	0.26
	S	0.57	0.21
	SE	0.58	0.25
Robot	N	0.70	0.23
	R	0.56	0.20
	RE	0.43	0.17

Two linear regression analyses were performed to test if there was a correlation between the recognition accuracy and the enjoyability profile. The first test investigated the difference across all conditions (see Table 17).

Table 17: Correlation between recognition accuracy and the eight concepts across all conditions.

Concepts	Correlation	Sig.	R^2
Entertaining	0.10	0.11	0.01
Motivating	-0.04	0.29	0.00
Challenging	-0.26	0.00	0.07
Novelty	-0.25	0.00	0.06
Concentration	0.02	0.41	0.00
User Control	0.16	0.03	0.02
Usability	0.19	0.01	0.04
Pride	0.01	0.46	0.00

Challenging, *novelty*, *user control* and *usability* were significantly, but weakly correlated with the recognition accuracy. The R^2 values were very low for all cases. This indicates that these variables are only weakly affected by the recognition accuracy. Scatter plots revealed that the scores for *challenging* and *novelty* were relatively high at low recognition accuracy and low for high recognition accuracy. The opposite effect was observed for *user control* and *usability*. Their scores were high at high recognition accuracy and low at low recognition accuracy.

The second linear regression investigated the correlation between recognition accuracy and the enjoyability profile depending on their type of embodiment and presence of emotional expressions (see Table 18).

Table 18: Correlation between recognition accuracy and the eight concepts in the screen (S, SE) and robot (R, RE) conditions. The bold style indicates significant correlations.

		Pearson Correlation		
Character	Concepts	Neutral	Character	E.Character
Screen	Entertaining	0.10	0.05	0.15
	Motivating	-0.22	0.27	0.15
	Challenging	-0.25	**-0.56**	**-0.42**
	Novelty	**-0.44**	-0.18	**-0.36**
	Concentration	-0.26	0.18	-0.05
	User Control	**0.50**	0.14	0.26
	Usability	**0.42**	0.09	0.23
	Pride	-0.12	-0.06	-0.04
Robot	Entertaining	-0.06	0.09	0.03
	Motivating	-0.30	-0.18	-0.29
	Challenging	0.09	-0.22	-0.30
	Novelty	-0.06	-0.10	-0.01
	Concentration	-0.07	0.01	0.15
	User Control	0.05	0.06	0.16
	Usability	0.19	0.20	0.20
	Pride	0.04	-0.04	-0.11

Usability, user control, novelty and c*hallenging* were significantly, but weakly correlated with the recognition accuracy in the screen conditions (N, S, SE). The significance disappeared for *user control* and *usability* in the presence of the character (S, SE), but remained for *challenging* and *novelty*. Scatter plots revealed the same direction of the correlation as mentioned before.

6.10.4. The negotiation performance effect

Two sets of linear regression analyses were performed to test the influence of the user's score on the enjoyability profile. The first set investigated the individual differences between the participants across all conditions (see Table 19).

Table 19: Correlation of the user's score and the eight concepts across all conditions.

Concepts	Correlation	Sig.	R^2
Entertaining	-0.04	0.32	0.00
Motivating	-0.06	0.21	0.00
Challenging	-0.13	0.05	0.02
Novelty	-0.21	0.00	0.04
Concentration	-0.12	0.06	0.02
User Control	0.29	0.00	0.08
Usability	0.15	0.03	0.02
Pride	-0.08	0.16	0.01

Table 20: Correlation between the user's score and the eight concepts in the screen (S, SE) and robot (R, RE) conditions. The bold style indicates significant correlations.

		Pearson Correlation		
Character	Concepts	Neutral	Character	E.Character
Screen	Entertaining	-0.23	-0.09	0.13
	Motivating	0.26	0.11	-0.20
	Challenging	0.01	0.17	-0.24
	Novelty	**-0.40**	-0.17	-0.25
	Concentration	-0.07	-0.05	**-0.34**
	User Control	0.06	0.17	**0.38**
	Usability	0.11	-0.02	0.20
	Pride	0.26	-0.30	0.03
Robot	Entertaining	-0.22	-0.23	0.18
	Motivating	-0.18	-0.06	-0.20
	Challenging	-0.11	-0.21	-0.24
	Novelty	**-0.32**	-0.11	-0.22
	Concentration	0.04	**-0.35**	-0.04
	User Control	**0.45**	**0.43**	0.26
	Usability	0.04	**0.33**	0.31
	Pride	0.07	**-0.38**	-0.25

N*ovelty*, *usability, user control* and c*hallenging* are significant but weakly correlated with the user's score. The R^2 values were very low for all instances (see Table 20). Scatter plots revealed that the score for *challenging* and *novelty* were high at a low user's scores and high low for a high user's scores. The opposite effect was observed for *user control* and *usability*. Their score was high at high user's score and low at low user's score.

Approximately the same pattern was observed for *usability* and *user control* across all conditions appeared in the robot conditions (R,RE). In the screen conditions, however, *usability* is no longer correlated significantly with the user's score and user control was only significantly correlated with the user's score in the emotional screen character condition (SE). The pattern for *challenging* and *novelty* disappeared.

There was no significant difference for the joint gain between the conditions (see Table 21).

Table 21: Joint gains in euro.

	Screen	Robot
Character	12022	12071.43
E. Character	12052	12001.79

6.10.5. The effect of the character's emotional expressions

Table 22 presents the mean values for the enjoyability profile in the four conditions. An ANOVAs was performed to test the influence of the addition of emotions to the characters. Emotional expressions had no influence ($F[7,18]=1.959$, $p=.119$) on the enjoyability profile in the Screen Character condition (S,SE), however in the robot condition (R,RE) it had a significant influence ($F[7,21]=2.749$, $p=.34$).

T-Tests were performed to check differences on the individual concepts. In the robot condition the score for *novelty* was significantly higher ($t[27]=3.451$, $p=.002$) in the emotional character (RE) condition (0.132) than in the character (R) condition (-0.093). The score for *concentration* was significantly lower ($t[27]=2.083$, $p=.047$) in the emotional character (RE) condition (0.1428) than in the character (R) condition (0.2842).

In the screen condition the score for *novelty* was significantly higher ($t[24]=2.274$, $p=.032$) in the emotional character (SE) condition (-0.021) than in the character (S) condition (-0.1415). The score for *usability* was significantly lower ($t[24]=2.635$, $p=.015$) in the emotional character (SE) condition (-0.1074) than in the character (S) condition (-0.0675).

Table 22: Mean z-Scores for the eight concepts in the different conditions (S, R, SE, RE).

	Screen Character		Robot Character	
Concepts	Character	E.Character	Character	E.Character
Entertaining	0.01	0.07	0.09	0.10
Motivating	-0.20	-0.11	-0.20	-0.32
Challenging	0.14	0.10	0.06	0.13
Novelty	-0.14	-0.02	-0.09	0.13
Concentration	0.19	0.14	0.28	0.14
User Control	-0.04	-0.02	0.00	-0.15
Usability	0.07	-0.11	-0.08	0.00
Pride	-0.06	-0.04	-0.08	-0.12

Adding emotions to screen or robot character had no significant influence on the negotiation behaviour of the participants.

ANOVAs and t-Tests were performed to investigate the influence of the emotional expressions on the user's score. No significant difference was found.

Table 23: The user's score in the different conditions (S, R, SE, RE) in Euro.

		Mean	Std. Deviation
Screen	Character	6176	744
	E. Character	6298	900
Robot	Character	6680	747
	E. Character	6555	635
Total	Character	6442	781
	E. Character	6434	775

6.10.6. The effect of the character's embodiment

An ANOVA was performed to test the influence of the embodiment on the user's score.

First, the influence on all conditions was tested. There was a significant influence ($F[1,104]= 6.745$, $p=.01$). The users scored significantly higher ($T[104]= -2.59$, $p=.011$) in the robot (6618 euro) condition (R,RE) than in the screen (6237 euro) condition (S,SE).

Next the influence of the embodiment was investigated separately for the non-emotional (S, R) and emotional characters (SE, RE). There is a significant difference ($F[1,51]= 6.04$, $p=.017$) in the non-emotional character condition (S,R). The users in the robot condition (R) scored

(6680 euro) significantly (t[51]= -2.45, p=.017) higher than in the screen (S) condition (6176 euro). In the emotional character conditions (SE,RE) only a trend is visible. The users scored higher in the robot (RE) condition (6555 euro) than in the screen (SE) condition (6298 euro).

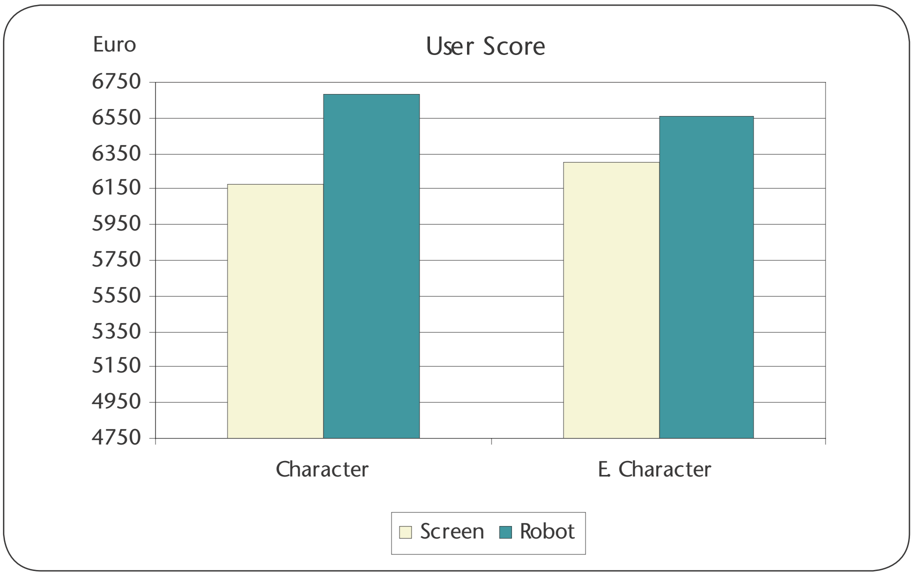

Figure 38: User's score in Euro in the different conditions (S, R, SE, RE). The users had an initial score of 4750 Euro.

ANCOVAs were performed to investigate the influence of the embodiment (screen or robot) on the enjoyability profile. The corresponding depending variable in the neutral condition (N) was used as the covariant. Only *novelty* was influenced significantly (F[1,103]=4.137, p=.04). The same effect was observed when looking at the character and emotional character conditions separately. Only *novelty* was influenced significantly (F[1,50]= 3.919, p=.05) in the emotion character conditions (SE,RE). However, a t-test revealed that the means (SE=-0.02, RE=0.13) were not significantly different (t[51]=1.237, p= .22).

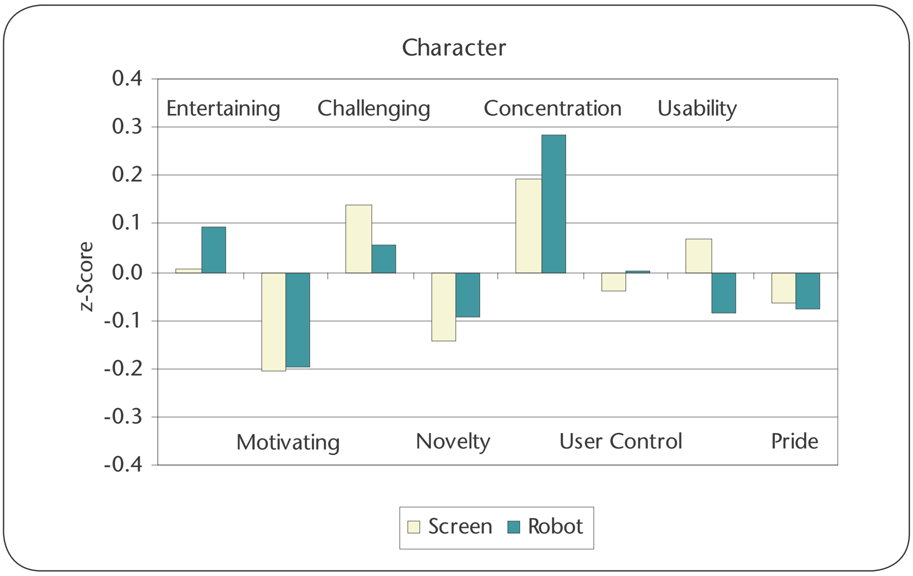

Figure 39: Mean z-Scores of the eight concepts across the character conditions (S, R).

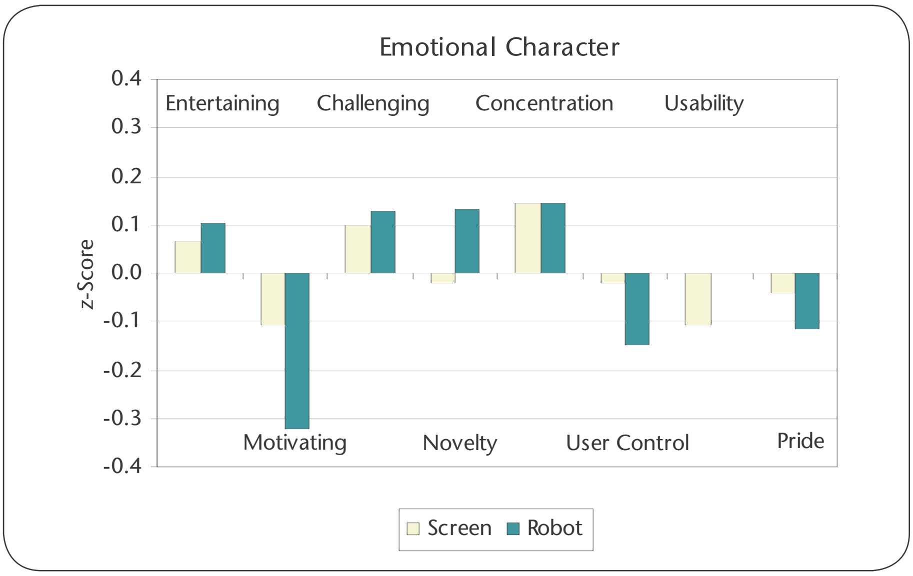

Figure 40: Mean z-Scores of the eight concepts across the emotional character conditions (SE, RE).

ANCOVAs were performed to test the influence of the embodiment on the negotiation behaviour of the participants. The corresponding depending variable in the neutral condition (N) was used as the covariant. None of the actions were significantly influenced. However, a general tendency that the participants used the *full offer* action more often in the robot condition than in the screen condition at the cost of all other actions can be observed.

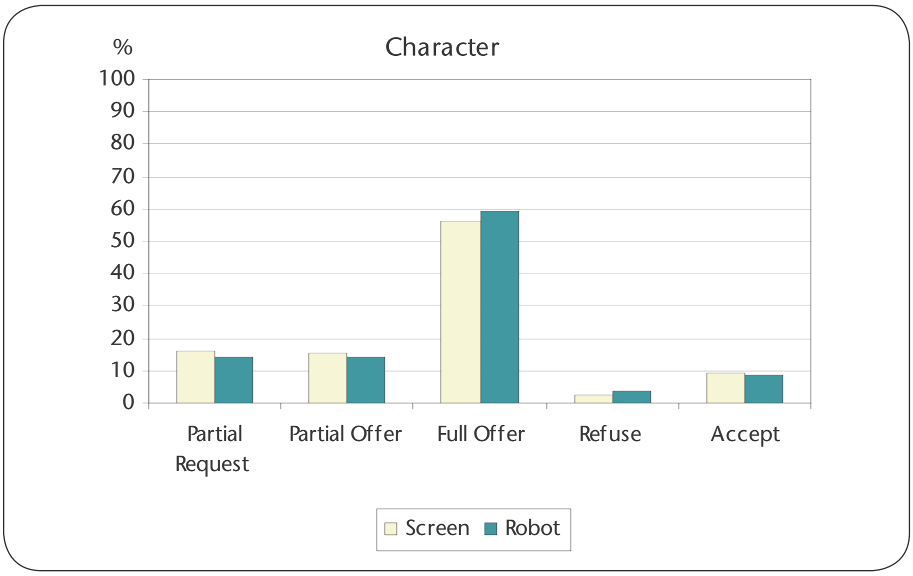

Figure 41: Negotiation behaviour in the character conditions (S, R).

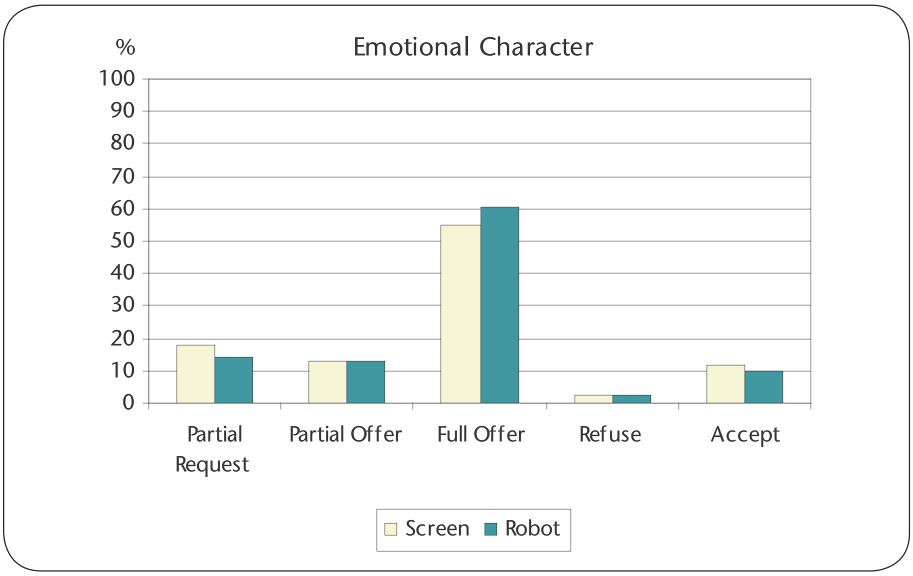

Figure 42: Negotiation behaviour in the emotional character conditions (SE, RE).

6.11. Discussion and conclusion

The speech recognition software used for this experiment showed two practical problems. First, the speech recognition in the robotic conditions (R,RE) decreased (see Figure 37) and second there was a considerable difference in recognition accuracy between the participants (see Table 16).

The decline in speech recognition accuracy in the robot conditions was most likely caused by the noise of the robot's motors. In the character condition (R) the motor for the panning of the head was randomly switched on and off. In addition to this, the motors of the eyebrow and lip were switched on and off irregularly in the emotional character condition (RE). It is to be expected that such irregular sounds are disturbing speech recognition software more than monotonous noise, such as a constant hum.

The considerable difference in speech recognition accuracy between participants might be related to the quality of the speech recognition program. It recognized only English and none of the participants was a native English speaker. The program might have had problems with the Dutch pronunciation of English language, even though all of them were fluent speakers.

To avoid the problems with the speech recognition accuracy it is recommendable to use the "Wizard of Oz" type of experiment in the future until the speech recognition software works considerably better.

Adding emotions to the character only affected measurements in the robot conditions and the effect size was rather small. The expressions did not influence the user's score even though the expression provided useful information for the negotiation task. A possible reason for this could be that the time the participants interacted with the character might have been too short to acquire a good understanding. Getting to know a human and learn his preferences and expressions takes considerably longer than the time available in the experiment. However, the emotional expressions certainly increased the novelty aspect.

In addition, an order effect was observed for the emotional expressions. It appears beneficial in terms of *user control* and *usability* to add emotions to a robot but taking them away afterwards has a negative effect.

The hypothesis that users perceive the interaction with a character that uses emotional expression more enjoyable than with a character that does not use emotional expressions should not be rejected. The results of the experiment suggest that the participants enjoyed the interaction with an emotional expressive character more than with non-expressive character, in particular if it is a robotic character.

The embodiment of the character (screen or robot) had no influence on the enjoyability profile, but affected the user's score. In the robot condition the participants acquired higher scores than in the screen condition. However, they did not change the frequencies of the different actions or the joint gain, which is the sum of the user' score and the character's score. This leads to the interpretation that the well-known *social facilitation* effect, which says that the mere presence of another person will motivate the people to exert more effort in a task was observed. The robot character appears to have a stronger social facilitation effect than the screen character and therefore the participants put more effort into the negotiation.

Another interesting effect of the embodiment was that the speech recognition accuracy in the robot condition was below the screen conditions, and even decreased from the robot character condition (R) to the emotional robot condition (RE). However, it did not affect the rating on *usability* and *user control*, which it does in the screen conditions (S, SE). It appears that the participants were more forgiving in the robot conditions. In the screen condition they might have considered the speech recognition accuracy as a challenge.

Considering all the measurements this study comes to the conclusion to reject the hypothesis that user will perceive the interaction with a robotic character more enjoyable than with a screen character. It appeared that the participants did not enjoy the interaction with robotic character more than with a screen character. The robotic character embodiment, however, has other interesting effects, such as the *social facilitation effect* and a higher forgiveness of the participants for the speech recognition errors. These effects alone might make it worthwhile to use a robotic character for the ambient intelligent home.

7

Conclusions

7.1. Summary

A salient feature of the ambient intelligent home of the future will be the natural interaction between the home and its inhabitants through speech. An embodied home character could be the social entity necessary to provide natural dialogues. This character would have to be able to utilize the full range of communication channels, including emotional expressions, to give intuitive and continuous feedback to the user. The emotional expressions are based on a certain embodiment and two types of embodiment are frequently found in literature and products: screen characters and robotic characters.

Based on this, the following three research questions were derived:

1. How convincing are the emotional expressions of machines?
2. Will the user perceive the interaction with a character that uses emotional expression more enjoyable than with a character that does not use emotional expressions?

3. Will the user perceive the interaction with a robotic character more enjoyable than with a screen character?

To answer the first research question a literature review on emotional expressions of humans and machines through speech, music and body language was performed. Furthermore, a model for the convincingness of emotional expressions, based on Fogg and Hsiang Tseng (Fogg et al., 1999), was developed and tested. The experiment investigated also if the type of emotion (happiness, sadness, anger, surprise, fear and disgust), knowledge about the source (human or machine), the level of abstraction (natural face, computer rendered face and matrix face) and medium of presentation (visual, audio/visual, audio) of an emotional expression influences its convincingness and distinctness.

The results of the experiment show that emotional expressions of machines, including screen characters, are perceived as convincing as emotional expressions of humans. Moreover, the results showed that abstracted embodied characters are able to express convincing emotional expressions. Fogg and Hsiang Tseng's original mode of convincingness was not supported by the data of the experiment and was therefore revised. The new model of convincingness suggests that an emotional embodied character should express the right emotion at the right time with the right intensity.

The results of this experiment in combination with the results of a review on existing characters were applied in the design of a home character and its emotional facial expressions. This home character was used in a second experiment to investigate the two remaining research questions.

The results of the second experiment suggest that the interaction with an emotional expressive character was perceived more enjoyable than the interaction with a non-expressive character in particular if the character was a robot. However, the effect size was not as strong as being hoped for. This does not necessarily mean that emotional expressions are not beneficial. It is possible that, similar to the pronunciation of a sentence, the correctness of it is not perceived, but a fault would have a strong negative effect. An experiment with matching and mismatching emotional expressions would be necessary to test this aspect.

The embodiment of the character (screen or robot) had no influence on the enjoyability of the interaction, but a *social facilitation effect* was observed. The participants put more effort into the negotiation with a robotic character than with a screen character. Moreover, the robot embodiment made the participants more forgiving for speech recognition errors. These effects alone might make it worthwhile to use a robotic character for the ambient intelligent home because the speech recognition and synthesis are unlikely to become free of errors in the near future.

7.2. Further research

The measurements of the enjoyability profile, that was used for the experiment in Chapter 6, showed a lot of variance. It would be necessary to further improve the questionnaire for the enjoyability profile before further usage and investigate the relationships between its eight different concepts. The unavailability of a validated tool for the measurement of enjoyability makes clear that this research area is just at its beginning. Without such a tool it will remain difficult to compare the results of the various characters that are currently under development and hence monitor the progress of this research area.

In addition, it would be important to test interface characters over a longer period of time. The results of this study indicate a novelty effect, which means that the participants liked the characters because they were new to them. If the experiment was repeated with the same participants throughout several weeks the novelty would wear off and different results for the other measurements of the experiment might be obtained.

It would also be important to investigate how much attention the participants pay to the different characters. A qualitative analyses of the video recordings of the participants in the robot condition revealed that the participants looked at the robot more often if it expressed emotions. The participants' attention was divided between the screen and the robot. To perform such a test in the screen character condition eye tracking would be required. The attention of the participant might be related to the enjoyability profile, the negotiation behaviour and the negotiation performance.

A closely related question is whether the enjoyability of an interface character varies depending on how the user interacts with it, for example through speech, writing or point-and-click. An interface character might be perceived differently than in this study if the user interacts with it only through speech, assuming that the speech recognition worked properly.

In this study, the emotional expression provided useful information for the negotiation; still they had no influence on the user's score. The question remains if and how interface characters can offer useful functionality and in which applications domains this might be appropriate. Besides the entertainment and education domain little is know if and where else they could be useful.

7.3. eMuu in the homelab

In the introduction of this thesis it was argued that the ambient intelligent home is an appropriate application domain for interface characters. Philips Research opened a first prototype of such an ambient intelligent home, called HomeLab, on April 24th, 2002 (Aarts, 2002). It is interesting to think about how eMuu would perform if it would be placed in the HomeLab today, because it reveals the unresolved problems of eMuu and the HomeLab.

Clearly, eMuu would not yet be able to fulfil all the functions described in the scenarios (see Chapter 1.1). One can distinguish between open problems of eMuu and of the HomeLab itself.

eMuu is currently only able to emotionally evaluate the actions and events that occurred during the trading of stamps. It needs a massive extension of its goals and standards to be able to, for example, evaluate the complex behaviour of humans.

In addition, it would be necessary for eMuu to be able to process previously unobserved situations. At this point in time it can only evaluate and act within the limited area of the stamp trading game. It would not even be able to evaluate a new stamp or trade it. This limitation is based on the use of a world model that requires knowledge about each and every object in it. However, it cannot be expected that the users enter information about every object in the home.

An adaptive algorithm would be able to generalize and therefore process previously unobserved situations. It would be able to evaluate and trade previously unobserved stamps. In the context of the home it would be able, for example, to generalize the object shoe and therefore evaluate the new pair that the user bought that day.

eMuu also needs a considerably better hardware, since LEGO might be suitable for the first prototype, but certainly not for everyday use. LEGO is too unreliable and weak to operate for a longer period of time.

The HomeLab itself needs additional functionality to be able to interact with the user as described in the scenarios in Chapter 1.1. Speech recognition and synthesis are already operational and the HomeLab can execute direct commands such as “light on”. However, the recognition accuracy has to improve, particularly if additional sources of noise, such as playing children, are present. Performing a complex conversation with the user, as described in the scenarios, also requires an additional dialogue manager that is not yet available. In addition, the speech synthesis needs to be able to express emotions. It would be unconvincing if eMuu displayed a certain emotional facial expression, but talked with a neutral voice. Moreover, the user would possibly interpret eMuu's missing self-revelation of its emotional state as a lack of sympathy and commitment. If, for example, eMuu suggests to watch a certain TV program with a complete neutral voice, the user might interpret it as eMuu does not care about the user or its own recommendation.

The user interprets the fact, relationship, appeal and self-revelation features of each utterance from eMuu independently of whether they were actually sent or not. It is not possible for eMuu not to communicate these features. It therefore appears beneficial for eMuu to actively take control of all four features while communicating a message to the user.

In addition the HomeLab needs to have world knowledge, such as the nutritious values of various types of food (see scenario "dinner" in Chapter 1.1), to make recommendation to the user. Currently, the HomeLab has no knowledge about the existence and qualities of chocolate. Considering the complexity of the world an advanced knowledge management system would need to be implemented.

The butler role of eMuu has specific consequences for the relationship and self-revelation features of its voice messages. The butler role itself defines the relationship feature. The butler takes a hierarchically lower position than the user. Humans tend to be especially sensitive to this feature of messages (Schulz, 1981) and therefore eMuu would have to pay special attention to it.

7.4. Emotion Model

This study based the emotion model of the characters on the OCC model (Ortony et al., 1988). This model is intended to provide a cognitive model of human emotions. With this model events and objects can be evaluated, which result in specific intensity levels for each of the 22 emotional states, such as hope, fear love and hate. However, the interaction of emotional states is missing in the model. If, for example, a person evaluated a certain situation with the result of being mainly angry and next something positive happens, then the joy of the second event should decrease the remaining anger of the first event. This interaction of emotions is not described in the OCC model, but is necessary for the development of emotional characters. In this study, the interaction of the emotional states was implemented by mutually counteracting the positive and negative emotional states. Happiness decreased sadness and anger and vice versa. The user might have, for example, refused a profitable deal for the character, which makes the character very angry. Next, the user offers a slightly profitable deal to the character, which results in an increase of the character's happiness. The character is, however, still angry from the previous event and therefore the character becomes not suddenly happy, but less angry. However, this study used only a simplified OCC model. The interactions of the 22 emotional sates of the original model might be more complex.

Another problem is the missing mapping of the OCC model's 22 emotional states to the character's behaviour, such as its emotional facial expressions and its actions. Only a limited number of basic emotional facial expressions exists and are recognized independent of the viewer's cultural background (Ekman et al., 1972). The 22 emotional states of the OCC model need to be mapped to the 6 basic emotional expressions proposed by Ekman. A one-to-one mapping of the 22 emotional state from the OCC model to 22 different emotional facial expressions appears to be impossible, because it would be very difficult to create distinctive facial expression for closely related emotion states, such as gratitude and gratification. Moreover, only one positive emotional facial expression exists: the smile. All positive emotional states of the OCC model need to be mapped to this single expression. The user, however, might very well be able to distinguish between the expression of love and pride with the help of context information (see Figure 43). Each expression appears in a certain context that provides further information to the viewer. As mentioned earlier, the user might interpret the smile of a mother next to her son receiving an academic degree as pride, but exactly the same smile towards her husband as love.

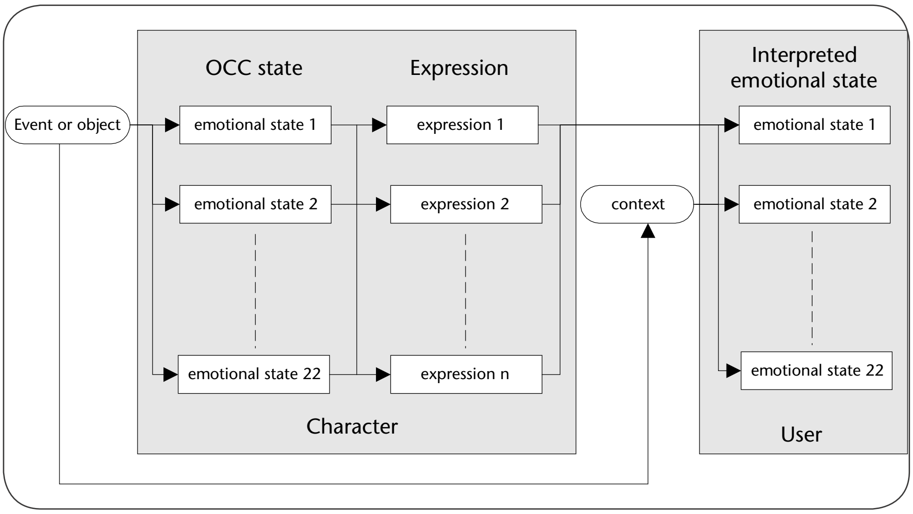

Figure 43: The process of emotional expression generation and the resulting perception

This mapping might be different for the different communication channels of the character. The distinct emotions that can be expressed through a face might be different from the ones that can be expressed through body language or speech. The mapping needs to be adjusted for each communication channel available to the character.

Another behaviour of the character that needs to be influenced by the emotional state is its actions. eMuu's actions were limited to the choice between five negotiation actions, such as making a full offer or refusing a deal. The mapping of the emotional state to negotiation behaviour was based on results of previous studies (Bazerman et al., 2000; Forgas, 1998; Carnevalve et al., 1986; Allred et al., 1997; Loewenstein et al., 1989). This theoretical foundation might not be available for every action that a character might be able to execute in the future and thus force the developer of the character to make up these mappings. This procedure has the intrinsic disadvantage that the developer might introduce a bias based on his or her own experiences and opinions. The character's behaviour should therefore be carefully tested to avoid disasters, such as the products of the Sirius Cybernetics Corporation (Adams, 1989). This fictional company creates lifts, doors, computers and androids with unbearable personalities. The android Marvin, for example, is manic-depressive and does not execute any task without grumbling about its misery.

Once the character's abilities to express emotions in all its communications channels and behaviours are known, the OCC model might be simplified to focus only on these emotional states. It is not necessary for the character to internally process events in the same way humans do. It is sufficient if the character's emotional model produces outcomes similar to those resulting from humans. Such a "Black Box" approach (Wehrle, 1998) simplifies the development process of characters, because a more simple emotion model requires less world knowledge from the character. This knowledge includes the meaning and emotional value of every relevant event and object in the character's world and has to be filled into the character's database by hand since an artificial character has no default world knowledge. The complexity of the emotion model therefore has a great influence on the development effort.

In general, the OCC model provides a good starting point to equip a character with an emotion model, but it falls short on suggestions on what to do with the internal emotional state. The mapping of the character's emotional state to its behaviour remains the responsibility of the developer of the character.

7.4. Emotional expressions

In this study the characters only expressed the three most distinctive emotions: happiness, sadness and anger. For a different application domain it might be useful for a one eyed character like eMuu to have a wider repertoire of expression. Figure 44 shows the six basic emotional expression proposed by Ekman (Ekman et al., 1972) applied to an archetypical one-eyed character. The happy expression is achieved by curving the eyebrow up and the lip down. For the sad expression the eyebrow and the lip is curved up. The angry expression uses a curved down eyebrow and a curved up lip. The surprised expression is created by opening the mouth, raising the eyebrow and curving it up. The fear expression is similar to the surprise expression, but it uses a less open mouth and an opened eye. The expression of disgust is difficult for this type of character because it does not have a nose. A raised nose and the resulting wrinkles next to the nose are essential features of human's expression of disgust. The next best thing is to use an iconic zigzag form for the mouth that often resembles disgust in comic figures.

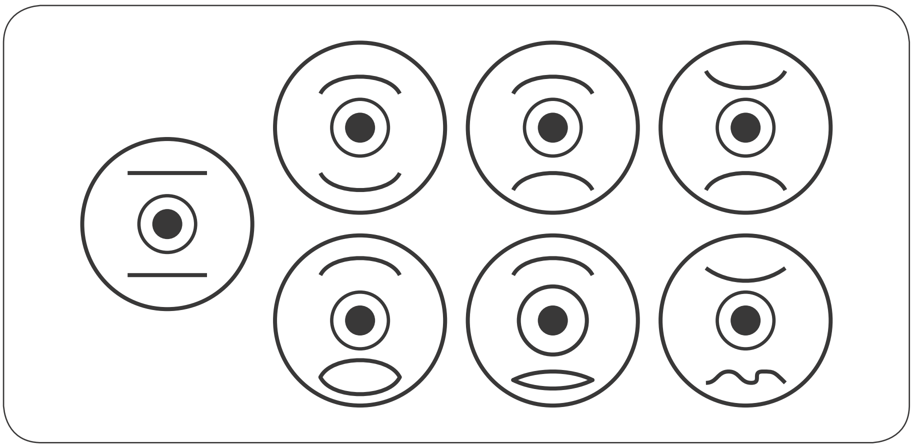

Figure 44: The six basic emotions of one-eyed characters. From left to right: neutral, happiness, sadness, anger, surprise, fear and disgust.

Many other characters, such as Kismet (Breazeal, 1999) or Pong (IBM, 2000) use two eyes and eyebrows. This enables them to imitate human faces slightly better. Figure 45 presents the basic configuration for eyebrows, eyes, nose and lips for an archetypical two-eyed character, based on the results of Chapter 2. The expressions are based on variation of the eyebrows, eyes, nose and lips. These variations are similar to the ones for one-eyed characters, the only difference being the expression for disgust that uses a raised nose, curved down eyebrows and an asymmetrical opened mouth.

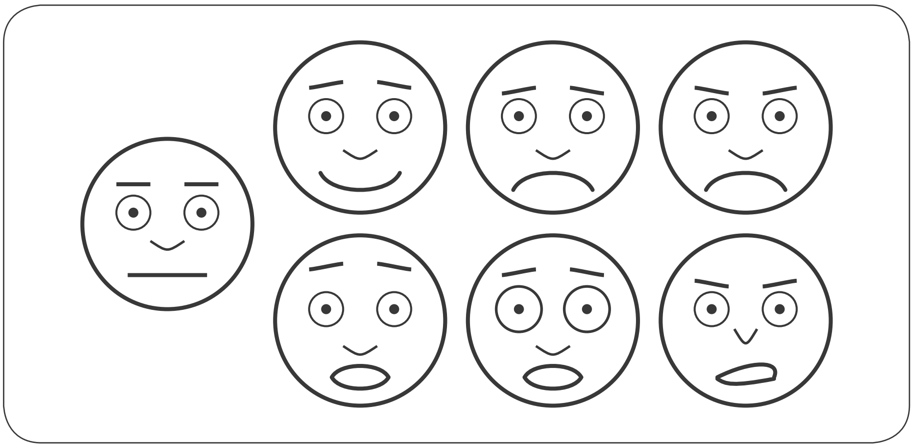

Figure 45: The six basic emotions of two-eyed characters. From left to right: neutral, happiness, sadness, anger, surprise, fear and disgust.

For both character types it is important to have various degrees of intensities for each of their expression in order to be convincing (see Chapter 2). In everyday life, humans are only rarely confronted with extreme situations that would evoke intense emotional expressions. Most of the time low or medium expressions are present. The same is likely to hold true for home characters. Most events and interactions with humans will result in low or medium intense emotional expression. Therefore special attention should be paid to the emotional expressions of characters at low intensity levels.

The results of Chapter 2 showed that the emotional expressions of happiness, anger and sadness are more distinct than surprise, fear and disgust. An earlier study (Bartneck, 2000) revealed that surprise is often mistaken for fear and vice versa. Disgust is often mistaken for anger, but not the other way around. The same effect can be observed in Figure 44 and in Figure 45. The expressions in the top rows are appear to be clearer and easier to identify.

In Chapter 3.3 various robotic characters are described. They can be roughly mapped into three groups: heads, pets and humanoids. It is intriguing that the robotic heads and the humanoid bodies have evolved considerably, but nobody combined the two of them yet. It appears to be an obvious move to mount a kismet-like head (see Figure 10) on an Asimo-like body (see Figure 13). An anthropomorphic character that is able to convey the complete range of emotional and conversational expressions can only arise through the combination of these two segments.

7.5. The future

Ultimately, it would be desirable to have a home character, such as the android Mr. Data of the TV series StarTrek (Paramount Pictures, 2002). His verbal skills are on the level of humans except for his problems with humour. His physical abilities, logical thinking and memory are superior to humans and he is not only able to work as the interface between the ambient intelligent home and the user, but also to support the user by cleaning the dishes. It would be almost lavish to deploy him exclusively for the tasks of butler. Many years will pass before such home characters will become widely available and maybe it will take as many years before Mr.Data will be able to feel emotions.

8

References

AAMAS2002 (2002). Workshop: Embodied conversational agents - let's specify and evaluate them. http://www.vhml.org/workshops/aamas/ [On-line].

Aarts, E. (2002). *Ambient Intelligence in HomeLab*. Eindhoven: Neroc.

Aarts, E., Harwig, R., & Schuurmans, M. (2001). Ambient Intelligence. In P.Denning (Ed.), *The Invisible Future* (pp. 235-250). New York: McGraw Hill.

Adams, D. (1989). *The Hitchhiker's Guide to the Galaxy*. (10th Anv ed.) New York: Harmony Books.

Allred, K. G., Mallozzi, J. S., Fusako, M., & Raia, C. P. (1997). The influence of anger and compassion on negotiation performance. *Organizational behavior and human decision process, 70,* 175-187.

Bandai (2000). Tamagotchi. http://www.bandai.com/ [On-line].

Barnes, A. & Thagard, P. (1996). Emotional decisions. In *Proceedings of the Eighteen's Annual Conference of the Cognitive Science Society , University of*

California (pp. 426-429).

Bartneck, C. (2000). *Affective Expressions of Machines*. Master Thesis Stan Ackerman Institute.

Bartneck, C. (2002). Marvin. http://www.bartneck.de/marvin [On-line].

Bartneck, C. & Okada, M. (2001a). eMuu - An Emotional Robot. In *RoboFesta Kansai* Osaka.

Bartneck, C. & Okada, M. (2001b). Robotic User Interfaces. In *Proceedings of the Human and Computer Conference (HC2001)* (pp. 130-140). Aizu.

Bazerman, M. H., Curhan, J. R., Moore, D. A., & Valley, K. L. (2000). Negotiation. *Annual review of psychology, 51,* 279-314.

Beach, L. R. (1997). *The psychology of decision making: people in organizations*. London: Sage.

Bell, G., Ling, d., Kurlander, D., Miller, j., Pugh, D., Skelly, T. et al. (1997). Lifelike Computer Characters: The Persona Project at Microsoft Research. In J.M.Bradshaw (Ed.), *Software agents* (pp. 191-222). London: AAAI Press.

Bornstein, M. H. (1982). *Development In Infancy: An Introduction*. New York: McGraw Hill.

Botelho, L. M. & Coelho, H. (1998). Adaptive agents: emotion learning. In *Workshop on Grounding Emotions in Adaptive Systems at the 5th International Conference of the Society for Adaptive Behavior (SAB'98)* Zurich.

Bouwhuis, D. G. (1991). Interactive Instructions as a Communicative Dialogue. In A. G. Elsendoorn & F. Coninx (Eds.), *Proceedings of the NATO Advanced Research Workshop on Interactive Learning Technology for the Deaf* (pp. 3-17). Berlin: Springer-Verlag.

Breazeal, C. (1999). Robot in Society: Friend or Appliance? In *Agents99 workshop on emotion-based agent architectures* (pp. 18-26). Seattle.

Brennan, S. E. (1985). The caricature generator. *Leonardo, 18,* 170-178.

Bresin, R. & Friberg, A. (1999). Synthesis and decoding of emotionally expressive music performance. In *Proceedings of the IEEE 1999 Systems, Man and Cybernetics Conference (SMC'99)* (pp. 317-322). Tokyo.

Bunt, H. C. (1989). Information dialogues as Communicative Action in Relation to partner Modelling and Information Processing. In F.N.Taylor & D. G. Bouwhuis (Eds.), *The Structure of Multimodal Dialogue* (pp. 47-73). Amsterdam: Elsevier Science Publishers.

C-Lab (University of Paderborn and Siemens) (2002). Mexi - A machine with emotionally extended intelligence. http://www.c-lab.de/home/de/news/2001/01-02-21-emotional_engine_lang.html [On-line].

Cahn, J. E. (1990). The generation of affect in synthesized speech. *Journal of american voice I/O society, 8,* 1-19.

Cañamero, L. D. (1997). Modeling motivations and emotions as a basis for intelligent behavior. In *Proceedings of the first international conference on autonomous agents* (pp. 148-155). Marina del Rey, CA.

Cañamero, L. D. & Fredslund, J. (2000). How Does It Feel? Emotional Interaction with a Humanoid LEGO Robot. In K.Dautenhahn (Ed.), *Socially Intelligent Agents: The Human in the Loop. Papers from the AAAI 2000 Fall Symposium* (Menlo Park: AAAI Press.

Carnevalve, P. J. D. & Isen, A. M. (1986). The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiation. *Organizational behavior and human decision process, 37,* 1-13.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, W. et al. (1998). Animated Conversation: Rule-based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. In M.Maybury & W. Wahlster (Eds.), *Readings in Intelligent User Interfaces* (San Francisco: Morgan Kaufmann Publishers.

Crowder, R. G. (1984). Perception of the major/minor distinction: I Historical and theoretical foundations. *Psychomusicology, 4,* 3-10.

Crucq, J. (2000). *Robotics* Eindhoven: Nat.Lab. Technical Note TN 200/454, Circulation Restricted.

CSLU (1999). CSLU Toolkit. http://cslu.cse.ogi.edu/toolkit/ [On-line].

Cunningham, J. G. & Sterling, R. S. (1988). Development change in the understanding of affective meaning in music. *Motivation and Emotion, 12,* 399-413.

Davis, J. B. (1978). *The psychology of music* . Stanford: Stanford University Press.

Dehn, D. & van Mulken, S. (2000). The Impact of Animated Interface Agents: A Review of Empirical Research. *International Journal of Human-Computer Studies, 52,* 1-22.

Diederiks, E. (2000). L-icons: Your Personal Friends in the Digital Television Era. In *Proceedings of the Philips UI 2000 Conference* (pp. 95-99). Eindhoven.

Diederiks, E. & Sluis, R. (2000). “Bello”, an Animated Character Facilitating Voice Control. In *Proceedings of the Philips UI 2000 Conference* (pp. 25-27).

Dijkstra, J. J. (1998). *On the use of computerised decision aids: An investigation into the expert system as persuasive communicator.* Ph.D. Thesis University of Groningen.

Doyle, P. (1999). When is a Communicative Agent a Good Idea? In *Workshop on Communicative Agents of the Third International Conference on Autonomous Agents* Seattle.

Ekman, P. (1973). *Darwin and facial expression : a century of research in review*. New York: Academic Press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face : guidelines for research and an integration of findings*. New York: Pergamon Press.

Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of personality and social psychology, 38,* 270-277.

Ekman, P. & Frieser, W. (1976). *Pictures of facial affects*. Palo Alto: Consulting Psychologist Press.

El Nasr, M. S., Ioerger, T. R., & Yen, J. (1999). PETEEI: a PET with evolving emotional intelligence. In *Proceedings of the third annual conference on Autonomous Agents* (pp. 9-15). Seattle.

Electronic Arts (2000). The Sims. http://thesims.ea.com/us/ [On-line].

Elliott, C. D. (1992). *The Affective Reasoner: A Process model of emotions in a multi-agent system*. Ph.D. The Institute for the Learning Sciences, Northwestern University, Evanston, Illinois.

Eruptor (2000). Portapets. http://www.eruptor.com/portapets/ [On-line].

Etcoff, N. L. & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44,* 227-240.

Feist, G. (1994). The affective consequences of artistic and scientific problem solving. *Cognition and emotion, 8,* 489-502.

Fogg, B. J. & Hsiang, T. (1999). The elements of computer credibility. In *Proceedings of CHI 99* (pp. 80-87). Pittsburgh.

Forgas, J. P. (1998). On feeling good and getting your way: Mood effects on negotiator cognition and bargaining strategies. *Journal of personality and social psychology, 74,* 565-577.

Frijda, N. H. (1969). Recognition of emotion. *Advances in Experimental*

Social Psychology, 4, 167-223.

Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.

Fujitsu (2002a). Chocoa. http://www.labs.fujitsu.com/free/chocoa/ [On-line].

Fujitsu (2002b). HOAP-1. http://pr.fujitsu.com/en/news/2001/09/10.html [On-line].

Fujitsu, I. (1997). Fin Fin. http://www.fujitsu.co.jp/jp/news/1997/Apr/finfin.html [On-line].

Gadanho, S. S. & Hallam, J. (1998). Emotion-Triggered Learning for Autonomous Robots. In *Workshop on Grounding Emotions in Adaptive Systems at the 5th International Conference of the Society for Adaptive Behavior (SAB'98)* Zurich.

Gerardi, G. M. & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception, 12,* 279-290.

Ginn, I., Heister, A., & Kohar, H. (1997). *Report of the Mediators Project. NL-Rep 6974* Eindhoven: Philips Research.

Grand, S., Cliff, D., & Malhotra, A. (1997). Creatures: Artificial Life Autonomous Software Agents for Home Entertainment. In *Proceedings of The First International Conference on Autonomous Agents* (pp. 22-29). ACM Press.

Hassenzahl, M. The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction,* (in press).

Heister, A. (1999). *Double Agent: design process of the anthropomorphic interface. NL-TN 354/98* Eindhoven: Philips Research.

Hollnagel, E. (1999). Keep cool: the value of affective computer interfaces in a rational world. In *Proceedings of HCI International '99* (pp. 676-685). Munich.

Honda (2002). Asimo. http://www.honda.co.jp/ASIMO/ [On-line].

Hoonhout, H. C. M. (2002). *Development of a rating scale to determine the enjoyability of user interactions with consumer devices* Eindhoven: Philips Research.

IBM (2000). Pong. http://www.almaden.ibm.com/almaden/media/image_pong.html [On-line].

IBM (2002a). Speech for Java. http://alphaworks.ibm.com/tech/speech

[On-line].

IBM (2002b). ViaVoice. http://www-3.ibm.com/software/speech/ [On-line].

Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.

Jordan, P. W. & Servaes, M. (1995). Pleasure in product use: beyond usability. In S.Robertson (Ed.), *Contemporary Ergonomics* (pp. 341-346). London: Taylor & Francis.

Jordan, P. W. & Servaes, M. (1998). Human Factors for pleasure in product use. *Applied Ergonomics, 29,* 25-33.

Jun Xiao, Stasko, J., & Catrambone, R. (2002). Embodied Conversational Agents as a UI Paradigm: A Framework for Evaluation. In *Proceedings of The First International Joint Conference on Autonomous Agents & Multi-Agent Systems. Workshop on: Workshop on Embodied conversational agents - let's specify and evaluate them!* Bologna.

Juslin, P. N. (1997a). Emotional Communication in Music Performance: A Functionalist Perspective and Some Data. *Music Perception, 14,* 383-418.

Juslin, P. N. (1997b). Perceived emotional expression in synthesized performances of a short melody: Capturing th listener's judgment policy. *Musicae Scientiae, 2,* 225-256.

Kabala, J. (2000). DEGOS: Interactive Visualization of a Self-Adapting System. In *Proceedings of the Philips UI 2000 Conference* (pp. 100-108). Eindhoven.

Kastner, M. P. & Crowder, R. G. (1990). Perception of major/minor: IV. Emotional connotation in young children. *Music Perception, 8,* 189-202.

Kellner, A., Martin, S., Philips, P., Portele, T., & Souvignier, B. (2000). Conversational User Interfaces for Consumer Electronics. In *Proceedings of the Philips UI 2000 Conference* (pp. 35-39). Eindhoven.

Kitano Symbiotic Systems Project (2002). Pino. http://www.symbio.jst.go.jp/~tmatsui/pinodesign.htm [On-line].

Kluger, A. N. & Adler, S. (1993). Person- versus computer-mediated feedback. *Computers in Human Behavior, 9,* 1-16.

Koda, T. (1996). *Agents with Faces: A Study on the Effect of Personification of Software Agents*. Master MIT Media Lab, Cambridge.

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of personality and social psychology, 61,* 743-754.

Laurel, B. (1991). *Computers As Theatre*. Boston: Addison-Wesley Publishers.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford: Oxford University Press.

LEGO (2002). Mindstorms. http://mindstorms.lego.com/ [On-line].

Loewenstein, G., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of personality and social psychology, 57,* 426-441.

Malone, T. W. & Lepper, M. R. (1987). Making learning fun: a taxonomy of intrinsic motivations for learning. In R.E.Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction* (Hillsdale, N.J: Lawrence Erlbaum.

Martinho, C. & Paiva, A. (1999). Pathematic agents: rapid development of believable emotional agents in intelligent virtual environments. In *Proceedings of the third annual conference on Autonomous Agents* (pp. 1-8). Seattle.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge: MIT Press.

Matsushita (1999). Kuma and Tama. http://www.matsushita.co.jp/corp/news/official.data/data.dir/en990324-2/en990324-2.html [On-line].

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago.

Menzel, P. & D'Aluisio, F. (2000). *RoboSapiens*. Cambridge: MIT Press.

Michita Imai (2002). Robovie. http://www.mic.atr.co.jp/~michita/everyday-e/ [On-line].

Microsoft (1999). Actimates: Barney. http://www.microsoft.com/Actimates/default.htm [On-line].

Microsoft (2000). Microsoft Agents. http://www.microsoft.com/msagent [On-line].

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review, 63,* 81-97.

Monk, A. (2002). CHI2002 Workshop: Funology: Designing Enjoyment. http://www-users.york.ac.uk/~djr14/CHI02_funology [On-line].

Mozziconacci, S. (1998). *Speech variability and emotions: production and perception*. Eindhoven: Technical University of Eindhoven.

Murray, I. R. & Arnott, J. L. (1992). Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the acoustic society of america, 93,* 1097-1108.

Nass, C. & Gong, L. (2000). Is Maximization or Consistency the More Social? The Case of Synthesized Voices and Faces. In *submitted to CHI 2000Conference* Den Hague.

Nass, C. & Reeves, B. (1996). *The Media equation*. Cambridge: SLI Publications, Cambridge University Press.

NEC (2001). PaPeRo. http://www.incx.nec.co.jp/robot/PaPeRo/english/p_index.html [On-line].

Nielsen, J. (1994). *Usability Engineering*. San Diego: AP Professional.

Nobe, s., Hayamizu, S., Hasegawa, O., & Takahashi, H. (1997). Are Listeners Paying Attention to the Hand Gestures of an Anthropomorphic Agent? An Evaluation Using a Gaze Tracking Method. In I.Wachsmuth & M. Fröhlich (Eds.), *Gesture and Sign Language in Human-Computer Interaction* (Heidelberg: Springer.

Norman, D. A. (1981). Twelve issues for cognitive science. In D.A.Norman (Ed.), *Perspectives on cognitive science* (pp. 265-295). Norwood: Ablex Publishing.

Nowlis, V. (1966). Research with the mood adjective check list. In S.S.Tomkin & C. E. Izard (Eds.), *Affect, cognition and personality* (pp. 352-389). New York: Springer.

Nunally, J. C. (1967). *Psychometric Theory*. New York: Mc-Graw Hill.

O'Reilly, W. S. N. (1996). *Believable Social and Emotional Agents*. Ph.D. Carnegie Mellon University, Pittsburgh, PA.

Oatley, K. & Jenkins, J. M. (1996). *Understanding emotions*. Oxford: Blackwell.

Okada, M. (2001). Muu: Artificial Creatures as an Embodied Interface. In *Proceedings of the ACM Siggraph 2001* (pp. 91). New Orleans.

Omron (2002). NeCoRo. http://www.necoro.com/ [On-line].

Ortony, A., Clore, G., & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurements of meaning*. Champaign: University of Illinois Press.

Paramount Pictures (2002). StarTrek. http://www.startrek.com/ [On-line].

PF Magic (2000). Petz. http://www.pfmagic.com/ [On-line].

Picard, R. (1997a). Hal's legacy: 2001's computer as dream and reality. In D.Stork (Ed.), *Hal's legacy: 2001's computer as dream and reality* (Cambridge: MIT Press.

Picard, R. W. (1997b). *Affective computing*. Cambridge: MIT Press.

Pixar (1998). Geri's game. http://www.pixar.com/shorts/shorts.html [On-line].

plan_b media (2000). imigos. http://www.planb-media.de/ [On-line].

Plutchik, R. (1980). *Emotion: a psycho evolutionary synthesis*. New York: Harper & Row.

Proudfoot, K. (2002). RCX Opcode Reference. http://graphics.stanford.edu/~kekoa/rcx/opcodes.html [On-line].

Rinkens, T. (2002). RCX Tools. http://rcxtools.sourceforge.net/ [On-line].

Rizzo, P. (1999). Emotional Agents for User Entertainment: Discussing the Underlying Assumptions. In *Proceedings of the International Workshop on Affect in Interactions held in conjuction with the AC'99, Annual Conference of the EC I3 Programme* Siena.

RoboFesta (2002). RoboFesta. http://www.robokansai.gr.jp/e_guide/Robofesta%20kansai%202001.htm [On-line].

RoboScience (2002). RoboDog RS-01. http://www.roboscience.com [On-line].

Roseman, I. J., Antoniou, A. A., & Jose, P. E. (1996). Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cognition and emotion, 10,* 241-277.

Russel, J. A. (1979). Affective space is bipolar. *Journal of personality and social psychology, 37,* 345-356.

Sanyo (2002). Sanyo T7S Type 1. http://www.sanyo.co.jp/koho/hypertext4/0203news-j/0325-1.html [On-line].

Scherer, K. R. (1979). Nonlinguistic vocal indicators of emotion and psychopathology. In C.E.Izard (Ed.), *Emotions in personality and psychopathology* (New York: Plenum.

Scherer, K. R. & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 4,* 331-346.

Schiano, D. J., Ehrlich, S. M., Rahardja, K., & Sheridan, K. (2000). Face to InterFace: Facial affect in (Hu)Man and Machine. In *CHI 2000 Conference Proceedings* (pp. 193-200). Den Hague.

Schlossberg, H. (1954). Three dimensions of emotion. *Psychological review, 61,* 81-88.

Schulz, F. v. T. (1981). *Miteinander Reden - Stoerungen und Klaerungen*. Reinbeck bei Hamburg: Rowolth Taschenbuch Verlag GmbH.

Sloman, A. (1999). Architectural requirements for human-like agents both natural and artificial. In K.Dautenhahn (Ed.), *Human Cognition And Social Agent Technology, Advances in Consciousness Research* (Amsterdam: John Benjamins Publishing Company.

Sloman, A. & Croucher, M. (1981). Why robots will have emotions. In *Proceedings IJCAI* Vancouver.

Smith, C. A. & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology, 48,* 813-838.

Solorzano, J. (2002). LeJos. http://lejos.sourceforge.net/ [On-line].

Sony (1999). Aibo. http://www.aibo.com [On-line].

Sony (2002). SDR-4X. http://www.sony.co.jp/en/SonyInfo/News/Press/200203/02-0319E/ [On-line].

Stienstra, M. & Hoonhout, H. C. M. (2002). TOONS Toys. Interaction toys as means to create a fun experience. In M. M. Bekker, P. Markopoulos, & M. Kersten-Tsikalkina (Eds.), *Proceedings of Interaction Design and Children* Eindhoven.

Studdard, P. (1995). *Representing Human Emotions in Intelligent Agents*. Master Thesis The American University, Washington DC.

Sun (2002a). Java 2 Platform. http://java.sun.com/java2/ [On-line].

Sun (2002b). Java Communications API. http://java.sun.com/products/javacomm/ [On-line].

Sun (2002c). Java Speech. http://java.sun.com/products/java-media/speech/ [On-line].

Sun (2002d). Java Swing. http://java.sun.com/products/jfc/tsc/index.html [On-line].

Thomas, F. & Johnson, O. (1981). *Disney animation: the illusion of life*. Walt Disney Productions.

Toshitada Doi (2002). Robodex. http://www.robodex.org/e/index.html [On-line].

Toyota (2002). POD. http://www.toyota.co.uk/infront/products/pod.htm [On-line].

Tsukuda Original (2002). Pino. http://www.tsukuda-original.co.jp/hp1116/omocha.html [On-line].

Wehrle, T. (1998). Motivations behind modelling emotional agents: Whose emotion does your robot have? In C.Numaoka, L. D. Cañamero, & P. Petta (Eds.), *Grounding Emotions in Adaptive Systems* (Zurich: 5th International Conference of the Society for Adaptive Behavior Workshop Notes (SAB'98).

ZMP (2002). Pino. http://www.zmp.co.jp/e/index.html [On-line].

9

Appendix

9.1. The instructions for the participants

Instructions

Welcome to this experiment about negotiation. We are interested in how people negotiate and are grateful that you participate in this experiment. The whole experiment will take approximately 1 hour. Besides your Thank-You-Present you have a chance to receive a bonus present, which will be handed out to the top 3 negotiators. See below for further details.

Outline of the experiment

After reading these instructions you have the chance to ask questions about the experiment. Afterwards you will start with a training negotiation session to familiarise yourself with the task and the interface. In a short pause you can again ask questions before you start with 3 negotiation sessions. After each session you will be asked to fill in a questionnaire.

Summary of the experiment:

Introduction Part

- Read the Instructions
- Ask questions
- Training session
- Ask questions

Main Part

- Negotiation Session 1
- Fill in questionnaire
- Negotiation Session 2
- Fill in questionnaire
- Negotiation Session 3
- Fill in questionnaire

The Negotiation Task

Introduction

Your task is to negotiate with a computer character to exchange international stamps. Your goal is to maximise your profit. The value (in Euro) that each stamp has for you is shown on a small value label on the bottom side of each stamp. This value might be different from your opponent's point of view. A certain stamp might be particularly valuable for you but worthless for your opponent and vice versa.

The sum of all the stamps in your possession is displayed in the score field at the bottom left of the board. You can increase your score by making profitable deals with your opponent. Since there are only 2 trading pads on each side you can negotiate about a maximum of two stamps from each side at a time. However, you can make multiple deals per session.

Your score is automatically saved and will be used to determine the top 3 negotiators, which will receive a bonus present.

The Board

The board on the screen (see figure below) has 2 parts divided by a line. The top part belongs to your opponent and the bottom part to you. Each part has 30 deposit pads, 2 trading pads, 4 action indicators and 1 turn indicator. In the beginning each party owns 20 stamps.

Procedure
Your opponent starts the negotiation round with moving stamps to the trading pads and choosing an action from the list below, which will be displayed on the action indicators in the middle of the board. Now it is your turn to move stamp(s) and choose one of the actions from the list below. Both, your opponent and you, can choose from the same list of actions below.

You move both, **the opponents and your own stamps**, by clicking once on it with the pen. If you click on a stamp in a trading pad then it will be moved back to the deposit pads of that side and when you click on a stamp in a deposit pad then it will be moved to a trading pad of that side. You can only move a stamp if there is a pad free in its destination (trading pad or deposit pad). If, for example, both of your trading pads are filled then you cannot move another stamp to it. You first have to move one stamp back to the deposit pads.

Your turn is completed when you speak out your desired action. You only need to speak the text on the action indicators, which will change its appearance to reflect your choice. The speech recognition software is not perfect and it might happen that your opponent did not understand your utterance. In that case your opponent will utter a short "hmmm?". Speak your utterance again and try to speak in the direction of the microphone with a natural loudness. When your opponent recognised your action correctly it starts with its turn.

If you and your opponent accepted a deal by choosing the "Make Deal" action on **the exact same stamps** in the trading pads, then the stamps in the trading pads are automatically exchanged and placed in deposit pads. At this time your score is automatically updated.

There is no limitation in the number of deals you can make, however, each negotiation session is limited in time.

Make Deal
You consider the stamps in the trading pads as a complete deal and would like to exchange them.

Requirements:

- A complete deal consists of at least one stamp on both sides.

Refuse Deal
You consider the stamps in the trading pads as a bad deal and do not want to exchange them.

Requirements:

- You can only choose this action if your opponent actually offered a complete deal, which is indicated by the his “Deal” action field being highlighted.

What do you offer?
You would like to know what your opponent would give you for the stamp(s) in your trading pad(s).

Requirements:

- There must be at least one stamp in your trading pads.
- At least one of the trading pads of your opponent must be free so that your request can be answered.

What do you want?
You would like to know what your opponent would want for the stamp(s) in its trading pad(s).

Requirements:

- There must be at least one stamp in your opponent’s trading pads.
- At least one of your trading pads must be free so that your request can be answered.

Any questions?

Instructions for the Questionnaire
You are going to fill in 1 questionnaire for each of the 3 negotiation sessions. Please rate your experience of only the **last** negotiation session. The term “opstelling” in the questionnaire refers to the **whole** negotiation system.

Use the cursor keys to choose your answer and the space bar to confirm your choice.

9.2. The infrared communication protocol

The protocol is designed to be flexible and extensible. It can, for example, be used for several RCXs and PCs and additional behaviour patterns can easily be added. The first byte of the protocol contains the destination of the message. All units in range will read a message send by one source. Therefore each message needs to have and identifier for whom the message is intended. Each unit has an internal identification number and only if it matches the message destination number the further content of the message is processed.

The next byte defines the action. In case of a request the RCX will return certain values. If an emotional state is send then the Eengine of the RCX will update and return the new estate, as mentioned above. The command action gives direct access to the motors, sensors and behaviour patterns. Byte 3-5 contains either specifications for the actions or the values for the emotional stimulus.

Table 24: The infrared communication protocol

Byte 1 Destination	Byte 2 Action	Byte 3 Data 1	Byte 4 Data 2	Byte 5 Data 3
1=RCX 1 (face)	1=request	1=e. state 2=alive 3=sensor values		
	2=e. stimulus	Happiness	Sadness	Anger
	3=command	1=motor a	1=stop 2=forward 3=backward	
		2=motor b	1=stop 2=forward 3=backward	
		3=motor c	1=stop 2=forward 3=backward	
		4=reset motors		
		5=behaviour pattern	1=look up 2=look down	
2=RCX 2 (head)	As above			
3=PC	As above			

The communication between the PC and the RCX has its own set of problems. First, the communication is unreliable. Therefore the robot character contained a missing response exception handling. Moreover, the PC was able to send any type and length of byte messages via the attached infrared tower, but LeJos on the RCX was only able to understand messages that comply to the OpCode standard (Proudfoot, 2002). This standard is used with LEGO's original firmware to send direct commands to the RCX, such as switch on motor A. LeJos receives this messages, but unlike the LEGO firmware, it does not execute the command. The infrared protocol required sending and receiving of 5 bytes and only OpCode 15/1d “Get versions” for requesting the ROM and firmware versions from the RCX has the required 5 byte parameters. Again, the OpCode is an empty skeleton that does not evoke any reactions in LeJos.

9.3. The dice game

The goal of the game is to maximize your points. Player A starts the game by rolling a 20 sided dice. In the first round, he has to pass the dice to the player B. Player B must decide either to accept the number and roll again or to pass. If he decides to pass he gets points, equal to the number on the dice and player A gets the double of that. If player B decides to roll the dice again than he must role more than the previous number. If he fails, player A gets the double amount of points of the last throw and player B gets nothing. If he succeeds it is player A’s turn, and so on.

If a player rolls a 20 then the opponent cannot roll more. Instead, he has to roll a 20 again. The starting player alternates (ABABABAB…).

10

Samenvatting (summary in Dutch)

Een markante noot in de muziek van de ambient intelligent home van de toekomst is de natuurlijke interactie tussen het huis en de bewoners door middel van spraak. Daarvoor is een sociaal wezen nodig dat de belichaming van het huis vertegenwoordigt, zodat een natuurlijke dialoog kan ontstaan. Dit personage moet alle facetten van communicatie kunnen gebruiken, en ook emoties kunnen uitdrukken, om intuïtieve en voortdurende feedback te kunnen geven aan de gebruiker. De uitingen van emoties vinden hun oorsprong in een bepaalde belichaming. Zowel in producten als in de literatuur komen we twee vormen van belichaming tegen, namelijk schermpersonages en robots. Uit deze gegevens hebben we de volgende drie onderzoeksvragen afgeleid:

1. Hoe overtuigend zijn emotionele uitingen van machines?

2. Vindt de gebruiker het prettiger te communiceren met een personage dat emoties uitdrukt, dan met een personage dat dat niet doet?

3. Vindt de gebruiker het prettiger te communiceren met een robot dat met een schermpersonage?

Om de eerste onderzoeksvraag te beantwoorden, hebben we een overzicht gemaakt van de literatuur met betrekking tot emotionele uitingen van mensen en machines door middel van spraak, muziek en lichaamstaal. Verder hebben we een model voor de overtuigingskracht van emotionele uitingen ontwikkeld en getest, gebaseerd op Fogg en Hsiang Tseng (Fogg et al., 1999). Het experiment onderzocht ook of de soort emotie (blijdschap, verdriet, boosheid, verrassing, angst of afschuw), kennis over de herkomst (mens of machine), het abstractieniveau (een natuurlijk gezicht, een computergegenereerd gezicht of een matrixgezicht) of het gebruikte medium (visueel, audio-visueel of auditief) van een emotionele uitdrukking invloed heeft op de overtuigingskracht en duidelijkheid.

De resultaten van het experiment laten zien dat emotionele uitingen van machines, inclusief schermpersonages, net zo overtuigend overkomen als emoties van mensen. Daarenboven kwam uit de resultaten naar voren dat abstract belichaamde personages overtuigende emoties kunnen uitdrukken. Het oorspronkelijke model voor overtuigingskracht van emotionele uitingen van Fogg en Hsiang Tseng werd niet door de verkregen gegevens ondersteund en werd daarom herzien. Het nieuwe model geeft aan dat een met emoties uitgerust personage de juiste emotie op het juiste moment met de juiste intensiteit moet gebruiken. De resultaten van het experiment, in combinatie met de resultaten van een analyse van bestaande personages, werden gebruikt om een huispersonage met emotionele gezichtsuitdrukkingen te ontwerpen. We hebben dit huispersonage gebruikt in een tweede experiment om de twee overblijvende vragen te onderzoeken. De resultaten van het tweede experiment geven aan dat de communicatie met een personage dat emoties uitdrukt als prettiger werd ervaren dan de communicatie met een personage dat dat niet doet, vooral als het personage een robot was. Het effect was echter niet zo sterk als we gehoopt hadden. Dit zegt niet noodzakelijkerwijze dat emotionele uitingen niet nuttig zouden zijn. Het is mogelijk dat, net zoals bij de uitspraak van een zin, correcte uitspraak niet wordt opgemerkt, maar fouten een sterk negatief effect hebben. Om dit te testen, zou een experiment nodig zijn met gepaste en ongepaste emotionele uitingen.

De soort belichaming (schermpersonage of robot) van het personage had geen invloed op hoe prettig de communicatie werd ervaren, maar een zogenaamd social facilitation effect werd wel waargenomen: deelnemers deden meer moeite voor de onderhandeling met robots dan met schermpersonages. Daarenboven waren de deelnemers eerder geneigd een robot te vergeven voor spraakherkenningsfouten. Deze effecten maken het de moeite waard om robots te gebruiken voor de ambient intelligent home, want waarschijnlijk zullen spraakherkenning en –synthese in de nabije toekomst nog niet vrij van fouten zijn.