Utrecht University
Faculty of Humanities
TNO Human Factors

Master of Science Thesis

How emotional should the iCat robot be?

A children's evaluation of multimodal emotional expressions of the iCat robot

by

Melanie Kroes

Utrecht, August 2008

Supervisors:

Dr. Ir. Gerrit Bloothooft
Utrecht University

Prof. Dr. Mark A. Neerincx
TNO Human Factors
Delft University of Technology

Drs. Rosemarijn Looije
TNO Human Factors

Contents

Contents i

1 Introduction 1
- 1.1 Research Question . . . 3
- 1.2 Relevance to AI . . . 3
- 1.3 Thesis overview . . . 4

2 The iCat robot 5
- 2.1 Capabilities . . . 5
- 2.2 Functionality . . . 6
- 2.3 Programming the iCat . . . 7

3 Emotions 9
- 3.1 Emotional expressions in the face . . . 10
  - 3.1.1 Emotions in human facial expressions . . . 10
  - 3.1.2 Emotions in artificial facial expressions . . . 11
- 3.2 Emotional expressions in speech . . . 13
  - 3.2.1 Emotions in human vocal expressions . . . 13
  - 3.2.2 Emotions in synthetic speech . . . 13

4 Experiment 17
- 4.1 Introduction . . . 17
- 4.2 Methods . . . 19
  - 4.2.1 Participants . . . 19
  - 4.2.2 Tasks . . . 19
  - 4.2.3 Conditions . . . 20
  - 4.2.4 Design . . . 21
  - 4.2.5 Procedure . . . 22
  - 4.2.6 Questionnaires . . . 23
  - 4.2.7 Data . . . 24

4.2.8 Materials . . . 25
4.2.9 Dependent variables . . . 26
4.3 Results . . . 27
4.3.1 Questionnaires . . . 27
4.3.2 iCat as electronic pet . . . 31
4.3.3 iCat as educator . . . 32
4.3.4 iCat as motivator . . . 33
4.4 Discussion . . . 34
4.4.1 Questionnaires . . . 35
4.4.2 iCat as electronic pet . . . 37
4.4.3 iCat as educator . . . 37
4.4.4 iCat as motivator . . . 38

5 Conclusions **39**
5.1 Suggestions for further research . . . 41

6 Acknowledgments **43**

Bibliography **45**

A Dialogs **51**

B Questionnaires **61**

C Questionnaire results **67**
C.1 Questionnaire B . . . 67
C.2 Questionnaire C . . . 70
C.3 Questionnaire D . . . 72

D Objective Results **73**
D.1 iCat as electronic pet . . . 73
D.2 iCat as educator . . . 75
D.3 iCat as motivator . . . 77

Chapter 1

Introduction

Nowadays, children like to play with electronic toys like computers and robots. But as children play with these toys, they develop their social skills less than when they play with other children [28]. Also, they do not get enough exercise, which could lead to the developing of obesity [11]. In this thesis, an electronic 'toy' is transformed into an electronic personal assistant for children. Such an assistant could have several functions which may include: learning children something about a topic are interested in, playing a quiz, motivating the children to do physical exercises, keeping a diary of the children's eating pattern in order to know whether they eat healthy enough, or remembering the children when it's time to go to bed. Additionally, such an electronic assistant is fun to interact with. These functions could be beneficial to children.

In order to be effective and appreciated, an electronic personal assistant should communicate in a natural way with its user [22]. To improve effectiveness and to make conversations between humans and electronic assistants more natural, the robot might need emotions, as emotions play a large role in everyday life [2, 17, 22, 24]. Almost all events in a human life triggers some emotion. For instance, if you hear that your cat has died, you will almost certainly be sad about it. Other events, like getting a birthday present, will result in being happy. Humans can show emotions in various ways, for instance in their facial and bodily expressions, but also in their speech.

In this thesis, it is investigated whether a personal assistant should show emotions in its facial and/or vocal expressions in order to be effective in its

communication with children. Several studies have already been done on emotional facial expressions of an electronic assistant [1, 3] and on emotional synthetic speech [5, 19, 26], but to combine emotional facial expressions with emotional speech is quite new. As an electronic assistant, the iCat robot was used. The iCat robot is a 38 centimeter high, cat alike robot, which is able to express emotions through its face and voice. The robot will be discussed in detail in chapter 2.

This thesis is part of the SuperAssist project [29]. This project develops personal assistants for patients and medical specialists in order to improve self care of the patient (see Figure 1.1). In the SuperAssist project, the user is called a 'patient', as this project focuses mainly on support for (chronically) ill people. Also, children and elder people are target groups of the SuperAssist project. In this thesis, the 'patient' is a healthy child of about eight years old. An electronic personal assistant could be beneficial to children in various ways (see first paragraph). Children of about eight years were tested in this study, as they become more and more independent from their parents at this age, and on the other hand still have enough imagination to think that the robot is intelligent. SuperAssist is partially financed by the Dutch government (SenterNovem). Project partners are Delft University of Technology (DUT), Leids University Medical Center (LUMC), research institute Netherlands Institute for Applied Science, Human Factors (TNO HF), and industrial partners Pemstar, Philips research, Science & Technology, and Sigmax PDA Solutions.

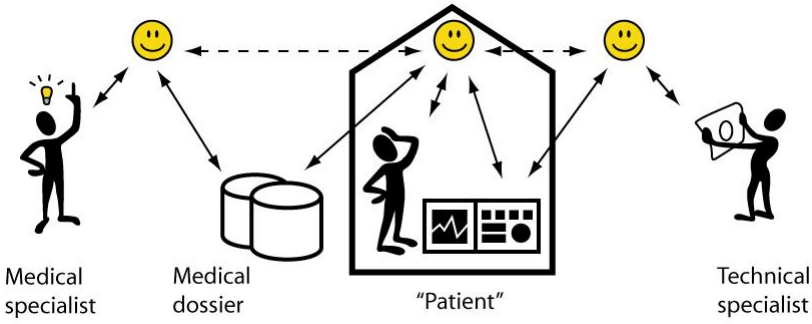


Figure 1.1: Interaction between patient and specialists as envisioned by the SuperAssist project. In this case, the 'patient' is a healthy child that interacts with the iCat robot. The 'medical specialist' and 'technical specialist' could for instance be the parents of the child, the family doctor or a teacher. In the so called 'medical dossier', not only medical information, but all kinds of information about the interaction between the robot and the child can be stored.

1.1 Research Question

The current study focuses on the iCat robot as an electronic embodied personal assistant. This robot was chosen, because it has shown good results earlier in the SuperAssist project [15, 16] and because this multimodal interface can show emotions both in its face and its voice. The research question of this thesis is the following:

> To what extent do emotions in facial and vocal expressions improve the iCat's effectiveness in convincing children to show certain behavior?

To study this question, an experiment was conducted with the iCat robot. Three conditions were compared with each other, namely a completely neutral iCat, an iCat that only showed emotions in its facial expressions and an iCat that showed emotions in its facial and vocal expressions. The iCat robot was assigned the roles of an *electronic pet*, an *educator* and a *motivator*. With each role of the iCat, participants had to perform certain tasks. The iCat as *electronic pet* was meant to see whether participants were able to recognize the iCat's emotional expressions. With the iCat as *educator*, the iCat told a story about which the participants had to answer questions afterwards. The iCat as *motivator* tried to motivate participants to undertake certain actions concerning exercising.

1.2 Relevance to AI

Artificial Intelligence is a very broad research area, which consists among others of psychology, computer science, philosophy and speech technology. This thesis includes at least three of those four topics. The psychology part of this thesis consists of emotional expressions and how to model them. In the past, emotion expression of humans has been studied extensively. This thesis builds on that information and to tried to base emotional settings for facial and vocal expressions on the way humans express emotions. The vocal expressions of the robot were fully artificial. An existing speech synthesis system was connected to an existing emotion editor in order to synthesize emotional speech. So, speech technology also plays an important role in this thesis. Finally, computer science is present in this thesis as an embodied robot was used which is a very useful device for implementing Artificial Intelligence.

1.3 Thesis overview

In the next two chapters, background theory for the experiment will be discussed. In chapter 2, the features and capabilities of the iCat robot will be presented. Chapter 3 is about emotions. It discusses two topics, namely the way in which humans express emotions in their face and voice, and how this can be translated into parameter settings for artificial faces and voices. Chapter 4 presents the current study (i.e. the methods, results and discussion). Finally, in chapter 5, general conclusions are drawn.

Chapter 2

The iCat robot

This study was conducted with the iCat robot. This robot was designed as a human-robot interaction research platform by Philips Research [23]. The iCat robot is a social interface which means that it can express itself emotionally through its facial and vocal expressions. There are two versions of the iCat robot, namely an embodied one and a virtual one. In this study, the embodied robot was used as previous research has shown that it is more attractive [15]. In this chapter, the physical capabilities and functionality within the scope of the SuperAssist project are described.

2.1 Capabilities

The iCat is a yellow, cat alike robot of 38 centimeter high (see Figure 2.1).

Figure 2.1: This picture shows the iCat robot. On the right, the front side of the robot is shown. On the left, the lower side of its back, where USB cables, power cables, microphone and earphones can be connected to the iCat, is shown.

The iCat robot has thirteen servo motors to control the following features:

- **2 x eyebrow:** the iCat robot can turn its eyebrows around a fixed point in the middle of the eyebrow. The left and right eyebrow can move independently from each other.

- **2 x eyelid:** the upper eyelids of the iCat robot can move in order to open and close its eyes. The left and right eyelid can move independently from each other.
- **3 x eyes:** the iCat robot's eyes can move horizontally as well as vertically. Horizontally, the left and right eye can move independently. However vertically, the eyes are both connected to one servo motor as it is very unnatural to be able to move the eyes vertically independently.
- **4 x lips:** each lip corner of the iCat robot is connected to a servo motor, which can turn up or down independently from the other lip corners.
- **1 x head:** with the servo motor connected to the head, the iCat robot can tilt its head up or down.
- **1 x body:** the iCat robot is able to turn its whole body to the left or right with the servo motor in its body.

Besides these servo motors, it has lights and touch sensors in both ears and on top of both paws and a speaker between its paws. Finally, it has a web cam installed in its nose.

With these features, the iCat is capable of expressing itself. Philips has implemented a number of emotional expressions which are based on studies by Ekman [9]. In a study by Bartneck, Reichenbach and Van Breemen the most extreme positions of the emotional animations made by Philips were extracted and it was tested how well these so called 'frozen' expressions were recognized [1]. Their results show that happiness and sadness were recognized best (90% each), followed by anger and surprise (80% each) and fear (30%).

2.2 Functionality

The focus of this thesis is on electronic personal assistants for children. With this target group in mind, the iCat robot could be used in several ways.

First, the iCat robot could remind children of things they have to do, such as going to school, making their homework and going to bed. Thus, the iCat functions as a sort of agenda or alarm clock. As children won't always listen

to their parents, their personal robot, which should be like a friend, may have more impact on them. Also, the parents can forgot some appointment their child has and then the iCat robot can remember both the child and parents of the appointment.

Second, the iCat robot could help the children to live healthy. It could keep a diary of what a child eats, drinks and how much it exercises. It could give advice how to improve the child's health, by giving the child (or his parent) advice about what the child could eat to stay healthy and by motivating the child to exercise. This could be very useful as children nowadays often don't eat healthy enough and don't get enough exercise, which both could lead to the developing of obesity [11].

Third, the iCat robot could learn children things about topics they are interested in, or help them with their homework. A child could pose questions to the robot, which it could answer. Or the iCat could give the child information on a certain topic. The robot could also test whether the child has learned something by playing a quiz. What has to be kept in mind for this function, is that the iCat robot needs a large database of information on topics the child could be interested in.

Fourth, the iCat robot should be fun to interact with. The functions described above are all very useful, but if the iCat robot does not have a fun element, children may loose interest over time and abandon it. Therefore, the iCat should also be capable of entertaining its user, for instance by playing games together.

Besides the functionality described above, many other things could be thought of. As long as the iCat has the appropriate physical capabilities, almost everything is possible.

2.3 Programming the iCat

The iCat is delivered with a program called OPPR (Open Platform for Personal Robotics^{TM}). With the OPPR software, the iCat's movements can be controlled. Included in OPPR is a library with standard actions (such as 'greeting and 'trying to get attention'), emotions (happy, angry, sad, etc.) and state transitions (like 'waking up' and 'going to sleep'). OPPR was controlled by the scripting language LUA.

In the software used for the current study, functionality of the iCat was defined in different agents, which are pieces of software that can act on their own or in combination with one another. For instance, a dialogExpertAgent controlled the dialog with the user. Communication between the agent software (in C#) and the OPPR animations and timing of the dialog is handled with LUA scripts.

This chapter described the capabilities, functionality and programming of the iCat robot. To properly function, the iCat robot should communicate with its user in a natural way [22]. One way in which the naturalness of the robot can be improved, is by adding emotions to its facial and vocal expressions. The next chapter describes how humans express emotions in their facial and vocal expressions and how this can be translated to emotional expressions of the iCat robot.

Chapter 3

Emotions

Emotions are very complex phenomena. It is difficult to find a good definition for emotion, as emotions are dependent on many things, such as how people feel (both mentally and physically), things that happen in the environment and other people's emotions. Another reason why emotions are complex, is the way people express them. Not only facial expressions are important for emotion, but also speech, attitude, body language and hand movements are used to express emotions [9].

Many studies on emotions can be found in the literature. Mainly, these studies focus on emotions in facial or vocal expressions. Although the number and names of emotion categories differ largely across these studies [6, 7, 10, 19, 21], these studies also show many similarities. All studies report on anger, fear, happiness and sadness as basic emotions. The 'Big Six' of basic emotions, defined by Cornelius are the previous mentioned (anger, fear, happiness and sadness) plus surprise and disgust [7]. Ekman also considers interest as a basic emotion [10]. In the current study, only the emotions that these studies agree on to be basic emotions (i.e. anger, fear, happiness and sadness) were used.

In this chapter, emotional expressions in facial and vocal expressions of humans is discussed. Humans can identify each others emotions well (around 80% correct for both speech an facial expressions [9, 25]). Children can also identify emotions quite well. At the age of two, they already reach above-chance levels for basic emotions (e.g. fear, happiness and sadness) and this develops as they grow older [32]. As humans (and also children) can identify each others emotions So, human emotional expressions could be a good

foundation for the definitions of artificial expressions. It is also discussed in this chapter how the facial and vocal emotional features of humans can be translated to artificial emotional expressions and how these expressions can be created.

3.1 Emotional expressions in the face

3.1.1 Emotions in human facial expressions

The way in which humans express emotions in their face has been studied extensively in the past, for instance by Ekman and Friesen [9] and Smith and Scott [27]. Smith and Scott made a literature overview on which parts of the face are involved in a number of emotions (see table 3.1).

	Facial action						
	Eyebrow frown	Raise eye-brows	Raise upper eyelid	Raise lower eyelid	Lip corners	Open mouth	Raise upper lip
Emotion							
Happiness				X	Raise	X	
Surprise		X	X			X	
Anger	X		X	X			
Disgust	X			X			X
Fear	X	X	X			X	
Sadness	X	X			Lower		

Table 3.1: Important components of some widely recognized facial expressions (adopted from Smith and Scott [27]). A cross (X) indicates that the specific facial action occurs while expressing a certain emotion. For instance, when someone is surprised, he will raise his eyebrows.

Obviously, these components are not the only things that occur in the face while expressing emotions. Many more things occur in a human face when emotions are expressed. For example, if someone raises his eyebrows in surprise, not only the eyebrows move, but also wrinkles appear in the forehead. Other movements in the face are not named at all by Smith and Scott. According to Ekman and Friesen, 'nose-wrinkling' is a likely movement in the face that occurs with expressing extreme disgust [9]. They also discuss that a specific emotion can be expressed in different ways. For instance with the emotion anger, the lips could be pressed together firmly or the mouth could be opened.

Thus, it is very hard to capture a complete and unambiguous set of the facial movements for each emotion. Generalizations always have to be made in order to be able to model something and modeling emotions is no exception.

3.1.2 Emotions in artificial facial expressions

How well emotions can be expressed by artificial faces is largely dependent on the physical aspects of the artificial face. Most artificial faces have the possibility to move parts of their face in the same way humans do. But not all of the human facial muscles could be built in. For instance, 'wrinkling' of the forehead or nose are features that are not often present in synthetic faces as this is very hard to model.

To focus more specifically on the current study, it is now discussed which facial actions from table 3.1 the iCat robot is capable of, in order to express emotion in its face. The iCat robot is capable of frowning, but it can not raise its eyebrows and it cannot have wrinkles in its forehead. Further, the upper eyelids can be moved, but the lower eyelids cannot move. Because it has no possibility of moving its nose, nose wrinkling is not possible. As these lip corners are connected to four servo motors, the corners of the lips cannot move. However, when the lips are in a smile, the lip corners seem to stand higher than when the lips are curved the other way around (see Figure 3.1). It is possible to open or close the mouth, although pressing the lips very hard together is also not possible.

Figure 3.1: This picture shows the most extreme lip positions of the iCat robot. On the left side of the picture where the iCat has its lips curved upward, the mouth corners seem to be lower than on the right, where the iCat has its lips curved downward.

Despite these limitations, it was possible to define emotional expressions for the iCat robot. Philips comes with a set of standard emotional expressions, which are based on studies by Ekman [9]. Animation for a number of emotions have been made. The most extreme positions of these emotions were stored as 'frozen' emotional expressions. A pilot study (see below) has been done to test how well Philips' standard emotions and some additional expressions were recognized.

Pilot study on facial emotional expressions

As there existed animations of emotional expressions and also frozen variants of these (all developed by Philips), a pilot study was conducted with children of eight and nine years old (mean age = 8.36, SD = 0.5) to determine which of these should be used in the actual experiment that was also conducted with children of the same age. In addition to the existing emotional expressions, some frozen emotional expressions made by Philips were modified by the experimenters to make them more similar to Ekmans study [9]. In total, sixteen videos were made of the iCat robot expressing the emotions happiness, fear, anger, sadness, disgust and surprise. The videos of the iCat robot were shown individually to the children on a 15 inch laptop screen. Each video lasted about ten seconds, after which the child had to choose which of the emotions (forced choice) was expressed. The results (see also table 3.2) show that for happiness and surprise, the modified versions of the frozen emotional expressions were recognized best. For fear, the original animation showed the best recognition rates. For anger, both the frozen expressions and the modified versions of them were recognized best. For sadness, the animation and the original frozen variant showed the highest recognition rates.

Emotion	Variant	mean recognition	Standard deviation
Anger	animation	0.444	0.511
Anger	frozen	0.833	0.383
Anger	modified	0.833	0.383
Disgust	animation	0.500	0.515
Fear	animation	0.389	0.502
Fear	frozen	0.111	0.323
Happiness	animation (exited)	0.333	0.485
Happiness	animation (laugh)	0.111	0.323
Happiness	frozen	0.167	0.383
Happiness	modified	0.667	0.485
Sadness	animation	0.667	0.485
Sadness	frozen	0.667	0.485
Sadness	modified	0.611	0.502
Surprise	animation	0.389	0.485
Surprise	frozen	0.222	0.428
Surprise	modified	0.500	0.515

Table 3.2: This table shows the recognition rates and standard deviations of the sixteen video clips in the pilot experiment. For each emotion, different variants were already made by Philips (the animations and frozen variants). Some emotions were modified, in order to adjust them to the parameters Ekman defined for emotional faces [9]. Only happiness did not have an animation. The animations 'exited' and 'laugh' were used instead and it was tested how much these animations were annotated as happy.

In the actual experiment, only the four most basic emotions (anger, fear, happiness and sadness) were expressed by the iCat, as these four emotions

were the four most basic emotions and could both be modeled in the face and the voice. For anger and happiness, the modifications of the frozen expressions were used and for fear and sadness the original animations made by Philips. As the results showed that for anger it didn't matter whether the 'frozen' or 'modified' version was chosen and for sadness it didn't matter at all, the variants 'modified' and 'animation' respectively were chosen as then two 'modified' and two 'animations' were used in the experiment.

3.2 Emotional expressions in speech

3.2.1 Emotions in human vocal expressions

As with emotional expressions in the face, also emotional speech has been analyzed extensively [6, 19, 26]. The speech parameters that are affected by various emotions are generally grouped in four categories: pitch, timing, voice quality and articulation [6]. Murray and Arnott [19] made an overview of which parameters affect certain emotions and how they do so, based on the literature (see figure 3.3). In this overview, the three parameters starting with 'pitch' belong to the group of pitch parameters, 'speech rate' is a timing parameter and intensity belongs to voice quality. Other research shows high correlations with this overview [6, 26].

	Anger	Happiness	Sadness	Fear	Disgust
Speech rate	slightly faster	faster or slower	slightly slower	much faster	very much slower
Pitch average	very much higher	much higher	slightly lower	very much higher	very much lower
Pitch range	much wider	much wider	slightly narrower	much wider	slightly wider
Intensity	higher	higher	lower	normal	lower
Voice quality	breathy, chest tone	breathy, blaring	resonant	irregular voicing	grumbled, chest tone
Pitch changes	abrupt, on stressed syllables	smooth, upward inflections	downward inflections	normal	wide, downward terminal inflections
Articulation	tense	normal	slurring	precise	normal

Table 3.3: Summary of how human vocal emotion affects various speech parameters (adopted from Murray and Arnott [19]).

3.2.2 Emotions in synthetic speech

Above, it was shown which parameters in human speech are altered to make speech sound emotional. To make synthetic speech emotional in the same way, a translation of these parameters into parameters that can be manipulated in synthetic speech should be made. Which parameters can be changed in the synthesis process, depends highly on the synthesis technique.

In this section, first the techniques for neutral speech synthesis are discussed and after that, the emotion modeling techniques are described.

Synthesis techniques

To synthesize speech, different techniques are possible. With all of these synthesis techniques, emotional speech synthesis is possible, but with some techniques it is more complicated than with others. Possible speech synthesis techniques are:

- **Formant synthesis:** with this technique, speech is build from a rule set [5, 6, 18, 20]. When it comes to adding emotions to the speech, there is a high degree of control on the parameters as for every parameter new rules can be defined. However, to find appropriate rules for each emotion can be very difficult.

- **Diphone synthesis:** speech is generated by concatenating diphones. Diphones are very small human speech units which consist of the second half of one phoneme and the first half of the following phoneme. To make the speech sound emotional, pitch and duration parameters can be modified. Other parameters, like voice quality and articulation, cannot be manipulated [26].

- **Unit selection synthesis:** this synthesis method generates speech by concatenating diphones and larger speech units from a database with human speech. Emotion modeling with this method is difficult, because the large speech units make it impossible to change synthesis parameters directly. Emotion can only be modeled by recording a complete set of units for each emotion.

- **HMM-based synthesis:** with this synthesis method, first context dependent Hidden Markov Models are trained with a natural speech database. In the synthesis part, the context dependent HMMs are concatenated according to the text to be synthesized. Then, spectrum and excitation parameters are generated from the HMM. Finally, the excitation generation module and the synthesis filter module synthesize speech waveforms using the generated excitation and spectrum parameters [30]. An advantage of this approach is that emotion can easily be modeled. The only thing that has to be done is to train the system with emotional speech data. This however, can be a time consuming process.

Emotion manipulation

The neutral speech that is synthesized with the techniques mentioned above, can be manipulated in several ways in order to let it sound emotional. Which of the emotion manipulation techniques are applicable, is highly dependent on the synthesis technique used. Possible emotion manipulation techniques are:

- **Modifying parameters:** the synthesis parameters can be modified by manipulation rules for every emotion. The rules are often based on human emotional speech parameters [19, 26]. This method is applicable when speech is synthesized with diphone [26] or formant synthesis [6, 18]. A drawback of this method is that it can be difficult to find the optimal parameter settings.

- **Emotional speech units:** With unit selection or diphone synthesis, speech units can be recorded separately for each emotion [31]. This is a very costly process, because a speech database should be generated for every emotion. On the other hand, the synthesized speech sounds very natural.

- **Copy synthesis:** Speech parameters could be extracted from a natural speech utterance and directly copied into the speech synthesis process of the same utterance (copy synthesis) [26]. An advantage of this method is that synthesis parameter modifications highly resemble the parameter modifications of human speech, but as every utterance that has to be synthesized should be first recorded from human speech, this is also a costly method.

- **Voice conversion:** A neutral utterance (natural or synthetic) can be modified in order to make it emotional with rules learned from human speech by Gaussian Mixture Models (for voice quality), duration trees (for durations) and Hidden Markov Models (for pitch contours). This method is for instance described and tested by Inanoglu et al. [12]. With this method, manual analysis of human speech signals, which must be done with the first method described above, is no longer needed. However, the system has to be trained with human speech for every emotion.

As can be read in section 4.2.7, the speech in this study was synthesized using diphone synthesis (with Fluency TTS [8]), in combination with modification of the parameters with an emotion editor called EmoFilt, developed

by Burkhardt [4]. Diphone synthesis was preferred over formant synthesis as diphone synthesis sounded more natural and as it was practically not possible to get access to emotion manipulation programs for formant synthesis. Although unit selection synthesis results in more natural sounding speech than diphone synthesis, diphone synthesis was chosen as emotions could more easily be modeled. It was also the only synthesis method for which all the tools were available to make the speech sound emotional. The emotion editor EmoFilt was chosen as it appeared from a previous study that this technique resulted in the higher percentages of correct recognition for emotional speech compared to another emotion editor and copy synthesis [14] and because it was the only system with which Dutch speech could be made emotional.

This chapter discussed the way in which humans express emotions in their facial and vocal expressions and how this can be translated into artificial facial and expressions for the iCat robot. The next chapter describes an experiment with the iCat robot in which it is investigated whether these emotional expressions can improve its effectiveness.

Chapter 4

Experiment

4.1 Introduction

In the experiment, it was investigated to what extent the iCat robot needs to express emotions in order to be effective. The research question was:

> To what extent do emotions in facial and vocal expressions improve the iCat's effectiveness in convincing children to show certain behavior?

To answer this question, three iCat conditions were evaluated, namely a neutral iCat, an iCat with emotional facial expressions and neutral speech and an iCat with emotional facial expressions and emotional speech. It was not possible within the time scope of the experiment to test a fourth condition, namely an iCat with only emotional speech. This condition, and not the condition with only emotional facial expressions, has been left out because it is expected that it shows less differences with the neutral iCat than the condition with only emotional facial expressions [3].

Children were asked to evaluate the conditions. The reason that children were tested, is that these are one of the target groups of the SuperAssist project and children can be assisted by the iCat robot in many ways as discussed in chapter 1, for instance with doing their homework. Children of eight and nine years old were recruited, as children at this age have sufficient language proficiency to understand the iCat and become more independent from their parents. On the other hand, children of this age still have enough imagination to believe the robot is really helping them instead of the people who programmed it.

Each participant evaluated two iCats in different conditions by doing certain tasks. With each task, the iCats played different roles. The iCat played the roles of an *electronic pet*, an *educator* and a *motivator*. With the iCat as *electronic pet*, participants had to annotate the emotions of the iCat robot in order to see whether the emotions were recognized well. The iCat as *educator* was developed to test whether participants were able to learn something from the iCat robot. To test this, the iCat explained the rules of the sport korfball and afterwards asked question about these rules. To see whether participants can be motivated by the iCat while doing exercises, the iCat as *motivator* was developed. The iCat tried to motivate the participants while they were collecting marbles from a basket and while doing a running task. Performance on all of these tasks was registered. The tasks are described in more detail in section 4.2.2.

Besides the tasks the participants had to do, they also filled in some questionnaires. These questionnaires were conducted to see what the participants thought of the robot and the tasks they had to do and which of the two iCats they preferred. Also questions about acceptance, empathy and trust were posed.

The following hypotheses were formulated for this experiment:

1. The more emotional the iCat robot is, the higher the scores in the questionnaires. It was expected that, with the more emotional iCat, participants would:

   a) like the iCat robot more;
   b) find the iCat robot more intelligible;
   c) like the tasks more;
   d) find the tasks easier;
   e) more frequently do the tasks again;
   f) more frequently work with the robot again (acceptance);
   g) find it friendlier (empathy);
   h) trust it more (trust).

   Furthermore, it was expected that:

   i) a difference was perceived between the iCat robots;
   j) the more emotional iCat robot was preferred over the less emotional iCat robot

2. With a more emotional iCat robot, it was expected that the performance of the children on the tasks was better:

   a) with the iCat as *electronic pet*, emotions would be recognized better and faster;

   b) with the iCat as *educator*, more questions would be answered correctly and questions would be answered faster;

   c) with the iCat as *motivator*, more marbles would be found and more steps would be taken.

4.2 Methods

4.2.1 Participants

In the experiment, children of eight and nine years of age were tested (mean age = 8.5, SD = 0.514). Eighteen children (10 male, 8 female) participated in the experiment. The children were recruited from a primary school in Soesterberg. Participants received a gift voucher and a photograph of themselves with the iCat robot.

4.2.2 Tasks

The experiment consisted of three parts. In each part, the iCat robot played a different role and the children had to perform different tasks. The three tasks in the experiment were the following:

iCat as electronic pet The iCat robot pretended to be the participant's pet and expressed its emotions during a story that was told by a storyteller (a male human voice). The iCat robot showed emotions by expressing sentences that were semantically emotional. Semantically emotional sentences were used, because in emotional speech, the semantics of the sentences are often not neutral either. Five students judged the text semantics of the sentences and only if at least four of them agreed on the semantics, the sentences were used for synthesis with the same emotion. In the experiment, after the iCat robot had said something, participants had to report its emotion. The participants had to annotate eight expressions (two in each emotion category anger, fear, happiness and sadness) of the iCat in a forced choice test with five emotion labels (anger, fear, happiness, sadness and neutral). A small fragment from this task is displayed below. The two dialogs for this task can be found in appendix A (in Dutch).

Storyteller: When you enter the living room, your robot cat Sanne is sitting on the window-sill, looking outside.
iCat robot: The red cat has died. I do terribly miss him.

iCat as educator The iCat robot explained the rules of the sport korfball and afterwards asked questions about those rules. The sport korfball was selected as this was not a very popular sport and therefore it was expected that the foreknowledge of the participants was about equal for all participants. The rules told by the iCat come from an instruction manual for the sport korfball by the Royal Dutch Korfball Union (in Dutch: KNKV), especially written for eight to ten year old children [13]. After each question, participants had to choose the correct answer from three alternatives in a forced choice test. Feedback about the answers was provided to the participants by the iCat robot. An example of a question, the possible answers and feedback is presented below. The two dialogs for the iCat as educator can be found in appendix A (in Dutch).

Question: You start the game as a striker. For how long will you be a striker during this match?
Answers: A, during the whole match. B, before the break 10 minutes and after the break also 10 minutes. C, only before the break. After the break, you will be defender.
Feedback (right answer): Very good!
Feedback (wrong answer): Too bad, the right answer was answer B.

iCat as motivator The iCat tried to motivate participants to exercise in two exercising tasks. The first task was to do a search task. The children had two minutes to collect as many blue marbles from a large basket filled with richly colored marbles. The second task was to take as many steps as possible during two minutes. These steps were registered by a step counter. The iCat tried to motivate the participants with statements like "keep up the good job". These statements were given at fixed times (each 15 seconds) and did not depend on the performance of the participant. All motivating statements can be found in appendix A (in Dutch).

4.2.3 Conditions

In this experiment, it was tested whether there were differences in the interaction of the participants with the iCat between different conditions. The three test conditions were:

- an iCat with a neutral facial expressions and neutral speech (*neutral*). This iCat does not adapt to the situation at hand.

- an iCat which shows emotions corresponding to the situation in its facial expressions. However, the speech of iCat robot was neutral all the time (*face only*).

- an iCat which shows emotions corresponding to the situation in its facial expressions and its speech (*speech and face*).

Natural behavior, like blinking with its eyes and nodding with its head was already implemented in the iCat robot, because otherwise the robot would be very unnatural. Especially in the *neutral* condition, it would be unpleasant to work with, as in that condition, it would stare at the participant all the time. Blinking and nodding thus occurs in all of the three conditions.

4.2.4 Design

Each participant evaluated two out of the three iCat conditions and performed all tasks with both conditions. The design is presented in Table 4.2.4. In this table, the letters A, B and C, stand for the different combinations:

A *speech and face* versus *neutral*

B *neutral* versus *face only*

C *face only* versus *speech and face*

To balance the conditions across the tasks, the order of the conditions was reversed in half of the cases. In Table 4.2.4, the order of these conditions was reversed in the cases where '$^{-1}$' was added. For instance, 'A^{-1}' stood for *neutral* versus *speech and face* instead of *speech and face* versus *neutral*.

The order of the three tasks was balanced across participants, because judgments made about the iCats during an earlier task could influence the judgments in later tasks. As there were three tasks, there were six possible orders to present the tasks. The minimum number of participants needed (eighteen) was calculated by multiplying the number of task orders (i.e. six) with the number of combinations of conditions (i.e. three).

pp no.	Task 1	Task 2	Task 3
	pet	educator	motivator
1	A	A^{-1}	A
2	B	B^{-1}	B
3	C	C^{-1}	C
	pet	motivator	educator
4	A^{-1}	A	A^{-1}
5	B^{-1}	B	B^{-1}
6	C^{-1}	C	C^{-1}
	educator	motivator	pet
7	A	A^{-1}	A
8	B	B^{-1}	B
9	C	C^{-1}	C
	educator	pet	motivator
10	A^{-1}	A	A^{-1}
11	B^{-1}	B	B^{-1}
12	C^{-1}	C	C^{-1}
	motivator	pet	educator
13	A	A^{-1}	A
14	B	B^{-1}	B
15	C	C^{-1}	C
	motivator	educator	pet
16	A^{-1}	A	A^{-1}
17	B^{-1}	B	B^{-1}
18	C^{-1}	C	C^{-1}

Table 4.1: Design of the experiment. The letters A, B and C, stood for the different combinations of conditions (see text) and when a '$^{-1}$' was added, the order of the conditions was reversed. The order of the tasks was balanced across participants.

4.2.5 Procedure

In the experimental setup, it was taken into account that the attention span of children is limited. Therefore, the experiment as a whole did not last longer than one hour. Within this hour, children received instructions, filled in the questionnaires (for the contents of the different questionnaires, see section 4.2.6), and the actual experiment took place. The procedure of the experiment was the following (although the order of the tasks and the iCat conditions were different for each participant):

- Introduction by both iCats (2 minutes)
- Instruction and questionnaire A (10 minutes)
- Electronic pet task (14 minutes)
  - Electronic pet task with *speech and face* iCat (4 minutes)
  - Questionnaire B (2 minutes)
  - Electronic pet task with *neutral* iCat (4 minutes)
  - Questionnaire B (2 minutes)
  - Questionnaire C (2 minutes)

- Educator task (16 minutes)
  - Educator task with *neutral* iCat (5 minutes)
  - Questionnaire B (2 minutes)
  - Educator task with *speech and face* iCat (5 minutes)
  - Questionnaire B (2 minutes)
  - Questionnaire C (2 minutes)
- Motivator task (12 minutes)
  - Motivator task with *speech and face* iCat (5 minutes)
  - Questionnaire B (2 minutes)
  - Motivator task with *neutral* iCat (5 minutes)
  - Questionnaire B (2 minutes)
  - Questionnaire C (2 minutes)
- Questionnaire D (5 minutes)

4.2.6 Questionnaires

During the experiment, four different questionnaires were filled in by the participants. The first questionnaire (A), was a general questionnaire at the beginning of the experiment. The participants were asked what their expectations were, what they thought a robot should be capable of and what their foreknowledge about the sport korfball was. This last question was posed in order to know whether the children already knew something about korfball as this could have influenced the results with the iCat as educator.

Questionnaire B was filled in by the participants after each dialog with the iCat. As there were three tasks and both of the iCats did all tasks, these questions had to be answered six times during the whole experiment (see 4.2.5). Questionnaire B consisted of five questions, which could all be answered on a five point scale:

1. How nice did you find the robot?
   (1 = very boring, 5 = very nice)
2. How intelligible was the robot?
   (1 = very bad, 5 = very good)
3. How nice did you find the task?
   (1 = very boring, 5 = very nice)

4. How difficult did you find the task?
   (1 = very difficult, 5 = very easy)
5. Do you want to do the task again?
   (1 = certainly no, 5 = certainly yes)

After each of the three tasks (and thus after each second dialog with the iCat robot), questionnaire C had to be filled in (see 4.2.5). This questionnaire consisted of the questions whether participants noticed a difference between the two iCats during the task just finished and which one of the iCats was preferred to do the task with.

Finally, at the end of the experiment, it was asked whether the participants did like to work with the robots and which of the two iCats was preferred in general (questionnaire D). Further, questions about acceptance, empathy and trust were posed for each of the robots. In each category, one question was selected from a questionnaire used by [15]. Answers could again be given on a five point scale. Also, some additional remarks could be made at the end. As this did not happen very often, the experimenter asked questions like 'which task did you like the most?' and 'what do you think of the appearance of the iCat robot?' to trigger the participants to say something.

The questions of all questionnaires can be found in appendix B (in Dutch).

4.2.7 Data

In the experiment, the emotions anger, fear, happiness and sadness were used, as these basic emotions are studied extensively [10]. A practical argument for choosing these emotions is that only these four emotions could be modeled both in the synthetic speech and facial expressions of the electronic personal assistant.

Generating emotional and neutral speech

As was already discussed in section 3.2.2, diphone synthesis was used to synthesize speech. The Dutch speech synthesizer Fluency was used to generate the speech [8]. The speech was in Dutch, because Dutch speaking children were tested. Fluency was chosen, as this was the only Dutch diphone speech synthesis system that could be combined with an emotion editor (to synthesize emotional speech) and which supported Speech API 5 (which was needed in order to perform lip synchronization with the iCat robot).

To synthesize the neutral speech from text, only Fluency was needed. The synthesis of emotional speech was done with Fluency in combination with the emotion editor EmoFilt, which manipulates pitch and duration parameters [4]. EmoFilt is chosen as emotion editor, as this showed the best results in a previous study that made a comparison of different methods for synthesizing emotional speech [14]. The emotional speech was generated as follows:

- Fluency converted text input to a .pho-file (a file with phonemes and its durations in neutral speech)
- This .pho file was loaded into EmoFilt and one of the basic emotions was chosen. EmoFilt converted pitch and duration according to the basic settings of that emotion.
- The settings of EmoFilt can be scaled from 0 to 200 percent of the original settings. For Dutch, the scaling to 40 percent of the original settings sounded the most natural to three experimenters (independently) and was chosen to modify the .pho files.
- The modified .pho files were loaded into Fluency again and synthesized with the Fluency voice 'Diana'. Diana was chosen as a voice, as it appeared from earlier research that the iCat was mostly judged to be female [15]. Fluency has only two female voices and of these Diana sounded the most convincing.

Generating Facial Expressions

Facial expressions of the iCat robot can be controlled easily with its animation editor, a program in which each movable part of iCat can be controlled by setting its parameters. For each emotion, standard settings could be chosen from a menu. It is also possible to define animations and save them for later use. A pilot study was conducted (see section 3.1.2) and the emotions with the highest recognition rates (i.e. the original animations for sadness and fear, and the modified versions of the still emotions for happiness and anger) were used for the actual experiment.

4.2.8 Materials

During the experiment, the participant and experimenter were seated in different rooms. The participant was seated in a room that resembled a living room. Two iCats were placed on a kitchen table, one connected to a

laptop, the other to a pc. Between the two iCats, a web cam was placed on the table and connected to another pc. Behind the seat of the participant, two video cameras were placed high at the wall (see Figure 4.1).

In the experimenter room, there was one video screen on which the input of the two video cameras was displayed. There was one pc, to which two monitors were connected, one for each iCat robot. The pc was connected to the iCat pc and the iCat laptop via a remote desktop connection such that the iCat robots could be controlled from the experimenter room..

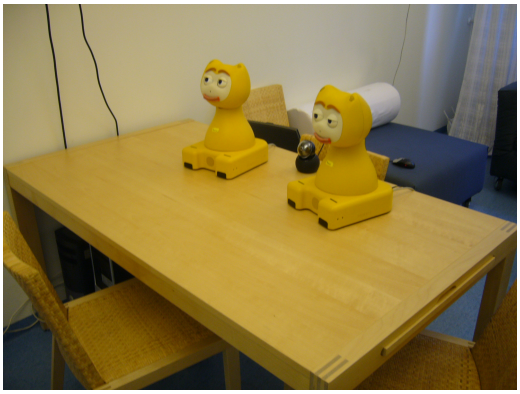


Figure 4.1: This picture shows the living room where the participant was seated during the experiment. The participant was seated on the chair at the left, in front of the two iCat robots and the web cam which stood on the table. Two pc's and a laptop were placed under and behind the table respectively and two video cameras were placed behind the participants, so that they were not distracted by them.

The experiment had a Wizard of Oz setup, which in this case meant that the input of the user (his or her speech) had to be typed in by the experimenter in the iCat program. The iCat robot does not yet have a speech recognition module implemented, and therefore it could not react directly to the participant's input.

4.2.9 Dependent variables

In all tasks, the answers to the questionnaires are used as dependent variables in the analysis. Besides these subjective judgments, also objective data is collected. With the iCat as electronic pet, the percentages of correct recognition of the emotions and the reaction times were analyzed. With the iCat as educator, the percentages of correctly answered questions and also the reaction times were analyzed and finally with the iCat as motivator, the amount of marbles collected and steps taken were z-scored and analyzed together.

4.3 Results

Results were analyzed with ANOVAs and Tukey HSD post hoc tests (if appropriate). The results were interpreted as if the experiment had a between subjects design. This, however, was not the case, as each participant evaluated two out of three conditions. Therefore, also T-tests were performed to compare the iCat conditions. but the T-tests results did not differ from the ANOVA results and are henceforth not reported.

4.3.1 Questionnaires

Questionnaire A

The participants came up with some interesting answers on the question what they thought that the robot should be capable of. One participant answered that she would take the robot to school to tell her all the answers to the questions of a test. Another participant said that he would like to play football with the iCat as keeper so that he would always score.

On the question about the foreknowledge on korfball, nobody answered that they were actually doing the sport korfball. Half of the participants answered that they had played it once or twice and the other half that they hadn't. Participants who already knew something about korfball did not answer more questions right than those who didn't ($F(1,178)=0.026$, $p=.87$).

Questionnaire B

For each question, all questionnaires B were first analyzed together in a Condition (3) x Task (3) ANOVA. Thereafter, these questionnaires were also analyzed for each individual task in a Condition (3) ANOVA. In this section, only relevant results are discussed. All results can be found in appendix C.

For question B2, significant results were found for Condition ($F(2,99)=10.416$, $p<.001$, see Table 4.2 for means and standard deviations) and Task ($F(2,99)=5.782$, $p<.005$). Post-hoc analyses showed that participants judged the intelligibility of the speech of the condition *speech and face* lower than the conditions *face only* ($p<.01$) and *neutral* ($p<.001$) and that the iCat was less intelligible as an educator than as a motivator ($p<.005$). For question B4, significant results were found for Task ($F(2,99)=5.605$,

p<.005). Here, post-hoc analysis showed that the educator task was judged more difficult than the motivator task (p<.005) and the electronic pet task (p<.05).

Condition	N	mean	SD
neutral	36	4,42	0,604
face only	36	4,22	0,681
speech and face	36	3,72	0,779

Table 4.2: means and standard deviations for intelligibility for each iCat condition on all tasks. The question was answered on a five-point scale.

Analysis of the questions of questionnaire B for each individual task with a Condition (3) ANOVA, showed significant results for question B2 (on intelligibility, see Table 4.3 for means and standard deviations) in the educator task ($F(2,33)=6.045$, $p<.01$) and in the motivator task ($F(2,33)= 3.307$, $p<.05$). With the educator task, the post-hoc analysis showed that the condition *speech and face* was less intelligible than the condition *face only* (p<.02) and *neutral* (p<.01). With the motivator task, the post-hoc analysis revealed that the *speech and face* condition was less intelligible than the *neutral* condition (p<.05).

		Electronic pet		Educator		Motivator	
Condition	N	mean	SD	mean	SD	mean	SD
neutral	12	4,42	0,669	4,17	0,557	4,67	0,492
face only	12	4,17	0,577	4,08	0,793	4,42	0,669
speech and face	12	3,92	0,669	3,25	0,754	4,00	0,739

Table 4.3: means and standard deviations for intelligibility for each iCat condition for each individual task. The question was answered on a five-point scale.

Questionnaire C

As with questionnaire B, all questionnaires C were also first analyzed together and then individually for each task . For question C1, "did you notice a difference between the two iCats with this task?", a Condition (3) x Task (3) ANOVA was performed, but no significant results were found, neither for the combination of conditions ($F(2,45)=0.530$, $p=.59$) nor for the tasks ($F(2,45)=1.894$, $p=.16$), but it is noteworthy that only six participants perceived a difference between the two iCats during the motivator task, while eleven participants perceived a difference between the iCats during the other two tasks (see Table 4.4. T-tests were performed to see whether there was a difference in the amount of choosing 'yes' or 'no'. In neither of the com-

binations of conditions, this difference was found (T(17)= -0.465, p=.65 for *neutral* versus *face only*, T(17)=0.000, p=1.00 for *neutral* versus *speech and face* and T(17)=0.939, p=.36 for *face only* versus *speech and face* (see Table 4.5 for the numbers of participants choosing the answers 'yes' or 'no').

Task	N	yes	no
Electronic pet	18	11	7
Educator	18	11	7
Motivator	18	6	12

Table 4.4: number of participants who did / did not perceive a difference between the two iCats with each task (generalized over condition combinations).

Conditions	N	yes	no
neutral versus *face only*	18	8	10
neutral versus *speech and face*	18	9	9
face only versus *speech and face*	18	11	7

Table 4.5: number of participants who did / did not perceive a difference between the two iCats with each combination of conditions (generalized over tasks).

Preference for one or the other iCat within a specific task (question C2) did not differ for the combinations of conditions *neutral* versus *face only* (T(17)=0.465, p=.64) and *neutral* versus *speech and face* (T(16)=-0.236, p=.82), but the *face only* condition tended to be chosen more often in combination with *speech and face* (T(16)=1.808, p<.1, see Figure 4.2). Within each task, there were no significant differences between the combinations of conditions on question C2.

Additionally, the preference of the participants who perceived a difference between the conditions in question C1 was analyzed. This did not lead to any significant results in preference for one or the other condition in each combination of conditions (p=1.00 for *neutral* versus *face only*, p=.76 for *neutral* versus *speech and face* and p=.13 for *face only* versus *speech and face*).

Questionnaire D

No significant differences were found in a Conditions (3) ANOVA on the questions about acceptance (F(2,33)=0.538, p=.59), empathy (F(2,33)= 1.269, p=.29) and trust (F(2,33)=0.040, p=.96), as all conditions scored very high on these questions (see Table 4.6).

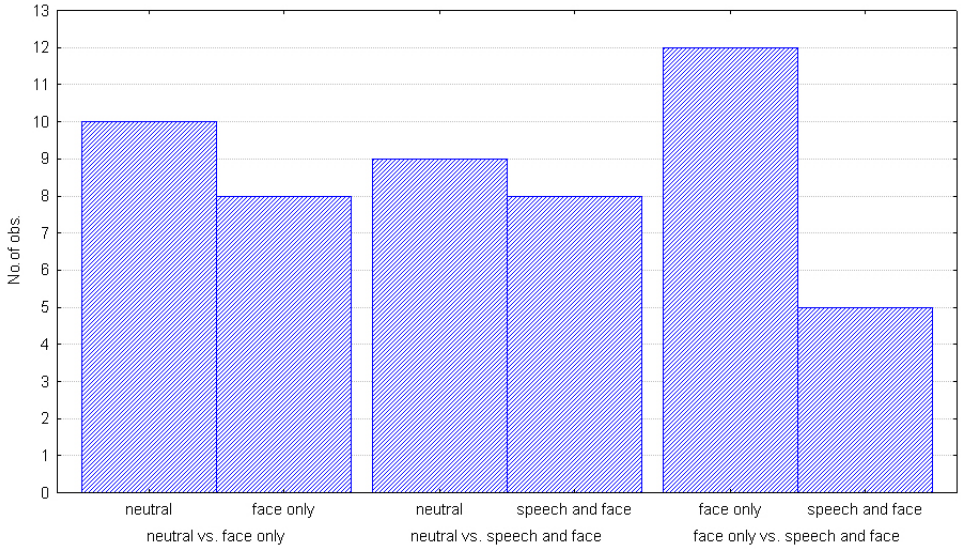


Figure 4.2: Number of participants that preferred one or the other iCat condition within a specific combination of conditions. Each of the three groups indicate one of the three possible combinations of conditions. Within each group, the number of participants that preferred one or the other condition was indicated. Only in the combination *face only* versus *speech and face* a trend was found that *face only* was preferred over *speech and face*.

On the question which condition participants preferred overall (question D8), the answers are presented in Table 4.7. These answers do not show noteworthy differences with the question about preference after each task (question C2).

		acceptance		empathy		trust	
	N	mean	SD	mean	SD	mean	SD
all conditions	36	4.64	0.593	4.83	0.447	1.97	0.810
neutral	12	4.67	0.492	5.00	0.000	1.92	0.900
face only	12	4.50	0,798	4.75	0.452	2.00	0.853
speech and face	12	4.75	0.452	4.75	0.622	2.00	0.739

Table 4.6: means and standard deviations for each conditions on the questions about fun, empathy and trust.

		preferred condition		
	N	*neutral*	*face only*	*speech and face*
neutral versus *face only*	6	3	3	-
neutral versus *speech and face*	6	4	-	2
face only versus *speech and face*	6	-	4	2

Table 4.7: Answers to question D8 ("which iCat do you prefer in general?"). This shows that in general none of the iCat conditions is preferred over another one.

The feedback of the children on the experiment was not analyzed, but some

remarkable answers were given. One participant was really concerned with possible applications for ill children and said that the iCat should need a jet pack to fly through a hospital to bring ill children their medicines. Other children were more concerned with the way the iCat robot should express emotion. One participant said that the iCat should have arms and legs in order to express emotions not only through its face and voice, but also via body language. Others also wanted the iCat robot to have arms or legs, but for completely different reasons, for example 'it needs legs to play football' or 'it needs arms to wave or shake hands with someone'.

The questionnaires did not reveal many significant results. Results found in the questionnaires are that the emotional speech was less intelligible than the neutral speech (question B2), that the educator task was more difficult than the motivator task (question B3) and that a trend was found that the *face only* condition was preferred over the *speech and face* condition.

4.3.2 iCat as electronic pet

Percentages of correct recognition

The mean percentage of correct recognition for all emotions together was 83.6%. Figure 4.8 shows a confusion matrix of the percentages of correct recognition for each emotion. Happiness was only confused with neutral. Anger, fear and sadness are sometimes confused with each other or with neutral.

		Perceived				
		Happiness	Fear	Anger	Sadness	Neutral
Intended	Happiness	**79.2%**	0%	0%	0%	20.8%
	Fear	0%	**81.9%**	5.6%	5.6%	6.9%
	Anger	0%	2.8%	**84.7%**	4.2%	8.3 %
	Sadness	0%	0%	5.6%	**88.7%**	5.6%

Table 4.8: Confusion matrix for all emotions. Sentences were intended to be happy, fearful, angry or sad and could be annotated with these labels as well as neutral. On the diagonal (in bold), the percentages of correct recognition are shown. In the other cells, the confusions with other emotions are shown. For instance, 20.8% of the sentences that were intended to be happy, were annotated as neutral.

The mean percentage of correct recognition in the condition *face only* was the highest (88.4%), followed by the condition *speech and face* (84.4%). Remarkable is that even the condition *neutral* had a percentage of correct recognition of 78.1%. In a Condition (3) ANOVA, it appeared that these differences were not significant ($F(2,285)=1.898$, $p=.15$).

Reaction times

The mean reaction time over all conditions was 5735 ms, with a minimum of 2856 ms and a maximum of 16038 ms. Three outliers (with reaction times more than three standard deviations higher than the mean reaction time) and one case in which no answer was given, were removed. The reaction times for the three conditions are shown in Table 4.9. No significant differences between the conditions were found in a Condition (3) ANOVA ($F(2,281)=0.293$, $p=.74$), but in a it was found that the reaction times are significantly shorter for correct answers than for incorrect answers ($F(1,282)=51.282$, $p<.001$).

Condition	N	mean	SD
neutral	12	5742	2239
face only	12	5859	2353
speech and face	12	5608	2182

Table 4.9: mean reaction times and standard deviations for each condition in the electronic pet task.

Reaction times and percentages of correct recognition in part one and two of the emotion recognition task were then compared in a Part (2) ANOVA to see if both parts had the same difficulty. It appeared that this was the case, as percentages of correct recognition ($F(1,286)=0.604$, $p=.44$) and reaction times ($F(1,282)=0.259$, $p=.61$) were not significantly different from each other.

No significant differences have been found between the iCat conditions with the iCat as educator. However, the reaction times were shorter for correct answers than for incorrect answers.

4.3.3 iCat as educator

Percentages of correctly answered questions

With the iCat as educator, the mean percentage of correctly answered questions was 69.4%. In the individual conditions, the percentages of correctly answered questions were as follows: in the *speech and face* condition 66.7% correct, in the *face only* condition 66.7% and in the neutral condition 75.0%. No significant differences were found in a Conditions (3) ANOVA on the amount of questions answered correctly ($F(2,177)=0.648$, $p=.52$).

Reaction times

It was observed that at first the reaction times did not significantly differ across the iCat conditions. But as the outliers (reaction times which were more than three standard deviations higher than the mean) were removed, the reaction times (see Table 4.10) differed significantly in a Conditions (3) ANOVA ($F(2,174)=9.950$, $p<.001$). Post-hoc analysis showed that in the *speech and face* condition, reaction times were significantly shorter than in both other conditions ($p<.001$). With the iCat as educator, a trend was found that correctly answered questions had shorter reaction times than incorrectly answered questions ($F(1,175)=2.800$, $p<.10$).

Condition	N	mean	SD
neutral	12	5938	1882
face only	12	5877	1682
speech and face	12	4685	1554

Table 4.10: mean reaction times and standard deviations for each condition in the educator task (outliers removed).

Furthermore, as with the iCat as electronic pet, the reaction times and percentages of correctly answered questions were both analyzed with a Part (2) ANOVA to see whether there were differences between part one and two of the task. Both for percentages of correct recognition ($F(1,178)=0.233$, $p=.63$) and reaction times ($F(1,175)=0.303$, $p=.58$), no significant differences have been found between the two parts of the task. Besides this, there was also no effect of foreknowledge about the sport korfball on the percentage of correctly answered questions ($F(1,178)=0.026$, $p=.87$).

With the iCat as educator, reaction times were significantly faster in the *speech and face* condition than in the other two conditions. Also, a trend was found that correct answers had shorter reaction times than incorrect answers.

4.3.4 iCat as motivator

Amount of marbles collected and steps taken

In the exercising tasks, the mean of marbles collected was 71.8 (SD = 12.24) and the mean of steps taken was 458.3 (SD = 131.8). These results have been analyzed together after z-scoring them, so that enough data was present in each condition. The mean z-scores for each condition are

shown in Figure 4.3.

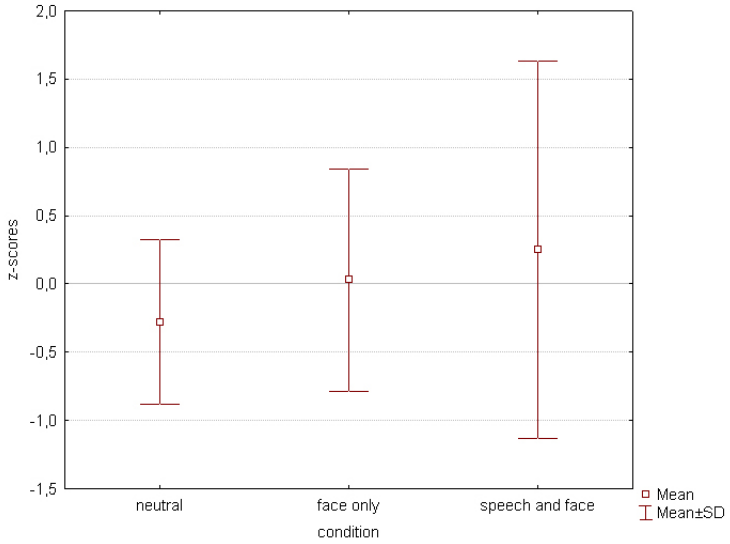


Figure 4.3: Z-scores (and standard deviations) on the exercising tasks. The results of collecting marbles and taking steps was z-scored in order to analyze these results together. The results lie in the expected direction (i.e. *neutral* scores lowest and *speech and face* highest), but results were not significant.

The z-scores were analyzed with a Conditions (3) ANOVA and no significant differences between the conditions were found ($F(2,33)=0.880$, $p=.42$), but the means for the conditions follow the hypothesis that the more emotional the iCat, the better the performance on the task.

4.4 Discussion

In the experiment, not many significant differences were found. It may be the case that the circumstances and tasks were not ideal for finding results between the different test conditions. The children were very motivated to participate in the experiment and therefore the emotions of the iCat could not motivate the participants more than they already were. Also, the test environment (i.e. a lab at TNO Human Factors) could have influenced the results as children were a bit aroused by this. Furthermore, the participants did like the tasks very much and did their best, even with the neutral iCat. Both on subjective answers about the robots and the tasks as well as on performance on the tasks, possibly a ceiling effect was reached and therefore it is almost impossible to find significant results.

4.4.1 Questionnaires

Questionnaire B

The first significant result found was not one that was expected, namely that the emotional speech of the *speech and face* condition was less intelligible than the neutral speech of the other two conditions. An explanation for this fact can be found in the parameter settings of the emotional speech. With the emotions anger, fear and happiness, the speech rate and pitch both go up compared to the neutral speech. It is likely that the faster speech rate easily could have caused the lower scores on intelligibility.

Looking at the intelligibility for each task individually, it was found that with the iCat as educator, also the emotional speech was judged less intelligible than the neutral speech. While playing the quiz, it is obviously important to understand the speech well and therefore children might have noticed a difference between the neutral and emotional speech. This was less the case with the other two tasks as the children could react directly or didn't react at all to what the iCat robot was saying. With the iCat as motivator the *neutral* condition was judged more intelligible than the *speech and face* condition, but the *face only* condition was not judged higher although it has the same speech (i.e. neutral) as the *neutral* condition. Apparently, the same speech has not been judged exactly the same although the differences between *neutral* and *face only* are not significant. Possibly the emotional movements of the head in the *face only* condition may make the speech a bit less intelligible which accounts for the fact that the difference between *face only* and *speech and face* was not significant.

Comparing the different tasks with each other, it was found that the quiz (with the iCat as educator) was judged more difficult than the other two tasks. This outcome is not very remarkable as with this task the participants had to remember about two minutes of explanations of the iCat robot and then had to answer questions. The other two tasks took less mental effort as participants could react directly to what the iCat was saying.

Questionnaire C

Questionnaire C revealed that participants did not often see differences between the two iCats they evaluated (question C1). This was especially the case with the motivator task. Twelve out of eighteen participants reported that they did not perceive a difference between the two iCats after this

task. This is in contrast with the other two tasks, where only seven out of eighteen participants reported that they did not perceive a difference. This could be explained by the fact that the two parts of the motivator tasks were already very different (i.e. collecting marbles and taking steps) and that it was therefore very hard to compare the iCats with each other. It was also observed that during the exercising tasks, the participants did not frequently look at the iCat during the task as they were busy with their task. With the other two tasks, the participants did look to the iCat more. An alternative explanation for the finding that participants often did not perceive a difference could be that they concentrated so much on the task the had to do, that they did not notice the differences between the two iCats.

As for question C2, which iCat was preferred, a preference for *face only* was found for the combination of *face only* versus *speech and face*. Between these two conditions, only the speech varied, so apparently neutral speech was preferred over emotional speech, possibly because it was more intelligible as stated in questionnaire B.

Questionnaire D

There were no differences between the conditions on the questions about acceptance, empathy and trust. To find differences on these categories, possibly more questions should be asked in each category. This has not been done during this experiment as the focus lay on the questions whether emotions could play a role in motivating children and therefore other issues like the intelligibility of the emotional speech were more important. So, to release the participants of too much question answering, the questionnaire about acceptance, empathy and trust was shortened. Another explanation for not finding significant differences here, could be that a ceiling effect was reached for these questions.

As in questionnaire C, which was filled in after each task, in questionnaire D the question "which iCat did you prefer in general?" also did not result in significant differences between the conditions. However, the condition *speech and face* has been chosen the least often, possibly because its emotional voice was less intelligible than the neutral voice in the other two conditions.

4.4.2 iCat as electronic pet

First of all, the percentage of correct recognition for the emotion recognition tasks is very high (83.6% correct). This could be the case because of the semantics of the sentences. All of the sentences in the experiment had emotional semantics and are synthesized neutrally and with the corresponding emotions. The high percentage of correct recognition for the *neutral* condition (78.1%) also supports the explanation that recognition was high because of the semantics.

As for the confusions between emotion categories (see Table 4.8): happiness was only confused with neutral and not with the other emotions (fear, anger and sadness). That happiness was not confused with fear, anger or sadness, could be explained by the fact that happiness is a positive emotion and fear, anger and sadness are negative emotions. Sometimes, fear, anger and sadness were confused with each other, which is also likely for the same reason, namely that they are all negative emotions. All emotions were sometimes confused with neutral, which could be explained by the fact that neutral both lies close to positive as well as negative emotions. Neutral is also an answer that was given often when the participants did not know the answer.

It was expected that the more emotions were shown, the higher the percentage of correct recognition would be. It was, however, the case that the percentage of correct recognition for *face only* was higher (although not significantly) than that of *speech and face*. As is argued before, the emotional speech was not always intelligible, and therefore it may have confused the participants a bit.

Reaction times were longer for incorrect answers than for correct answers. An explanation for this is that the participants hesitated and then randomly chose one of the answers which was more likely to be incorrect than to be correct.

4.4.3 iCat as educator

Although the percentages of correctly answered questions did not differ significantly, it is remarkable that the percentage of correctly answered questions in the *neutral* condition was somewhat higher than in the other two conditions. Maybe this was the case, because in the other two conditions the emotional expressions the head movements were a bit distracting.

It was found that the reaction times were significantly shorter for the *speech and face* condition than for the other two conditions. The emotional expressions aroused the participants and triggered them to answer the questions faster.

As with the iCat as electronic pet, it seems that the reaction times for correctly answered questions was shorter than for incorrectly answered questions, although with the iCat as educator only a trend was found.

4.4.4 iCat as motivator

The performance in the exercising tasks was the lowest in the *neutral* condition and the highest in the *speech and face* condition. Although the differences were not significant, the order of conditions found here is as expected; the more emotions are shown by the iCat, the better the performance on the task. As with all differences which were not significant, more participants need to be tested to see whether this result will become significant.

Chapter 5

Conclusions

The hypotheses of this thesis were:

1. The more emotional the iCat robot is, the higher the scores in the questionnaires. It was expected that, with the more emotional iCat, participants would:

   a) like the iCat robot more;

   b) find the iCat robot more intelligible;

   c) like the tasks more;

   d) find the tasks easier;

   e) more frequently do the tasks again;

   f) more frequently work with the robot again (acceptance);

   g) find it friendlier (empathy);

   h) trust it more (trust).

   Furthermore, it was expected that:

   i) a difference was perceived between the iCat robots;

   j) the more emotional iCat robot was preferred over the less emotional iCat robot

2. With a more emotional iCat robot, it was expected that the performance of the children on the tasks was better:

   a) with the iCat as *electronic pet*, emotions would be recognized better and faster;

   b) with the iCat as *educator*, more questions would be answered correctly and questions would be answered faster;

c) with the iCat as *motivator*, more marbles would be found and more steps would be taken.

From the results, it appeared that all iCat conditions were judged very positively on most questions in the questionnaires and that a ceiling effect was reached for most of the subparts of hypothesis one. Only on intelligibility and preference for one or the other iCat robot, significant results were found, but these were not supporting the hypothesis. For intelligibility, it was the case that the emotional speech of the *speech and face* condition was less intelligible than the neutral speech of the other two conditions. The 'preference'-question did not support the hypothesis, as a trend has been found that the face only condition was preferred over the *speech and face* condition.

As for the performance on the tasks, it has been found that participants answered the questions with the iCat as educator significantly faster with the *speech and face* condition than with the other two conditions. Additionally, although the result was not significant, an upgoing line was found with the motivator task from the condition *neutral* to *face only* to *speech and face*. Some support has thus been found for hypothesis 2 b) and c). The fact that no support for hypothesis 2 a) has been found could also be caused by a ceiling effect, because even in the *neutral* condition, the recognition of emotions was already very good, possibly because of the text semantics.

The research question of this thesis was:

> To what extent do emotions in facial and vocal expressions improve the iCat's effectiveness in convincing children to show certain behavior?

We have found positive as well as negative evidence. For the second hypothesis, some positive support was found, but for the first hypothesis none, and even some negative support was found. Still, while no definite answer is provided, some interesting accentuations can be made. The emotional speech of the robot was less intelligible than the neutral speech during this experiment, but if improvements could be made to the emotional speech (for instance by adjusting the speech rate), it might be found that emotion does improve the iCat's effectiveness. Furthermore, the ceiling effect, which occurred in most questions, might have occluded significant differences between the iCat conditions. If this could be avoided, for instance by testing the participants for a longer period or in various sessions, differences

might be found between the iCat conditions. Henceforth, as improvements could be made on the emotional iCat robot and the way of testing (see below), it is expected that the emotions of the iCat robot might improve its effectiveness.

5.1 Suggestions for further research

Some other suggestions for further research are:

- **Improve intelligibility of emotional speech** As the emotional speech was not always intelligible (or at least less intelligible than the neutral speech), improvements on intelligibility should be made. This could be done by slowing down all speech a bit. Then, also the neutral speech has to be slowed down, because the parameter settings of the emotional speech are defined relative to neutral speech.

- **Test more participants** If more participants were tested, it would be more likely to get significant results. In this study, eighteen participants were tested and as each participant evaluated two out of three conditions, only twelve data points were collected in each condition. This could be a reason that some results were not significant.

- **Test participants more than one timen and in a familiar environment** If there were more experiment sessions, it would be more likely that participants get used to work with the iCat robot. Then maybe not everything is fun and nice any more. Also, if tests were done in a familiar environment, the participants may be less aroused and motivated about participating. This could take away the ceiling effect in the questionnaires.

- **Do other tasks** As children were tested, the tasks were designed to be at least a bit of fun. However, as everything was much fun to them, no differences were found between the conditions. Maybe other tasks which are more difficult or less fun could be done so that the iCat really could motivate the participants.

- **More differences between conditions** To avoid that no differences are perceived between the robots, the conditions itself could be made more different. With the current study, for instance the robot blinked with its eyes and nodded now and then in all conditions (also in the *neutral* condition). If these natural movements are removed from the

neutral condition, there is already a larger difference between the neutral and the emotional iCats. Also could be thought of overacting the emotions as actors often do in films or plays, to make the emotional iCat even more emotional, but perhaps this will make the iCat less natural.

Chapter 6

Acknowledgments

First of all , I would like to thank the people who took care of my daily support at TNO Human Factors: Judith Kessens, Mark Neerincx and Rosemarijn Looije. I would also like to thank Willem Melder, Khiet Truong, Bert Bierman, Jeroen Jonkheer and Jan van de Kooij, who helped me very well getting the hardware and software ready for my experiment. Also thanks to Edwin Driessen, for being the storyteller in the 'iCat as electronic pet'-part of the experiment, Vincent de Lange, for helping me very much with programming the iCat and to all children for participating in my experiment. Without the help of Basisschool De Postiljon, this study would not have been a success, so I would like to thank them too. Off course I would like to thank the other interns at TNO Human Factors (Coraline Barends, Hanna Muis, Ksander de Winkel, Marloes van Welie, Karin Gerrits, Thijs Buurman, Thijs Brandsma, Koen Stegeman and Piet Schellens), who made my daily routine much more fun and who would always listen when something did not go as planned.

From Utrecht University, my thanks go to my first supervisor Gerrit Blothooft, who always had useful input for my experiment and was critical about the way of testing. Further, I would like to thank Noor Blaauw and the administration office for helping me with all the paperwork that comes with writing a master's thesis.

Finally, my thanks go to all the people who supported me in other ways to get my thesis finished: my parents, for calling now and then to ask how everything went, my friends for being there when I needed them and last, but certainly not least my boyfriend Tom. I would like to thank him for

supporting me in all ways he could, from calming me down when I was stressed to helping me with my statistical analyses. Many thanks to all of you.

Bibliography

[1] Christoph Bartneck, Juliane Reichenbach, and Albert van Breemen. In your face, robot! the influence of a characters embodiment on how users perceive its emotional expressions. In *Proceedings of the Design and Emotion*, Ankara, 2004. [cited at p. 2, 6]

[2] Joseph Bates. The role of emotion in believable agents. *Communication of the ACM*, 37(7):122–125, 1994. [cited at p. 1]

[3] Cynthia Breazeal. *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. PhD thesis, Massachusetts Institute of Technology, May 2000. [cited at p. 2, 17]

[4] Felix Burkhardt. http://emofilt.syntheticspeech.de/ (consulted on February 11th, 2008). [cited at p. 16, 25]

[5] Felix Burkhardt and Walter F. Sendlmeier. Verification of acoustical correlates of emotional speech using formant-synthesis. In *Proceedings of the ISCA Workshop on Speech & Emotion*, Belfast, Northern Ireland, 2000. [cited at p. 2, 14]

[6] Janet E. Cahn. Generating expression in synthesized speech. Master's thesis, Massachusetts Institute of Technology, May 1989. [cited at p. 9, 13, 14, 15]

[7] Randolph R. Cornelius. *The Science of Emotion. Research and Tradition in the Psychology of Emotion*. Prentice Hall, Upper Saddle River, NJ, 1996. [cited at p. 9]

[8] Arhtur Dirksen. Fluency text to speech. http://www.fluency.nl (consulted on February 11th, 2008). [cited at p. 15, 24]

[9] Paul Ekman and Wallace V. Friesen. *Unmasking the face*. Prentice Hall, New Jersey, 1975. [cited at p. 6, 9, 10, 11, 12]

[10] Paul Ekman, Wallace V. Friesen, and Phoebe Ellsworth. *Emotion in the human face: Guidelines for Research and an Integration of Findings*. Pergamon Press, New York, 1972. [cited at p. 9, 24]

[11] James O. Hill and John C. Peters. Environmental contributions to the obesity epidemic. *Science*, 280:1371–1374, 1998. [cited at p. 1, 7]

[12] Zeynep Inanoglu and Steve Young. A system for transforming the emotion in speech: Combining data-driven conversion techniques for prosody and voice quality. In *Interspeech 2007*, Antwerpen, Belgium, September 2007. [cited at p. 15]

[13] Koninklijk Nederlands Korfbalverbond. Korfbal, speel 't volgens de regels. http://www.knkv.nl/web/show/id=157092, consulted on 02-02-2008. [cited at p. 20]

[14] Melanie Kroes. Emotional speech synthesis: a comparison of different methods. December 2007. [cited at p. 16, 25]

[15] Vincent de Lange. icat as personal assistant for children with diabetes. Master's thesis, Delft University of Technology, 2007. [cited at p. 3, 5, 24, 25]

[16] Rosemarijn Looije. icat for you. a comparison between different personal e-health assistants. Master's thesis, Rijksuniversiteit Groningen, 2006. [cited at p. 3]

[17] Michael Mateas. An oz-centric review of interactive drama and believable agents. In M.J. Wooldridge and M. Veloso, editors, *Artificial Intelligence Today: Recent trends and developments*, pages 297–398. Springer Berlin/Heidelberg, 1999. [cited at p. 1]

[18] J.M. Montero, J. Gutiérrez-Arriola, S. Palazuelos, E. Enríquez, S. Aguilera, and J.M. Pardo. Emotional speech synthesis: From speech database to tts. *ICSLP*, 3:923–926, 1998. [cited at p. 14, 15]

[19] Iain R. Murray and John L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2):1097–1108, February 1993. [cited at p. 2, 9, 13, 15]

[20] Iain R. Murray and John L. Arnott. Implementation and testing of a system for producing emotion-by-rule in synthetic speech. *Speech Communication*, 16:369–390, 1995. http://www.computing.dundee.ac.uk/staff/irmurray/hamlet.asp. [cited at p. 14]

[21] Andrew Ortony and Terence J. Turner. What's basic about basic emotions? *Psychological Review*, 97(3):315–331, July 1990. [cited at p. 9]

[22] Byron Reeves and Clifford Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge University Press, 1996. [cited at p. 1, 8]

[23] Philips Research. http://www.research.philips.com/technologies/syst_softw/robotics/index.html (consulted on March 12th, 2008). [cited at p. 5]

[24] Daniel Rousseau and Barbara Hayes-Roth. Interacting with personality-rich characters. Technical Report KSL 97-06, Knowledge Systems Laboratory, Department of Computer Science, Stanford University, Stanford, CA, September 1997. [cited at p. 1]

[25] Marc Schröder. Can emotions be synthesized without controlling voice quality? *Phonus 4, Research Report of the Institute of Phonetics, University of Saarland*, pages 37–55, 1999. [cited at p. 9]

[26] Marc Schröder. *Speech and Emotion Research: An Overview of Research Frameworks and a Dimensional Approach to Emotional Speech Synthesis.* PhD thesis, University of Saarland, 2003. [cited at p. 2, 13, 14, 15]

[27] Craig A. Smith and Heather S. Scott. *The Psychology of Facial Expression*, chapter A Componential approach to the meaning of facial expressions. Cambridge University Press, 1997. [cited at p. 10]

[28] Kaveri Subrahmanyam, Robert E. Kraut, Patricia M. Greenfield, and Elisheva F. Gross. The impact of home computer use on children's activities and development. *The Future of Children: Children and Computer Technology*, 10(2):123–144, 2000. [cited at p. 1]

[29] SuperAssist. http://mmi.tudelft.nl/superassist (consulted on Februari 4th, 2008). [cited at p. 2]

[30] K. Tokuda, H. Zen, and A. Black. An hmm-based speech synthesis system applied to english. In *IEEE TTS Workshop*, Santa Monica, CA, 2002. [cited at p. 14]

[31] Oytun Turk, Marc Schröder, Baris Bozkurt, and Levent M. Arslan. Voice quality interpolation for emotional text-to-speech synthesis. In *Proceedings of Interspeech 2005*. Interspeech, 2005. [cited at p. 15]

[32] Sandrine Vieillard and Michèle Guidetti. Children's perception and understanding of (dis)similarities among dynamic bodily/facial expressions of happiness, pleasure, anger, and irritation. *Journal of Experimental Child Psychology*, (In press). [cited at p. 9]

Appendices

Appendix A

Dialogs

On the next few pages, the dialogs of the experiment are presented. As the experiment was in Dutch, the dialogs are also in Dutch. Every sentence in the dialog is preceded by the emotion of the iCat during that sentence. It depends on the condition of the iCat (i.e. *neutral*, *face only* or *speech and face*) if and how this emotion is shown.

The sentences in the pet dialogs that are not preceded by an emotion and printed in italic were not expressed by the iCat, but by a male storyteller. Comments are printed in bold and between brackets. Feedback which had to be provided by the participant is printed in bold.

Petdialog 1

neutral	Bij dit onderzoek doe ik net alsof ik jouw huisdier ben.		
happy	Ik ben net een echte kat, want ik eet brokjes en houd van spelen en slapen.		
neutral	Zometeen wordt er een verhaal verteld. Als ik dan iets zeg, moet jij daarna zeggen of ik blij, verdrietig, bang, of boos was, of dat ik geen emotie had.		
happy	Als je antwoord goed is, dan krijg je een punt. Probeer zoveel mogelijk punten te scoren.		
neutral	Nadat je een antwoord hebt gegeven, gaat het verhaal verder.		
neutral	Snap je wat je moet doen?		
	Ja	happy	Okee.
	Nee	neutral	Dan mag je je vraag zometeen stellen. **(proefleider komt binnen en kind krijgt extra uitleg.)**
neutral	Zullen we beginnen met het verhaal?		
	Ja	happy	Okee.
	Het is ochtend en je wordt wakker. Het zonnetje schijnt en je ziet dat jouw robotkat Roos in het zonnetje ligt.		
happy	Wat is het toch fijn om in het heerlijk warme zonnetje te liggen. **(kind rapporteert de emotie)** *Als ze jou ziet, komt Roos meteen op je af en begint ze om je heen te draaien.*		
sad	Ik wil buiten spelen, maar mijn baasje heeft per ongeluk het kattenluikje dichtgedaan. **(kind rapporteert de emotie)** *Je doet het kattenluikje voor haar open om haar naar buiten te laten. En kort nadat je haar naar buiten hebt gelaten, komt ze alweer naar binnen rennen.*		
fear	Ik durf absoluut niet meer naar buiten. Daar zit een grote enge hond. **(kind rapporteert de emotie)** *Dan gaat Roos in haar mandje liggen slapen. Je wilt haar niet storen en gaat met iemand buiten spelen. Als je s middags terugkomt, ligt Roos nog steeds te slapen. Je zet je nieuwste cd op en gaat dansen.*		
angry	Kan de muziek uit? Ik wil graag slapen. **(kind rapporteert de emotie)** *Je doet de muziek maar weer uit en gaat een boterham eten. Daarna wil je moeder gaan stofzuigen.*		
fear	De grote gele stofzuiger is eng. Hij maakt zoveel herrie. **(kind rapporteert de emotie)** *Roos vlucht naar boven. Als je later ook naar boven gaat, zit Roos boven aan de trap al op je te wachten.*		
sad	Ik voel me zo ontzettend alleen. Helemaal niemand wil met me spelen. **(kind rapporteert de emotie)** *Je gaat met Roos spelen. Even later is het eten klaar. Jullie eten vandaag zalm. Dat heeft Roos natuurlijk meteen door. Als jullie zitten te eten, komt ze naar de tafel.*		
angry	Ik vind het heel stom! Ik krijg alleen maar droge kattenbrokken. Ik wil ook een stuk vis. **(kind rapporteert de emotie)** *Eigenlijk mag het niet, maar stiekem geef je Roos toch een klein stukje vis.*		
happy	Dat is een lekker hapje eten. **(kind rapporteert de emotie)** *Tevreden gaat Roos in haar mandje liggen slapen.* **(iCat valt in slaap)**		

Petdialog 2

happy	En nu ben ik jouw huisdier.		
neutral	We gaan weer precies hetzelfde doen, alleen wordt er nu een ander verhaal verteld en zeg ik ook andere dingen.		
neutral	Weet je nog wat je moet doen?		
	Ja	happy	Okee, maar voor de zekerheid leg ik het toch nog even uit.
	Nee	neutral	Okee, dan leg ik het nog een keer uit.
neutral	Zometeen wordt er een verhaal verteld. Als ik dan iets zeg, moet jij daarna zeggen of ik blij, verdrietig, bang, of boos was, of dat ik geen emotie had.		
happy	Als je antwoord goed is, dan krijg je een punt. Probeer zoveel mogelijk punten te scoren.		
neutral	Nadat je een antwoord hebt gegeven, gaat het verhaal verder.		
neutral	Is het nu duidelijk wat je moet doen?		
	Ja	happy	Okee.
	Nee	neutral	Dan mag je je vraag zometeen stellen. **(proefleider komt binnen en kind krijgt extra uitleg.)**
neutral	Zullen we beginnen met het verhaal?		
	Ja	happy	Okee.
	Het is ochtend. En Als je in de kamer komt, zit Sanne in de vensterbank naar buiten te kijken.		
sad	De rode kater is dood. Ik mis hem verschrikkelijk. **(kind rapporteert de emotie)** *Dan ziet Sanne ineens een andere kat door de tuin heen lopen, die je nog nooit eerder hebt gezien.*		
angry	Wat doet die kat in de tuin? Ik zal hem snel wegjagen. **(kind rapporteert de emotie)** *Sanne rent naar buiten en jaagt de kat weg. Trots komt ze weer naar binnen. Even later ga je met je vader boodschappen doen.*		
fear	Blijf alsjeblieft hier om me te beschermen. Ik wil niet alleen thuisblijven. **(kind rapporteert de emotie)** *Eigenlijk heb je helemaal geen zin om mee te gaan en je vader gaat alleen boodschappen doen. Jij blijft bij Sanne en aait haar een beetje.*		
happy	Je bent een heel lief baasje voor mij. **(kind rapporteert de emotie)** *Je blijft haar dus maar een beetje aaien, maar je merkt dat Sanne eigenlijk van je schoot af wil springen. Je houdt haar tegen.*		
angry	Laat me direct gaan. Ik wil geen minuut langer op schoot zitten. **(kind rapporteert de emotie)** *Je laat haar dus maar weer gaan en gaat bij een vriendje of vriendinnetje spelen. Als je aan het eind van de middag thuis komt, wil je weer met Sanne gaan spelen. Ze wil alleen niet spelen en ze blijft eigenwijs in haar mandje liggen.*		
fear	Ik voel me niet lekker. Ik hoop niet dat ik naar de dierenarts moet. **(kind rapporteert de emotie)** *Je aait Sanne en geeft haar een pilletje waarvan ze beter zou moeten worden. Sanne slaapt. En als je na een uur weer bij haar komt, springt Sanne uit haar mandje.*		
happy	Ik voel me wel weer goed. **(kind rapporteert de emotie)** *Nou, gelukkig maar. Je was al bang dat je met Sanne langs de dierenarts zou moeten. En dan is het etenstijd. Als jullie aan het eten zijn, komt Sanne naar je toe.*		
sad	Ik heb ook honger, maar de brokjes zijn op. **(kind rapporteert de emotie)** *Je vult de bak met brokjes bij. En Sanne, die eet ze lekker op.* **(iCat valt in slaap)**		

Quizdialog 1

happy	Heb je zin om een quiz te spelen?		
	Ja/Nee		
happy	Het wordt een hele leuke quiz. Hij gaat over de sport korfbal.		
neutral	Weet je al iets over korfbal?		
	Ja	happy	Okee.
	Nee	neutral	Okee.
happy	Ik zal eerst even zeggen wat we gaan doen.		
neutral	Ik ga zometeen de regels van korfbal uitleggen. Jij moet dan goed opletten en proberen te onthouden wat ik allemaal vertel.		
neutral	Daarna stel ik een vraag, en geef 3 antwoordmogelijkheden. Jij moet zeggen welk antwoord goed is.		
neutral	In totaal stel ik 5 vragen.		
happy	Voor iedere vraag die je goed beantwoordt, krijg je een punt. Probeer zoveel mogelijk punten te scoren.		
neutral	Heb je de uitleg begrepen?		
	Ja	happy	Okee.
	Nee	neutral	Dan mag je je vraag zometeen stellen. **(proefleider komt binnen en kind krijgt extra uitleg.)**
happy	Ben je klaar voor de quiz?		
	Ja	happy	Okee.
neutral	Hier komt de uitleg over korfbal. Let goed op.		
happy	Korfbal is een teamsport voor jongens en meisjes.		
happy	Je speelt op een veld met twee vakken en in ieder vak staat een korf.		
happy	Een team bestaat uit vier jongens en vier meisjes.		
happy	Vaak zijn er ook wisselspelers.		
	(pauze van 3 sec)		
happy	Een wedstrijd duurt twee keer twintig minuten.		
happy	Samen probeer je doelpunten te maken en tegendoelpunten te voorkomen.		
neutral	Voor de rust mag je tien minuten aanvallen en tien minuten verdedigen en na de rust ook.		
neutral	Je wisselt in de rust van korf.		
	(pauze van 3 sec)		
angry	Een aanvaller mag niet op de korf schieten als je hem verdedigt.		
angry	Verdedigen doe je met 1 hand omhoog.		
angry	Je moet ook dicht genoeg bij je tegenstander staan.		
angry	Een jongen mag alleen een jongen verdedigen en een meisje alleen een meisje.		
angry	De speler met de bal mag maar door 1 andere speler verdedigd worden.		
	(pauze van 3 sec)		
neutral	De lijnen van het veld horen eigenlijk niet bij het veld.		
sad	Als je op de lijn staat met de bal in je handen, is de bal uit het veld.		
sad	De tegenstander mag dan ingooien.		
neutral	Je mag wel op de lijn staan als je de bal niet vast hebt of aanraakt.		
	(pauze van 3 sec)		
neutral	Als een aanvaller op de korf wil schieten en een verdediger maakt een overtreding, dan krijgt de aanvaller een strafworp.		
happy	De aanvaller mag dan, zonder dat iemand hem verdedigt, proberen te scoren op 2,5 meter van de korf.		
neutral	Alle andere spelers moeten 2,5 meter afstand houden van de korf en de strafworpnemer.		
	(pauze van 3 sec)		

angry	De scheidsrechter is iemand die er op let dat iedereen goed de regels toepast, bijvoorbeeld dat je niet loopt met de bal.		
angry	Als er een overtreding wordt gemaakt, legt de scheidsrechter het spel even stil.		
neutral	De scheidsrechter kan je dan uitleggen waarom hij heeft gefloten.		
neutral	Als de scheidsrechter niet weet wie als eerste de bal heeft gepakt, dan geeft hij een opgooibal.		
	(pauze van 3 sec)		
neutral	Dat was de uitleg.		
neutral	We gaan nu naar de vragen. Let goed op, want hier komt de eerste vraag.		
neutral	Vraag 1: Je begint de wedstrijd als aanvaller. Hoe lang mag je deze wedstrijd aanvallen?		
neutral	A, de hele wedstrijd		
neutral	B, voor de rust 10 minuten en na de rust ook 10 minuten.		
neutral	C, de hele eerste helft. De tweede helft moet je verdedigen.		
	B	happy	Goed zo.
	A/C	sad	Jammer, het goede antwoord was B.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 1')
neutral	Vraag 2: Wanneer mag je niet op de korf schieten?		
neutral	A, als je te ver van de korf af staat.		
neutral	B, als je verdedigd wordt.		
neutral	C, als je al 3 doelpunten hebt gemaakt.		
	B	happy	Goed zo.
	A/C	sad	Jammer, het goede antwoord was B.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 2')
neutral	Vraag 3: Mag je met 1 of 2 voeten op de lijn staan?		
neutral	A, nee, dat mag niet.		
neutral	B, ja, op de lijn mag wel, maar er overheen niet.		
neutral	C, ja, dat mag, als je de bal maar niet vast hebt of aanraakt.		
	C	happy	Goed zo.
	A/B	sad	Jammer, het goede antwoord was C.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 3')
neutral	Vraag 4: Wanneer is het een strafworp?		
neutral	A, als een verdediger een overtreding maakt, terwijl de aanvaller op de korf wil schieten.		
neutral	B, altijd als een verdediger een overtreding maakt.		
neutral	C, altijd als een aanvaller een overtreding maakt.		
	A	happy	Goed zo.
	B/C	sad	Jammer, het goede antwoord was A.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 4')
neutral	Vraag 5: Wanneer geeft de scheidsrechter een opgooibal?		
neutral	A, als hij niet weet wie de bal het eerste had.		
neutral	B, na een overtreding.		
neutral	C, als de bal uit het veld geweest is.		
	A	happy	Goed zo.
	B/C	sad	Jammer, het goede antwoord was A.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 5')
neutral	Dat was de eerste quiz.		

Quizdialog 2

happy	Zullen we nog zo'n leuke quiz doen?		
	Ja/Nee		
happy	Ik weet zeker dat je het leuk zult vinden.		
happy	Ik zal voor de zekerheid nog even uitleggen wat we gaan doen.		
neutral	Ik ga zometeen andere korfbalregels uitleggen. Jij moet dan goed opletten en proberen te onthouden wat ik allemaal vertel.		
neutral	Daarna stel ik een vraag, en geef 3 antwoordmogelijkheden. Jij moet zeggen welk antwoord goed is.		
neutral	In totaal stel ik 5 vragen.		
happy	Voor iedere vraag die je goed beantwoordt, krijg je een punt. Probeer zoveel mogelijk punten te scoren.		
neutral	Is de uitleg duidelijk?		
	Ja	happy	Okee.
	Nee	neutral	Dan mag je je vraag zometeen stellen. **(proefleider komt binnen en kind krijgt extra uitleg.)**
happy	Ben je er klaar voor?		
	Ja	happy	Okee.
neutral	Hier komt de uitleg over korfbal. Let goed op.		
happy	Als je een korfbalwedstrijd gaat spelen, kom je eerst met je team bij elkaar.		
happy	De coach vertelt hoe je moet spelen en waar je op moet letten.		
	(pauze van 3 sec)		
happy	Bij het begin van de wedstrijd lopen alle spelers het veld in en geven de tegenstanders in hun eigen vak een hand.		
happy	Ze wensen de tegenstander een prettige wedstrijd.		
happy	De thuispartij krijgt de bal en mag beginnen.		
neutral	Na de rust krijgt de bezoekende partij de bal en mag de tweede helft beginnen.		
	(pauze van 3 sec)		
happy	Je maakt een doelpunt als de bal van boven helemaal door de korf heen gaat.		
neutral	Na ieder doelpunt neemt de tegenstander de bal uit in het midden van het veld.		
angry	Volg de beslissingen van de scheidsrechter goed op.		
neutral	Winnen is leuk, maar verliezen moet je ook leren.		
sad	Ook al heb je verloren, dan nog kun je met je team een goede wedstrijd gespeeld hebben.		
	(pauze van 3 sec)		
neutral	Bij korfbal is het de bedoeling dat je samenspeelt om een doelpunt te kunnen maken.		
angry	Je mag niet lopen met de bal en alleen spelen mag ook niet.		
angry	Je moet de bal overgooien, dus je mag de bal niet aan een ander geven.		
	(pauze van 3 sec)		
neutral	Als je een overtreding maakt, krijgt de tegenstander een vrije worp.		
neutral	Iedereen moet bij een vrije worp 2,5 meter afstand houden van de speler met de bal.		
sad	Je mag niet meteen uit een vrije worp op de korf schieten.		
neutral	Als je team de bal over de lijn speelt, krijgt de tegenstander de bal.		
	(pauze van 3 sec)		
angry	Bij korfbal is ruw spel niet toegestaan.		
angry	Daarom mag je de bal niet uit de handen van de tegenstander slaan, je tegenstander niet aanraken en niet te dicht bij staan.		
angry	Wat ook niet mag, is de bal pakken als je valt of op de grond ligt.		
angry	En de bal met je voet of been spelen mag ook niet.		
neutral	Dat heet voetbal.		
	(pauze van 3 sec)		
neutral	Dat was de uitleg.		

neutral	We gaan nu naar de vragen. Let goed op want hier komt de eerste vraag.		
neutral	Vraag 1: De wedstrijd begint. Wie krijgt de bal?		
neutral	A, de partij die uit speelt.		
neutral	B, de partij die thuis speelt.		
neutral	C, de aanvoerder van de partij die thuis speelt.		
	B	happy	Goed zo.
	A/C	sad	Jammer, het goede antwoord was B.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 1')
neutral	Vraag 2: Wanneer telt een doelpunt?		
neutral	A, als de bal van boven naar onder helemaal door de korf is.		
neutral	B, als de bal van onder naar boven helemaal door de korf is.		
neutral	C, in beide gevallen, dus A en B zijn allebei goed.		
	A	happy	Goed zo.
	B/C	sad	Jammer, het goede antwoord was A.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 2')
neutral	Vraag 3: Mag je tijdens de wedstrijd de bal aan iemand geven?		
neutral	A, Nee, dat mag nooit.		
neutral	B, Ja, als je niet anders kunt.		
neutral	C, Ja, dat mag altijd.		
	A	happy	Goed zo.
	B/C	sad	Jammer, het goede antwoord was A.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 3')
neutral	Vraag 4: Wanneer wordt er een vrije worp gegeven?		
neutral	A, Als de scheidsrechter fluit.		
neutral	B, Als je een strafworp hebt gemist.		
neutral	C, Bij een bewuste overtreding in het aanvalsvak.		
	C	happy	Goed zo.
	A/B	sad	Jammer, het goede antwoord was C.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 4')
neutral	Vraag 5: Wanneer is het voetbal?		
neutral	A, Als de bal op je bovenbeen of knie komt.		
neutral	B, Als de bal op je onderbeen of voet komt.		
neutral	C, Als de bal op je been, knie, of voet komt.		
	C	happy	Goed zo.
	A/B	sad	Jammer, het goede antwoord was C.
	(opnieuw)	neutral	Ik zal de vraag nog een keer herhalen.
			(terug naar 'Vraag 5')
neutral	En dat was de tweede quiz.		

Exercisedialog 1

happy	Houd jij ook zo van knikkeren?		
	Ja	happy	Dan vind je deze oefening vast leuk.
	Nee	neutral	Jammer, maar deze oefening ga je vast wel leuk vinden.
neutral	Ik heb hier een grote mand met knikkers, en die mag jij zo gaan uitzoeken.		
neutral	Het is de bedoeling, dat je in twee minuten zoveel mogelijk doorzichtige blauwe knikkers, zoals deze, die hier voor mij ligt, uit de volle mand haalt, en in de andere mand doet.		
happy	Voor iedere goede knikker krijg je een punt.		
sad	Maar als je een verkeerde knikker in de andere mand hebt gegooid, gaat er juist een punt af.		
neutral	En je mag steeds maar 1 knikker tegelijk pakken.		
happy	Ik zal je aanmoedigen als je bezig bent.		
neutral	Ik zeg ook wanneer je mag beginnen en wanneer je moet stoppen.		
neutral	Snap je wat je moet doen?		
	Ja	happy	Okee.
	Nee	neutral	Dan mag je je vraag zometeen stellen. **(proefleider komt binnen en kind krijgt extra uitleg.)**
neutral	Als ik zometeen start zeg, dan mag je beginnen.		
neutral	Ben je er helemaal klaar voor?		
	Ja	happy	Okee.
happy	Start.		
	(pauze van 15 sec.)		
happy	Je doet het goed. Ga zo door.		
	(pauze van 15 sec.)		
sad	Kan je niet wat sneller?		
	(pauze van 15 sec.)		
happy	Wauw. Jij hebt al veel knikkers gevonden.		
	(pauze van 15 sec.)		
happy	Nog 1 minuut.		
	(pauze van 15 sec.)		
angry	Sneller. Sneller.		
	(pauze van 15 sec.)		
happy	Nog 30 seconden.		
	(pauze van 15 sec.)		
happy	Het gaat goed. Nog even volhouden.		
	(pauze van 15 sec.)		
happy	Stop maar. Je hoeft geen knikkers meer te zoeken.		

Exercisedialog 2

happy	We gaan nog een andere oefening doen.		
neutral	Je hebt net een stappenteller gekregen, en daar gaan we zometeen een oefening mee doen.		
neurtal	Je mag namelijk als ik het zeg, achter je stoel gaan staan, en dan 2 minuten lang proberen zo hard mogelijk te rennen, terwijl je toch op dezelfde plek blijft.		
neutral	Je hoeft dus niet de hele kamer rond te rennen.		
happy	Elke stap is 1 punt waard, dus ren zo hard mogelijk voor zoveel mogelijk punten.		
happy	Ik zal je aanmoedigen als je bezig bent.		
neutral	Ik zeg ook wanneer je mag beginnen en wanneer je moet stoppen.		
neutral	Is dat duidelijk?		
	Ja	happy	Okee.
	Nee	neutral	Dan mag je je vraag zometeen stellen. **(proefleider komt binnen en kind krijgt extra uitleg.)**
neutral	Ga maar vast achter je stoel staan.		
neutral	Als ik zometeen start zeg, dan mag je beginnen.		
neutral	Ben je er helemaal klaar voor?		
	Ja	happy	Okee.
happy	Start.		
	(pauze van 15 sec.)		
angry	Sneller. Sneller.		
	(pauze van 15 sec.)		
happy	Je bent goed bezig. Ga zo door.		
	(pauze van 15 sec.)		
happy	Wat kan jij hard rennen.		
	(pauze van 15 sec.)		
happy	Nog 1 minuut.		
	(pauze van 15 sec.)		
sad	Kan je niet wat sneller?		
	(pauze van 15 sec.)		
happy	Nog 30 seconden.		
	(pauze van 15 sec.)		
happy	Je bent er bijna.		
	(pauze van 15 sec.)		
happy	Stop maar. Je hebt het goed gedaan. Je mag nu weer gaan zitten.		

Appendix B

Questionnaires

This appendix shows the questionnaires the participants had to fill in. These questionnaires are in Dutch as the participants were Dutch children who did not speak English. Questionnaire A is a general questionnaire which was conducted at the beginning of the experiment. Questionnaire B was conducted after each dialog with the iCat robot to see what participants thought of the robot and the tasks they had to do. After each of the three tasks (and thus after each second dialog), questionnaire C was filled in to see whether participants perceived a difference between the iCat robots and which of the two was preferred to do the task with. At the end of the experiment, questionnaire D was conducted. Questions about acceptance, empathy and trust were posed. Participants were also asked if they enjoyed working with the robots, which iCat was preferred in general and they could make some additional remarks about the experiment.

Questionnaire A

1. Wat is je naam?
   .....
2. Wat is je leeftijd?
   ..... jaar
3. Ik ben een
   jongen/meisje.
4. Stel je voor dat een robot alles kan wat je maar bedenken kan. Wat zou jij dan willen dat een robot kan doen? Je mag meerdere hokjes aankruisen bij deze vraag.

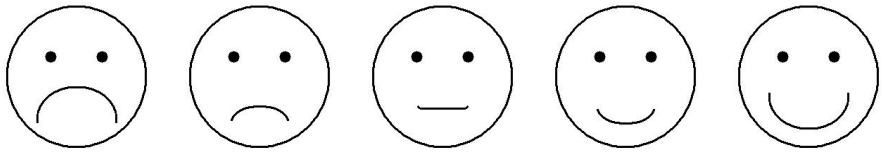
   [ ] Praten

   [ ] Een spelletje spelen

   [ ] Als wekker of agenda gebruiken

   [ ] Als huiswerkhulpje gebruiken

   [ ] Iets anders, namelijk
   .....
5. Hoe leuk denk je dat het werken met het robotje wordt? Je mag maar 1 antwoord kiezen.



   O heel stom O stom O gewoon O leuk O heel leuk
6. Heb je wel eens korfbal gespeeld?

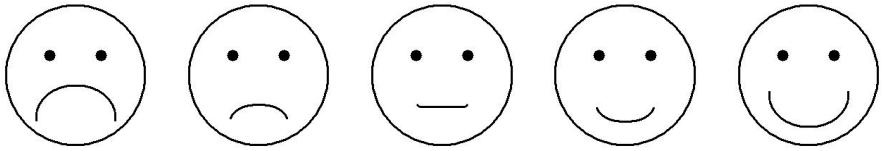
   O Ja, ik zit op korfbal.

   O Ja, ik heb wel eens korfbal gespeeld.

   O Nee, ik heb nog nooit gekorfbald.

Questionnaire B

1. Hoe leuk vond je deze robot?

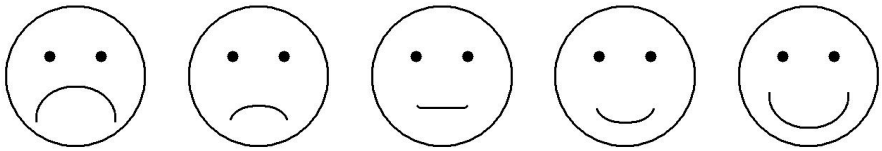


O heel stom	O stom	O gewoon	O leuk	O heel leuk

2. Hoe goed kon je deze robot verstaan?

O erg slecht	O slecht	O redelijk	O goed	O erg goed

3. Hoe leuk vond je het herkennen van emoties / de quiz / de oefening?



O heel stom	O stom	O gewoon	O leuk	O heel leuk

4. Hoe moeilijk vond je het herkennen van emoties / de quiz / de oefening?

O erg moeilijk	O een beetje moeilijk	O niet moeilijk en niet makkelijk	O een beetje makkelijk	O erg makkelijk

5. Zou je nog een keer emoties willen herkennen bij de robot? / Zou je nog een keer een quiz willen doen? / Zou je de oefening nog een keer willen doen?

O zeker niet	O ik denk het niet	O misschien	O ik denk het wel	O zeker wel

Questionnaire C

1. Merkte je een verschil tussen de twee robots waarmee je het herkennen van emoties / de quiz / de oefeningen hebt gedaan?

   O ja O nee

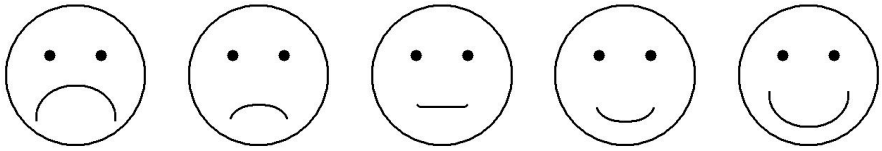
2. Zo ja, wat denk jij dat het verschil was?
   .....

3. Welke van de twee robots vond je het leukst?

   O Roos O Sanne

Questionnaire D

1. Hoe leuk vond je het om met de robots te werken?



O heel stom O stom O gewoon O leuk O heel leuk

Deze vragen gaan alleen over Roos

2. Zou je vaker met Roos willen werken?

O zeker niet	O ik denk het niet	O misschien	O ik denk het wel	O zeker wel

3. Vind je Roos vriendelijk?

O erg onvriendelijk	O een beetje onvriendelijk	O niet vriendelijk, niet onvriendelijk	O een beetje vriendelijk	O erg vriendelijk

4. Denk je dat Roos een geheim van jou zou doorvertellen aan een ander?

O zeker niet	O ik denk het niet	O misschien	O ik denk het wel	O zeker wel

Deze vragen gaan alleen over Sanne

5. Zou je vaker met Sanne willen werken?

O zeker niet	O ik denk het niet	O misschien	O ik denk het wel	O zeker wel

6. Vind je Sanne vriendelijk?

O erg onvriendelijk	O een beetje onvriendelijk	O niet vriendelijk, niet onvriendelijk	O een beetje vriendelijk	O erg vriendelijk

7. Denk je dat Sanne een geheim van jou zou doorvertellen aan een ander?

O zeker niet	O ik denk het niet	O misschien	O ik denk het wel	O zeker wel

8. Welke van de twee robots vond je het leukst?

O Roos O Sanne

9. Heb je nog opmerkingen over het experiment?
.....

Appendix C

Questionnaire results

C.1 Questionnaire B

On the following pages, the means, standard deviations and ANOVA results for the questions of questionnaire B are displayed. The results for the combination of all tasks is presented in Table C.1 to C.6. Thereafter, the results for each individual task are displayed. In the ANOVA tables, the significant results are indicated in bold.

All tasks											
		Question B1		Question B2		Question B3		Question B4		Question B5	
	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
all conditions	108	4.57	0.550	4.12	0.745	4.53	0.647	3.69	1.156	4.23	1.064
neutral	36	4.53	0.560	4.42	0.604	4.39	0.728	3.64	1.073	4.08	1.156
face only	36	4.67	0.535	4.22	0.681	4.67	0.535	3.69	1.215	4.33	0.862
speech and face	36	4.53	0.560	3.72	0.779	4.56	0.652	3.75	1.204	4.28	1.162

Table C.1: Means and standard deviations on the five questions in questionnaire B for all tasks together. All questions, which can be found in appendix B, were answered on a five-point scale.

Factor	*df* effect	*df* error	F	p
Condition	2	99	0.751	=.47
Task	2	99	0.842	=.43
Condition x Task	4	99	0.751	=.55

Table C.2: ANOVA results on the means of the question B1 with the design Condition (3) x Task (3)

Factor	*df* effect	*df* error	F	p
Condition	**2**	**99**	**10.416**	**<.001**
Task	**2**	**99**	**5.782**	**<.005**
Condition x Task	4	99	0.647	=.63

Table C.3: ANOVA results on the means of the question B2 with the design Condition (3) x Task (3)

Factor	*df* effect	*df* error	F	p
Condition	2	99	1.659	=.20
Task	2	99	0.677	=.51
Condition x Task	4	99	0.513	=.73

Table C.4: ANOVA results on the means of the question B3 with the design Condition (3) x Task (3)

Factor	*df* effect	*df* error	F	p
Condition	2	99	0.087	=.92
Task	**2**	**99**	**5.605**	**<.005**
Condition x Task	4	99	0.238	=.92

Table C.5: ANOVA results on the means of the question B4 with the design Condition (3) x Task (3)

Factor	*df* effect	*df* error	F	p
Condition	2	99	0.523	=.59
Task	2	99	1.038	=.36
Condition x Task	4	99	0.020	=.99

Table C.6: ANOVA results on the means of the question B5 with the design Condition (3) x Task (3)

iCat as electronic pet											
		Question B1		Question B2		Question B3		Question B4		Question B5	
	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
all conditions	36	4.50	0.609	4.17	0.655	4.47	0.736	3.83	0.941	4.11	1.036
neutral	12	4.33	0.651	4.42	0.669	4.17	0.937	3.75	0.866	3.92	1.165
face only	12	4.58	0.669	4.17	0.577	4.75	0.452	4.00	0.953	4.25	0.754
speech and face	12	4.58	0.515	3.92	0.669	4.50	0.674	3.75	1.055	4.17	1.193

Table C.7: Means and standard deviations on the five questions in questionnaire B for the electronic pet task. All questions, which can be found in appendix B, were answered on a five-point scale.

Question	*df* effect	*df* error	F	p
B1	2	33	0.660	=.52
B2	2	33	1.833	=.18
B3	2	33	2.005	=.15
B4	2	33	0.270	=.76
B5	2	33	0.324	=.73

Table C.8: ANOVA results on Condition (3) ANOVAs for the five questions in questionnaire B with the electronic pet task

iCat as educator											
		Question B1		Question B2		Question B3		Question B4		Question B5	
	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
all conditions	36	4.56	0.504	3.83	0.811	4.50	0.609	3.19	1.191	4.14	1.099
neutral	12	4.67	0.492	4.17	0.577	4.42	0.669	3.25	1.138	4.00	1.279
face only	12	4.58	0.515	4.08	0.793	4.58	0.515	3.08	1.311	4.25	0.866
speech and face	12	4.42	0.515	3.25	0.754	4.50	0.674	3.25	1.215	4.17	1.193

Table C.9: Means and standard deviations on the five questions in questionnaire B for the educator task. All questions, which can be found in appendix B, were answered on a five-point scale.

Question	*df* effect	*df* error	F	p
B1	2	33	0.755	=.48
B2	**2**	**33**	**6.044**	**<.01**
B3	2	33	0.214	=.81
B4	2	33	0.074	=.93
B5	2	33	0.153	=.86

Table C.10: ANOVA results on Condition (3) ANOVAs for the five questions in questionnaire B with the educator task

iCat as motivator											
		Question B1		Question B2		Question B3		Question B4		Question B5	
	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
all conditions	36	4.67	0.535	4.36	0.683	4.64	0.593	4.06	1.170	4.44	1.054
neutral	12	4.58	0.515	4.67	0.492	4.58	0.515	3.92	1.165	4.33	1.073
face only	12	4.83	0.389	4.42	0.669	4.67	0.651	4.00	1.206	4.50	1.000
speech and face	12	4.58	0.669	4.00	0.739	4.67	0.651	4.25	1.215	4.50	1.168

Table C.11: Means and standard deviations on the five questions in questionnaire B for the motivator task. All questions, which can be found in appendix B, were answered on a five-point scale.

Question	*df* effect	*df* error	F	p
B1	2	33	0.868	=.43
B2	**2**	**33**	**3.307**	**<.05**
B3	2	33	0.075	=.93
B4	2	33	0.253	=.78
B5	2	33	0095	=.91

Table C.12: ANOVA results on Condition (3) ANOVAs for the five questions in questionnaire B with the electronic pet task

C.2 Questionnaire C

The following tables show the results of questionnaire C. First, the results for the combination of all tasks are shown and thereafter the results for each individual task are displayed.

All tasks							
		Difference (C1)		Preference (C2)			
	N	Yes	No	*neutral*	*face only*	*speech and face*	none
all conditions	54	28	26	19	20	13	2
neutral vs. *face only*	18	8	10	10	8	-	0
neutral vs. *speech and face*	18	9	9	9	-	8	1
face only vs. *speech and face*	18	11	7	-	12	5	1

Table C.13: Answers to the questions of questionnaire C, which consisted of the questions whether participants did or did not perceive a difference between the iCats and which of the two iCats was preferred. This table shows the results for each condition on all tasks together. Because each participant evaluated two of the three conditions, the third condition could not be chosen (a '-' is shown in the table). Two participants couldn't choose between the two iCats (shown in the rightmost column of this table).

Factor	*df* effect	*df* error	F	p
Combination	2	45	0.530	=.59
Task	2	45	1.894	=.16
Combination x Task	4	45	1.326	=.27

Table C.14: ANOVA results on the means of the question C1 with the design Combination (3) x Task (3)

	Difference (C1)		Preference (C2)	
Combination	T	p	T	p
all combinations	0.27	.79		
neutral vs. *face only*	-0.46	.65	0.45	.65
neutral vs. *speech and face*	0.00	1.00	-0.24	.82
face only vs. *speech and face*	0.94	.36	1.81	.09

Table C.15: T test results for Difference (question C1) and Preference (question C2) for all tasks. For Difference (displayed on the left) a T test was employed to investigate whether participants chose 'yes' significantly more or less than 'no'. For Preference, a T test was employed to investigate whether participants preferred one of the conditions significantly over another.

iCat as electronic pet							
		Difference (C1)		Preference (C2)			
	N	Yes	No	*neutral*	*face only*	*speech and face*	none
all conditions	18	11	7	6	7	5	0
neutral vs. *face only*	6	5	1	3	3	-	0
neutral vs. *speech and face*	6	3	3	3	-	3	0
face only vs. *speech and face*	6	3	3	-	4	2	0

Table C.16: Answers to the questions of questionnaire C for the electronic pet task. Because each participant evaluated two of the three conditions, the third condition could not be chosen (a '-' is shown in the table).

	Difference (C1)		Preference (C2)	
Combination	T	p	T	p
all combinations	0.94	0.36		
neutral vs. *face only*	2.00	.10	0.00	1.00
neutral vs. *speech and face*	0.00	1.00	0.00	1.00
face only vs. *speech and face*	0.00	1.00	0.79	.46

Table C.17: T test results for Difference (question C1) and Preference (question C2) for the iCat as electronic pet. For Difference (displayed on the left) a T test was employed to investigate whether participants chose 'yes' significantly more or less than 'no'. For Preference, a T test was employed to investigate whether participants preferred one of the conditions significantly over another.

iCat as educator							
		Difference (C1)		Preference (C2)			
	N	Yes	No	*neutral*	*face only*	*speech and face*	none
all conditions	18	11	7	6	7	5	0
face only vs. *neutral*	6	2	4	3	3	-	0
neutral vs. *speech and face*	6	4	2	3	-	3	0
face only vs. *speech and face*	6	5	1	-	4	2	0

Table C.18: Answers to the questions of questionnaire C for the educator task. Because each participant evaluated two of the three conditions, the third condition could not be chosen (a '-' is shown in the table).

	Difference (C1)		Preference (C2)	
Combination	T	p	T	p
all combinations	0.94	0.36		
neutral vs. *face only*	-0.79	.46	0.00	1.00
neutral vs. *speech and face*	0.79	.46	0.00	1.00
face only vs. *speech and face*	2.00	.10	0.79	.46

Table C.19: T test results for Difference (question C1) and Preference (question C2) for the iCat as educator. For Difference (displayed on the left) a T test was employed to investigate whether participants chose 'yes' significantly more or less than 'no'. For Preference, a T test was employed to investigate whether participants preferred one of the conditions significantly over another.

iCat as motivator							
		Difference (C1)		Preference (C2)			
	N	Yes	No	*neutral*	*face only*	*speech and face*	none
all conditions	18	6	12	7	6	3	2
neutral vs. *face only*	6	1	5	4	2	-	0
neutral vs. *speech and face*	6	2	4	3	-	2	1
face only vs. *speech and face*	6	3	3	-	4	1	1

Table C.20: Answers to the questions of questionnaire C for the motivator task. Because each participant evaluated two of the three conditions, the third condition could not be chosen (a '-' is shown in the table). Two participants couldn't choose between the two iCats (shown in the rightmost column of this table).

	Difference (C1)		Preference (C2)	
Combination	T	p =	T	p =
all combinations	-1.46	.16		
neutral vs. *face only*	-2.00	.10	0.72	.46
neutral vs. *speech and face*	-0.79	.46	-0.45	.70
face only vs. *speech and face*	0.00	1.00	1.50	.21

Table C.21: T test results for Difference (question C1) and Preference (question C2) for the iCat as motivator. For Difference (displayed on the left) a T test was employed to investigate whether participants chose 'yes' significantly more or less than 'no'. For Preference, a T test was employed to investigate whether participants preferred one of the conditions significantly over another.

C.3 Questionnaire D

The following tables show the results of the questions D2 to D7. The results of questions D1, D8 and D9 were discussed in chapter 4.

		acceptance (D2,D5)		empathy (D3,D6)		trust (D4,D7)	
	N	mean	std.dev.	mean	std.dev.	mean	std.dev.
all conditions	36	4.64	0.593	4.83	0.447	1.97	0.810
speech and face	12	4.75	0.452	4.75	0.622	2.00	0.739
face only	12	4.50	0,798	4.75	0.452	2.00	0.853
neutral	12	4.67	0.492	5.00	0.000	1.92	0.900

Table C.22: Means and standard deviation of the answers given to questions D2 to D7 on acceptance, empathy and trust.

Question	*df* effect	*df* error	F	p
acceptance (D2,D5)	2	33	0.538	=.59
empathy (D3,D6)	2	33	1.269	=.29
trust (D4,D7)	2	33	0.040	=.96

Table C.23: ANOVA results on Condition (3) ANOVAs for the questions on acceptance, empathy and trust in questionnaire D.

Appendix D

Objective Results

D.1 iCat as electronic pet

The tables on the following pages show the means, standard deviations and ANOVA results of the electronic pet task. First, the recognition rates are shown and thereafter the reaction times.

	N	Mean	SD
all conditions	288	0.836	0.370
neutral	96	0.781	0.415
face only	96	0.884	0.320
speech and face	96	0.844	0.365

Table D.1: Mean recognition rates and standard deviations in the electronic pet task.

	df effect	*df* error	F	p
Recognition rate	2	285	1.898	=.15

Table D.2: ANOVA result on Condition (3) of the mean recognition rates in the electronic pet task.

	N	Mean (ms)	SD
all conditions	284	5735	2253
neutral	95	5742	2239
face only	94	5859	2353
speech and face	95	5608	2182

Table D.3: Mean reaction times and standard deviations for each condition in the electronic pet task. Outliers were removed.

	df effect	*df* error	F	p
Reaction times	2	281	0.293	=.74

Table D.4: ANOVA result on Condition (3) of the mean reaction times in the electronic pet task.

	N	Mean (ms)	SD
correctly annotated	239	5351	2017
incorrectly annotated	45	7778	2361

Table D.5: Mean reaction times and standard deviations for correctly and incorrectly answered questions in the electronic pet task. Outliers were removed.

	df effect	*df* error	F	p
Reaction times	**2**	**282**	**51.282**	**<.001**

Table D.6: ANOVA result on Correct (2) of the mean reaction times in the electronic pet task.

D.2 iCat as educator

The following tables show the proportions of questions answered correctly and the mean reaction times in the educator task in combination with the ANOVA results.

	N	Mean	SD
all conditions	180	0.694	0.462
neutral	60	0.750	0.437
face only	60	0.667	0.475
speech and face	60	0.667	0.475

Table D.7: Mean proportions and standard deviations of questions answered correctly in the educator task.

	df effect	*df* error	F	p
Recognition rate	2	177	0.648	=.52

Table D.8: ANOVA result on Condition (3) of the mean proportion of question answered correctly in the educator task.

	N	Mean (ms)	SD
all conditions	177	5507	1797
neutral	59	5938	1882
face only	60	5877	1682
speech and face	58	4685	1554

Table D.9: Mean reaction times and standard deviations for each condition in the electronic pet task. Outliers were removed.

	df effect	*df* error	F	p
Reaction times	**2**	**174**	**9.950**	**<.001**

Table D.10: ANOVA result on Condition (3) of the mean reaction times in the educator task.

	N	Mean (ms)	SD
correctly answered	125	5362	1794
incorrectly answered	52	5855	1772

Table D.11: Mean reaction times and standard deviations for correctly and incorrectly answered questions in the educator task. Outliers were removed.

	df effect	*df* error	F	p
Reaction times	2	175	2.800	=.08

Table D.12: ANOVA result on Correct (2) of the mean reaction times in the electronic pet task.

D.3 iCat as motivator

The tables below show the means, standard deviations and the ANOVA result of the z-scored results in the motivator task.

	N	Mean	SD
all conditions	36	0	1
neutral	12	-0.281	0.604
speech and face	12	0.253	1.381
face only	12	0.027	0.814

Table D.13: Z-scores and standard deviations on the combined parts (i.e. collecting marbles and taking steps) of the motivator task.

	df effect	*df* error	F	p
Reaction times	2	33	0.880	=.42

Table D.14: ANOVA result on Conditions (3) of the z-scores in the motivator task.