Using a social robot as facilitator in usability-tests with children

Verschoor, Y.J.P.

Award date:
2007

Using a social robot as facilitator in usability-tests with children

Yvo Verschoor
Student Technology and Society, Human Technology Interaction
August 2007

Supervisors:
dr. W.A. IJsselsteijn
dr. ir. P. Markopoulos Msc.
B. de Ruyter

Faculty of Technology Management
Technical University Eindhoven
Philips Research Eindhoven

Preface.

While I already oriented myself for a while in the beginning of this year it became time to choose for a graduation subject and project for my study "Innovation sciences". I preferred something in which I could experiment with a "new" product or method, something with "direct human-technology-interaction", and anything in which I could observe people. Wijnand IJsselsteijn advised me to have a talk with Panos Markopoulos, a friend working in the User Centered Engineering group at the Faculty of Industrial Design. He proposed a project with "evaluation methods, the iCat and children". It was a nice mix of what I consider typically "technology and society", and I looked forward to an interesting explorative time. This also involved the discovering unknown parts of the self, I worried a lot about me going along with children. It was something I had very little experience with.
Looking back I can say that this was really a nice part of the project, going around with the kids was not that hard, and brought in quite some smiling moments.
I would like to thank all the children and their parents who participated in this research. And of course I acknowledge my supervisors for their support and advice during the project.
I started the project with some intentions and ideas. During the project some of those changed. Some voluntary, others were more or less forced due to the circumstances. I suppose this takes part of any graduation-project process.
One of the things I preferred to do, is to generate data which is measurable by numbers, something what can be done in real usability testing. But this can only be performed when the evaluation method has reached a 'mature' status and remains fixed over the tests. I consider the evaluation method as usable now, but it took longer than expected to reach this status. Therefore I had to adapt my way of working in the end to allow for analysis in a more qualitative way.
Despite the unplanned circumstances, the process towards the results that are available now, have been a very nice learning experience for me, and hopefully also for others.

Contents

Summary

Usability testing is nowadays common practice in the design of complex consumer products. It provides a methodological way to discover errors or to find strengths and suggest improvements in a product design when a user uses it. The procedures to elicit these errors from the design have been subject for many studies in the past two decades. However when it comes to children's products, these procedures give problems, because children behave differently, and children's products have other purposes.

This thesis describes the way an iCat can be used in usability testing with children. For that purpose a setup for testing has been developed, and after a pilot eight tests have been carried out to further investigate the potential of the method.

It can be concluded that a social robot can be used as an evaluator in a usability test with children. Also the children consider it as 'fun' to interact with the iCat, while playing or testing a computer game.

This conclusion is based on the data gathered from eight trials of the method. While proceeding with these trials, problems with its execution were identified, and improvements were implemented and tried out.

1. Introduction

1.1 Usability testing

Usability testing is meant for getting a conception how well a more or less complex product can be used by the intended user. It is part of a design philosophy called User Centered Design, which states that the needs, wants and limitations of the end user should be taken into account when a new product is developed. A user interface of a computer program, or the layout of a remote control should be designed around how people tend to utilize it rather than that the user should adapt to an existing usage paradigm. For designers it is hard to understand the intended user at forehand by only using their intuition. Therefore they need to test their design several times using a multi-stage problem solving process.

Although the earliest forms of usability testing stem from the 40's, the real pioneering work in making guidelines and procedures have been done in the 70's and 80's. Since then, usability testing has become a common practice as a part of the design process for complex products in the past two decades. In 1986 Donald Norman published a first version of a book titled "The Psychology of Everyday Things" (Norman 1988). Based on his earlier scientific work, this book contains examples of bad and good design, and offers rules that designers can use to improve the usability of everyday objects such as cars, computers, telephones and doors. Jacob Nielsen is an other influential scientist who has published several books and numerous articles around this subject, but most are in the field of human-computer-interaction and website interfaces. Amongst others he wrote "Usability Inspection Methods" (Nielsen 1994) in which he describes several approaches in the ways usability can be assessed.

Usability test procedures aim to detect problems, or unforeseen and unwanted side-effects in a product or prototype. When detected, the designer of the product can make improvements and alterations to the design. For new and complex products this can be done several times. A typical usability test follows a fixed procedure. The user has to use the product in a realistic way, for instance by giving him or her a typical task to perform. An observer looks at the steps that have been gone through, and tries to discover whether they comply with the designers intention. Also the observer may have a role as facilitator, reminding the user to think aloud. It has to be mentioned that this is just one of the usability evaluation methods. Other methods include constructive interaction, focus groups, expert reviews, heuristic evaluations, post task

interviews the use of guidelines. Which of those is most appropriate depends on several factors. Several studies in the past focused on this matter, but mainly focused on usability test methods for adults products. (for instance John 1996, Hartson 2003). Usability tests procedures are nowadays well evolved through continuing research in test methodologies. Although there is growing interest, usability testing with children and with children's products have not been investigated very extensively or in a methodologically sound manner (Donker & Markopoulos, 2002). These point to interesting gaps to explore since usability testing with children brings along specific and not so trivial issues.

1.2 Usability testing and children.

There are some important differences between adults and children. Children have a smaller attention span, they have a smaller ability to verbalize their thoughts and emotions. And due to less developed cognitive and physical capabilities, they have a smaller ability to carry out tasks. In usability test sessions there is a pronounced asymmetry between the adult facilitator, and the child. This can influence the interactions between the facilitator and the child. The result could be that the child will agree with suggestive questions, and the child might have some fear of making mistakes. Furthermore the child may think that his knowledge is known by the adult anyway, so silence fits best in that case.

Shyness is an other important characteristic of the younger age. As with most patterns of child behavior, it is partly the temperament children are born with and partly the way children are raised. Anyhow, children in a usability test will be confronted with a facilitator they don't know. This may trigger possible shyness even more. For usability tests with children one way to go is to select children who are not shy. Hanna et al (2004) selected children who were characterized by their parents as not shy.. Although this is a rather rough selection method, it may make sense. However one should realize that the products tested are meant for a broader range of personalities,

One of the main goals of usability testing is, to test the product in an as realistic as possible situation. As with all measurement, there will always be an influence of the measurement device itself. But in the case of usability test methods and children, the effect of the probe, so the facilitator, will be much greater due to the just described differences.

Using think aloud as method to discover usability problems can be a very valuable method (Nielsen 1993) . Other research showed that, when used with children, think aloud protocols prove to lead to more found usability problems than with other protocols such as interviewing or questionnaires (Donker & Markopoulos 2002). In a slightly different follow up study however, the difference is less pronounced. but when it comes to the number of verbalizations a think aloud protocol is still better. (Baauw & Markopoulos 2004). Think aloud protocols do require extra cognitive effort though. For children this may be too much. When children are continuously reminded by an adult to keep thinking aloud, eventually this becomes an unpleasant experience.

1.3 Alternative approaches

Alternative approaches have been proposed to help overcome the burdens of conventional usability tests with children. Not aimed at usability testing but rather at trying out alternative interview techniques for early design was the "Mission from mars" workshop (Dindler et al, 2005). It intended to create a narrative space in which conventional cultural expectations were temporarily bypassed. Within this research one tried to gather user requirements for a new concept of a electronic schoolbag. Therefore they needed insight in the children's life, for instance their attitude towards personalization, order and social relations. First, in an ingenious manner a convincing story was brought in which the children were told to be able to talk to a Martian in orbit. The nice aspect of this narrative space was that the Martian, a researcher with an unrecognizable, altered voice, could be fully naive about schoolbags in a believable way. A problem with this technique is that some deception takes place, which poses ethical problems. However the same study showed that whether children believe the story or not, does not have to be problematic.

An example close to the current research was a pilot study executed by Barendregt (2006) in which the child was engaged in communication with a hand puppet operated by the experimenter. Although the general idea was promising conceptually, its actual implementation appeared to be problematic since the real facilitator had to multi-task, and was not a trained puppeteer. This led to interactions of variable quality. Furthermore the children tended to keep talking to the facilitator directly, and not to the puppet.

Looking at the ideas behind the methods just described on can conclude that children are willing to talk to a strange voice, and children can interact with a puppet and a puppeteer. If the puppeteer could be on a greater distance, it could take away the tendency that the child wants to interact with the puppeteer player directly. Therefore it would be interesting to use some sort of marionette. a remote controlled puppet. In this way a deceptional story is avoided, and a natural distance can be created.

Within this graduation project I explored the idea of a computerized puppet in the form of the nowadays available 'social robots'. A social robot may act and react in a more standardized way, thereby decreasing the subjective influence of the human facilitator. Furthermore it may be easier for the child to communicate on an equal level. So, the use of a social robot as facilitator might improve usability testing with children. The number of verbalizations can be an indicator of this improvement, as well as the manner wherein the children will experience the test.
But the main question will be, what are the requirements for a usability test setup for children. when an iCat is used as social robot.

1.4 Goals

This research is intended to investigate whether and how a social robot can be used as a proxy for an interviewer during the usability testing of edutainment PC applications with children testers. Ideally it would be useful to know how this could be done for a very general situation, for instance, "a technological product" (which needs a design-rationale), a more or less known instance of a social robot, and children of a varying age-group. This is a partially explorative study, and to take away too much possibly dependent variables, some choices have to be made. Amongst other variables these involve the choice of the robot, the way it is programmed, and for instance the age-group. For this research an iCat of Philips research will be used, and the age-group will be children between 6 and 8 years old. The setup should be designed to be ran with a single facilitator. Depending on the outcomes, other research can be performed to see if some findings can be stated in a more generalized way.
Also the primary goal is to see how children will behave when they interact with the social robot. For the purpose of this study it is already known how they interact with a human being as facilitator. Direct within or between-subject comparisons can be

made in a follow up research as this requires a fixed test setup and possibly more test-subjects.

There will be aimed to get answers on the following two questions:

1. In what manner will children interact with a social robot as facilitator and will the data be usable to report issues in a product ?
2. Will children experience a usability test session as fun when a social robot in the usability test is used as facilitator?

Since this is the exploration of what might become a new evaluation method, an other aim of this project is to be alert for any other aspects that add to the learning experience.

2.4. Thesis outline

This chapter mentioned the questions to be answered, and the aims of this exploration. In chapter 2 the interaction between the three key elements in the setup will be discussed in more detail, namely the product on test, the user, so the child, and the facilitator, so the social robot. Furthermore the boundaries within this research are set. Chapter 3 informs about the iCat as user interface research platform, and focuses on the technological issues. Chapter 4 describes the method used and the steps taken to find answers on the research questions. Chapter 5 discusses the results of using the method.

2. Actor analysis

2.1 Actors

Within the setup of a usability test 3 key elements can be recognized as actors. Namely, 1, the user or participant who tests the product. 2, the test facilitator, whether is a real person, or a social robot in a Wizard of Oz situation. and 3, the product to be tested.

The product on test can usually not be seen as a real actor. In a broad definition an actor is a participant in an action or process. In this study an interactive computer game is used. Nevertheless, within the game, one or more figures can be seen as actors and they interact with the user in a pre-programmed manner. The way the actors interact is pictured in figure 2-1.

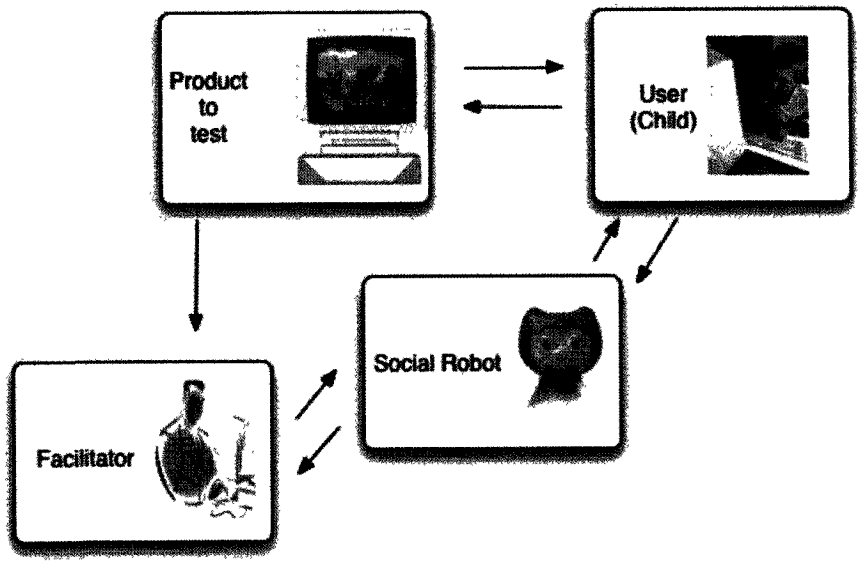


Figure 2-1: Actors in the proposed setup.

2.2 Children in usability studies

Children have different capacities and interests than adults. These differences express themselves in the smaller attention span, a smaller ability to verbalize their thoughts and emotions, and a smaller ability to carry out tasks. Moreover they are developing their skills, and their ability to carry out tasks changes a lot over the years. In this study children between 6 and 8 years were selected to participate. This is the age were they enter elementary school, and start to learn arithmetic's and start to learn reading and writing. They can maintain doing tasks for a reasonable time-span, but this

depends on the time of the day as well. Preschool children (ages 2 to 5 years) have a too short attention span to do the type of testing as intended. When it comes to think aloud sessions in usability testing then it is possible to do so with 6 to 8 year olds. But they will have more problems to express themselves than older children, mainly because of their verbal skills and vocabulary. Therefore the age group of 6 to 8 years is interesting, because when the robotic intervention method works, it probably will work for older ages as well.

The goal within this study is to find out whether the children experience the test with the presence of the robotic facilitator as fun, or as pleasurable. More in general it is interesting to notice how their reactions on the social robot will be. It might be possible that the robot will be a distraction in the setup. Or there might be other issues that can not be foreseen.

2.3 The product to be tested

Usability test methods can be used for a wide range of products. The evaluation method proposed in this research is focused on computer programs, and one educational game in particular. A social robot will be used as facilitator, and this is a non-mobile device. For this reason it is most convenient to choose for a product that can not be relocated in the space. When it comes to choosing a product to try in a new evaluation method, it should have some errors in the design, but it should better not be in the earliest phase of the design. In such a case too many errors would occur to run the procedure smoothly, and it would be harder to determine whether the errors would be detected by the test, or would have been detected by a simple observation on its own.

Several games have been tried out to see if they could be used within this study. It appeared that there are a lot of games without a clear indication for the intended age. Also many games do not have high system requirements. On the contrary, running them on modern Windows XP computers gives troubles due to incompatibilities with user accounts and device drivers. This are not the type of errors of interest so a requirement for the game is that it should run smoothly on a modern computer.

The computer program that will be used is an educational game meant for the first year of the elementary school. The title is Rabbit Robbie, Group 3 (Mindscape 2003) .

The game itself is already on the market for a few years, this means it probably has passed several design iterations already. This makes finding usability problems harder, or they are harder to relate to design errors.
The game narrates about a land named cloudland. For some reason all kinds of raingear comes falling down the sky, and this causes problems. The goal in the game is to find the cause and solve the problem. This has to be done by solving several puzzles and sub games which all have a educational value. Some let the child do simple arithmetic's, other games let the child play with words and letters.
An important aspect of the game is that is has several figures who are talking to each other and to the child. The several game screens are alternated with short cartoons that need attention in order to solve the puzzle. Therefore the game is a interaction partner in the setup as well.
The goal within this part of the research is to determine whether the data can be used to report issues in the product.

2.4 The facilitator

One of the interests in improving existing usability test methods, is the quest for making the procedures more efficient. For example the reasoning behind the number of tests that should be executed to obtain sufficient feedback for the designer has been subject for research. Only 5 tests per iteration would be enough, the number of new found problems would not increase much more when more tests are executed (Nielsen et al 1993) . A conventional usability test carried out with video observations or a think aloud session can be done by one facilitator. The whole procedure can be split up in single task actions, which can be carried out sequential. When a second facilitator helps with starting the videos this is only convenient, but not strictly necessary. Because of the explorative nature of this research, and the view that it does not deviate extremely from a conventional setup, it has been tried to design the procedure in such a way that only one facilitator is needed. When it proves to work, then carrying out the study with an second facilitator or helper will only add to the quality of the outcomes and the smoothness in which the procedures can be carried out.
The experimenter in this thesis had no practical experience in usability testing. This means that the skills required for this ability had to be learned along the way.

So, the number of facilitators will be kept fixed. Question is, in what way the proposed way of testing differs from conventional testing, especially from the perspective of the evaluator. And what other difficulties will the facilitator have to face.

2.4.1 Social robotics.

In a broad definition a robot is a machine capable of carrying out a series of actions automatically. There are still discussions about what characteristics are required to define a machine or device as a robot. But when they aim on a more or less specific goal, and look like a human or animal there is less debate. A robot is called a Social Robot, when it can interact and communicate with humans by following the behavioral norms expected by the people with whom it is intended to interact. At least this is the definition used by Bartneck and Forlizzi (Bartneck 2004). The definition proposed by Dautenbahn and Billard Dautenbahn is somewhat different and more specific. They state: Social robots are embodied agents that are part of a heterogeneous group.(...) They are able to recognize each other, and engage in social interactions. They posses histories (have experience), and explicitly communicate with, and learn from each other. (Dautenbahn 1999)
In this thesis the term Social Robot will be used for an entity in some form of physical embodiment with programmable and automated functions. It has the ability to communicate with humans, and to show its emotional state.

Within this research the iCat of Philips Research, or "Interactive Cat" is used as the social robot. The iCat can move its head, eyes and mouth to create gestures and emotional expressions. The iCat also has the capability to sense its environment via several sensors, including two microphones, touch sensor (in its ears and claws), and a camera. It is programmable and depending on the way it is programmed, it can simulate some degree of intelligence. The program and thereby the intelligence resides on a separate PC used to control the iCat.

Actually the iCat has been marketed as a user-interface robot, or as an experimentation platform for Human-Robot Interaction research. Since August 2005 it is available for research laboratories and universities. A part of the current research

of the iCat is aimed at improving its capabilities, mainly by programming additional software to add to the surrounding software. In other research the iCat has been used as a personal TV assistant, as a e-health assistant (Looije 2006), as a relational agent for older adults (Bickmore 2005, Heerink 2006a), or in a dialogue system (Foster 2006). In many research projects the iCat is used to compare the influence of a varying personality (i.e. Looije 2006, Heerink 2006a, Heerink 2006b). Many studies apply the iCat also in a Wizard of Oz type of setup (i.e. Looije 2006, Poel 2006, Bartneck 2007). A wizard of Oz setup means that the iCat has to be perceived to the participant as an autonomous character, while it is connected to a hidden operator who is controlling its behavior.

The requirements for the computerized 'puppet' come down to, 1. creating a believable experience towards the child, and 2, having means to control the robot in a flexible manner. The iCat meets these requirements, and is for this purpose the best option available today.
Some studies focused on the believability of this social robot, for instance by adding gaze behavior to the iCat (Poel 2006). All studies concerning believability have been performed with adults. One should be aware that for children, these conclusions regarding believability do have to apply. The requirements for a believable experience can be relaxed, since children are more prone to believe. Also children of the age group between 6 and 8 years still have a rich imagination, and most of them still believe in fantasy figures such as Santa Claus. They can believe more easy in a story or concept that actually is not true.

3. Technology

3.1 Capabilities of the iCat.

The iCat is a robot for research purposes. It has been developed as flexible platform for robot-user interface experiments.

The hardware consists of 13 servo's to control its neck, eyes, eyelids, eyebrows, and mouth. For signaling it has colorleds in it ears and paws. It has two microphones in its feet. By using the right software, one can detect the direction of sounds in this way. Also the sound-data can be used to recognize voices or commands. It has a proximity sensor in one of its feet to detect presence on a certain distance. It has a camera in its nose. Again by using the right software one can use this camera in a flexible manner. For instance to detect faces, or even the eyes within faces. When this data is feedback to its eyes, head and neck of the iCat, it will be able to gaze at a target, and follow its movements.

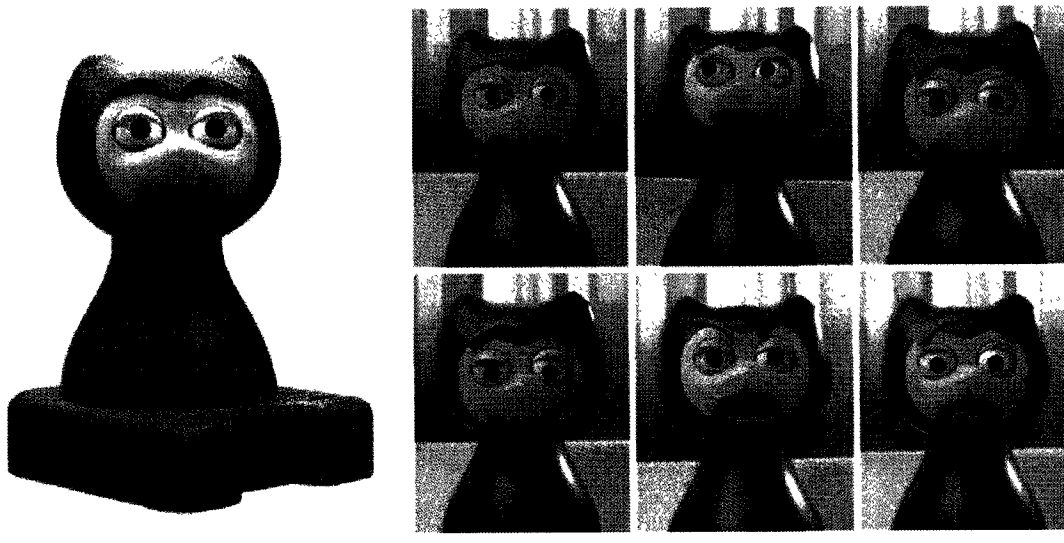


Figure 3-1: The iCat and the many facial expressions

Along with the package there comes software for speech synthesis. The words produced can be easily linked to the lips of the face in order to obtain lip-synchronization. The package comes with a so-called animation editor which makes it possible to program a sequence of movements, or speech utterances via a graphical user interface.

The internal hardware components are all standard USB devices internally connected to a USB hub. The iCat can connect to a PC running windows XP via one USB connection, the supplied drivers make it plug-and-play.

3.2 Software.

The iCat comes along with the necessary drivers, and OPPR software. OPPR is the acronym for Open Platform for Personal Robotics. The architecture part of the OPPR system contains a software component model that allows users to develop software using building blocks that can be re-used easily. The modules within this model are DML modules which stands for Dynamic Module Library. Each module has an input or an output port depending on its purpose. The DML framework and the data it exchanges is supported by a virtual parallel machine running on the background. New or special purpose modules can be written using C++. Along with the OPPR package comes the scripting module. It allows to use a scripting language named LUA in order to create dialogues. On the internet a community board is available to let iCat users help each other with their projects.

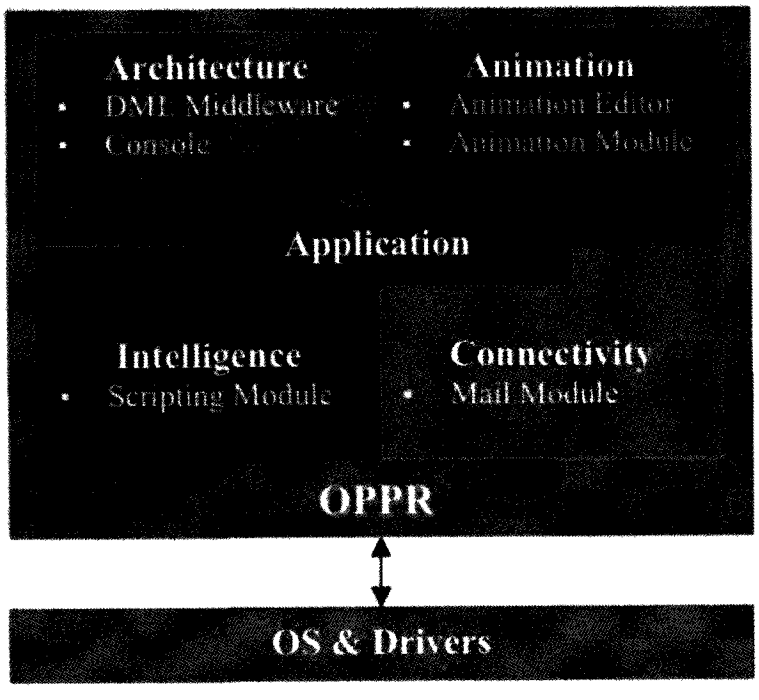


Figure 3-2: The structure of OPPR.

3.3 Required capabilities for this research

Using the C-language and the flexible modular design allow for potentially very advanced interactions with the iCat. One can conclude that it's capabilities may only be limited by the sophistication of it's hardware, and the programming skills of its programmer. However, programming special applications will take some time also for skilled programmers, especially when DML modules are not readily available.

For the purpose of using the iCat as facilitator the required programming skills depend on the unanswered question how sophisticated a social robot should be. The guess in this exploration was that it does not have to be very sophisticated. So one can then keep using LUA. There are a few limitations in that case though. A major part of the interaction will take place via speech. When the cat should be programmed as if it is a puppet, it is required to let its lips move while it speaks. The mentioned speech synthesis software has that capability but is meant for the English language. Furthermore there have been messages within the iCat research community that it needs improvement regarding the quality of the speech synthesis.
So, this module can not be used. One can program predefined words and sentences though. And one can make a command for moving its lips. In practice it is difficult to make this happen, because of some delay between the speech and the lips-move-command.

In this research two LUA scripts have been used. One is to create a more or less automated interactive introduction. The other is to be able to send commands and movements to the iCat via a user interface. This interface was and is subject for improvement along the sessions.

Appendix 2 shows a screenshot of the interface used in the last session of the experiment.

4. Method

4.1 Design.

This research is setup mostly as a case study. It should be seen as an evaluation and exploration of a potential new research method. Within this study, with eight sessions were held. Each session can be considered as a case. After each session the participants had to fill in a survey. Because of the fact that the participants are children, this survey is applied in an other manner than when applied to adults. The children were not able to read the texts within reasonable time. Therefore the questions were read out aloud. In this way the survey gets the form of a structured interview, in which most answers are closed.

Each experiment could be considered as a separate case, but the conclusions regarding the requirements are based to the method as a whole. The conclusions regarding the answered questions, are based on both the outcome of the questionnaire, but also by the way the children expressed their thoughts.

4.2 Materials & setting

The basics of the setup built and are described in this section. The parts mentioning the iCat and surrounding elements are in italics, since this is where this setup deviates from a conventional usability test setup.

The setup involves:

- A controlled lab-facility.

In a controlled lab facility all means are available to observe and record human behavior. In usability labs this also involves a space that looks like a living room so the participants can feel comfortable. We used the KidsLab for this purpose. This is a usability lab dedicated for children. So it's furniture is smaller, and the colorful paintings and objects make it look like a kids-room.

Within the living-room space of the lab-facility there are:

- The personal computer on which the software to be tested runs.
- A stable chair for the child to sit on

The chair needs to be of an appropriate height, and should be fixed because children tend to move a lot.

- *The iCat placed next to the computer,*

The iCat is turned towards the child in front of the computer.

- All other means to record all behavior (i.e. camera's, microphones)

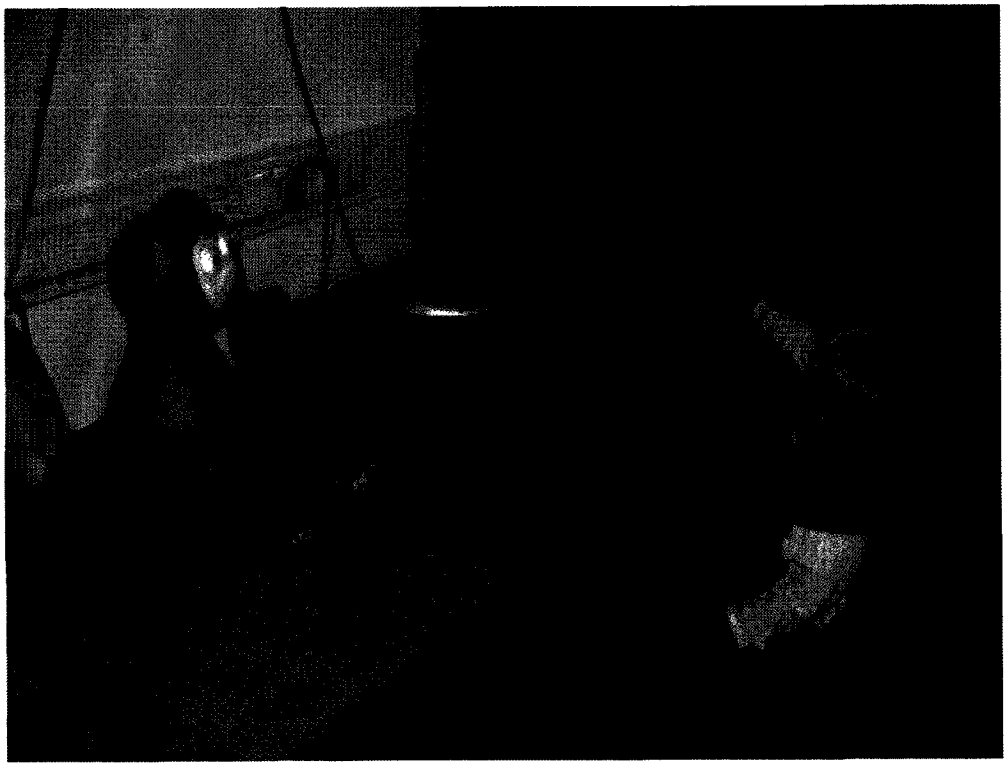


Figure 4-1: iCat and child in the user-space of the Kidslab

The icat is turned with towards the child, also its footing points towards the child. In this position it suggests that it has full attention. One could also move its neck via the software. A disadvantage is that with the current software it is not easily possible to use an off-set. The animations created in the animation editor only allow absolute positions. One can put the iCat closer towards, or farther away from the computer in order to vary its presence. During each session this variable remains fixed since the iCat is a non-mobile robot.

Within the control room of the lab-facility there are:

- All means to monitor the child's behavior real-time

  In the Kidslab there are one-way windows, and extra camera's for this purpose.

- All means to record relevant behavior.

  (i.e. audio-video recording and mixing equipment)

A computer to control the iCat or social robot with.

A microphone

A computer to alter the voice of the facilitator in such a way that it is comprehensible but not recognizable.

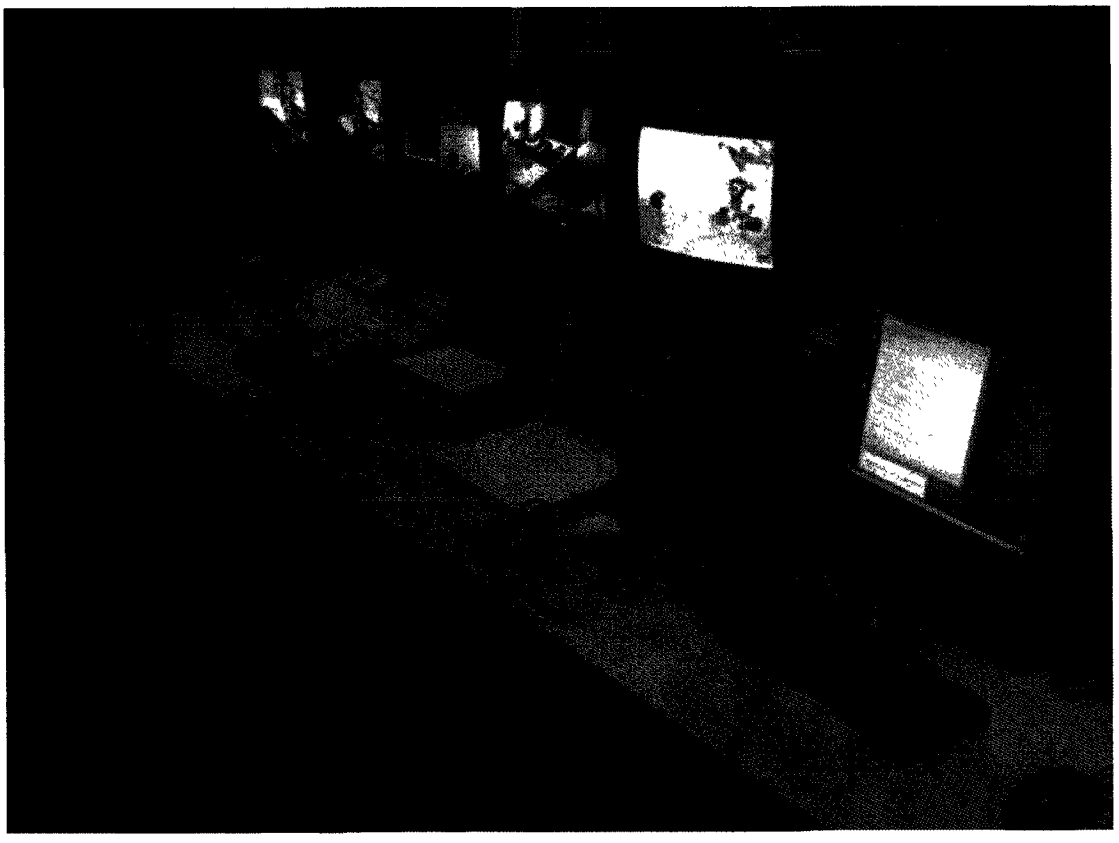


Figure 4-2. The control room of the Kidslab

4.3 A typical test session

This section describes what happens during a typical test session. It only describes the steps around the session, not the data gathered with the session itself.

The tests sessions are run along a fixed protocol. This means, as fixed as possible, but also as fixed as appropriate. When the protocol is followed in a rigorous manner, this ensures that no properties will change that might influence the outcome. But applying strict rules is sometimes difficult and may not be needed. An issue is that the people in the test (both parents and children) can not be lead along an always the same scenario. A detailed version of the last version of the protocol can be found in the appendix 1 of this thesis. In short it involves the following: A technical & goods check. Before the test, all computers and programs should be up and running, including the iCat introduction program. After arrival of the participants in the

building it is important to create an atmosphere they can trust and feel relaxed in. In practice it comes down to offering drinks, try to level with them by doing some chit-chat, and by showing them around in the building and the lab. Almost always the parents accompany with the children, and one has to keep them involved too.
After entering the Kidslab, and telling about the purpose of the room, the child has to take place in front of the computer. In the same time the parents are asked to read and fill in the consent form.
The child is told about the game, and its purpose. Within Robbie Rabbits cloudland, rain gear comes down the sky and this has to be stopped by solving puzzles. We will play for an hour, and the child should just try to play the game as good as possible. Then by a touch command the iCat wakes up, and introduces himself as "Lou". Lou can be asked questions but Lou tends to be very sleepy just like any other cat. Lou is also sleepy or shy when adults are around. That's why the parents and the real facilitator will hide behind the wall. Lou likes helping with games though. Since the preprogrammed mode of the iCat is short, the cat tends to fall asleep already at this point. The child is ensured that when it gets too hard, or when there are other problems it can always come around, the doors are open. After this introduction the game will be started, and the child is left alone.
Within this particular game the characters in the game are introduced in the first minute This gives the facilitator the time to start the iCat control program.
Also the parents are told that they can watch the process, but should try not to interfere.
During the game, the facilitator has to monitor two observations screens, and has to control the iCat. Apart from that the facilitator has to keep in mind any special behavior he intended to give to the iCat. This is exacting work, but add just an extra visual cue to a puppeteers role. Also when extra behaviors are added to the iCat the consequence of each button should be clear. A sheet with these behaviors should be ready, since one tends to forget during the session.
The control consist of two ways of acting. Namely when no action is required, the cat should stay passive, but alive. So it should be moving during the whole session. When action is required, the cat should be alert and reacting to the circumstances. In that instance the facilitator can speak through its speaker, and move its mouth, as if it is a puppet.

After an hour, the cat should fall asleep, and in an appropriate part of the game the facilitator will enter the room.
The survey is filled in, together with the child. During this step an immediate impression is obtained about how the session was perceived.
After this the child is given a little present, and the child and the parents are thanked for their cooperation.
The basics of the overall test setup did not change while proceeding from one session to the other. But the way the iCat could be controlled, and the scripts for standardized reactions did. This is described in chapter 4.7.2

4.4 Participants

The children in this research were in the age group of 6 till 8 years old. The were all recruited in and beyond the circle of acquaintances of the experimenter. First the parents were told about the project, and asked whether they, and their children would be interested to participate. Most of the children did not know each other, in one case they were brothers, in an other case they were friends. In cases were it was needed there were told not to tell others other about the tests. None of the children had used the game "Robbie Konijn groep 3" before. For recruiting parents with children a pro-active attitude is needed. They are willing to cooperate in this kind of research but usually have busy schedules. The tests have been scheduled with a preference for the morning, as much as possible. The disadvantage of running the test in the afternoon is, that the child might already have been fatigued by the other activities during the day. Such has consequences for attentiveness and mood.
Most sessions were done during school weeks and this leaves lesser choice for planning in the morning.

4.5 Tasks

All children were given the same assignment. This included: "Play the game", "Try to solve Cloudland's problem (of 'falling umbrella's')". The playing of the game lasted for approximately one hour. One hour is enough to collect a large amount of data. And it is not too long, since playing such a game can be tiresome. In this manner the sudden mood-swings that can occur when children get tired are avoided.

4.6 Acquired data

The data collected in this research consist out of four parts. The video data, the audio data, the next-to-video impressions of the observer, and the answers given in the survey

The audio/video stream have been analyzed by watching the videos, and annotate them using the timestamps as anchors. In this way particular findings can be easily traced back for further analysis. Related to the timestamps, the voices of the child and the robot have been transcribed. And the behaviors and all other mentionable expressions have been recorded separately. After transcribing the dialogues and behaviors can be traced back as showed in figure 4-3.

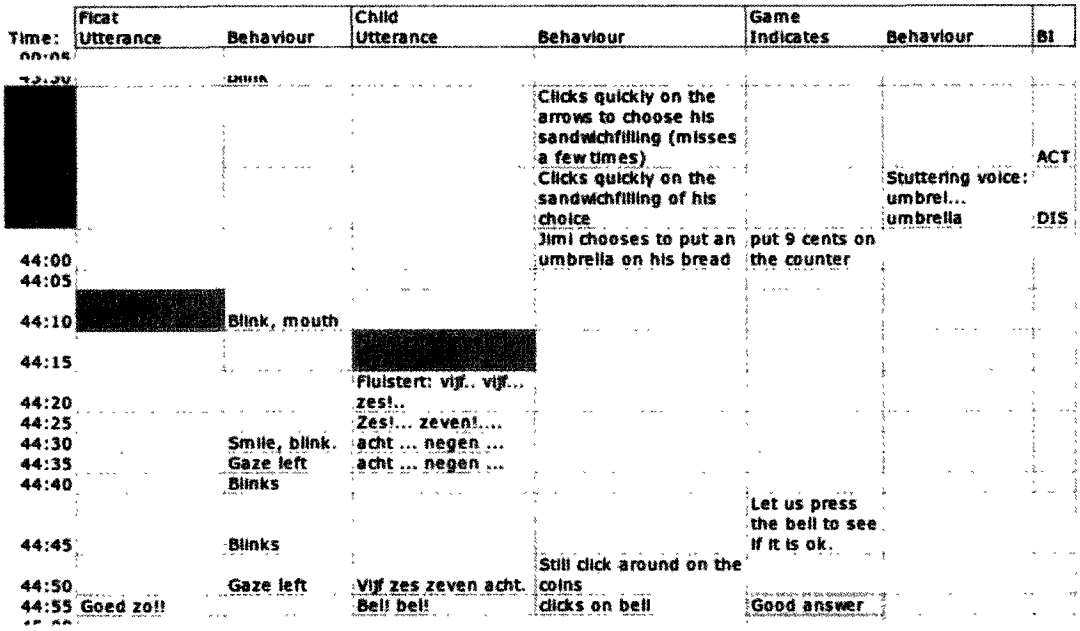


Figure 4-3 Partial screenshot of one of the transcription sheets

Although this research was not aimed to detect which, or how many, usability problems occurred with the game, the data can be used for that. The last column of the transcription sheet was meant to note breakdown indicators.

4.6.1 The survey

After interacting with Lou and playing the game for an hour, the children were presented a short questionnaire. The final version of this survey, presented them 12 questions. Those included (translated from Dutch): 1. I find it fun to play this game, 2, I find the figures in this game nice to look at, 3, I enjoyed talking with Lou, 4, I find

that Lou was good in helping me, 5 I find Lou pretty smart, 6 I think I would like to play this game more often. 7 If I could change something in this game, it would be.... 8, I have other things to mention, namely: 9: I like to use a computer, 10 I played quite often with the computer already.

The questions were formulated as in first person singular, and the children were asked to relate it to the smileyometer. The smileyometer shows five smiley faces varying from a unhappy smile to a very happy smile.

The questions of most interest were those related with the iCat (in this research named "Lou"). The first and the second question were put on this place to check whether the children would understand the answering scale, and just to make them used to filling in a survey. Asking 12 items on the end of a possibly tiring game session was expected to be problematic. Therefore the questions of lesser interest were put on the end. The content did not change over the last six sessions apart from layout changes. The final version of this survey can be found in appendix 3

4.6.2 Measures for the sessions

As mentioned the data consists of four parts. Next to the audio and video streams and the survey there are other impressions and observations that come along with the test. The audio and video streams are likely to resemble the video footage of conventional usability tests. But the in-test impressions of the evaluator may deviate a lot because the way of interacting with the child is indirect and rather different. Eight sessions with changing setups have been executed.

While each change in the setup was contrived from an as good as possible considerations, in practice many other variables still determine the difference between a successful session and a less successful session. So this success relies on both demonstrable causes and less predictable factors.

A simple measure for the success of a test is to see if the session can be done without disturbance halfway the test. So, the child should interact with both the game and the robot, and no-one else. An other measure for success is the amount of interaction between the robot and the child. One should be aware, that the test is not meant to communicate as much as possible. Rather should the child just use the product, and all exchanged information should relate to this product. Another indicator for a session can be the way it was executed according to the plan or protocol. Sessions with children will always lead to unexpected events. The question is whether the

unexpected events were manageable, and were of minor importance in relation to the events that happened in accordance to the protocol.

4.7 Systematic approach.

There are two important aspects that tend to change while performing the research steps. On the one hand, the experimenter learns about doing the tests as a whole. This is a change that exists even when there would be no variation in the setup. On the other hand the experimenter learns from the changes he makes in the setup. It is important to note that both learning experiences can be valuable, but, they need to be kept separate while doing the analysis. Learning effects of the experimenter are on a personal level, they will not be the same for everyone. Learning effects due to changes of the setup should be experienced in a similar way, when someone else is doing the experiment.

For every planned change in the setup, the experimenter has to deal with certain constraints. A change takes time to implement. Furthermore the change should be possible within the limits of the system and the available resources. One should to take into account that there will be a remaining difference between a wished change, and the change that actually can be performed, especially in the beginning of the research.

A basic part of the setup did not change. This is the idea that has been used to start with. This involves: - Using a 'usability-test' setup in a fixed controlled environment, - Using a computer to control the iCat, - Using some means to monitor and produce spoken questions, prompts and help from the control room. - The game.

The changes in the setup involved: changes in ways how to control the iCat, changes in the protocol, changes in what to do when the children do not proceed while playing the game. The changes in how to control the iCat involved mostly changes in the user interface. Based on learned experience, animations linked to buttons were added, and the introduction program was adapted and extended. The changes in the protocol involved additions, due to previously unforeseen events.

4.7.1 The first test.

The first test was meant to get used to a usability testing and to get a notion whether the intermediary evaluator-idea would work at all. In absence of a functioning iCat, a fluffy-toy with a built in speaker has been used. In order to make it look "alive" some

motor in side the fluffy could let it move or shake a little. The lack of autonomy made it hard to put stress on the introduction of Ebby. When a problem occurred in the game to be tested, the child did not engage in interaction with Ebby, instead it stood up and searched for help from his father or the facilitator. Later on some interaction succeeded which was promising. After a functional iCat arrived this device has been used for all sessions. The succeeding 8 tests have been done in a systematic way and will be portrait in the next section.

4.7.2 The subsequent tests

Session 1

In the first session Ilse, a 8 year old girl interacted with the iCat. The script involved: A simple "help her out" in case of trouble and try to keep the iCat alive with the buttons in the interface. I programmed a basic interface for sending movement commands to the iCat.

Because of a PC problem I used my notebook for running the game. I used a PC for doing the voice morphing, with "Morphvox Pro". This gives a delay in the order of one second. One of the environmental circumstances that should be taken into account is that the usability lab can also be reserved and used by other people. Several times problems were caused by unseen changes in the room. Namely rewired cables, switched lab PC's, changed mixer settings. A complete check of any link in the technical part of the setup should resolve this.

Camera's were not in the position were I left them before. Repositioning with the remote controlled servo's gives noise, Ilse noticed that. The scan converter was not available, this could be solved sufficiently (but not ideally) by pointing one camera to the screen.

Ilse is eight years old, she is the daughter of a friend of mine. Ilse is very patient, listens to all the characters and stories in the game, and proceeds well.
In the reading parts, she reads aloud (whispering but audible)
Typical phrases from the transcript were: This, this, this and then... this. (While pointing in the puzzle, showing a planned solution).
The sound coming from the iCat is for some reason not good, it is hard to understand.
The only change from a pre-Kidslab test-setup is the used PC speaker. The iCat

movements are clearly audible, and distract Ilse in the beginning. There is hardly any other interaction in this session.
The mother is along in the control room to watch the whole process and enjoys it. In the end we fill in the questionnaire together. The mother gives suggestive hints while going through the questions. The last questions are hard to answer due to the answering scale.

While analyzing the video, usability problems could be identified using the coding scheme of DEVAN (Vermeeren 2002) . Also some behaviors that might occur more often have been noted. Those can be used to respond to behaviors in a standard manner. Actions and changes after this session: Added camera position checks in the protocol.
Altogether the session went not so well, especially due to the lack of interaction..
When I ask my participant and her mother if she liked the everything fortunately their overall impression of the experience is different, much more positive.
An other important observation is that keeping an eye on everything including the control of the iCat is a rather intensive task.
Items to change or add: Change speaker, improve sound quality.
A change in the survey, due to gained insight.

Session 2
Intended interaction with the iCat: Help out in case of trouble in the game. "Think along"
This time a normal PC is used for running the game. This only makes a difference from a ergonomic point of view.

The mother has an other child with her, she decided not to wait in the control room, but to wait in the hallway. This made the session a bit easier for me.
While I did the introduction the other Casper's brother runs freely through the rooms, including the control room. "mammy, I see you!" This distracts a lot, since it might reveal what we are doing, so I have to remove him in a friendly manner, from this room. The introduction should be held as structured as possible, since the iCat is an eye catcher, and is easily awakened (by a touch command). What I am afraid for happens, Caspers brother starts the iCat, and I have to improvise.

Casper tells that he plays with computers once in a while, and that he plays games on the internet. He likes puzzles.
I explain about that the iCat likes it when Casper thinks aloud. Casper acknowledged that he will try.
I have to re-run a version of the control software, it takes a while before I can enter the Facilitator iCat program.
The sound is clearly better understandable now, but the level of the game sound is overruling it somewhat. Door is open, voices are coming from a neighboring room. Casper seems not distracted, but it interferes in the recorded sound. Within the control room the monitor sound is not clear, therefore reacting on Casper's whispering thoughts is hardly possible. Due to delayed feedback speaking with the iCat is hard. The sound is harsh sometimes.
A typical phrase in the out of the transcription (translated from Dutch):
(Casper is stuck at a screen and sighs)
Lou: Ehhh, I think I can help you in this
Casper: Well, help me then!
Lou: Ehhh, but I have to think about this as well!
(...) What if you place your cursor on the right?
Lou: Do you see your cursor becomes green?
Casper: Huh?
Casper: Can you repeat that, I could not hear you!
Lou: Click on the left... yes!

The sound is better now, at least there is quite some interaction between the iCat and the child in this session. Therefore it is a better session. Again the work in the control room is experienced as rather busy. The fact that there is delay in the sound makes the stress worse.
The microphones in the Kidslab should be pointed better towards the child.
Maybe an other voice morphing program should be used. Also I do not trust the microphone used for the voice morphing.

Session 3

This session differs quite a lot from the others in the sense that it was not taken in the Kidslab. The Kidslab was not available at the moment it was needed, but I did not

want to cancel the date. So I gathered all required equipment for a portable usability test setup, and I reserved two suitable rooms. I used a mirror and a web cam to monitor the room and the game. As a change in the setup I used an other microphone and other application for the voice morphing. This time the adapted helium voice of Garageband v5.0 is used.
Intended interaction with the iCat: Help out in case of trouble in the game. "Think along", encourage when puzzles in the game are solved.

Sara was very shy. Her father joined the introduction. But after the game started we left for the "control room", a room next to the room this session was held in. At that point Sara did not show that shy anymore, after discovering that her name was not in the list (children are supposed to fill them in) she began talking. Nevertheless it was hard to obtain some interaction. Later on it appeared to be a sound problem again. It seemed the sound was ok in the control room. There was good feedback from the iCats voice. But this was not the same as monitoring the living room space. Proceeding through the screens of Robbie Konijn went slow. Sara shows quite tired, but still she seems very patient. Still I interrupted to see if I can help her out. After 20 minutes suddenly the game disappears, Sara accidentally touched the power/standby button of the computer by her feet. When the sound problem is discovered there is more and better interaction after all. Sara is only six years old, and is not yet capable of the game were money has to be counted.
A phrase from the transcript:

Sara: This much? is this nineteen cents?
Lou: Ehm, no, not yet, add a bit more!
Sara: Heejj (seems to understand what her task is)

Several things went wrong in this session, they did no go according the protocol. This is due to the improvised setup. Still it was possible to engage in interaction, after some problems were solved. Thus, and important conclusion is: A controlled and fixed environment is convenient, but probably not strictly necessary. Sara felt comfortable in this room.

Session 4

Intended interaction with the iCat: Help out in case of trouble. "Think along", and try to encourage to think aloud. Proceed with improved sound, adjust directional microphone. I adjusted the interface to provide more "stay alive" actions.

The father came with 4 children, 2 of them were in the right age-group. To cope with the expected business I arranged an extra computer with (other) games, so they could entertain themselves in an other room, while the session with the first participant was proceeding. This first participant was Daphne, she was 7 years old. Daphne knows about "Robbie Konijn" but she has not played with "Fun in cloudland" yet. She has seen Lou already since she saw a picture of him on my website. I discovered an error in the introduction of the iCat, due to a notorious bug in the animation editor. Daphne does not seem to be bothered by this. The sound is smooth, even the breathing of Daphne can be heard. But Daphne is very quiet in this game, she just proceeds, and does a good job in playing the game. The sound of the iCat is loud and clear, louder than the game sound, this might be a bit too much.

In the post task interview it becomes clear that Daphne knows she didn't talk to Lou, she says because she found everything easy in the game, she did not have to think often. She would have liked to have talk with Lou, with while playing the game she experienced Lou more as disturbing.
For the next session which follows directly after this session it is only decided to be more active in finding moments to prompt for thinking aloud

Session 5

The intended interaction with the iCat resembles the interaction of the previous setting. All other conditions are kept the same as well. Only the sound level is lowered. This session is also observed by an other colleague of the UCE group.

Michael is the second participant today and a friend of Daphne. In the introduction I ask him his age (7), and also in which group he attends on primary school, since the game is intended for "groep 3". I guess, "which group do you attend?, groep 3? no? Groep 4? 5? He says group 5. Then I realize he is attending a Belgium school, and he does not know about the Dutch naming of the level of education. Within the little talk and post task interview I detected some more desirable answers.

During the introduction he asks whether we are controlling Lou. My answer is honest but also complicated and thereby vague, and the introduction is quickly continued. After starting Robbie Konijn Michael shows to be a bit bored by this game. He also keeps on moving. The sound has good feedback but seems to soft. After a 15 minutes, my colleague interrupts the session to increase the sound level again.
After this the interaction goes much better. Michael seems not very motivated and tries to trick the iCat by blocking the screen. In the post task interview he says that he discovered that Lou became more silent when he did that.
In a few cases Michael answers aloud, when a character asks something in the game.

An other not expected problem shows up. My throat is dry and my voice becomes hoarse. If I keep on talking, the high pitched and friendly voice becomes a very different scary voice. I can only avoid this by speaking with a high pitched voice myself. This is fairly unnatural and also stressful. with the already loaded task as facilitator.
Michael was moving a lot, it was the first time the office chair was giving too much freedom. For the further tests a fixed chair should be used.

Session 6
The intended interaction with the iCat for this session was: Help out in case of trouble. "Think along", and try to encourage to think aloud. Because I have a wish to be able to help better, I want to be able to give suggestions. Therefore some basic knowledge about control elements is needed. So, what are "a cursor", left, right, in accordance of a guideline of Hanna . This is added to the introduction.

With the session with Lars I manage to send his father to the control room before I give the introduction. Lars is attentive and listens good to my introduction. He mistakes when it comes to right/left differences, but I explain what I mean.
After starting the game, at first the interaction with the iCat is rather low. Lars seems to be quite occupied with the game. After a "Lars? do you still like the game? " he just turns his head. The level of the sound is increased a bit. Lars tends to whisper his thoughts. Later on also verbal responses to Lou are collected although still scarce.
After:
Lou: How many marble do we need?

Lars: (after a few seconds) eleven!

In a dull moment of the session I command the iCat to fall asleep. At first this is not noticed. Then Lars notices this, and waves to the iCat. Later on in the bread shop:

Lou: It seems hard to continue here, shall I help you?

Lars: I'll succeed! (Goes on counting, 1, 2, 3, ...)

Having the iCat not too much in the line of sight has the advantage that when there is attention for the iCat, this is more visible, Still for the following tests the iCat is going to be placed more near to the computer.

Session 7

For the next sessions it is decided to let the iCat be more close to the line of sight of the child.

Vincent is a cheerful boy. Soon after I leave him, he tries to wake up Lou by touching his feet. Because I'm still starting up the program this does not work, and Vincent seems somewhat disappointed. Later on, after the game proceeds, the interaction with the iCat starts right away. Vincent is somewhat impatient, and eager to click on everything. Silently he whispers some thoughts in the game. At a certain point within the game he makes a mistake. He fills in, "A cat produces eggs"

Lou: Nooo, do cats produce eggs?? Then he plays on and after a few seconds:

Vincent: ... No, you are right ... Clearly Vincent is very occupied with the game, answering Lou just comes on a second place.

Lou: What does produce eggs?

Vincent: A chicken!

Lou: very good!!

While filling in the questionnaire Vincent makes a remark about my writing. It's too shifted. Furthermore he seems very well capable of understanding the smiley-o-meter, he tends to give the "not super" answers too.

Session 8

For the following test I want to be able to help the child in case it gets stuck somewhere. So some remote controlling software is installed and tested. Starting the

game with a network connected gives a problem though, Apparently there is some verification, and without tricks remote control software can not be used.

The introduction is more chaotic, because the sister and the father are staying in the room. At a certain point I'm ready to start the game, but the sister is still there and the Father has a phone call with somebody. Of course I have to entertain the children in the meantime.
Marco begins communicating with the iCat already at the name input screen. Therefore I wake up Lou a bit earlier. I can help Marco by telling him were he can find the start button. Later on in the game he keeps on asking more or less trivial questions. An other issue raised, one that was only discovered after analyzing the video. The mouse seemed to have problematic behavior. The video showed that Marco did not notice that he moved the mouse to much too the corner of the table. When the same camera as the monitoring camera would have been used, then this problem would not have occurred. Still I expect that Marco is somewhat less experienced than others in these sessions.
After 24 minutes Marco gets stuck somewhere, and I have to interrupt to help him out. At that moment I still do not discover what is wrong with his mouse.
A typical phrase from the transcription:
Marco: Where should I be for the bell?
Lou: Did you discover Hugo already? Hugo!, the small box!
Marco: Left or right?
Lou: Do you see the box?
Marco: I see a box! (Hoovers over it)... should I click?
Lou: Yes, try to click!
Marco: How should I get it? (The bell [Y])
Lou: Just try!

Some possibility to help in a more structural way can be handy. A remote control software package could be a solution, but just adding an extra mouse with longer cable probable works even better and is more reliable. Only it should not be touched in the control room, in any other situation than when it is needed.

5. Results

8 Sessions have been performed, with children in the age group of 6 till 8 years. 3 girls and 5 boys were involved, with an average age of 6,9 years. Some were about to proceed to group three of the Dutch education system, others were about to proceed to group 4. 2 children were attending a primary school in Belgium.

The data generated by the survey should be interpreted carefully. While adding and changing elements in the subsequent tests, there was the intention to let the tests run better. This means that the changes in results can be due to several factors.

When there was the intention to let the test run better, one would expect increasing average values of the measures along the sessions. There have been too little sessions to draw solid conclusions on this basis. In appendix 4 shows the coded results of the questionnaire.

All children enjoyed their session. They enjoyed playing this game, only one child was not enthusiastic about this game, it was too easy. More important, all children liked to talk with Lou. In one instance there was hardly any talking, still the child liked the idea.

The video footage is usable for detecting usability problems. All sessions have been transcribed for verbal utterances, and parts of the footage have been examined closer for usability problems. Some parts of this footage are comparable, for instance because all children visited the central screen next to "cloudland". It is hard to say whether the between sessions change behavior, caused the detection of more or less problems in the sub game. The observations suggest that other variables, such as the personality of a child, or its temporal mood have a greater influence.

For the facilitator it is needed to have some experience with usability testing at forehand. With the current setup the facilitator has to keep an eye on the child and what it is doing in the game, and it has to control the iCat with the user interface on the screen. This is a tall order. Besides the facilitator has to cope with circumstances that can not be fit within a protocol. This means a flexible attitude is required, and one should not be afraid to improvise. Towards the participants both children and parents the facilitator should have a fair judgment of the human nature. When the session has

to be interrupted, a second facilitator is needed in order to keep the robot moving, in order to preserve the believability of the setup.

The robot should be able to meet the following two requirements: 1. It should be able to act autonomous for at least short periods of time. 2. It should be able to be controlled in a convenient manner by the facilitator. For research purposes this demands a robot that is programmable, such as the Philips iCat.
The autonomy is needed to facilitate the current script, and to make it believable that the robot has a life of its own. The convenience of control is needed to lower the load on the facilitator.
The sessions that have been executed and analyzed suggest that the richness of the experience does not have to be very high. The latest version of the software did not include face tracking, or technology to move the eyes in a more natural manner. Still the impressions of the sessions prove that the children are prone to believe that it is some sort of autonomous robot. So, the robot does not have to exhibit all humanlike social characteristics, given by existing guidelines for social robots.

Because of the fact that the child is engaged with the game and all other impressions it is likely that the surroundings of the test environment have not too much influence on the test. But for the setup at least two separate rooms should be available, the facilitator has to hide, and use his voice to help, or to remind the child to think aloud. Usually these conditions can only be met in a usability lab with the right equipment. When the session is to be held elsewhere, two spaces are required. Furthermore one should have a camera for monitoring the child, and a camera or scan converter to monitor the screen.

6. Discussion

The setup as how it is tested can be seen as a real life situation. However due to the design of this particular study important variables have been kept fixed. This raises the question how the setup would work if those variables would be changed.
Believability is not a big issue in this setup. But when this setup would be used for an older group of children, believability might become a problem. Part of this setup is a story. Namely, Lou is a robot that can think along, and can help when the game gets too difficult. Although the facilitator plays a role in this, it is not conveyed how important this role is. An educated adult will see through this setup quickly because it is less naive towards this kind of research. Older children are in their development to grow less naive just like adults, and they will be more discerning toward what they are told to believe.
An open, and important question is, how this method will work with an other kind of product, for instance with a product that is in an earlier phase of its development. When a product is in an earlier phase of its development, the chances are high that the child will get stuck somewhere in its task. At that moment the facilitator should have means to help the child out, so the session does not have to be interrupted.
As the executed sessions show the behavior of the Robot do not have to be very sophisticated in order to accomplish a acceptable conversational partner. During the session the child does not look at the robot for long periods of time because it is too occupied with the task in the game. Maybe, when the robot would have a very realistic and autonomous behavior, it could divert the attention of the child from the game to the robot. In that case the robot becomes a far more interesting object to play with than the game to be tested.

While testing a computer product, the child has to remain seated to control this computer. This is convenient because the iCat stayed on a fixed position in respect with the child. When other non-fixed products have to be tested, such as a handheld computer game, the way the child interacts with the iCat may be quite different.

It is feasible to use a social robot as a indirect facilitator. This conclusion can be used to try out many other variations in the setup. There have been studies that let the iCat maintain a certain behavior such as more or less social behavior (

7. Conclusion

Along the sessions that have been run, it can be concluded that the children are well willing to engage in question - answer communication with the social robot. The task, namely playing the game takes a part of their attention as well, so they do not always pay attention to the robot. How much attention the robot will get depends on how often it tries to initiate contact, but also it depends on which product is tested, and the developmental stage this product is in. The stress of the search for usability problems will be on the analysis of the video data. Detecting major usability problems on the fly is possible, but, due to the load on the facilitator there will be hardly any time for that. The video footage however, is very well suitable for further analysis, be it by the use of own observation schemes or by using specialized logging software.

The children also experience this way of usability testing as fun. The answers on the questionnaire show this view, and the less structured 'buzz' around the sessions do indicate this as well.

The setup relies on several technical links which should function without problems. The human facilitator is extended with a computer to control the iCat, a computer to alter his voice, and screens to monitor the test. This proved to be feasible, but a lot of unforeseen issues can show up. Those do not relate to the product to be tested but rather the have to do with the usability test method itself. All further exploration of this method should be put in optimizing the way a test is run.

8. References

Baarda, D.B. & Goede, M.P.M. de (1996). Methoden en Technieken: Praktische handleiding voor het opzetten en uitvoeren van onderzoek. Houten: Stenfert Kroese.

Baauw, E., Markopoulos, P, (2004) A comparison of think aloud and post-task interview. (2004) Proceedings IDC 2004, June 1-3, 2004, Maryland, USA, , ACM Press, 115-117.

Bartneck C., and Forlizzi, J. (2004). A design centered framework for social human-robot interaction, Proceedings Ro-Man 2004, 591-594.

Dindler, C., Eriksson, E., Iversen, O.S., Ludvigsen, M., Lykke-Olesen, A. (2005). Mission from Mars – A Method for Exploring User Requirements for Children in a Narrative Space. Proceedings of Interaction Design and Children, IDC 2005, Boulder, Colorado, USA, Jun. 2005.

Donker, A., & Markopoulos, P. (2002). A comparison of think-aloud, questionnaires and interviews for testing usability with children. People and Computers XVI - Memorable Yet Invisible, Proceedings of HCI 2002, edited by X. Faulkner, J. Finlay, and F. Detienne (London: Springer), pp. 305-316.

Druin (ed.) (1999) The design of children's technology. San Francisco, CA: Morgan Kaufmann, 1999

DAUTENHAHN, K. (1999), Embodiment and Interaction in Socially Intelligent Life-Like Agents, In: C. L. Nehaniv (ed): Computation for Metaphors, Analogy and Agent, Springer Lecture Notes in Artificial Intelligence, Volume 1562, Springer, pp. 102-142

Hanna, L., Risden, K. & Alexander, K.J. (September 1997). Guidelines for usability testing with children. Interactions, 9-14.

Heerink, M., Kröse, B., Evers V., Wielinga, B., (2006), "Studying the acceptance of a robotic agent by elderly users," International Journal of Assistive Robotics and Mechatronics, vol. 7, 2006.

John, B.E. and Marks, S.J. (1997) Tracking the effectiveness of usability evaluation methods. Behaviour and Information Technology, Vol. 16, no. 4/5, 188-203.

Markopoulos, P. & Bekker, M.M. (2002). How to compare usability testing methods with children participants. In M.M. Bekker, P. Markopoulos & M. Kersten-Tsikalkina (Eds.), Interaction Design and Children, pp. 153-158. Eindhoven: Shaker Publishing.

MindScape, (2003), Robbie Konijn Groep 3: Pret in de Wolken (Robbie Rabbit, Group 3: Fun in the Clouds) [Computer Software] Mindscape.

Norman, D. A. (1988) The psychology of everyday things. New York: Basic Books.

Nielsen, J. (1993). Usability Engineering. Boston, MA: Academic Press.

Nielsen, Jakob, and Landauer, Thomas K. (1993)"A mathematical model of the finding of usability problems," Proceedings of ACM INTERCHI'93 Conference (Amsterdam, The Netherlands, 24-29 April 1993), pp. 206-213.

Nielsen, J, & Mack R.L. (1994) Usability Inspection Methods, John Wiley & Sons, New York.

Vermeeren, Arnold P. O. S., Bouwmeester, Karin den, Aasman,
Jans and de Ridder, Huib (2002) 'DEVAN: a tool for detailed video analysis of user test data', Behaviour & Information Technology, 21:6, 403 - 423

Appendix XXXX

This is a protocol which describes the way an iCat can be applied for a usability test. In the tests that have been performed during my graduation project, the game "Robbie Konijn" has been used as a standard piece of software to put on the test. In principle this can be any other product (as this is the purpose of the evaluation method). Therefor the names of the iCat, and the product are shown in italics.

Protocol for the test.

Technical checklist:
- Electriciity in Kidslab up.
- iCat Control PC up and running
    - iCat **intro** program runs
    - VNC viewer program standby
- KidsLab game PC up and running
    - Game is standby
    - VNC server program is running
- Voicemorph computer is up and running.
    - Voicemorph program is up and running
- Videorecording computer is up and running
    - Video recording software should be standby.

Soundcheck:
- Voicemorph same level as pre-programmed samples
- Gamesound on a moderate level, should be lower than the voicemorph level.
- Microphone should receive the childrens sound clearly. In the livingroom of the Kidslab should one directional microphone be pointed to the child.
- Monitor level in control room should be OK.
- Keep the doors closed, as other sounds may interfere.

Imagecheck:
- Check the position of the camera's in advance. Repositioning makes audible noise
- Scanconverter should work fine (VGA signal in control room)

Check other goods:
Fresh drinks for the children are available?
Little presents to give away afterwards are available?
Consent forms are available? (on the table)
Ball-point pen is available? (for the consent form)
Questionaire is available? (on the table)
Interaction script ready? (In control room)
Color felt-tip pens are ready? (for the questionaires)
Glasses of water for the facilitator (against a hoarse voice!)

Protocol:

Requisitories / assumptions:

The parents have been informed about the procedure in an earlier contact. So they know what will happen. They agree with the fact that a video recording will be made.

After the participants have arrived:
- Ask them if they want something to drink. Offer different juices to the kids.
- Try to level with the child, by asking simple questions and do some chit-chat.
- Show them around in the building and the room, also show some other projects. (Show other rooms, but not the controlroom)

- At this moment start the videorecording.
- Tell them again about the procedure (assure them, nothing bad will happen) .
The procedure: The talk about the procedure encloses the following:
About what to do:
- The kid is going to play an educational computergame. (tell what the game is about)
- The kid should just try to play as good as possible, but if it seems difficult, thats not a problem.
- Tell the that the name of the iCat is "*Lou*" (just name him *Lou*, avoid "iCat").
- We test the game, we do not test the kid.
- Tell the kid that it will not hurt the feelings of the evaluator (*Lou* or myself) if they do not like the program
- We (myself and *Lou*) appreciate it when the kid thinks aloud. "ehhmmm" is good!
- We are free to stop at any time.
- Tell the child that everything is being videotaped, and tell them why.

About the cat (*Lou*)
- The Cat (*Lou*) can be asked for help (but don't expect too much!!)
- The Cat (*Lou*) is shy and sleepy when we (parents/investigator) is around. Therefore we will hide behind a door.
- The Cat (*Lou*) appreciates it when the kids thinks aloud. So, "uhhhmmms" are allowed!.

What else:
- Tell the kid that he/she will be alone. But we are just around the corner!.
- Ask some basics about the computer and the present screen. This involves:
Point me the cursor
Do you know the difference between left and right?
Tell about the game Robbie Konijn, it about raincoats and umbrella's. The task is to decrease the problem of the falling umbrella's
- The room has camera's, the images will not be used outside this room.
- Give the consent form to the parent.
- Awake the cat, and watch its movements.... It takes 17 seconds for the intro. After the long blink, it will show a happy, sad and surprised face.(takes 13 seconds).

it will fall asleep soon after that...
- Start the game.
- Go with the parent to the other room
- stop the intro program.
- start the WOZ program, let the cat get awake soon, but not in an important part of the game (preferably just after the introduction)

The test itself
Let the child play the game
Depending on the sort of interaction you want:

use the iCat to get the kid more talkative...

Remind them of thinking aloud, give prompts ("vertel wat je denkt", "wat denk je?" etc.)

Do not let them keep silent too long

This should be covered in the interaction script. .

Help them out if they ask for it and if they have tried it themselves

If it takes too long to perform a task, help the children to complete the task

This can by done by using VNC or using an extra mouse

Stop the test after an hour, or when specific tasks are fulfilled (In the case of Robbie Konijn stop after an hour)

Try to stop at an appropriate moment, so for instance when a new game screen showed up.

After the test:

Thank the child

Fill in the extended questionnaire during the remaining structured "interview"

Explain them about the "smileyometer"

Ask if they can read the questions themselves let them try. They will be proud to show if they can, although it might go very slow.

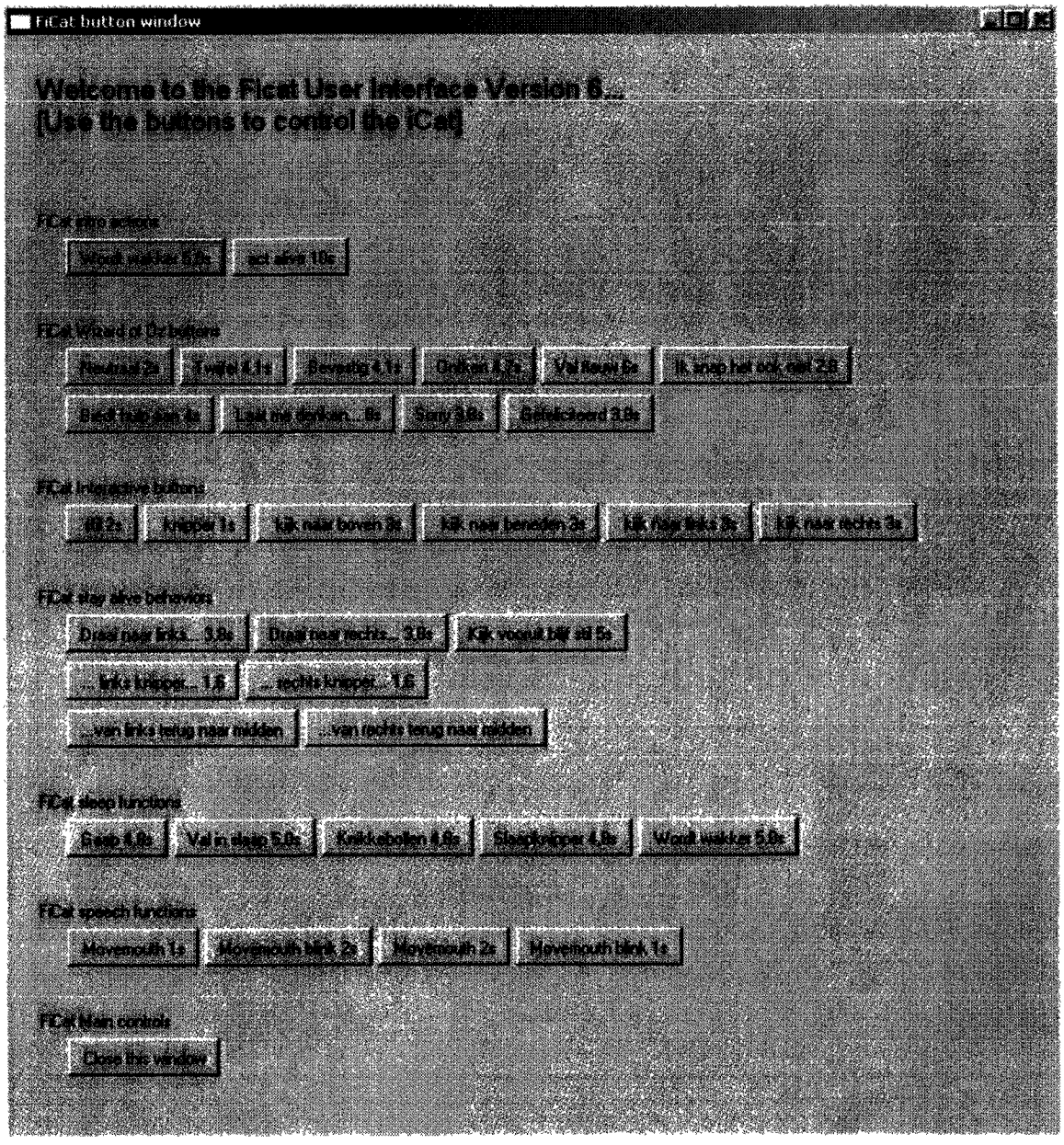
It is no problem to read the questions aloud after this. Try to stay neutral and not suggestive (ofcourse!)

Tell the children not to talk about this evaluation to other children, because some of them still have to perform the test as well.
Thank them again, and tell they have been of great help, give them a small present show this.
Show and walk along with them to the exit of the building.
Give parking exit cards to the parents, if needed.

- stop the videorecording.

Practicalities when Robbie Konijn as a game is used.
In the current setup Robbie Konijn does not install correctly when the PC has no administrator rights.

Appendix 2:
Screenshot Ficat User Interface v6.



Appendix 3: Questionnaire after session

Vragenlijst pagina 1/3

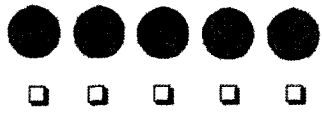
Vragenlijst v7 *** Samen invullen... ***

1. Ik ben een: Meisje / Jongen

2. Ik ben: _____ Jaar

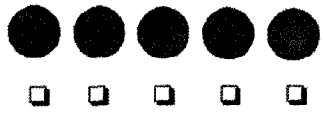
---

3 Ik vond het spel leuk om te spelen.



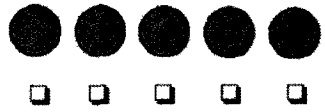
---

4 Ik vond de figuren in het spel mooi om te zien.



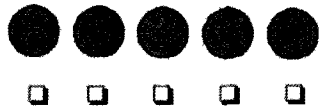
---

5. Ik vond het leuk om met lou te praten.



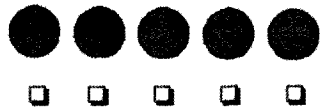
Vragenlijst pagina 2/3

6 Ik vond dat Lou me goed hielp.



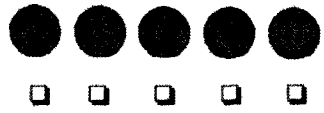
---

7 Ik vond Loe best wel slim.



---

8 Ik denk dat ik dit spel vaker zou willen spelen.



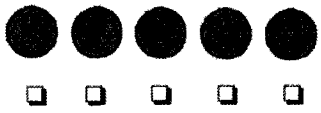
---

9 Als ik iets in het spel mocht veranderen, dan zou ik:

Vragenlijst pagina 3/3

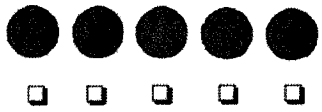
Algemeen: Houding

11. Ik vind het leuk om met een computer te werken!



Algemeen: Ervaring

12. Ik heb al veel met computers gespeeld.



Dat was het! Bedankt voor het meedoen!