Design and evaluation of an avatar-mediated system for child interview training
Abstract
There is an apparent problem with children being abused in different ways in their everyday life and the lack of education related to these issues among working adults in the vicinity of these children, for example as social workers or teachers. There are formal courses in child interview training that teach participants how to talk to children in a correct manner. Avatar-mediation enables new methods of practicing this communication without having to involve a real child or role play face-to-face with another adult. In this study it was explored how a system could be designed in order to enable educational practice sessions where a child interview expert can be mediated through avatars in the form of virtual children. Prototypes were developed in order to evaluate the feasibility of the scenario regarding methods for controlling the avatar and how the avatar was perceived by the participants. It was found that there is a clear value in the educational approach of using avatar-mediation. From the perspective of the interactor it was found that using a circular radial interface for graphical representation of different emotions was possible to control a video-based avatar while simultaneously having a conversation with the participant. The results of the study include a proposed design of an interface, description of underlying system functionality and suggestions on how avatar behavior can be characterized in order to achieve a high level of presence for the participant.

Key words
Avatar, avatar-mediated, child interview training, communication, control, co-presence, emotion, facial expression, interaction, interface, non-verbal communication, presence, realism, realistic, social presence
# Table of contents

## 1 Introduction
- 1.1 Background ........................................ 1
- 1.2 Purpose and goal ..................................... 1
- 1.3 Problem domain and research questions .......... 2
  - 1.3.1 Interaction aspects ............................. 2
  - 1.3.2 Avatar behavior ................................ 2
  - 1.3.3 Research questions ............................. 2
- 1.4 Limitations ........................................... 3
- 1.5 Outline of the thesis ................................ 3

## 2 Theory
- 2.1 Mediated communication ............................. 4
- 2.2 Avatar-mediated communication .................... 4
  - 2.2.1 Definition ....................................... 4
- 2.3 Examples of applications for avatar-mediated communication ........................................... 4
- 2.4 Human communication in mediated environments ......................................................... 5
  - 2.4.1 Non-verbal communication ...................... 5
  - 2.4.2 Facial expression and human emotion .......... 6
- 2.5 Interaction aspects .................................. 6
  - 2.5.1 Multiple levels of control and automation .... 6
  - 2.5.2 Behavior states .................................. 6
- 2.6 Social presence and co-presence ................... 7
  - 2.6.1 Definition ....................................... 7
  - 2.6.2 Measuring social presence and co-presence ... 8
  - 2.6.3 Avatar level of detail and realism .......... 8
- 2.7 Interface design ..................................... 9
  - 2.7.1 Design principles ............................... 9
  - 2.7.2 Interfaces for controlling moods in avatars ... 9

## 3 Methodology
- 3.1 The design process .................................. 11
  - 3.1.1 Interaction design ................................ 11
  - 3.1.2 Co-design ....................................... 11
- 3.2 Literature review .................................... 12
- 3.3 Establishing requirements ........................... 12
  - 3.3.1 Stakeholders ..................................... 12
  - 3.3.2 Initial requirements ............................ 13
- 3.4 Evaluation of Prototype 1 .......................... 13
  - 3.4.1 Training sessions ............................... 13
  - 3.4.2 User questionnaire ............................. 13
  - 3.4.3 Post session interview with the expert ........ 14
- 3.5 Evaluating Prototype 2 ............................. 14
4 Prototype design process

4.1 Developing Prototype 1

4.1.1 Concept description

4.1.2 Video sequences

4.1.3 Identified general goals of the system

4.1.4 Identified prototype technical requirements

4.1.5 Prototype presentation

4.2 Evaluating Prototype 1

4.2.1 Trial 1: Description

4.2.2 Trial 1: User survey results

4.2.3 Trial 1: Identified problems from the interviewer's perspective

4.2.4 Trial 2: Description

4.2.5 Trial 2: User survey results

4.2.6 Trial 2: Identified problems from the interviewer's perspective

4.2.7 Trial 1 and 2: Identified problems from the expert's perspective

4.2.8 Findings from the prototype evaluations

4.3 Developing Prototype 2

4.3.1 Avatar video sequence requirements

4.3.2 Interface requirements

4.3.3 Extending emotions and behaviors

4.3.4 Control of behavior

4.3.5 Interface approaches

4.3.6 Prototype presentation

4.4 Evaluating the first iteration of Prototype 2

4.4.1 User test description

4.4.2 User test results

4.5 Developing the second iteration of Prototype 2

4.5.1 Prototype presentation

4.6 Evaluating the second iteration of Prototype 2

4.7 Technical implementation

5 Results

5.1 Prototype 1

5.2 Prototype 2

5.2.1 Avatar behavior

5.2.2 User interface

5.3 Research questions

5.3.1 Research question 1

5.3.2 Research question 2

6 Discussion

6.1 Reflecting on previous research
6.2 Interpreting the data collected 32
6.3 Possible limitations to the study 32
6.4 Future work 33
  6.4.1 Behavior and emotions 33
  6.4.2 Evaluating competence outcomes 33
  6.4.3 Further evaluations 33
  6.4.4 Sound processing 33
  6.4.5 Hardware controllers 33

7 Conclusion 34

References 35

Appendix I
  Appendix A Interview with Pär Stihl I
  Appendix B Trial 1 questionnaire VI
  Appendix C Trial 2 questionnaire VIII
  Appendix D Selected interface design alternatives IX
Table of figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adapted from Plutchik’s wheel of emotions</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Concept overview</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>Overview of the prototype</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>Interface of Wirecast with avatar sequences</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>Skype view from the expert’s perspective</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>Trial 1 survey results</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>Interface of Wirecast with avatar sequences</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>Skype view from the expert’s perspective</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>Trial 2 survey results</td>
<td>21</td>
</tr>
<tr>
<td>10</td>
<td>Radial interface</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>Vertical sliders</td>
<td>26</td>
</tr>
<tr>
<td>12</td>
<td>Radial interface</td>
<td>26</td>
</tr>
<tr>
<td>13</td>
<td>Sliders interface</td>
<td>26</td>
</tr>
<tr>
<td>14</td>
<td>Second iteration of the prototype interface</td>
<td>28</td>
</tr>
<tr>
<td>15</td>
<td>Final prototype interface: interactors's perspective</td>
<td>29</td>
</tr>
<tr>
<td>16</td>
<td>Final prototype interface: participant's perspective</td>
<td>30</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Background

There are many reported cases of children being abused by family members or adults they know and have close relations with (Rädda Barnen, 2012). It is believed that the amount of hidden statistics in this area is large. According to a study less than 10% of abused children talk to a professional about their situation (Rädda Barnen, 2012). The study also reports that approximately 15% of school children in Sweden state that they have been physically abused at home. A report by the Swedish government agency Barnombudsmannen (2012) indicates that children feel like they are not being listened to about their situation.

Teachers have a unique possibility to spot and engage with these potential situations since they meet children every day. Unfortunately research findings show that children feel that teachers do not have enough knowledge and skills to help them (Bäckman & Forslund, 2013). This is most likely due to the teacher not being used to the situation and not knowing how to act correctly or not knowing which signals to look for. The same study shows that there are almost no direct discussions being held with the affected child. There is a need for competence among the school teachers, and they need to learn the communication skills required for these types of situations. Teachers are also insecure about reporting such incidents to the social services.

At The Institute of Police Education at Linnaeus University methods for child interview training are taught. A standard web-based video conferencing system is being used to conduct role-playing scenarios where one person acts as the child and another as the child interviewer. It is shown to a large degree that there is a value in these types of training scenarios according to Stihl. However, the platform used (Adobe Connect) is not fully suited for the purpose as it is essentially merely a video-based conversation between two adults. The visual and auditory appearance of the person role-playing a child is normal and unchanged. Therefore, there are limitations in terms of perceived realism from the participants perspective according to Stihl.

1.2 Purpose and goal

The work in this thesis consists of research and development around the design of a system for child interview training. The contribution of this system would ideally be that the participant can practice on how to look for the communication signals of an abused child. The participant could learn how to interact and ask relevant questions in a correct manner directly to the child. Rather than seeing and talking to an adult as in the previously described solution used at the institute of Police education, it is proposed that a system based on avatars has the potential to be more effective.

An avatar can be used as a virtual character representation which can be controlled by a real person (from now on referred to as interactor). In a sense the interactor’s behavior is mediated through an avatar. Apart from being widely used for entertainment purposes avatars are also used in different ways in fields such as education (Nagendran et al., 2013), medical settings (Andrade et al., 2010; Andrade et al., 2011) and remote communication (Vilhjálmsson, 2003). The proposed system is based around an avatar, which is being controlled by an expert in the field of child interview training. The goal is to increase the realism of the training scenario and contribute to its effectiveness in terms of learning. Increased knowledge and confidence among teachers also increases the possibilities for children to communicate about complicated matters such as abuse.

---

1 Pär Stihl, teacher at The Institute of Police Education, interview 11 November 2014
While in the scenario for police education, which was previously described, video conferencing was the used method for role-playing a child, the situation can be seen as not being authentic because of the evident adult who is sitting in front of the camera. One solution could be to only rely on audio for communication, however it is fair to assume the conversation will not be as rich as with video being enabled.

The research work presented in this thesis comprises a description of the design process of two different prototypes developed in close collaboration with an expert on child interview training. A working prototype of the system was developed and evaluated with a group of social workers in Sweden. Even though they were not school teachers, they are part of a group who very often encounter children. It is argued that the work in this thesis could benefit any target group of people who in their line of work meet children on a daily basis.

Based on the findings from the evaluation of the working prototype an interface design was developed and evaluated together with the expert and additional users.

1.3 Problem domain and research questions

1.3.1 Interaction aspects
Since the interactor will need to sustain a conversation while simultaneously controlling the avatar behavior the interface needs to be effective in terms of required user interaction. Vilhjálmsson (2003) highlights the importance of finding a balance between what an avatar can express and how much burden is put on the user in terms of control. In normal face-to-face conversation it is uncommon that one’s behavior requires a conscious effort. Vilhjálmsson (2003) introduces the concept of automated augmentation, where a system is used to control certain aspects of avatar behavior in order to reduce the overhead of interaction required from the user.

The research focus in this thesis is on how the person being mediated interacts with the mediating layer, i.e. how the avatar is being controlled and how the interactor can control the avatar in an efficient low-workload manner.

1.3.2 Avatar behavior
How an avatar is behaving and is being experienced are key factors when realism is the goal. In a sense the goal is to make the participant feel like the avatar is a real person. In the field of avatar-mediated communication this phenomena is referred to as social presence or co-presence (Biocca, Harms & Burgoon, 2003). The research focus is also on exploring how the participant talking to the avatar experiences the mediation, in order to identify which aspects that can lead towards a more realistic experience in the training scenario.

1.3.3 Research questions
The scope of the research is defined with two research questions.

• Research question 1: Which interaction aspects can contribute towards a low effort for the interactor in a video-based avatar-mediated system for child interview training?

• Research question 2: Which aspects of the avatar behavior can lead to increased experience of co-presence for the participant?
1.4 Limitations

The proposed solution consists of a custom-made training platform using avatar-mediation for practicing child interview training. An expert may control an avatar in a way that can be compared to the Wizard of Oz-approach\(^2\) (Rogers, Sharp & Preece, 2011). Hence, there will be no aspects of Artificial Intelligence (A.I.) involved. There are plenty of research challenges related to how an avatar can communicate and adapt to a user’s behavior autonomously (Nagendran et al., 2013). However, for this thesis the goal is to have both the expert and the teacher engaged in a spoken conversation. Any aspects regarding real time analysis of the interactor’s intentions are beyond the scope of this research. At the time of writing this thesis one can assume that a real person can communicate with unmatched levels of realism and intelligence compared to that of an A.I.-based avatar, resulting in being able to adapt more naturally to a conversation (Blascovich et al., 2002; Ahn et al., 2012).

Controlling the avatar with total freedom in terms of movements will not be taken into account. Research about how to use face recognition and motion tracking technology for these types of applications is not an area that this thesis approaches. Nagendran et al. (2013) describe a novel approach for controlling an avatar using special hardware devices which enables the use of head tracking and alternative input methods. This could possibly result in an efficient interaction method. However, in the scope of this thesis, no special hardware outside of what is found in a standard laptop or tablet computer is required. The incentive to this is due to the possible spread of the system to a wide range of users and therefore, accessibility and a low threshold of getting started are key factors.

1.5 Outline of the thesis

This thesis is divided into seven chapters. After the introduction, Chapter 2 contains theory related to avatar-mediated communication, human communication in mediated environments, interaction aspects, social presence and co-presence and finally interface design.

Chapter 3 presents the methodology used for the study. This includes a description of the design process, the literature review, how requirements were established and how the prototype was evaluated together with users.

In Chapter 4 the prototype design process is described. The description of the process contains motivation to how the prototypes were formed and description of the iterations and evaluations.

Chapter 5 presents the findings from the design process, which are the lessons learned from designing and evaluating a prototype application used in a course with 15 participants and the interface that is developed with the expert needs in focus.

Chapter 6 consists of a discussion of the results with a reflection back on previous research, interpreting the data collected, identifying possible limitations of the study and a suggestion of further investigations that could be done.

Finally, Chapter 7 presents the conclusions that are drawn from the work in this thesis.

\(^2\) An experiment where a user believes to be interacting with an autonomous system which is actually operated by a hidden human being.
2 Theory

2.1 Mediated communication
Mediated communication can refer to any human communication which uses some sort of information technology. Computer-mediated communication is a subset of this area and it is a way of achieving human communication with the use of computers over local area networks and over the internet (Simpson, 2002).

2.2 Avatar-mediated communication

2.2.1 Definition
An avatar is a mediation of a human or artificial intelligence entity in a virtual environment and comes in the form of a visual embodiment. In the field of A.I. an avatar can also be categorized as an embodied conversational agent (Cassell, 2001), which appears as a graphical representation, most often in human form, for the user to interact with.

Avatars can also come in the form of mediation of a human user interacting and communicating with another human in real time. The avatar can change the way the other part is experiencing the user. Historically avatars have been used in virtual environments such as graphical chat, multiplayer games, online learning and military training simulations (Vilhjálmssson, 2003). Depending on the application an avatar can come in the form of a 2D or 3D representation and in varying levels of detail and realism in terms of appearance and behavior.

Bailenson and Blascovich (2004) describe the characteristics of avatar-mediated communication:

Using an avatar to interact with another person is qualitatively different from other forms of communication, including face-to-face interaction, telephone conversations, and videoconferencing. Via an avatar (which is constantly redrawn in real-time), interactants possess the ability to systematically filter their physical appearance and behavioral actions in the eyes of their conversational partners, amplifying or suppressing features and nonverbal signals in real-time for strategic purposes. In other words, interactants can use avatars to bend, twist, or entirely remove truth from a conversation.

(Bailenson & Blascovich, 2004:9)

Even though Bailenson and Blascovich (2004) present the possibility of removing truth as an ethical problem there are evident benefits of this when avatars are employed in context of simulation and training where role-playing is a key part and truth is not part of the equation.

Bailenson et al. (2006) define an avatar as a digital model representing a user either in terms of looks or behavior. Furthermore, a certain type of avatar can be considered being a puppet. A puppet can for instance have a high behavioral similarity (it moves and behaves similar just as the interactor does) while at the same time having a low form similarity (not looking anything like the user) (Bailenson et al., 2006).

2.3 Examples of applications for avatar-mediated communication
Today one of the most common uses of avatars is within entertainment applications such as games. For instance in massively multiplayer online role playing games where the player interacts with the virtual environment and have social interactions with other
users through a virtual character. This character can be considered an avatar that becomes a mediation of the user’s self.

Aside from entertainment purposes avatars can play a key role in applications for simulations and learning. Andrade et al. (2010) explore how an avatar-mediated training system can be used in a medical setting for practicing on how to deliver bad news to a patient. Delivering bad news is something that causes stress among medical personnel and the way that it is communicated can heavily affect the well-being of the patient. The scenario is customizable in terms of environment and the appearance of the patient and it also lends itself to fit with distance learning methods. Results from the study reported that while the participants did not feel as the use of avatars could completely replace the experience of real patient interactions, it provided a viable educational approach. After the experiment the participants reported higher levels of self-efficacy in delivering bad news.

Nagendran et al. (2013) propose a novel system for interacting with virtual students in a classroom setting. The purpose is to let the users of the system practice the situation of being a teacher in front of a class while addressing different types of student behavior. The developed system builds on the notion of digital puppetry. The virtual student avatars can be controlled in large detail manually using special hardware but the system is also capable of showing different pre-recorded idle behaviors such as act bored and look attentive without the needed active interaction from the interactor.

Ang et al. (2013) explore how an avatar-based system can support the communication among young users in brainstorming and negotiation tasks. Avatar-mediated communication is compared to video-mediated communication and it is found that the participants liked and trusted their partner to a notable higher degree in the case of an avatar being used. In the study this result is correlated to the reduced perception of status, power and social norms.

2.4 Human communication in mediated environments

2.4.1 Non-verbal communication

In face-to-face conversations human beings are rather effortlessly using several communication channels. Speech, hand gestures, facial expression, body posture, orientation and eye-gaze can be considered different channels in which communication can take place (Cassell & Vilhjálmsdóttir, 1999; Louwerse et al., 2007). Furthermore, Cassell and Vilhjálmsdóttir (1999) argue that the channels can be complementary such as being able to signal information and give feedback in another channel while the other part is talking. This type of simultaneous use of multiple channels is by Cassell and Vilhjálmsdóttir (1999) referred to as a multimodality.

Turn-taking management is another important aspect of conversations within mediated environments (Tanenbaum et al., 2014). In order to have a natural conversation where people do not talk at the same time there needs to be a way of managing the turn-taking. For instance a person can request to talk or give the word to another person. These behaviors of turn-taking often happen in a non-verbal manner using body language and eye-gaze (Tanenbaum et al., 2014).

Non-verbal communication is an essential component in human communication (Patterson, 1990). It may also be an important basis for social presence (explained in chapter 2.6) within mediated communication (Biocca, Harms & Burgoon, 2003). Communication in general is believed to deliver higher levels of perceived social richness the more modalities that are being used (Short, Williams & Christie, 1976; Rice, 1993; Sallnäs, 2004).
2.4.2 Facial expression and human emotion

Facial expression is an essential part in how people show emotions in face-to-face communication. Whittaker (2003) argues that communicating one’s facial expression is beneficial when social interaction is the goal. Furthermore, it is suggested that communication for purely task oriented purposes might actually benefit from an absence of visual features.

According to Ekman (1992a) there are six basic emotions which can be universally understood from observing someone’s facial features: happiness, surprise, fear, sadness, anger and disgust. Ekman (1992a) argues that more emotions exist but that they are combinations of the six emotions previously mentioned. Furthermore, Ekman (1992b) presents the idea of emotion families, meaning that a basic emotion does not have one unique facial characteristic but rather being comprised of many different facial features. Plutchik (1991) on the contrary argues that there are eight emotions, the six previously mentioned and interest and contentment in addition. In Plutchik’s model the emotions can be combined creating a total of 24 emotions that each can vary in a degree of intensity.

2.5 Interaction aspects

In his dissertation Vilhjálmsson (2003) argues that in avatar-mediated systems there should be a balance between what is possible to express and how much burden is put on the user in terms of control.

2.5.1 Multiple levels of control and automation

Building on previous research by Blumberg and Galyean (1995) Vilhjálmsson (2003) presents ideas of multiple levels of avatar control. It might not be feasible to rely on a human interactor to control an avatar all the way down to the finest detail in movement and behavior. Therefore, based on the circumstances an interactor might want to transfer control to the system.

Vilhjálmsson and colleagues conducted an experiment in which different levels of avatar control were compared (Vilhjálmsson, 2014). Fully manual, hybrid and fully automated. In a situation where the avatar was fully automated (in a Wizard-of-Oz manner, although without the knowledge of the participant) the participant reported higher level of perceived naturalness, expressivity and conversational control. Vilhjálmsson (2003) argues that when users are freed from the overhead of management of non-verbal behavior they can concentrate more on the actual conversation. The results also showed that participants that were using an automated behavior method recalled more of the conversation afterwards.

Blumberg and Galeyan (1995) present a solution for an interactor to control an animated character and divide actions into three levels: motivational level, task level and direct level. Depending on the situation an interactor might benefit from choosing a behavior for either one of the levels; resulting in not having to micromanage the control of the behavior more than needed.

2.5.2 Behavior states

In their avatar-mediated communication system Nagendran et al. (2013) developed animations for 3D-avatars that applied to certain behaviors like look attentive and act bored with the system applying transitions as needed. Vilhjálmsson (2003) argues that in order to be perceived as realistic the animated behavior of an avatar should stay in motion even after a certain user input has ceased. These two aspects relate back to the notion of delegated control to the system as previously described.
Furthermore, Nagendran et al. (2003) highlight the issue with transitioning between different states. In their application the avatars have certain key-poses that make it possible to transition between different animated behavior states.

2.6 Social presence and co-presence

2.6.1 Definition

The term social presence originates from the field of social psychology and was first introduced as a way of evaluating the telecommunications media (Short, Williams & Christie, 1976). It has since then been further defined and classified (Biocca, Harms & Burgoon, 2003) and is widely adopted in research on computer-mediated communication such as video conferencing, telepresence, virtual environments and avatar-mediated communication.

In the field of human-computer interaction social presence is being used as a measure of how interfaces affect and shape the perception of the experience of being with another. Thus, social presence is closely tied to avatar form and behavior and is an important aspect since it is a way of transferring the experience of being with another person (Bailenson et al., 2006).

The general definition of social presence is according to Biocca, Harms and Burgoon (2003:456) “the sense of being with another”, where “another” can be another person or an avatar which is a representation of that person. In some cases it could also mean having access to an artificial intelligence that is in control of an avatar. Furthermore, Biocca, Harms and Burgoon (2003) deepen the definition by classifying social presence into three different categories that are explained in the following subsections.

2.6.1.1 Co-presence

According to Biocca, Harms and Burgoon (2003) co-presence is a measure of to which degree another part is experienced as being there. Therefore, it is to be defined on a continuum rather than being a binary measure.

- **Sensory awareness of the embodied other** - representation of the other through the different senses.
- **Mutual awareness** - the user is aware of the mediated other and vice versa.

2.6.1.2 Psychological involvement

Merely observing that an avatar is present is not enough to be engaged psychologically. The avatar needs to communicate some sort of intelligence, be it artificial or human. Biocca, Harms and Burgoon (2003) make the connection to a virtual environment that may be populated by many virtual characters but if they are not controlled by an intelligence they will be regarded more as sculptures than presence of intelligence.

- **Sense of access to intelligence** - as suggested by Biocca (1997), being able to make a model of the intentions of the other is a fundamental element of social presence. Furthermore, Biocca (1997) states that social presence is activated the moment the user observes that an entity has minimal levels of intelligence in reactions to its environment and the user.
- **Salience of the interpersonal relationship** - to be able to be psychologically involved on a level that goes beyond merely giving attention to the other part (Short et al., 1976).
• **Intimacy and immediacy** - feeling psychologically involved with the other to such a degree that relationships can be created.

• **Mutual understanding** - being able to reach an understanding of the other part and project a sense of self through the restrictions of the medium.

### 2.6.1.3 Behavioral engagement

Depending on the complexity of a medium it can be possible to communicate behaviors to a varying degree. Due to the development in computer mediated communications it is possible to go beyond the possibilities that voice and text communication once made possible. Newer mediums and virtual environments can for instance allow the transmission of behavior such as eye contact, turn-taking and non-verbal mirroring.

### 2.6.2 Measuring social presence and co-presence

Biocca, Harms and Burgoon (2003) give a thorough overview of how social presence and co-presence are measured across different platforms for mediated communication. The main problem that is highlighted is that social presence and co-presence have their foundation in a quite diverse set of variables. It can for instance be a measure of an actual interaction, such as how a conversation was perceived. Whereas in other cases it can be a measure of the communicative properties that are enabled by certain technical aspects of a medium. For example how well a phone conversation was perceived compared to a conversation augmented with additional modalities such as video.

Within research of avatar mediation it is common to focus on the properties of the medium when measuring social presence. Biocca, Harms and Burgoon (2003:469) argue that social presence should be regarded as a phenomenological state: a subjective experience where the aspects of medium, knowledge of the other, content of communication, environment and social context should be taken into consideration.

If a system that uses avatar-mediation is to be designed it could be argued that it is logical to measure the aspects connected to avatar-mediation rather than the effects of video conferencing in general. It is probable that there is an overlap as a user’s experience of an interaction with an avatar is likely to be affected by the technical aspects which are similarly present in normal video conferencing applications. Having a clear sound and a well functioning video stream are important in order for the user not to be distracted and lose focus on the actual conversation with the avatar.

### 2.6.3 Avatar level of detail and realism

Kang, Watt and Ala (2008) compare the social presence of avatars with varying levels of realism, from video to animation to simple sketches. It is found that avatars with low level of realism will impair social co-presence and social richness.

Bente et al. (2008) conducted a study comparing differences in social presence between text chat, audio, video chat and avatars. In the study it was shown that audio, video chat and avatars outperformed text-chat in regards of aspects of communication richness. It is also notable that while there were no large benefits found in perceived presence between audio and video, there were also no substantial differences between the use of video chat and avatar-mediation. Bente et al. (2008) suggest that this indicates that there is no significant information loss when communicating through avatars compared to standard video communication.

Blascovich et al. (2002) report that an avatar with a high level of behavioral realism resulted in a higher level of co-presence among the participants. More in detail Blascovich et al. (2002) argue that the aspect of mutual eye-gaze between the avatar and participant was the factor that led to this result.
2.7 Interface design

In a user-centered design process it is naturally of most importance to consider factors such as goals, tasks and needs of the different end-users of the design (Gould, Boies & Ukelson, 1997). Likewise the choice of approach for interface design is dependent on such user factors. Rogers, Sharp and Preece (2011) argue that the interface design should be based on the requirements derived from the targeted user experience goal and that the context of the user affects how well a certain interface is suited for a specific task.

2.7.1 Design principles

Design principles can be used in order to design for a certain user experience. Rogers, Sharp and Preece (2011:25) define design principles as “generalizable abstractions intended to orient designers towards thinking about different aspects of their design”. Furthermore, Rogers, Sharp and Preece (2011:26) present five core design principles of interaction design:

- **Visibility** - Functions need to be visible in order for them to be used. Functions that are out of sight for the user are less likely to be used.

- **Feedback** - Information about the result of a certain action needs to be sent back to the user. This information can come in different forms such as visual, audial, verbal or tactile.

- **Constraints** - Not overloading the user with choices regarding possible interactions can enable a more effective interaction to take place. An interface could for instance be designed in a way that presents a minimum of what is necessary at a certain moment.

- **Consistency** - Being consistent about the mode of operation in interface interaction is argued to support the learning process. A user should not have to remember arbitrary interactions for similar operations.

- **Affordance** - A physical object has an affordance. It is instinctively known when an object should be grasped, pulled or pushed. Screen-based user interfaces are virtual but they have perceived affordances. Norman (1999) argues that these affordances become more of a convention that is learned over time.

2.7.2 Interfaces for controlling moods in avatars

Bittorf and Wuethrich (2012) compared different interface approaches for controlling the mood of an avatar. Their approach was adapted from the model of Plutchik (see figure 1) and translated into direct interaction with a mouse, making it possible to click the different areas in the model in order to alter the state of the mood of the avatar. The model was also adapted to physical devices such as a midi keyboard and a sound mixer. In the study they found that the tangible approach with the sliders performed better overall followed by the mouse when it came to ease of use, intuitivity, learning curve and expectations.

The participants in the study were asked to choose a corresponding behavior to a conversation that was read to them, rather than having to improvise and sustain a simultaneous conversation themselves.
Figure 1: Adapted from Plutchik’s wheel of emotions
3 Methodology

3.1 The design process

The work in this thesis originates from the perspective of interaction design in order to explore issues related to usability and user experience. The process is following widely adopted principles for user-centered-design in order to reach system- and usability goals. Gulliksen et al. (2003) identify key principles for user-centered systems design where user focus and active user involvement are two of the central principles. Gulliksen et al. (2003) argue that user focus is an essential part in a development process and should be sustained in order to have a focus on the users’ goals, tasks and needs. Active user involvement is accomplished by having representative users participating at the start of and throughout the development process. Users can be involved in the project on various levels, from direct involvement in the development project to contributing to organizational development and designing work practices. A concrete plan on how to involve users should be made for the beginning of the project.

3.1.1 Interaction design

Rogers, Sharp and Preece (2011:15) define the process of interaction design as consisting of four basic activities that are to be iterated until the desired goal is reached.

These activities are:
1. Establishing requirements
2. Designing alternatives
3. Prototyping
4. Evaluating

These four activities were used as the basis for the design work described in this thesis. Several iterations were completed for each step in the design in order to reach the design goals and requirements that were defined.

3.1.2 Co-design

In the process of co-design the traditional user becomes a co-designer and the role of the designer becomes more that of a facilitator that lets the users develop solutions for themselves (Sanders & Steppers, 2008). A system does not only have one single user type. Therefore, different stakeholders are engaged in the process and this is regarded as beneficial in order to make sure the concerns of the several users are taken into consideration.

Co-design is also a process known for being used mainly in educational contexts. Penuel, Roschelle and Shechtman (2007) define co-design as:

[…] a highly facilitated, team-based process in which teachers, researchers, and developers work together in defined roles to design an educational innovation, realize the design in one or more prototypes, and evaluate each prototype’s significance for addressing a concrete educational need.

(Penuel, Roschelle & Shechtman, 2007:53)

Furthermore, Penuel, Roschelle and Shechtman (2007) stress that this process it not about making teachers fully fledged designers that will create the solution but they can help construct the key challenges and the educational needs that the designer must address. The teacher can also provide a platform for testing the innovation in their classrooms and give input on improvements.
In this study both the expert and additional participants representing users were involved. The expert was used to capture perspectives regarding the requirements, goals, usability and feasibility of the prototype. In Table 1 an overview of the main co-design sessions that were conducted together with the expert is shown. A group of social workers were involved as users in order to collect impressions regarding the behavior of the avatar and how they experienced the training sessions in regard to technical and content aspects during the development of the first prototype.

During the development of the second prototype apart from the expert a group of employees at the department of Media Technology participated in evaluating the different interface approaches.

<table>
<thead>
<tr>
<th>Date</th>
<th>Phase</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 Aug</td>
<td>Project requirements</td>
<td>Overall goals and needs of the project</td>
</tr>
<tr>
<td>15 Aug</td>
<td>Prototype 1 requirements</td>
<td>Sketching and iterating design alternatives</td>
</tr>
<tr>
<td>21 Aug</td>
<td>Prototype 1 requirements</td>
<td>Sketching and iterating design alternatives</td>
</tr>
<tr>
<td>22 Aug</td>
<td>Video production</td>
<td>Video content planning</td>
</tr>
<tr>
<td>25 Aug</td>
<td>Video production</td>
<td>Filming video content</td>
</tr>
<tr>
<td>10 Sep</td>
<td>Video production</td>
<td>Editing and validating video sequences</td>
</tr>
<tr>
<td>7 Oct</td>
<td>Building Prototype 1</td>
<td>Discussing methods for child interviews</td>
</tr>
<tr>
<td>11 Nov</td>
<td>Evaluating Prototype 1</td>
<td>Discussing outcomes of user trials</td>
</tr>
<tr>
<td>2 Dec</td>
<td>Building Prototype 2</td>
<td>Goals and requirements for Prototype 2</td>
</tr>
<tr>
<td>15 Dec</td>
<td>Evaluating Prototype 2</td>
<td>Evaluation of interface design</td>
</tr>
<tr>
<td>18 Dec</td>
<td>Evaluating Prototype 2</td>
<td>Evaluation of interface design</td>
</tr>
</tbody>
</table>

Table 1: Participatory co-design sessions with the expert

3.2 Literature review

The following search keywords and strings were used in the literature review.

Avatar, avatar-mediated, communication, control, co-presence, emotion, facial expression, interaction interface, non-verbal communication, presence, realism, realistic, social presence,

3.3 Establishing requirements

3.3.1 Stakeholders

Interactor (expert) - The interactor is considered being an expert on child interviewing. This person has the theoretical and practical knowledge required in order to be able to conduct training sessions with students from the perspective of the child. A key person who is used as a basis for defining the requirements for the role of the interactor in the design process is a person for The Institute of Police Education in Växjö, Sweden. He is at the time of writing a teacher in this field and has years of experience in different teaching methods for putting theory of child interview training into practice.

Interviewer (student) - The interviewer is the person who is having a conversation with the avatar. This role could be anyone who in their profession has daily contact with children. For instance social workers, police personnel and teachers.

Organization - The organizational goal is that the personnel should be able to increase their competence in child interview training in an efficient manner regarding time and
resources. An employee should be able to gain valuable practical experience while practicing with an avatar instead of practicing on children in a real life situation.

3.3.2 Initial requirements
First initial goals and requirements were established by:
- Involving the expert in the design process through interviews and active participation in co-designing the system (see appendix A for interview).
- Researching related matters connected to avatar-mediated communication
- Defining the goals of the different direct users and stakeholders of the system (interactor, participant, organization)
- The goals and requirements for prototype 1 and 2 are explained in chapter 4.1 and 4.3 respectively.

3.4 Evaluation of Prototype 1
3.4.1 Training sessions
The first functional prototype of the system was tested individually with 15 participants over a period of four weeks. These participants were in their profession active social workers. They were all female and between 28 and 52 years old with an average being 38 years of age.
Each participant interacted with the avatar during two different occasions, resulting in 30 sessions in total. During the sessions the avatar was being controlled by the expert. All sessions were recorded with audio and video for educational purposes within the course but also for future evaluative purposes.

3.4.2 User questionnaire
After each session the participants filled in a questionnaire answering questions related to their experience during the session (see appendix B and C). The questionnaire collected both qualitative and quantitative data regarding the users’ interaction with the avatar. A five point Likert scale and free text answers were used. The questions in the questionnaire given during the evaluation of the prototype were adapted from Biocca et al. (2003) and used for different ways of measuring the perceived presence of having a conversation with the avatar.

The statements that follow were given to the participants to answer.

I consider that:
- I felt aware of another part being present in the conversation
- I perceived it as the other part was reacting to what I said
- I perceived the other part in the conversation as being a real person
- I perceived the other part in the conversation as being a real child
- seeing the video sequences of the child helped me feeling immersed in the training scenario
- hearing the voice of the child helped me feeling immersed in the training scenario
- seeing the facial expressions of the child helped me in understanding how the child was communicating
• seeing the body language of the child helped me in understanding how the child was communicating
• to practice interview technique in this way was useful for me overall

The following questions were also given.

• Disregarding the subject and details of the conversation. How well do you regard that the technology enables realistic conversations with a virtual child?
• What worked well and what didn’t work well? What do you think could have been done better?
• According to you, was there anything specific that made you regard the training situation as being unrealistic or not believable?

3.4.3 Post session interview with the expert
After conducting interview sessions using the avatar during 30 interview sessions the expert was interviewed. This interview served as a follow-up conversation with the purpose of examining if there had been any changes to the initial requirements, and to collect the impressions of the prototype and trials in general. See appendix A for a transcription of the interview.

The following topics were covered.

• Goals
• How this type of application can support the user in training
• System functionality
• Goals of the organization
• Which behavior should be possible to express with the avatar
• Evaluation of prototype functionality
• Impressions from interacting with students through the avatar
• General impressions from the trials

3.5 Evaluating Prototype 2
Prototype 2 was developed from the refined requirements that were part of the outcomes of the evaluation of Prototype 1. This interactive interface prototype had no back-end functionality as the main focus was on evaluating different interface alternatives in order to have an effective interaction from the perspective of the interactor. The prototypes were not evaluated with the social workers that participated in previously described trials but rather with the expert and six other people working at the Media Technology department at Linnaeus University. The reason for not involving the social workers were mainly due to them not being in the target group for acting as interactors. Furthermore, the main purpose of evaluating Prototype 2 was to validate the interface approach with additional users and not only the expert.
4 Prototype design process

The prototype design process is described in two phases.

- *Prototype 1* - the development and evaluation of a functional prototype built with existing tools and applications. This prototype was used to enable the expert to conduct avatar-mediated training sessions with students.

- *Prototype 2* - the development and evaluation of an interface and system design description based on the findings of evaluating prototype 1.

4.1 Developing Prototype 1

4.1.1 Concept description

The interactor has access to an interface which enables him or her to control an avatar. The interface presents a selection of different behavioral states. These states are based on pre-recorded video sequences of real children. Using these states the interactor can act out different behaviors while having a spoken conversation in real time together with the child interviewer in training. The voice of the interactor is changed by giving it a higher pitch as to make it more similar to the voice characteristics of a child. The child interviewer is presented with a video stream that is controlled by the interactor, and the interactor sees the interviewer through the web camera feed. See figure 2 for the overview of this concept.

![Figure 2: Concept overview](image)

4.1.2 Video sequences

Two children were individually recorded on video. They were instructed in how to act in order to capture different behavioral states of the avatar. These states were consisting of different moods: happy, sad and neutral. Different behaviors were also recorded: talking, not talking, walking in and out of the picture, sitting down and standing up, demonstrating getting a hit in the face and on the leg. These sequences were edited into individual sequences in order for the interactor to be able to control the avatar behavior in real time.

4.1.3 Identified general goals of the system

Following the interviews with the expert a set of general goals of the system were identified. These were:

- The possibility to conduct structured child interview training without involving real children.

- A realistic experience that resembles talking to a real child in order for the expert to give feedback on how students behave in a child interview situation.
• Possibility to control an avatar that through its behavior and appearance is perceived to provide the same amount of presence as a real child over video chat is considered to be the ultimate design goal.

• An interface through which the expert can control the avatar behavior while having a simultaneous conversation without feeling encumbered.

### 4.1.4 Identified prototype technical requirements

Similarly to what is referred to regarding the general goals of the system, the initial interviews with the expert allowed us to devise the following set of technical requirements:

• *Controlling the avatar* - In order to control the behavior of the avatar the user needs some way of controlling the pre-recorded video sequences. In other words, a way of acting as though the avatar is an extension of one’s self. This is done through expression of basic behavior and emotions.

• *Video chat* - The training scenario needs some way of transmitting the appearance of the avatar. The scenario also needs to be possible to conduct on distance. Therefore, video chat functionalities are needed.

• *Sound processing* - The sound needs to be processed to some degree. It is assumed that using one’s regular adult voice will break the illusion of talking to a child.

• *Recording* - Being able to record the sessions is a requirement since post-session reflection is a critical component of the educational activity.

### 4.1.5 Prototype presentation

The prototype was developed based on the previously mentioned requirements and was built using existing tools and applications (see figure 3 for an overview).

The expert controls the video stream by selecting video sequences in Wirecast (see figure 4). This video signal is transmitted to Skype as a virtual web camera feed (see figure 5 for the perspective of the expert).

In order to process the speech the audio signal from the microphone input goes through a pitch filter in Garageband and is transmitted into Skype as a virtual microphone input using the application Soundflower.

In order to document the tests for further analyzing and evaluation Quicktime is used for making screen and audio recordings of the sessions.

![Diagram of prototype setup](image)

Figure 3: Overview of the prototype
4.2 Evaluating Prototype 1

Over a period of four weeks the prototype was evaluated with active social workers studying a course on child interview training. Within the context of the course the purpose of the activity was to provide the opportunity for the students to have individual training sessions together with the expert. In these training sessions the students were given the opportunity to turn their theoretical knowledge into practice by having a conversation with the avatar.

4.2.1 Trial 1: Description

The first round of sessions consisted of 15 users each having a one-on-one interview conversation with the avatar. The interview scenario revolved around a fictional 12-year old girl who has problems in school and at home. It was the task of the interviewer to ask questions in a correct manner according to guidelines in child interview techniques.

Each session was approximately between 30 to 40 minutes in duration. The sessions were conducted on distance and facilitated by the expert. All sessions were screen recorded for the purpose of later analysis for this study as well as for training purposes within the educational course.
4.2.2 Trial 1: User survey results
After completing their interview the students were asked to fill in a questionnaire regarding their experience of the conversation with the avatar. See figure 6 for an overview of the survey results and appendix B for the complete free text answers that were given.

Some key points can be summarized into the following points:

- **Technological experience** - Approximately half of the participants had no previous experience with video chat applications such as Skype. One person reported not being comfortable being on camera. There were no indications to the notion that the system itself was hard to use from the interviewer’s side, meaning either using Skype or the process of talking to the avatar.

- **Presence** - 14 users agreed or strongly agreed that they perceived the avatar as a real child.

- **Presence** - 11 users agreed with the claim that the facial expressions of the avatar helped them feeling immersed in the training scenario and 3 were neutral to this statement.

- **Presence** - 12 users agreed that seeing the body language helped feeling immersed in the training scenario and only one person reported that body language of the avatar not being important at all.

- **Presence** - Video and sound - All users agreed regarding the claim that seeing the video of the avatar as well as hearing the voice had positive effects regarding immersion in the training scenario.

- **Overall feasibility of the approach** - When asked about the general usefulness of the approach that they had experienced during the session 11 users strongly agreed and 3 users agreed that it is beneficial for practicing child interview training. One person reported being neutral to this statement.

- **The potential of the technology** - The users were asked to disregard the details of conversation and rate how well the technology itself enables realistic conversations with a virtual child. On a scale of 1-5 where 5 meaning that it fully enables, 14 users rated it at 4 and 1 user rated it 5.
Following quotes are selected from the survey. Overall the participants were positive to the approach. The situation was perceived as rather realistic.

"I think that it worked well overall. You get a more realistic conversation when you see a child and hear a voice.” (User 2, 2014)

"It felt like a real conversation with a real child, this helped with feeling immersed in the situation.” (User 5, 2014)

"The conversation felt realistic.” (User 15, 2014)

Most of the negative comments that surfaced were mostly due to technical aspects.

"The image was interrupted at several occasions and it even happened that the voice was "choppy" or disappeared. I was uncertain if the child sat still and was silent or if it was technical problems.” (User 10, 2014)

"Having technical problems initially made me lose my focus” (User 15, 2014)

"I think it worked well. Even though it was hard to hear. There was also a certain delay in the mouth movements and in the speech that I heard. But when I disregarded of this it felt like a real conversation with a child.” (User 1, 2014)

4.2.3 Trial 1: Identified problems from the interviewer's perspective

- Delay in the conversation - Five of the 15 users reported to have experienced a delay in the conversation. The cause for this issue is not entirely established. The issue could be of technical character although the delay one generally experiences over Skype video chat should be negligible. A hypothesis is that it
could be related to the difficulty to pick up subtle details in non-verbal communication (see chapter 2.4.1) when speaking to the avatar. Predicting when the avatar is going to speak could be problematic since the turn-taking is not as natural as in normal face-to-face conversations. When analyzing the recordings this reported delay is not apparent and the fact that the expert expressed that he did not experience this delay further supports this hypothesis.

- **Audibility** - Five users reported that at times it was difficult to hear what was said. Again, it could be a technical issue related to the volume and the users not knowing how to increase it. Another possibility could be that using the pitch effect resulted in a less clear and audible voice.

- **Voice quality** - It was reported to be experienced as strange in some cases. Worth noting is that it was reported that over time the participants got used to hearing the voice.

- **Technical interruptions** - Three users reported that other technical-related problems such as interrupted calls had disturbed them in terms of immersion in the training session.

### 4.2.4 Trial 2: Description

The second trial was completed using the same participants and in the same manner. The scenario revolved around a ten year old boy with concentration difficulties common among young people having a ADHD diagnosis. The interviewer had to gain an overview about the boy’s situation at home and in school. At times during the interview when the boy was distracted he would stand up and walk away from his chair and look at other objects in the room. In order to gain useful information the interviewer had to communicate with the boy in an appropriate manner. See figure 7 for a screenshot of the interface with the sequences of the boy avatar.

![Figure 7: Interface of Wirecast with avatar sequences](image-url)
4.2.5 Trial 2: User survey results

In the second round 15 people participated (although one person did not answer the survey). There were no radical differences in the results when compared to the first round (see figure 9 for the results of trial 2). As the situation is very similar, albeit with a different avatar, these results were to be expected.

<table>
<thead>
<tr>
<th>Survey Question</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I felt aware of another part being present in the conversation</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I perceived it as the other part was reacting to what I said</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>I perceived the other part in the conversation as being a real person</td>
<td>4</td>
<td>9</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I perceived the other part in the conversation as being a real child</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seeing the video sequences of the child helped me feeling immersed in the training scenario</td>
<td>12</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hearing the voice of the child helped me feeling immersed in the training scenario</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seeing the facial expressions of the child helped me in understanding how the child was communicating</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Seeing the body language of the child helped me in understanding how the child was communicating</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>To practice interview technique in this way was useful for me overall</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 9: Trial 2 survey results
Following quotes are selected from the survey.

Some users reported having a more positive impression after this round.

“The child felt more real this time compared to the first. Maybe because I did not think of the third part at all.” (User 3, 2014)

“The voice is still a bit strange but I did not think about it in this conversation.” (User 3, 2014)

“It felt more natural than the first time. The technology worked better in this conversation and I was less nervous.” (User 10, 2014)

"Once again I think this is a very good way of training the technique of conversing with children" (User 12, 2014)

One user still had problems with the concept of having a conversation using video.

“Sometimes I have a hard time focusing and I'm disturbed by my own picture and "all the technology", but to a lesser degree than comparing to the first conversation, so it is probably a thing of habit.” (User 6, 2014)

Two users had similar feedback regarding the behavior of the child in the regard that it did not act as they expected.

"The child was very "easy talked" for being a child with ADHD and asperger, which was beneficial for me today, but it does not feel so realistic. I had expected much worse!” (User 14, 2014)

“It felt realistic. Even though the child maybe was a bit too "easy talked". But that is probably good since you get the possibility to practice the whole discussion and not get stuck. It is as realistic as possible.” (User 4, 2014)

“It is hard to interpret facial expressions” (User 11, 2014)

4.2.6 Trial 2: Identified problems from the interviewer's perspective

- **Facial expressions** - In one case it is reported that the facial expressions are hard to interpret. In the data (see figure 9) there were also some signs to that the users were more ambivalent to the facial expressions compared to other aspects.

- **Avatar behavior** - The child is in multiple cases reported as being too cooperative for having an ADHD diagnosis.

- **Technical aspects** - The black background is reported to cause some friction in the regard that the child is talking about the environment even though the interviewers are not able to see it.

4.2.7 Trial 1 and 2: Identified problems from the expert’s perspective

- **Expressing behavior and mood** - The expert is feeling restricted in terms of what is possible to express. Having emotions restricted to happiness, sadness and neutral is reported to be too crude of an approach. Being able to express emotion and behavior in finer detail is requested by the expert.

- **Choosing video sequences** - In the initial prototype the interactor manually chooses the video file that is showing the sequence of the emotional state. The
expert could choose the order of the sequences as seen in the interface (see figure 7), albeit in a quite arbitrary representation. Even though the sequences that are shown are previews in the form of thumbnails it might not be clear enough to identify the correct sequence, resulting in the interactor being required to read a quite small text of the sequence name.

- **High workload** - While the system is reported to being somewhat manageable to use there is a lot that is required by the interactor in order to control the avatar while having a conversation and conducting a training session. Monitoring the avatar and the interviewer, choosing suitable video sequences, following and adapting to flowchart-based script (outside of the application) all contributes to the total workload that has to be sustained. In the current state the expert also needs to manually take care of screen-recording the session.

- **Repetitive video material** - It is described by the expert that there is too small variation in the different clips. The visualization of the different states are always the same as the video sequences are being repeated.

### 4.2.8 Findings from the prototype evaluations

The study showed a rather positive result in terms of feasibility of the approach. Overall the participants reacted positively to the activity and they reported to a large degree that it felt like talking to a real child. Some users stated that the activity in the second trial felt more natural compared to the first time they interacted with the avatar. The expert who also has experience from previous training sessions without the use of avatar technology reported that he experienced that there is a noticeable higher degree of realism in this type of approach compared to traditional video conferencing solutions without the use of avatars. Furthermore, he reported that the prototype provides an easy way of at least achieving a situation that is close to realistic. As an instructor he experienced that the participants are present in the situation meaning that they were engaged and were listening to the child. He thinks that it is possible to see differences in people’s behavior and how they are handling the interview. The participants have reported that they are reflecting on different aspects they are doing wrong in the interview. Thus, it could be claimed that the prototype provides the possibility for self reflection within the area of child interview training.

From the usability perspective of the expert there is a need for making the interface more effective in terms of control. Sometimes the wrong signals are being communicated by mistake. While the prototype provides a somewhat crude control of different emotions and behaviors the results show that there is potential in this approach.

From the usability perspective of the interviewer the system does not require more interaction than having a regular Skype video call. This design goal should be kept in order to not introduce any more complexity.

### 4.3 Developing Prototype 2

New requirements were identified based on the findings from the evaluation of the first prototype.

#### 4.3.1 Avatar video sequence requirements

The expert expressed the need for making it possible to express a wider range of moods and behavior.
• A new set of behaviors based on the six basic emotions which can be identified through facial expressions according to Ekman (1992a).

• More variation in the behavior. Having a certain avatar state being represented by a single sequence made it easily identifiable as a sequence that is being repeated. A certain state could be recorded and saved into several versions in order to not look exactly the same.

4.3.2 Interface requirements

• Controlling the attentiveness of the avatar based on previous methods of avatar behavior states proposed by Nagendran et al. (2013) (further described in chapter 2.5.2).

• Controlling six different emotions, attentiveness and additional behaviors in a structured manner.

• Being able to control the behavior states in a high level manner. Effectively using a much larger library of video sequences while requiring only the bare minimum of interaction from the user.

• A certain level of automation in order to eliminate unnecessary interactions.

• The interface should be developed by applying the design principles described in chapter 2.7.1.

4.3.3 Extending emotions and behaviors

The emotions possible to express are extended from merely covering happiness and sadness to the six basic emotions universally understood from observing facial expression (Ekman, 1992a). These states: happiness, surprise, fear, sadness, anger and disgust, are to be recorded with a child acting out the behavior to a varying level of emotion ranging from (in the case of happiness): slightly happy to happy to very happy. This procedure is to be repeated for every mood state.

In addition a neutral state is required just as included in Prototype 1. The neutral state could be used when the interactor does not want to communicate any specific mood. Building on the approach of key-poses presented by Nagendran et al. (2003) these sequences could also act as transitional states when going between the different emotions.

In conversations with the expert an idea about controlling the avatar’s attentiveness surfaced: enabling the avatar to go from not being focused or present in the conversation to being fully focused and engaged. Similarly this avatar behavior was implemented by Nagendran et al. (2003) within a system for teacher training purposes. The behavior of attentiveness can be communicated by having the video recorded child adapt different facial expressions and postures. Attentiveness can also be communicated with eye-gaze by having the child look away from the camera or straight into it. Looking into the camera could have the effect of the interviewer experiencing as if eye contact is made with the avatar. Mutual eye-gaze is found to be beneficial for perceived co-presence (Blascovich et al., 2002). Even though the eye-gaze would only be experienced by the participant this is not regarded as an issue since it is the interviewer’s perceived presence of the avatar that is prioritized.

In practice having three different levels for each of the six emotions and the attentiveness variable in addition will result in four sequences for the non-talking
behavior and talking behavior respectively for a total of eight sequences per emotional state. If the aspect of randomness and variation are to be introduced in order not to repeat the same sequences again this number is multiplied in addition. Assuming that at least three variations of each sequence are needed, this would result in having a library of 24 sequences per emotion resulting in 456 video sequences in total. This larger selection of video sequences presents new challenges in terms of interface design. It would no longer be feasible for the interactor to manually select individual sequences.

4.3.4 Control of behavior
As described in chapter 2.5.1 automation of non-verbal behavior is a worthy goal. However if A.I. is not to be used for controlling the non-verbal behavior an approach that requires user interaction is needed. A goal could be to keep the interaction to a minimum by not overloading the user with choices, making it only possible to select high level behavior and delegating much of the control to the system by having the system automatically selecting different video sequences.

The starting point for mapping the emotions to the avatar is founded on research on human emotion by Ekman (1992a, 1992b). These emotions are universally understood from observing one’s facial features. The goal is to be able to control the six different emotions and the level of attentiveness. When dividing each emotion into three different levels of intensity (e.g. slightly happy, happy and very happy) and including the neutral state the user is left with 19 possible choices in terms of controlling the emotion. For the sake of simplicity of interaction single basic emotions are to be used. In reality emotions can consist of combinations of multiple emotions (see chapter 2.4.2) but this is not taken into account at this state of the prototype.

Some automation should be done by the system. It is argued when the interactor goes from one emotion to another the system should automatically transition accordingly by blending the sequences in an appropriate manner in order not to switch between two states in an instant, which would be likely to be perceived as unrealistic.

4.3.5 Interface approaches
In order to evaluate how to control the avatar behavior two different approaches were developed. This development was the result of an iterative process of designing different interface alternatives and can be seen in appendix D. One approach is a radial interface inspired by Plutchik’s (1991) wheel of emotions (described in chapter 2.7.2) but simplified to include Ekman’s (1992a) six emotions and a neutral state. In figure 10 the radial interface is demonstrating a situation where an avatar-state of “very happy” is chosen and active. Another approach that was evaluated was to control the same emotions but with the use of vertical sliders (see figure 11). This approach was believed to be regarded among the users as being a more established way of interacting.
4.3.6 Prototype presentation
An interactive web-based prototype was developed which enabled a way of evaluating which interface interaction approach was more preferred among users. The prototype does not include any connection to video sequences or back-end programming. The prototype can be interacted with using the mouse as well as keyboard. Although the keyboard is only used to indicate whether the avatar should be speaking or not using a push-to-talk approach. Two versions were developed, one using a radial interface (see figure 12) and one using vertical sliders (see figure 13).

4.4 Evaluating the first iteration of Prototype 2
The two approaches were tested with the expert and six additional users that had no prior experience of using the prototype or any connection to child interview training.

4.4.1 User test description
The users were introduced to the project and were shown the previous prototype as well as recordings from the trials with the social workers. The participants in the test were tasked with evaluating which interface they would prefer using if they were to control the avatar. The interface was also evaluated on an iPad even though the main focus was to develop for a scenario where a laptop is used. The reason for this was to find if there were any indications to users being in favor of a touch interface, which could be a possible focus for future work.

4.4.2 User test results
Five out of seven (including the expert) were in favor of prototype interface 1 that used the radial representation (see figure 12). Furthermore, when given the choice between
laptop and iPad five out of seven chose the iPad in favor of interacting with mouse and keyboard. Interestingly the expert chose the radial interface on the laptop as his first choice. The motivation behind this was that he preferred pressing a physical key down in order to activate the talking state.

Some users reacted to that the clickable target area of the radial interface is quite small, at least in the case of the first emotion level (e.g. slightly happy). It is likely that this low-intensity state would be used more than the extreme emotions (e.g. very happy) therefore, it could be argued that this area should be made slightly bigger.

The question of having to alternate focus between the different areas of the prototype was raised by the expert and other participants. Being able to control the emotions while still having a good understanding of which expression is being transmitted is of most importance according to the expert.

The slider is a widely adopted user-interface element. While the radial approach itself might not have the same high level of affordance as sliders it was reported in several cases as being easier to understand. The sliders were in several cases reported as being more cognitively demanding.

The use of a slider for attentiveness was reported as being easy to understand, but it is uncertain how necessary it is since it only has four different levels. It could possibly be replaced by buttons for a more effortless interaction.

4.5 Developing the second iteration of Prototype 2

From the findings of the previous iteration of developing and evaluating the user interface new and additional requirements were identified.

The interface should:

- provide a better awareness of the state of the avatar by giving it more screen estate.
- incorporate buttons instead of a slider for attentiveness but while still keeping it clear that it is a continuum.
- include slightly modified clickable areas according to how much they are used, effectively increasing the area for the first two levels of emotion.
4.5.1 Prototype presentation
The requirements described in chapter 4.5 are implemented and presented in figure 14.

Figure 14: Second iteration of the prototype interface

4.6 Evaluating the second iteration of Prototype 2
This version of the prototype was reported by the expert to be an improvement from the previous. The one suggestion for improvement given is that the buttons for attentiveness could be more distinct and grouped together, giving a more uniform appearance with the radial interface.

4.7 Technical implementation
A possible way to implement the technology would be to use WebRTC (W3C, 2013). WebRTC enables real-time communication between different browsers and platforms without the need for plugins. This technology is for instance used by Google for enabling in-browser video conferencing in the Google+ social networking service. Being able to use the system without installing special software other than a WebRTC compatible browser is seen as being beneficial.
5 Results

5.1 Prototype 1
What was found from the first series of evaluations of the working prototype was that the training scenario itself was overall perceived by the participants as a useful tool for practicing child interview techniques. The results from the evaluation of the initial prototype also served as a starting point for developing the custom interface for the expert and a description of the underlying system design and avatar behavior.

5.2 Prototype 2

5.2.1 Avatar behavior
The avatar behavior is heavily based on the different facial features which are universally understood. Each of the six different emotions can be controlled in three different degrees. In addition it is possible to control for the attentiveness of the avatar. These emotions and behaviors are to be made up from a library of 456 video sequences.

5.2.2 User interface
It is possible to interact in a manner that is reported by the expert as not being more complicated than using the initial prototype which included only a handful of behaviors. From an interaction perspective this indicates that it is a feasible approach to make. See figure 15 and 16 for a final version of the user interface.

![Figure 15: Final prototype interface: interactor’s perspective](image)

The interface for the participant shows the avatar as a major part in addition to a preview of their web camera stream that is transmitted to the expert.
5.3 Research questions

5.3.1 Research question 1

The first research question: “Which interaction aspects can contribute towards a low effort for the interactor in a video-based avatar-mediated system for child interview training?” was formulated in order to explore how the interactor could be able to sustain a conversation while controlling the avatar behavior. The question is answered by covering the following aspects that emerged from the results of all the empirical work done.

- **Minimal interaction effort** - Maintaining a high quality of expression while not requiring more interaction than what is absolutely necessary from the user.

- **Minimal movement** - The user should also be able to control the emotions without being required to move around in the interface too much. One hand can be held on the push-to-talk button and one hand can be used for the mouse which interacts in the right side of the interface. Using buttons instead of sliders makes the interaction faster.

- **Effective interaction using interface elements** - The interface approach using direct interaction by clicking a button on the screen is perceived as more effective compared to the using of sliders.

- **Automation** - By delegating parts of the avatar control to the system less micro-management is required from the interactor. The interactor only needs to select a high level behavior and the system selects the correct video sequences accordingly. Furthermore, transitions and randomization is offloaded to the system. Ideally, these aspects would also contribute to a more realistic experience for the interviewer as also indicated by previous research on avatar behavior automation (see chapter 2.5.1).
• **Representation of emotions** - using a graphical representation (for instance a circular radial interface) that easily can be interpreted contributes to a lower effort in controlling emotions of avatars.

• **Monitoring of avatar appearance** - getting instant feedback for which behavior that is transmitted is important in order for the interactor to act accordingly.

### 5.3.2 Research question 2

The second research question: “**Which aspects of the avatar behavior can lead to increased experience of co-presence for the participant?**” was formulated in order to find an answer to which ways the avatar behavior can contribute to the perceived realism of the interaction with the avatar.

• **Audio and video** - The combination of audio and video can lead to a higher sense of co-presence.

• **Facial expression and body language** - There are indications that facial expression and body language is beneficial for the participant to feel immersed in the conversation, although a few people were neutral to this statement. It is very likely that the use of such few sequences were not contributing as much to presence as one would hope. The expert also expressed his view that he felt restricted in terms of emotions that he could express and at times the emotion did not match up to what was said in the conversation. Having a larger selection of possible expressions is likely to contribute more to the perceived co-presence.

• **Technical stability** - While not being in immediate relation to avatar behavior, the conversation needs to take place on a stable platform without technical issues that interfere with the experience. Whenever technical interruptions occurred during the trials users reported losing focus on the conversation and as a result feeling less immersed.
6 Discussion

6.1 Reflecting on previous research
The use of video-based avatars is not as common as using animated 2D or 3D characters as avatars. One could argue that it is easier to achieve a realistic avatar appearance by filming real people compared to modeling 2D or 3D avatars which have the same level of realistic appearance. An apparent downside is that video content is more complicated to adjust post recording and is naturally less dynamic. Aspects of the appearance such as mouth movements and transitions between different states cannot be achieved as easily. However, the findings of how to control the avatar could as well be applied within approaches of systems using animated computer graphics.

The findings support the concept of automation and delegation of avatar behavior to the system as proposed by Vilhjálmsson (2003). This concept appears to be a key feature for sustaining a conversation while simultaneously controlling the behavior of an avatar. Ideally one would want it to be possible to express the slightest detail in avatar behavior and it could be argued that it is an oversimplification to isolate emotions to merely six since the research says that emotions can be made up out of multiple ones (Ekman, 1992b). However, just as suggested by Blumberg and Galyean (1995) and Vilhjálmsson (2003) one needs to find a middle ground between required interaction and what is possible to express. As also suggested in this thesis, a minimal interaction effort is naturally a worthy goal.

6.2 Interpreting the data collected
The users participating in the initial trials were overall very positive to the approach. However, they had not used avatar-based applications before and approximately half of the participants even had no previous experience of using video chat. It could be the case that the participants were overly positive just for the sake of getting the opportunity to practice child interview training over the internet, while a person with more technical experience would have been more critical to how the avatar was appearing.

Some results from the first trial were not as positive but it is also likely to have been misinterpreted. For instance when asked if they were aware of the presence of another part in the conversation a number of participants indicated to be between neutral or strongly disagreeing with this statement. It is likely that some of these participants did not understand the question correctly, as they in fact had a spoken conversation with the avatar, and arguably this would be an apparent sign of another part present in the conversation.

6.3 Possible limitations to the study
Something that could be argued as being one of the weaker parts of the study is that throughout the design process it was heavily relied on one expert as the single authority on child interview training. Arguably involving more people in the role as experts would have been preferred. The result is a design which to a large degree is the product of investigating one person's needs and goals with the system. Although it is likely that this design also could be useful for others. The design was also evaluated with other users, albeit no other experts on child interview training. However, it is also likely that there could have been a slightly different design end-result if more child interview experts had participated in the co-design process.
6.4 Future work

6.4.1 Behavior and emotions
Introducing a new set of behaviors and emotions present challenges in terms of working with an actor that has to perform the different behaviors in a correct manner. It puts responsibility on the actor and it would be necessary to evaluate if the recorded sequences are interpreted as intended. This could be done using widely adopted methods for measuring how people react to facial expressions (Ekman, 1992a).

6.4.2 Evaluating competence outcomes
It could be explored whether participants using the system really improve their actual competence in child interviewing.

6.4.3 Further evaluations
In the future the system could be evaluated with a larger group of additional participants in order to explore the experiences among users from different demographics and professions.

6.4.4 Sound processing
In its current state the prototype uses a typical pitch effect, which, while acceptable in execution, could be improved upon. Further research efforts could be put towards exploring how to process the voice in order to achieve a more childlike characteristic. Since voices are highly unique and can vary in pitch and characteristic it could also be explored how the pitch effect can be adapted to suit different voice types.

6.4.5 Hardware controllers
It could be explored how certain hardware controllers (preferably commonly available and affordable) could be used to control the avatar. Gamepads could for instance possibly be useful for controlling the avatar behavior without having the interactor take the focus off the avatar and the participant. In addition using such hardware could enable motion control and haptic feedback, which, when used in an intuitive way, possibly could be beneficial for achieving low-effort interaction.

It could as well be explored how a tablet could be used together with a computer in a second screen approach, effectively having the control interface on the tablet while seeing the participant and avatar on the larger screen.
7 Conclusion

The purpose of the thesis was to explore how a system using avatar-mediation can be designed to enable a realistic educational scenario for practicing child interview training. While still at the early stages of development, the prototype developed and used in the study has shown that it is possible to achieve a solution that is perceived as useful and valuable in terms of the possibilities that it enables for training interviewing techniques. Even with just a handful of video sequences which illustrate different avatar behaviors it is possible to a large degree to reach the experience of talking to a child. The avatar-mediated conversation is reported by the participants as overall being a very feasible approach for training. Furthermore, the expert is of the opinion that it is a big leap forward in terms of usefulness compared to earlier trials using video chat without avatar technology.

The study also shows that with the proposed design it is possible to substantially extend the video material (approximately multiplied 75 times) with negligible change in complexity in terms of interaction needed from the interactor.

Experiencing the appearance of the avatar with audio and sound simultaneously is most likely to contribute to the overall perceived presence. In itself the audio and video components are not perfect, as the video is not a fully correct representation and achieving a fully realistic pitching of the voice was out of the scope of the research. However, when combining the two modalities it indicates that the combination of audio and video results in a rather realistic experience. An advantage over some avatar-based applications is the use of real time voice communication rather than using pre-recorded or computer generated voices. Having a spoken conversation in real time with another person, albeit with a pitching effect applied, is likely to be beneficial when realism is to be considered, as compared to hearing pre-recorded sequences or reading text chat generated by an artificial intelligence.

It is shown in this thesis that using avatar-mediation in child interview training has an apparent benefit in terms of providing a training scenario that is perceived as being realistic. The case presented is for a very specific purpose but the design could be applied tentatively within educational contexts where human communication is central.

The system that has been described builds on the notion of augmenting human abilities rather than using technology as a replacement for human interaction. The abilities that are gained when mediating one’s self through an avatar enables new ways of interacting and working which was impossible not long ago. The use of avatar-mediation can hopefully be beneficial to society in terms of a contribution to a more effective educational training of child interviewing, resulting in more widespread knowledge and preparedness for talking to children about sensitive subjects in real life.
References


Appendix
Appendix A Interview with Pär Stihl

Vad skulle du säga är det huvudsakliga målet med verktyget och hur det ska stödja dig?

Ja precis, det är ju liksom att skapa en flexibel träningsplattform för alla grupper som har barnsamtal. Så att de kan genomföras på ett så bra sätt som möjligt, och att verktyget ska skapa den övningsmöjligheten för att kunna genomföra strukturerad träning utan att barn utsätts för det här övningsmomentet.

Om du ser från studentens perspektiv. Hur tycker du att verktyget ska kunna stödja studenten?

Det ska stödja så till vida liksom att det ska vara lättillgängligt och att de ska få en så realistisk upplevelse som möjligt så att de faktiskt upplever som att de sitter och pratar med ett riktigt barn och därigenom få möjligheter att få väldigt handfast och exakt feedback på det så de faktiskt säger i samtalet. Inte vad de skulle vilja säga utan det de säger när de sitter och pratar. det blir den stora skillnaden när man faktiskt ser sig själv och hör sig själv säga, och hur agerar jag då. Det blir den stora skillnaden och att det blir inspelat på ett enkelt sätt.

Så det är viktigt att det blir just inspelat så att de kan ha den här reflektionen?

Det är mycket centrat att det blir inspelat så att du både ser dig själv och hör dig själv när du kan reflektera tillbaks på vad du faktiskt har sagt och den situationen då är realistisk när du faktiskt upplever samtalet och kunna reflektera i efterhand då.

Om man ser till verksamhetens mål? nu är det ju lite olika målgrupp som jag har förstått. Vad skulle du säga är verksamhetens mål med att ha tillgång till ett sånt här verktyg. Vad ska det leda till?

Ja alltså, oavsett verksamhet så är det av stort värde att samtal med barn genomförs på ett korrekt sätt så att barnets rättigheter tillgodoses och inte skadas av en bristfällig intervjuare. oavsett om det är en skola där man uppfattar missförhållanden, socialförvaltningens utredning eller en brottsutredande polis. att man då kan göra att man inte skadar barnets berättelse genom att bemöta barn på felaktigt sätt, ställa ledande frågor, i övrigt hantera liksom på ett ofördelaktigt sätt sett ur barnets perspektiv.

Om jag säger att ett best-case scenario skulle vara att efterlikna ett skypesamtal till exempel med ett riktigt barn, hänger du med om det? skulle det ge något ännu mer?

Om vi lyckas med att vi får barnet att uppleva, då är det min uppfattning absolut att vi har en träningssituation som hänger mycket hög kvalitet för barnsamtalsträning.

Vad skulle du vilja uttrycka genom avataren om du fick välja helt fritt.

Om vi går vidare och tänker på att du har gjort mellan 15 och 20 sessioner nu. Vad skulle du säga är som funkar bra med systemet?

Det som funkar bra är att man faktiskt kan skapa den här situationen och att man kan få sätta upp de här förutsättningarna och som att du faktiskt med väldigt enkla sätt kan få en väldigt verklighetsnära en situation som de uttrycker att det faktiskt känns väldigt, man upplever det som instruktör att de är närvarande och de engagerar sig i samtalet och lyssnar på barnet och man ser skillnader i olika personers beteende hur de hanterar de samtalet och det blir väldigt tydligt när man ska kan följa hur lägger du upp samtalet, vilket struktur har du förmåga, kan du följa de centrala frågeställningarna som barnet kommer fram med, så att det blir väldigt tydligt, de uttrycker de själva också att redan efter det andra samtalet så kan säga att efter nu hör jag när jag gör detta igen, och det är en optimal träningseffekt egentligen, när du själv gör självreflektioner, att du ser att du vet vad du gör för fel, så att det är ett gott tecken detta tycker jag att man gör så.

Du gjort liknande utbildningar tidigare, men med andra verktyg.

Ja precis, vi hade det här med videokonferensmiljön när man faktiskt var en vuxen människa. Men det blir liksom faktiskt en vuxen människa som sitter där med en vuxen röst och du sitter mitt emot den andra personen och det krävs väldigt mycket att skådespelare och du kommer inte bakom den här känslan att du faktiskt ser ett lite barn. så det har sina påtagliga begränsningar där, men på vissa formella bitar så kan du ju träna dig men detta är en helt annan nivå som man höjer, alltså realistiska upplevelserna av samtalet.

Du sa att du kände skillnad själv, att de blev engagerade, förstod jag dig rätt där?

Min upplevelse är att de blir engagerade och man ser på deras känslolductory och deras minspel och deras sätt att prata så känns det verkligen som att de är inne och då kommer vi in på den här biten att vikten av att man kan kontrollera lite mera finstämt av barnets känslolductory. och det skulle man kanske kunna göra genom att trigga vissa sekvenser att inte lägga in exempelvis som barnet är neutralt då ska det inte komma in något leende, du ska kanske plocka bort det i bildlopparna. Och så är det också de här övergångarna mellan bilderna. Sitter man exakt här så nästa sekvens så sitter jag exakt så också. och sen kan jag börja röra mig, men att man har en så ganska så stillasittande
så att man inte hoppar. Då kan man göra även övergångarna. de blir ganska så fina, framförallt på pojken han satt mer stilla. De övergångarna blir bättre när man kan göra den delen.

**Vad fungerar mindre bra. Du vill kunna utrycka dig lite mer finstämt med känslorna där**

Finstämt med känslorna, bättre övergångar, och jag skulle vilja ha haft ytterligare lite variationer, att de sitter, för nu kör jag så här: ”mmmm” jag gör liksom ett uttryck med stängd mun eftersom att barnet inte sitter. då skulle jag gärna vilja att man sa: ”jaaa” jämför att man säger någonting som att man ser att munnen rör sig. Det hade varit en del för nu blir man lite mer begränsad jag skulle kunna få igång samtalet ytterligare mer att kunna säga men då kopplar jag inte över till att prata normal för då blir det lite konstnat och då använder jag sådana ljud för att visa att jag håller med.

**Jag läste i enkäten att en del rapporterar om det är något teknikstrul eller om du sitter och väntar på om de ska säga någonting. Kan det vara svårt att förmedla den informationen. Det kanske är en icke-verbal kommunikation som man vill förmedla på något sätt?**

Allting handlar om att kunna hantera de här finstämda delarna av samtalet. Att man kan få det till att man skulle kunna ha en snabbfunktion där när man bara gjorde korta kommentarer. Då blir det att jag inte gör de kommentarerna pga. av att jag inte vill byta bild för att kunna fokusera på att göra det med lite mer med tonfall. Som i då fall ska förstärka det de själva har sagt så att de förstår att man är med.

**Några rapporterar in att det är en fördröjning. jag tror inte att det är en riktig fördröjning men de upplever det som att ni ibland pratar i mun på varandra.**

Jag upplever det inte som det när jag pratar med dem.

**Min hypotes är att de inte riktigt kan läsa kroppsspråket, att nu ska han säga någonting eller liksom det bara kommer kanske.**


**Känner du dig begränsad på några fler sätt som du inte hunnit beskriva än?**


**Med bra ljud menar du tekniskt med volym?**
Ja bra ljud så att man är hörbar för en själv och andra. Det är den här tekniska stabiliteten som det krävs lite mycket av. t.ex. skype kopplar ner lite då och då. Det är ju tekniska bitar som eftersom att vi har den lösningen vi har, vilken är fantastisk i sig själv men om fler ska kunna använda den, just användarvänligheten så…

Håller du med om att det är enkelt att kontrollera avatarens beteende samtidigt som måste göra alltid, jag tänker att det kanske blir mycket att tänka på? du ska ha en konversation samtidigt som du måste styra och så vidare.

Det krävs ganska mycket nu i dagsläget. nu behöver jag titta på avataren, intervjuaren, och vilja rutor jag ska titta, och jag ska titta egentligen på den som intervjuar då och det minspelet och så mina tankebubblor då. Jag har den här visuella tankemässiga stöden så att man ska få någon struktur. Om man är ovan vid detta och lite teknikrädd så i dagsläget i den här prototypmiljön då krävs det ganska mycket om man ska känna att man kan driva detta på ett bra sätt och verkligen hur man då ska kunna moderera barnets sätt och de olika variationerna i kombination med att både att då göra en trovärdig del.

Skulle du tro att du vill ha det inbyggt i programmet på något sätt så att du kan sätta upp olika scenarier och beroende på det så presenteras olika klipp till exempel? så att du kan skräddarsy till dig själv innan. Eller är det bättre att hålla det generellt?

Om man har ett scenario är det viktigt att det är flexibelt tror jag. Rätt som det är ska du vara instruktör i ett sammanhang så har du en del som du tänker att det här skulle jag vilja ha in. Det är klart, kan du skriva in det låt säga att du har motsvarande tankebubblor på en skräm så att du flexibelt det är textutor i respektive bubbla om du då kan ha den då och kan klicka på den linjen. Ja det är möjligt att du visuellt skulle kunna hitta det. det är klart då får du ett moment till, nu sitter jag med penna och bockar av, men det är bra, det känns som en påtaglig hjälp när man arbetar med det. för att kunna hitta alltså ett sätt som man kan få struktur och lätt att driva samtalet, om jag börjar prata så markerar jag och bockar av allteftersom och sen kan jag landa ut någonting annat eller så kan jag hoppa ut i någon annan ruta och linje som jag tycker att jag vill. Och att jag kan fylla på med... ja vad heter de. ja de heter Bengt och Paul som idag då. Då kan jag fortsätta med det i någon ruta så att jag vet vad jag sagt.

Men kan det också vara dumt att styra det för mycket, för då blir det så som att du har ett spår som du går och kan inte riktigt avvika från det.


Har det ändrats någonting i kraven som du hade från början?

Det är det här med finstämdheten och känslouttrycken återigen som man känner, att man vill kunna motsvara fullt ut. Det är det mest tydliga som känns för stunden. Att
kunna variera mer saker men det får inte vara så avancerat att det krävs en treårig utbildning.
Appendix B Trial 1 questionnaire

Vad fungerade bra och vad fungerade mindre bra? Vad anser du skulle kunna göras bättre?

• ”Tyckte det fungerade bra. Dock lite svårt att höra samt en viss fördröjning på munnens rörelser och talet jag hörde. Men när jag bortsåg från det kändes det som ett riktigt samtal med ett barn.”

• ”Jag tyckte att det överlag fungerade väl. Får en mer verklighetstroget samtal då man ser ett barn och hör en röst. Har lättare att uppfatta kroppsspråket och tonläge i samtalet. Svårt ibland att uppfatta vad som sades och fördröjningen gjorde att jag pratade i mun på barnet.”

• ”Tekniken begränsar. Något suddig bild och viss fördröjning.”

• ”Bra att barnet inte var så följsamt. Då måste man vara kreativ och tänka efter samt kunna följa barnet i vad det säger. Så är det ju iibland i verkligheten. Vi har en agenda medan barnet pratar om något helt annat. Om vi följer barnet kan vi ändå nå fram tänker jag, även om det inte blir enligt vår mall.”


• ”Bildens frös vid flera tillfällen och det hände även att rösten hackade eller försvann. Jag var osäker på om barnet satt still och var tyst eller om det var tekniska problem.”

• ”Jag tycker att det är ett utmärkt sätt att träna på intervjuteknik. Det som jag upplevde annorlunda var att det virtuella barnet gick direkt till den substantiella delen och det brukar inte vara så i verkligheten. Sedan kände jag att det virtuella barnet inte reagerade på mig utan bara ställde frågor som hon var programerat till. Jag förväntade mig mer reaktioner men det sker såklart bara i verkligheten.”

• ”Tycker att det överlag fungerade bra. Lite hackigt och svårt att höra barnets röst ibland. Bra är att se bild på barnet, det gör det hela mer verkligt.”
"Rösten var lite konstig men förutom rösten var barnet mycket verklighetsbaserad. Kändes som att prata m riktigt barn.”

"Jag tycker att det fungerade bra över lag, problem med tekniken men sånt händer ju. Tycker att man snabbt kom in i att m,an talade med ett barn. !"

"Personligen mkt svårt för att beförra mig framför kameran, men det ligger ju på mig. Fördröjningen försvara emellanåt, men inget jätteproblem”

"Samtalet kändes realistiskt. Teknikstrul inledningsvis gjorde att jag tappade fokus.”

Enligt dig, var det något specifikt som gjorde att du uppfattade momentet som icke- trovärdigt eller realistiskt?

"Realistiskt blev det då barnet svarde och rörde sig som ett barn gör. bra tycker jag.”

"Nej.”

"Nej, egentligen inte. Mest att man kom på ibland att det egentligen är någon vuxen som talar. Men det var lått att leva sig in i det.”

"Jag tyckte att tekniken strulade lite. Det var fördröjning och bilden hackade.”

"Att bilden bröts,”

"De tekniska svårigheterna sänkte samtalet en del och gjorde att jag fick en känsla av att det "inte var på riktigt". Annars tycket jag att detta var ett jättebra sätt och jag är imponerad av bilden på barnet och dess rörelser.”

"Nej, det kändes trovärdigt.”

"Jag tyckte att vissa svar sken igenom att det inte var ett riktigt barn men det gör inget i det stora hela.”

"Rösten kändes inte helt trovärdig först, men det glömdes bort efter ett tag”
Appendix C Trial 2 questionnaire

Vad fungerade bra och vad fungerade mindre bra? Vad anser du skulle kunna göras bättre?

• "Barnet kändes mer verligt denna gången jämfört med första. Kanske för att jag inte tänkte på den tredje parten alls.

• "Rösten är fortfarande lite konstig men jag tänkte inte på det i detta samtal"

• "Kändes verkligt. Dock kanske barnet var lite för "lätt pratat". Men det är nog bra då man får möjlighet att träna hela samtalet och inte fastnar. Det är så verkligt som möjligt."

• "Har ibland svårt att fokusera och störs av min egen bild och "all teknik" men mindre än vid första samtalen så det är nog till viss del en vanesak också."

• "Kändes mer naturligt än första gången. Tekniken fungerade bättre vid detta samtal och det var mindre nervöst."

• "Tyvärr hade vi störande ljud i form av borrande byggnadsarbetare i huset som gjorde att samtalen inte blev bra. Men det ligger ju utanför vad ni kan påverka."

• "Åter igen tycker jag att detta är ett väldigt bra sätt att träna tekniken att samtala med barn."

• "Barnet var mycket "lättpratat" för att vara ett barn med ADHD och asperger, vilket var till fördel för mig idag, men inte känns så verkligt. Hade förväntat mig mycket värre!"

Enligt dig, var det något specifikt som gjorde att du uppfattade momentet som icke-trovärdigt eller realistiskt?

• "Bilderna är realistiska."

• "Bakgrunden är svart så det är svårare att leva sig in i rummet. Barnet berättade dock vad som fanns i rummet vilket gjorde det lite mer verkligt."

• "Barnet är ovanligt samarbetsvilligt"

• "Svårt att tolka mimik"
Appendix D Selected interface design alternatives

These selected design alternatives were part of the design process for the user interface and illustrates the progression of the prototype.