Intelligent, socially oriented technology

Projects by teams of master level students in cognitive science and engineering

Editors

Christian Balkenius

Agneta Gulz

Magnus Haake

Birger Johansson

Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.)
Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154.

ISSN 1101-8453
ISRN LUHFDA/HFKO--5070--SE

Table of Contents

Attention in Cognitive Robotics 1
Karl Drejing, Erik Largerstedt, Karl Nordehammar and Lars Nyström

Does the appearance of a robot affect attention seeking behaviour? 11
E. Lindstén, Tove Hansson, R. Kristiansson, Anette Studsgård, Lars Thern

AGDA: Automatic generous/defence alternator 25
Carl Bjerggard, Tarik Hadzovic, Johan Hallberg and Olle Klang

Goal Emulation in a Drummer Robot 39
Hossein Motallebipour, Björn Andersson, Emil Gunnarsson, Filip Larsson and Jim Persson

Ankungen, ankmamman och krokodilen - en jämförelse mellan konsekvensfeedback 49
och rätt/fel-feedback i ett lärspel för förskolebarn
Ann-Louise Andersson, Caroline Alvtegen, Dan Thaning, Emil Nikander, Karl Hedin Sånemyr

Freds Bondgård: Ett av de första mWorldspelen för förskolebarn 61
Emelie Brynolf, Axel Hammarlund, Marcus Malmberg, Caroline Rödén, Magnus Walter

Lekplatsen – Ett lek-och-lärspel där förskolebarn lär sig grundläggande matematik.................................73
genom att lära ut till någon annan
Simone Andersson, Sofie Eliasson, Mattias Götling, Erik Jönsson, Jonas Svensson

Worlds' first "Teachable Agent"- game outside the mathematics and science domain 85
The Secret Agents
Axel Forsberg, Morten Hougen, Linus Håkansson and Johanna Truong

Developing an educational game 109
In the subject of history and with a teachable agent
Jesper Funk, Johan Lilja, Johan Lindblad, Sofia Mattsson

Metaphorical gestures, Simulations and Digital characters 125
Martin Viktorelius, David Meyer, Lisa Eneroth

Size certainly matters – at least if you are a gesticulating digital character: 141
The impact of gesture amplitude on information uptake
Mette Clausen-Bruun, Tobias Ek, Tobias Andersson and Jakob Thomasson

A Troublemaker as a Teachable Agent: how learning is affected by an agent with resistance.................153
Christoffer Andersson, Simon Johansson, Anders Persson och Adam Wamai Egesa

Exploring the Use of a Teachable Agent in a Mathematical Computer Game for Preschoolers 161
Erik Anderberg, Anton Axelsson, Sanne Bengtsson, Maja Håkansson, Lisa Lindberg

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Attention in Cognitive Robotics

Karl Drejing Erik Lagerstedt Karl Nordehammar Lars Nyström

The goal of this project was to model how humans read attention. By using a Microsoft Kinect device, a robot consisting of 3 servo engines and an interface called Ikaros we aimed to make the robot point at objects that were attended. Using theories about attention as a backbone for programming, our initial proposal was that gaze-direction and gestures are a way to tackle this problem. What was later implemented was a simplification of gaze- direction, namely head-orientation, and the prototypical gesture namely ″pointing″. This report describes the how we solved the task and also pro- vides a foundation for future directions regarding this project.

1 Introduction

Attention in humans is a product of many top-down and bottom-up processes [12]. The complexities of attention, as well as many other cognitive phenomenons, encourage exploration whether be it by empirical experimentation, modeling or some other means. Our groups task is to make a system that understand what humans attend. We will therefore examine and apply theories of attention. One bottleneck in this endeavour is that we have a unimodal system; we only work with visual stimuli. It is obvious that the attention system in humans is multimodal, but this does not make models of unimodal attention less interesting. This report will start of by defining attention, and then look at some attention models in cognitive robotics. Based on the theoretical findings, we create and implement modules in a computer program (Ikaros; [1]) to control a robotic arm. This Attention-Laser-Mechagodzilla-Arm (ALMA) is then to engage in joint attention [4] with the user for it to be able to point at an attended object. Furthermore, we will present an experiment regarding visual attention in humans. We also discuss the work so far and examine successes and pitfalls that we have encountered during the work process.

2 Theory

2.1 Attention Models in Cognitive Robotics

A general, broad term, definition of attention is that it is the ability to focus on one stimulus or action while ignoring other stimuli [12]. With regard to cognitive robotics, there are several models that can be applied to simulate this attention processes. Below we will review some of the major common characteristics of each attention model and some of the computational advantages and disadvantages of each model.

Saliency Operator: The most salient pixel of an input image corresponds to the area where attention should be directed [2]. An example of this is a red ball in front of a white wall; the red ball will be more salient and hence attended (the saliency method proposed by [15]). In order to prevent the same location from

being attended many times, it is important to implement some kind of Inhibition of Return (IOR) [2]. If no such method is implemented, the same salient stimulus will be attended at all time, which can lead to undesirable side effects.

Covert Shift to Focus: Here one assumes that neither the head nor the eye moves [2]. This mechanism in primates is commonly called covert attention [12]. This implies that 1) the retinal image is constant (i.e. the image remains un- changed until some object enters the scene), 2) because of the constant retinal input, the implementation of IOR is simplified and 3) the saliency of the scene image remains unchanged after each attentional shift [2]. But, the limitation of a stationary head (camera) makes it difficult to compare the model with the per- formance of a human (since the human gaze is rarely constricted to a stationary point in space).

Bottom-up and Top-down Analysis: While early cognitive robotics focused on bottom-up, or stimulus driven, models of attention [16], later models also incorpo- rated top-down, or mind driven, models with the goal to mimic human behavior to a greater extent [18]. Using this processes in joint operation the resulting at- tention rely both on the saliency of objects and the top-down guidance of visual search. Therefore it is necessary to implement programs that decide whether to search and/or explore the input scene (top-down) and programs that can handle the necessary input (bottom-up).

Off-Line Training for Visual Search: When a system is to recognize an object through top-down processes in visual search there is often training phase prior to the performance of the system [2]. In this training phase, the system learns an ob- jects features and saliency through repeated learning trails. The final performance relies heavily on the training phase (number of trails, quality of the trails etc.).

Space- and Object-Based Analysis: While humans have some notion of a con- cept of objects (that is: a fork has many integral parts that makes it a fork, and we have a conceptual understanding about the fork) the early bottom-up approaches of visual attention investigated objects at the pixel level [2]. This approach does not take concepts into account, i.e. they are not modeled in the system. Lately, evidence from the fields of cognitive neuroscience and psychology suggest that there exist objects which are the main scope of attention selection, although there are only a handful of researchers that have implemented these two approaches in

concert (e.g. [3]).Before talking about the implementation and how these processes relate to our model we will review two major attention cues, that is gestures and gaze direction.

2.2 Attention and Gestures

According to [7] there are two kinds of deictic gestures that form indices to indi- vidual objects; these are directing-to and positioning-from. The first mentioned is a strategy identified by [17] which humans use to refer to an object in a visual space is, and that is pointing. [17] made an experiment in which they examined whether spatial description helped a listener to accurately determine where atten- tion is directed. They found that there is a significant difference between when there is no pointing and spatial description is provided vs when no such description is provided (mean correct answer 0.93 vs 0.63). No significant change was noted when there was pointing vs the spatial condition (0.93 vs 0.90). Hence, pointing without verbal spatial information is fairly accurate when it comes to determining where attention is located. This study was restricted to 12 stimulus at a time.

2.3 Tracing Visual Attention in Humans

We assume that the gaze of a person indicates attention towards an object or area. To make certain that is the case we rely on theories about Eye-tracking. Exper- iments by [20] suggest that one must restrict visual processing to one item at a time. That is, several items cannot be processed simultaneously without redirect- ing the gaze; one must restrict input in order to act upon it. [8] hypothesized that the study of eye movements opens a window to monitor information acquisition. Commonly used measures to infer cognitive processes from eye movement includes, but are not limited to, duration of the fixations one makes, frequency of fixations and scanning path [19]. Hence, the gaze of an individual is an appropriate measure to infer what that individual is attending.

2.4 Joint Attention and Common Ground

Joint attention is, briefly, the clue to intentional communication [5]. To properly model attention it is important that both agents in the attentional process share the same perceptual discrimination. In turn this discrimination should guide and form a basis for an ascription of perceptual content. This is a linguistic model of joint attention elaborated by [4] but is in no way restricted only to auditory input. [4] put forth three conditions that must be fulfilled if joint attention is to emerge: the speaker and hearer must 1) attend to each others state of attention 2) make attention contact, and 3) alternate gaze between each other and target object. We clearly see that condition 3 is tied to [20] and [8] theories of visual attention in humans. Furthermore, several authors in [4] have refined condition 1. They argue that in addition to 1 the attentional agents should also grasp that the attentional process is directed at objects, event or other entities. This implies that there is some form of coupling between attention and intention; there is an intention behind the behaviour that is attention. [4] also state that while there is a close connection between attention and intention, it is not the case that higher- order thoughts about the intention is induced in the receiver (i.e. the speakers intention is not fully treated as a higher-order thought in the receiver). Although it is necessary for two agents to access to the same perceptual properties when engaging in joint attention, it is not sufficient. [4] argue that it is necessary to take into account how the agents manage to identify the perceptual qualities of a stimulus in a similar way to focus on a common cause.

In his PhD thesis, [6] elaborate more on why common ground is important when it comes to human-robot coordination. According to [6],common ground is the shared knowledge between individuals which allow them to coordinate com- munication in order to reach a mutual understanding. One can also see optimal common ground for communication as when the collective effort for some individ- uals to reach mutual understanding is as minimized as possible [6].

2.5 Top-Down Control

Considering the theories of attention proposed above, it is important when im- plementing to consider in what order and how the system should treat inputs to

produce some output. Top-Down control can be defined as the system which with the use of knowledge, expectations and current goals drives, which in our case is attention [10]. In addition [10] describe some common conceptualizations of top- down processing. One method is to have a database with all relevant objects in it. The system can then discriminate these objects from other, irrelevant, objects. Another method is a context based. [10] exemplify this by describing a system that is looking for a person in the street. If the whole input image consist of a street and some part of the sky, the sky is ignored and only the street is attended. One could argue that we want are implementing the first of these examples. As we discuss below, we make use of gestures, eyes and heads to achieve object recog- nition. Parameters for these are already incorporated in the visual search and are hence like the first method described.

Although, based on the findings of [4] and [6] we also must construct a top- down model that makes a user understand what ALMA is supposed to do. It is crucial that an individual, who have not prior knowledge about ALMA, when interacting with it understands what it does. Else one cannot argue that there is joint-attention and especially no common ground.

The top-down control will be discussed in more detail below in the Module section.

2.6 ALMA and Attention Models

The goal that ALMA is supposed to achieve is to point at a target, among a finite number of targets, that someone in front of it is attending. The two main cues here are gaze direction and pointing a finger at some object. The above methods of discerning bottom-up attention are, in our case, somewhat inadequate or should at least be viewed with modification.

First of, ALMA get visual input via a Kinect device. This device sends RGB inputs and depth[1] of each pixel to Ikaros. In this image, our modules try to discern whether a finger is present or not, and the vector of that finger (i.e. where do the finger point). A hand detection algorithm is used for this purpose. This resemble the Object-based approach described above, where elements are combined to distinguish some object in the visual frame.

Second, ALMA does not work with a Saliency Operator. There are no modules present that use the technique of saliency proposed by [15]. Earlier modules which were implemented distinguished the closest pixel as an indication to where someone was pointing (a gross simplification in the early stages of development), which could be seen as depth saliency. The saliency operator is unlikely to be implemented because Ikaros can distinguish targets via bar codes (bar code cubes), which we use as targets of attention.

Third, the camera (Kinect device) that is used to gather visual information is immobile, making ALMA an overall poor model of human visual attention. As discussed above, the static head is not a natural state of any primate and hence one must be extremely careful when mapping visual attention from ALMA to human visual attention.

Forth, as for now the bottom-up top-down process in our model is very limited. The system actively searches for patterns and when such a pattern is found it is treated accordingly. There are no motivation, prioritizing or other top-down modules involved.

Fifth, a necessary aspect is to achieve joint attention. To answer that question, we will examine how the implementations relate to the joint attention process proposed by [4]. The extended version of the first condition in [4] is considered not to be fulfilled. That is because there has to be intention behind the behavior that is attention. At its current state ALMA posses a very low degree of intentional processes. The only intention there is to find the closest pixel in the picture and/or detect a face. Condition two [4] is partially fulfilled because of the feedback the user get from the robotic arm. If an agent is pointing in a direction, ALMA will do the same and hence create a feedback-loop between itself and thee agent. Hence there exists some form of attentional contact. The third condition [4] does not exist at all, due to the static nature of the Kinect. Is it possible to achieve joint attention? It is our opinion that ALMA can achieve this sufficiently. Once an adequate top-down model is implemented the intentional aspect of ALMA will increase significantly, possibly achieving intention behind its behavior that an

agent can grasp. Furthermore, the user will always get feedback from the arm, meaning that the feedback-loop (compare to the gaze alternation condition in join attention) will always exist.

3 The Experiment

3.1 Goal

To discern how accurate humans are when it comes to telling where attention is located. This is later to be compared with ALMAs performance.

[1]It is important to know how the depth is actually obtained from the Kinect. It projects a grid of IR-"dots" in front of the IR-camera. It then look at the echo emitted from these dots. In this manner, the Kinect can discern at what distance each of these dots are in relation to the device itself.

3.2 Participants
12 participants in pairs of two participated in the experiment (i.e. n=6).

3.3 Materials
Six multicolored cubes, measuring 1.5cm x 1.5cm x 1.5cm were used.

3.4 Method
Two participants sat opposite to each other. The mean distance from one par- ticipant to the table was approximately 1m. One of the participants was labeled instructor and the other labeled Robot (ALMA). The instructors role was to either gesture or look at some of the target cubes (described below) and the Robots role was to discern what cube was being attended. Before each trail, in all conditions, the Robot closed his or her eye to avoid cheating. When the eyes were closed the experiment leader choose a cube numbering 1 - 6 from the Robots left. Across trails, the same cube index was used, resulting in a predefined sequence of cubes that should be attended.

Condition 1 (C1) In this condition the instructor was told to point at a cube. After the Robot had made its decision of what cube was attended a new trial began. There were two sub-conditions in this condition. 10 trails were made with5 cm between the cubes, and 10 trails with 0 cm between the cubes.

Condition 2 (C2) In this condition the instructor was told to look at a cube. After the Robot had made its decision of what cube was attended a new trial began. There were two sub-conditions in this condition. 10 trails was made with 12 cm between the cubes, and 10 trails with 3 cm between the cubes

Condition 3 (C3) In this condition the instructor was told to look at a cube with the head rotated approximately 22,5 degrees x-wise to the left and right re- spectively. After the Robot had made its decision of what cube was attended a new trial began. There were two sub-conditions in this condition. 10 trails was made with 12 cm between the cubes (5 trails with the head rotated to the left and 5 to the right), and 10 trails with 3 cm between the cubes (5 trails with the head rotated to the left and 5 to the right).

3.5 Analysis
T-test was performed, comparing Condition 3 to Condition 1 and 2.

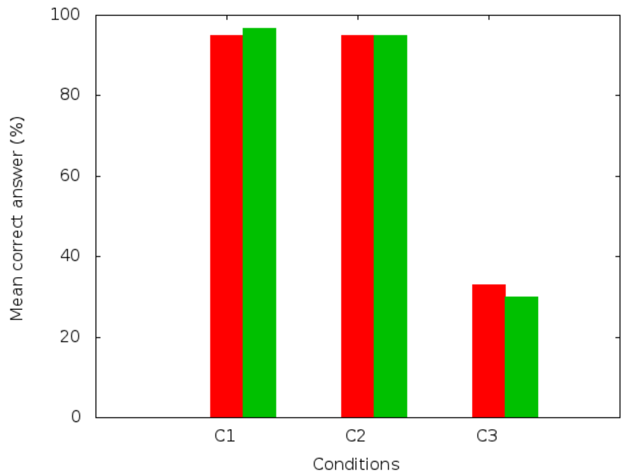


Figure 1: This graph depict the mean correct answers across conditions. The red bars correspond to 5cm (C1) and 12cm (C2 and C3) respectivly and the green bars correspond to 0cm (C1) and 3cm (C2 and C3) distance respectively.

3.6 Results
The mean correct answers across conditions are shown in 1. For convenience we only present t-test scores comparing C3 to C1 and C2. We also take the liberty of not presenting intra-condition t-test score, since n is very low and the mean correct answers are equal or close to equal within each condition. Comparing C3 to C1 resulted in a significant difference ($t(5)=2.115$, $p < 0.05$) and C3 to C2 also resulted in a significant difference ($t(5)=2.100$, $p < 0.05$).

3.7 Conclusion
Our results suggest that humans are rather bad at discerning where attention is located if they get insufficient information. In C3 part of the face is obfuscated and hence they participant did not get sufficient information. This resulted in a correct response frequency that was slightly better than chance. Furthermore, we see that the distance between objects seem to have little impact on the response frequency. It is also worth to note that these conclusions are drawn from a sample of n = 6 and should be interpreted with caution.

4 The Project

We will start this section by a short introduction of the Ikaros environment followed by examining past solutions. Explaining how and why we implemented them and if we failed at that effort and, in that case, why we failed. In addition, we explain the current successful implementations and how they work. Each of the solutions are explained module wise with an explanation following each. Stemming from the above theory, we came up with some initial solutions and how we were going to tackle the problem of reading attention (or more specifically: joint attention). These developed over time, giving birth to more efficient modules and/or scrapping the ones in the initial development phase. It is our intention to present the modules in a chronological order, starting with the ones that were first implemented.

4.1 Ikaros

The platform we use to control the robot is called Ikaros [1] and was developed with the intent that researchers should have an easy interface for neuromodeling and robotics. Basically, Ikaros provides an environment in which modules can be implemented and connected to each other. Each module has one or more inputs and outputs. Hence, module A can receive an input from some arbitrary module and produce an output.

4.2 Previous Solutions and Modules

4.2.1 Previous Solutions

Our previous solutions was focused on gestures (pointing) and gaze-direction. Al- though we wanted to focus on gestures, we did not implement a gesture recognition algorithm, but instead focused on the gaze-direction. The gesture that we did fo- cus on was simplified into calculating the closest point to the Kinect device and then, by a top-down process, direct the robotic arm so that it pointed to the cube closest to that point. Gaze-direction took up a lot of time and was later abandoned becasue of various reasons explained below. A review of the module schematic in the Ikaros environment can be found in picture 2.

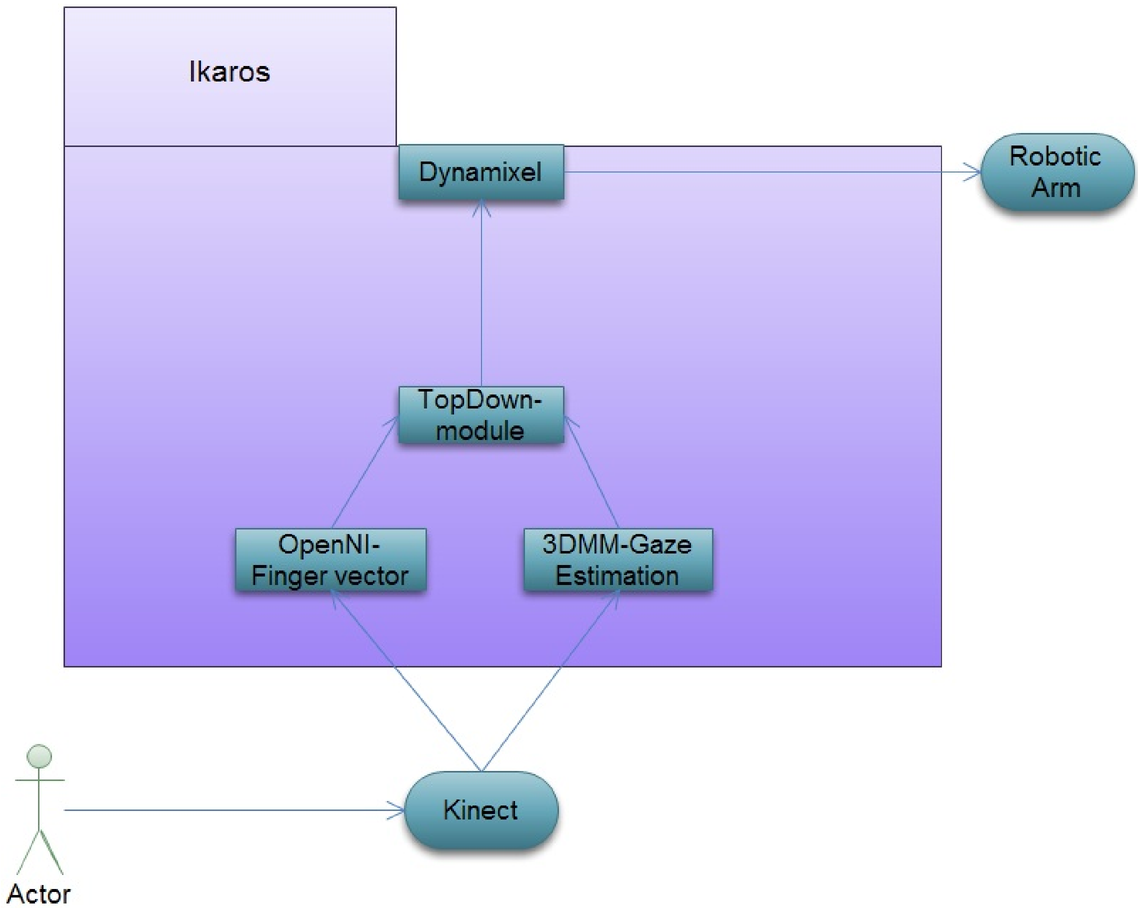


Figure 2: This is how the initial idea for the general structure of ALMA's modules. Every rectangular box represent a module, the rounded boxes are hardware and solid lines represent connections between modules.

4.2.2 Previous Modules

ClosestPoint This module track the closest pixel to the Kinect and return that pixels coordinates. This module was written in order to do two things. 1) we wanted to get familiar with the Ikaros environment and 2) we wanted to focus on the gesture part of the theory presented above. This module was used for a long time and worked with (surprisingly) high accuracy. However, it was abandon because it did not have the necessary requirements to represent a pointing gesture.

ArmPoint A module that was written after ClosestPoint. This module is, as of today, still in use. This module converts coordinates of points in space to angles

for the servo engines. If it only gets x- and y-coordinates, it can use a depth image to find the z-coordinate.

MarkerTracker - This modules is in the standard Ikaros Library and en- ables the reading of bar codes. This module has been used to get ids' for the cubes in order to have something to attend; we do not make use of novel objects and higher-level bottom-up processing to identify classic objects such as red balls.

3DMM-module - This module existed only in theory, but it was a main focus for a long time. Based on the work of [11], who used a 3D-Morphable-Model to estimate gaze-direction, we collected research about the means to implement such a solution. The algorithm that [11] propose creates a 'standard' face from a library of faces and paste it on a face that is currently visible via the Kinect device. There is also a gaze-estimation algorithm which work in concert with the 3D-face model. The advantage is that the gaze-estimation is very accurate even when a user move the head. The disadvantage is that no source code was avail- able. We tried to contact the researchers, but they replied that they sadly could not provide us with the code. After some time we deemed this too be to much of a challenge to implement and hence abandoned it and searched after other solutions.

4.3 Current Solutions and Modules

4.3.1 Current Solutions

The overall structure of out current solution can be divided in five major departments: 1) The recognition of a hand, 2) The recognition of a closest point, 3) A bar code id memory, 4) a head pose recognizer and, 5) A priority module. The first four structures feed information to a priority module further up in the mod- ule hierarchy. The recognition of the hand is based on the OpenNi framework and identifies a hand when it finds a certain gesture (waving). This gesture is, as described above, not a part of our basis of reading attention. It is rather a necessary evil that was needed in order for the OpenNi framework to recognize any hand at all. Unfortunately, there was no time to circumvent procedure. Our aim was to extract a skeleton of the hand in order to calculate a pointing vector, but the skeleton of the fingers could not be extrapolated (the software lacked the necessary

tools for this action), so we decided to revive the closest point module. Because we found the center of the hand, and provided the closest point was a finger (which as we explained above had previously worked with high accuracy), we could calculate a pointing vector using these two coordinates. There is an obvious drawback to this: that the closest point must the finger pointing at an object in order for this module to work.

The bar code id memory was implemented to solve a recurring issue, that is that the bar code ids were lost from time to time (depending on lighting). As soon as a bar code is visible, its coordinates is stored and only updates if the same barcode id coordinates change. This addition greatly stabilized the system.

The intention was to have our own 3DMM-module in this final version. With inspiration from [11] this should be accomplished by finding two vectors, namely the gaze vector and the head pose vector. In concert, these should create a more accurate estimation for where the gaze of a user was located. Also, it would in- crease the accuracy if one eye in the image was lost due to head rotation or other reasons.

The gaze direction module (EyeDir, described below) proved to perform accuratly, estimating if a user looked right/middle/left, but did so only in optimal lighting conditions. Hence, that solution was abandond and the whole gaze direction estimation was simplified into a head pose estimator. This estimator is built on the OpenNi environment, which takes data from the Kinect depth sensor and calculates a head direction.

The last department is the priority module which get fed data from these other departments. There are two versions of this module, the reason being that the project has evolved. Our initial project description was to model reading atten- tion in cognitive robotics. This can be interpreted as either locate where attention is directed or do that with a more robot-user-interaction approach. The later is based on theories of common ground, i.e. that there has to be a mutual under- standing of what the user and the robot intends. If one were to module a top-down process in this manner, it would have to include some type of feedback when the robot is in doubt of where attention is located. Therefore, it was proposed that if such a situation arise, then the arm would be directed to point at the users eyes. But the former interpretation of the project provided a different interpretation, namely

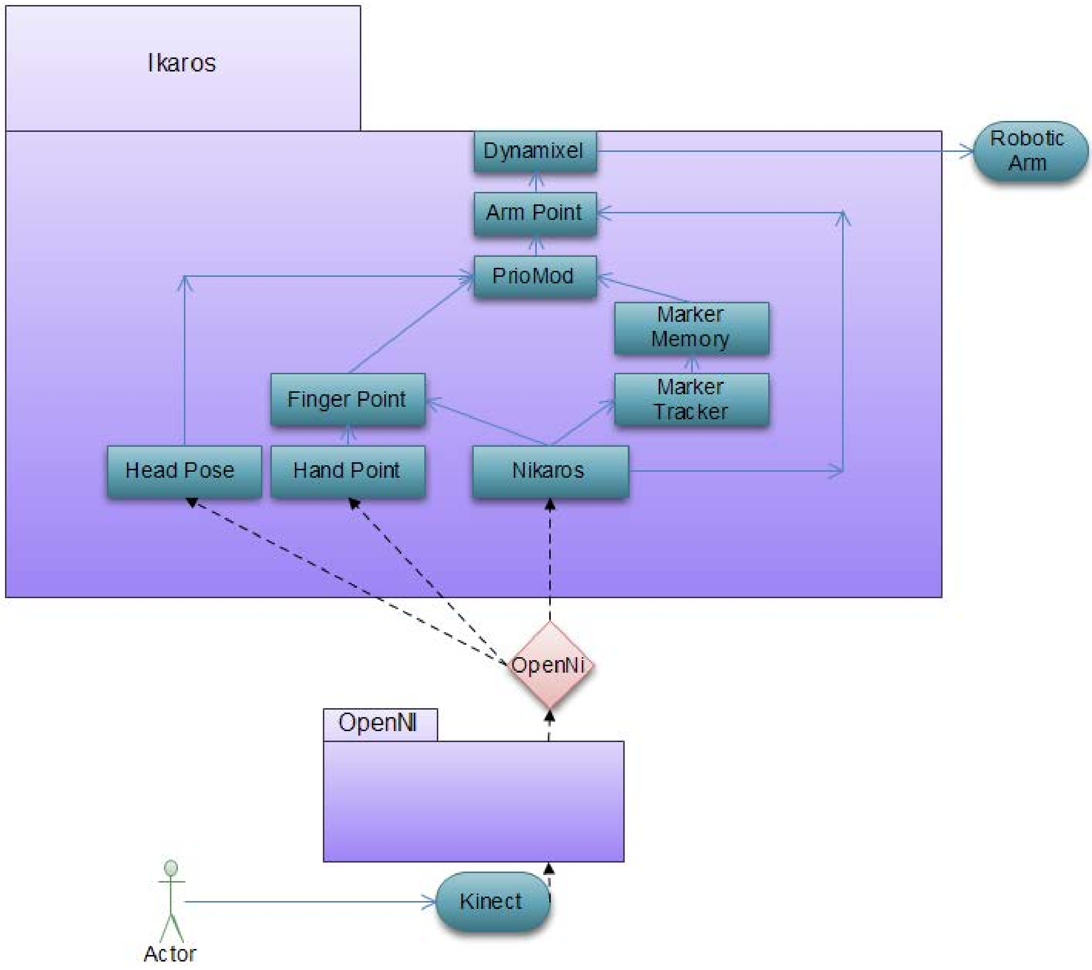


Figure 3: A depiction of all modules in the final version. Every rectangular box represent a module, the rounded boxes are hardware and the diamond is a software. Dotted lines represent dependencies and solid lines represent connections between modules.

that this is not what the project was about; it was about where attention was located. With that interpretation, the arm should only point at the latest attended object. Further details about this module is shown below.

All modules, and there respective inputs and outputs are seen in 3.

4.3.2 Current Modules

Nikaros - The problem that we encountered was the communication between OpenNI and the Ikaros environment. OpenNI is structured around production nodes, while Ikaros consist of interacting modules. In addition, the two system uses different drivers for the Kinect. That implies the if one called OpenNI func- tions in Ikaros then Ikaros ability to read data from the Kinect was disabled. What we did was to create a module (Nikaros) that uses OpenNI to get data from the Kinect. OpenNI uses something the call 'contexts' where all the nodes are registered. What we needed to do was to share a common context to all Ikaros modules. This was accomplished by creating a class that was not an Ikaros module, but rather a singleton class, where all Ikaros modules can get the same instance of the OpenNI context. OpenNi is a class, and it is an interface between the external library OpenNi and Ikaros.

EyeDir Instead of using the 3DMM model discussed above we created our own gaze-estimation module. This module gets 50pixelx50pixel sized pictures of eyes from Ikaros face detection module. By looking at average pixel values, standard deviations of pixel averages and intensity gradients in the image, it is possible to find features such as the pupil and the sclera. By looking at their relative position and size, it is possible to find the gaze direction. Our module was only developed to a point where it was reasonably good at deciding whether someone is looking left, right or straight. To do this, it demanded quite specific light conditions, which is the reason why the development halted.

HeadPose The 3DMM model used a 3D face which in concert with eye es- timation algorithms made gaze-estimation more accurate. Based on this idea, we thought that we could do something similar and hence we constructed this module. Using the OpenNI framework, this module finds a head and two points (the head center and head front). By using those two points it creates an output vector that indicate which way the head is pointing. By combining this module and the output of the EyeDir module we get a more accurate gaze-estimation. For example: If the head is rotated in a way that makes one eye invisible to the Kinect, this module compensates for that by instead using the head direction vector. It is noteworthy to say that it does have its flaws, such as pointing a little to the side of the target. Acknowledgements to [9] for developing this code.

HandPoint - This module uses OpenNI to find and track a hand. The user has to wave to the Kinect for the module to detect the hand. When the hand is detected, the module will track it. The output is a point on the hand, commonly in the center.

FingerPose - Gets the hand coordinates from HandPoint. It then crops out a square around these coordinates, and finds the pixel closest to the Kinect.These two points will often lie on approximately the line of a the pointing gesture. In practice this is not actual pointing gesture. Such a gesture would have been the base and the top of the finger. The resulting pointing estimation depends on where HandPoint detect the hand and also provided that the finger tip is the closest point relative to the Kinect.

MarkerMemory - This module stores the information from Ikaros marker- Tracker module. This means that Ikaros remembers the markers, even if they are occluded.

SetKinect - Is a model from before the days of OpenNi. It is used to set the LED and the motor position of the Kinect.

PrioMod - This is the top-down module. The purpose of this module is to col- lect all the cues, and as output send the coordinates of the attended point. It takes the tracked markers at input which will be the possible attendables. The other input are preferably the coordinates of one point, or the coordinates of one pair of points. PrioMod has two non-trivial functions. The first of them, PrioMod::dist(), takes the coordinates of two points in three dimensionally space, and returns the distance between them. The other function, PrioMod::line–PlaneIntersect(), takes the coordinates of three points. The third point lies on a plane, parallel to the lens of the Kinect. The other two points lie on a line. The function returns the smallest distance on the plane between the third point and the intersection between the line and the plane. PrioMod uses these two functions to determine what its input points at.

5 General Discussion

Our task was to create a system that could estimate where an agent directed his/her attention. This was done by using the unimodal Kinect device and theories ranging from joint attention [5], [4], gestures [17] and visual attention [20]. Humans sees this task as very easy, natural and trivial but as we delved deeper into the issue, it was clear that this phenomenon is not easy at all. All the confounders that exist in the environment can be as easy as ”Does she point at the painting or the wall it is put on?”. As the project progress, this type of problem has become extensively clear. Our end product can, with moderate accuracy, determine at what (predetermined) object a person point and if the head also attend that object.

An obvious limitation of our solution is the moderate accuracy. The problem lies in the top-down process; as standalone implementations, the gesture and head- pose work very well, but when those inputs are combined the chance of error increases. One explanation for this is that there are no modules that refine the resulting output that comes to the top-down module. Imagine that we have a scenario where a finger points to location X but also slightly to location Y, and we have no module between the pointing module and the top-down module to refine the signal so that the resulting output is ″Definatly pointing at X″. This is what happens in our case; the signals are not refined enough to generate an accurate result, which leads to an erratic behaviour in the top-down module. If refinement modules were introduced, we believe that this error would greatly decrease.

On the bright side, the implementation is not lighting dependant. We use only the depth sensor to gather data, and that happens to be an IR-sensor, which means that our system works well in complete darkness (provided that the attendable objects are visible and recognized before the environment turn dark).

There are several other limitations to our implementation that is worth noting, and is also interesting for further research, such as the ideas at the start of the project. Our inital ambition was that several people would be able to stand in front of ALMA and she would accuratly discern where attention was located; currently, this is not the case. We believe that the system cannot handle several agents at once. We say ″believe″ because we have not tested such a situation. This belief is based on the current performance and on how the code actually looks. The head-pose module is a prime example for why we believe that this is the case: it cannot handle more than one face at once. It only takes the latest recogonized face and treats this as the only face. Here, there is ample with room for further development.

There is also the possibility of other types of gestures, such as a open hand compared to traditional pointing. As of now, open hand gestures are possible for the system to recognize but have less accuracy than traditional pointing (one finger streched out from the base of the hand). The system is also limited when it comes to pointing at several objects at once. The system cannot identify several hands pointing at different objects (it is restricted to one hand only). The interesting question one must ask oneself in such a situation is: Where is attention actually located? Our implemented solution is that it is the latest updated inference, from gesture to object, is the correct one, but this question is open for further discussion.

Although our system may not be perfect it is pointing in the right direction. With more time and more code it would be possible to do a lot more, and that is why we have great hopes about what people interested in developing this further can do, and hopefully will do, to create a better theoretical framework and better applied robotics in the future.

References

[1] Balkenius, C., Morn, J., Johansson, B., and Johnsson, M. (2010) Ikaros: build- ing cognitive models for robots, Advanced Engineering Informatics, 24, 40-48, 2010.

[2] Begum, M.; Karray, F.; , "Visual Attention for Robotic Cognition: A Survey," Autonomous Mental Development, IEEE Transactions on , vol.3, no.1, pp.92- 105, March 2011

[3] Begum, M., Mann, G., Gosine, R., and Karray, F. (2008) Object- and space-based visual attention: An integrated framework for autonomous robot. IEEE/RSJ International Conference on Robots and Systems, 301-306.

[4] Brinck, I. (2004). Joint attention, triangulation and radical interpretation: A problem and its solution. Dialectica, 58, 179-206.

[5] Brinck, I. (2001). Attention and the Evolution of Intentional Communication. Pragmatics and Cognition, 9, 255-272.

[6] Brooks. A., G. (2007) Coordinating Human-Robot Communication. PhD the- sis, Massachusetts Institute of Technology.

[7] Clark, H. (2003). Pointing and placing. In S. Kita (Ed.), Pointing. Where language, culture, and cognition meet (pp. 243268). Hillsdale NJ: Erlbaum.

[8] Just, M., Carpenter, P. (1980). A theory of reading: From Eye Fixation to Comprehension. Psychological Review, 87, 329.

[9] Fanelli G., Weise T., Gall J., Van Gool L., Real Time Head Pose Estimation from Consumer Depth Cameras. 33rd Annual Symposium of the German Association for Pattern Recognition (DAGM'11), 2011

[10] Frintrop, S., Rome, E., and Christensen. H. (2010). Computational visual attention systems and their cognitive foundations: A survey. ACM Trans. Appl. Percept. 7, 1, Article 6.

[11] Funes Mora, K., and Odobez, J. (2012). Gaze estimation from multimodal Kinect data. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 25-30.

[12] Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (2009). Cognitive neuro- science: The biology of the mind. New York: W.W. Norton.

[13] Geary, D. C., and Huffman, K. J. (2002). Brain and cognitive evolution: Forms of modularity and functions of mind. Psychological Bulletin, 128(5), 667-698.

[14] Itti, L., and Koch, C. (2001). Computational Modelling of Visual Attention. Nature Reviews Neuroscience, 2(3), 194-203.

[15] Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Transaction on Pattern Analysis and Machine Intelligence. 20(11), 12541259.

[16] Koch, C., and Ullman. S. (1985). Shifts in selective visual attention: Toward the underlying neural circuitry. Human Neurobiology, 4, 219227.

[17] Louwerse, M., and Bangerter, A. (2005) Focusing attention with deictic ges- tures and linguistic . XXVII Annual Conference of the Cognitive Science So- ciety, 2123.

[18] Navalpakkam, V. (2006-10). Top-down attention selection is fine grained. Journal of vision (Charlottesville, Va.), 6(11), 4-4.

[19] Schulte-Mecklenbeck, M., Khberger, A., and Ranyard,R. (2011). The role of process data in the development and testing of process models of judgment and decision making. Judgment and Decision Making, 6, 733739.

[20] Wolfe, J. (2000). Visual attention. Seeing, 2, 335.

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Does the appearance of a robot affect attention seeking behaviour?

E. Lindstén, T. Hansson, R. Kristiansson, A. Studsgård, L. Thern.

ABSTRACT

In order to examine human attention seeking behaviour we conducted two experiments for this project. Our first experiment examined which gestures humans perform when trying to obtain attention from another human being. The second experiment examined the effects different appearances of a robot might have on human behaviour, when trying to obtain the robot's attention. The two different robot appearances consisted of a hi-fi looking robot and a lo-fi looking robot. We found that eye contact is crucial to attract attention from a desired human interlocutor as well as a robot, regardless of its appearance. We also found that certain gestures, such as turning hand, pointing, waving and leaning forward attract attention. Findings from our first experiment was used for implementing the robot used for our second experiment. The second experiment concluded that the different appearances do not affect attention seeking behaviour, but that the appearance is important regarding which robot is preferred. In a situation like the one we set up, humans seem to prefer lo-fi pleasant looking robots to hi-fi complex looking robots. It also seems we can confirm previous studies concerning the importance of a match between robot appearance and robot behaviour.

INTRODUCTION

Appearance matters. In robot interactions too. Various studies (Robins et al. 2004, Syrdal et al. 2006, Walters et al. 2007) show that the appearance of a robot affect how humans interact with and perceive the robot. However, previous studies (e.g., Johnson, Gardener & Wiles 2004) have also shown that humans interacting with a computer will behave as if it was another human, the theory called *the media equation*, regardless of the appearance. With regard to the media equation, we set out to examine whether a robot's appearance will affect human attitude enough to differ the gestures and strategies used trying to get a robots attention.

In this first section we present relevant theories on robot appearance as well as theories on human-robot interaction. We end the introductory sections by presenting an experiment which we conducted prior to the current experiment, and our hypotheses for the current experiment.

Figure 1: Faces in places

Source: http://makedo.com.au/inspiration-insight/faces-in-places.html

Judging a robot by its cover

In the process of designing a robot or humanoid, studies has often been conducted in order to examine the expectations and

Figure 2: The robot

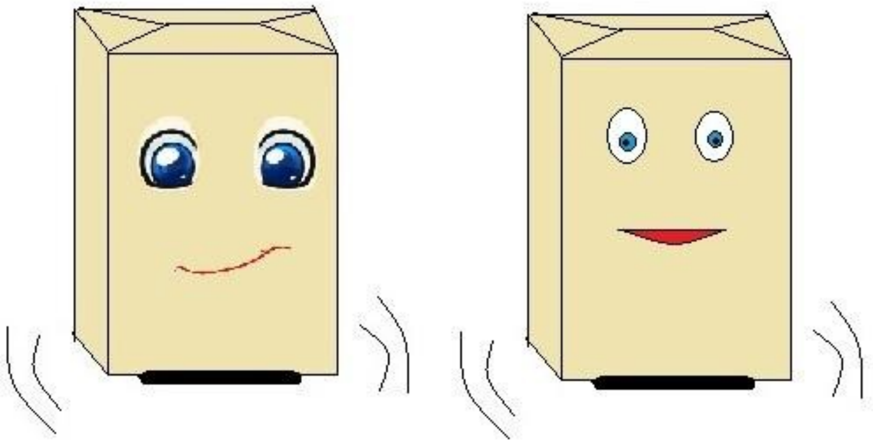


Figure 3: Interfaces

assumptions for the appearance of robots (Han, Alvin, Tan & Li 2010; Kahn 1998; Nomura, Kanda, Suzuki & Kato 2005). The reason for collecting such information is to be able to design a robot in a way that makes the future interlocutor want to interact. This presupposes that different appearances have different effects on the interaction. Even though humans are known to see 'faces in places', the phenomenon pareidolia (see figure 1), we do not believe the design of our robot (see figure 2) leeds to any face interpretation which make the examination of the effect of the appearance much more interesting.

Khan (1998) examined with a questionnaire survey what would be the preferred appearance of a service robot, but encountered problems; people tend to prefer different looks for various purposes. Human-like appearances are not always the most preferable. People tend to respond more positively when they perceive a match between the appearance of the robot and the behaviour of the robot. If a robot appears intelligent by looks, and turns out less intelligent, humans are disappointed and show less interest to proceed in interaction (Goetz, Kiesler & Powers, 2003). Comparisons of human impression concerning two humanoids (one more human than the other) and one human behaving the same way in interactions, show that the human gave the worst impression. This was believed to occur due to the humans lack of conventional behaviours we expect of an everyday interaction with other humans, such as "a particularly welcoming attitude such as a smiling face, a casual introduction, or conversation about common interests" (Kanda, Miyashita, Osada, Haikawa & Ishiguro, 2005:6). Robots behaving the same way was accepted.

When examining the benefits of a human-like appearance for a robot one has to consider an appropriate match between the appearance and the behaviour. If "a robot looks exactly like a human and then does not behave like a human, it will relate the robot to a zombie. This will cause the human robot interaction to breakdown immediately" (Han, Alvin, Tan & Li, 2010:799). Guizzo (2010) explain how the line separating pleasant from unpleasant in robot design is delicate as illustrated in the graph of the uncanny valley (figure 3).

In order to design a successful appearance for a robot, it does not seem necessary to make it more humanlike, attractive or playful in general. Rather it should be designed to match the users expectations concerning the robot's capacities and functions. This is suggested to increase users' will to co-operate (Goetz, Kiesler & Powers, 2003).

With this in mind, we design another look for the robot, which consists of a brown paper bag with "cute" eyes and a simple mouth (see figure 4). To show the effect of eyes and mouth the same paper bag, but with a different look is also shown.

The interlocutor and how s/he is perceived obviously affect what we talk about. Similarly how we talk also depends on who we are talking to. Baby talk, as the term suggests, is how we talk to babies. Baby talk (also referred to as infant-directed speech, parentese etc.) is defined by being high and gliding in pitch, and having shortenings and simplifications of words (Thieberger Ben-Haim et al. 2009). Though it is called baby talk it is not only used when interacting with babies, people also tend to use baby talk when talking to dogs (Mitchell 2001), when talking to chimpanzees (Caporael et al. 1983), and even when talking to elderly nursing home residents (O'Connor & Rigby 1996). We therefore hypothesize that the use of baby talk is based on an evaluation of the cognitive level of the interlocutor. That is, when we feel that our interlocutor is below our own cognitive level we will resort to

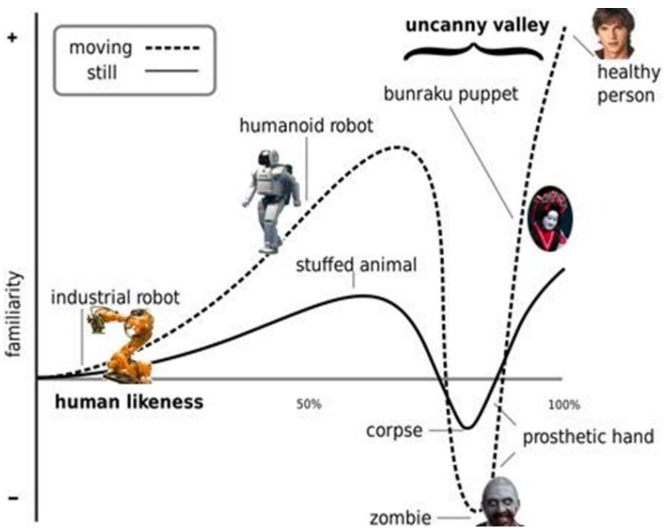


Figure 4: Uncanny valley

Source: http://spectrum.ieee.org/automaton/robotics/humanoids/040210-who-is-afraid-of-the-uncanny-valley

Figure 5: Mother interacting with baby
Source: http://nashvillepubliclibrary.org/bringingbookstolife/2012/08/09/kiss-your-brain/

Figure 6: The social robot, Leonardo
Source: http://web.media.mit.edu/~coryk/projects.html

baby talk to make sure that our utterances will be understood. Caporael et al. (1983) indeed did show that the elderly with a lower functionality score showed greater liking for baby talk. Figure 5 shows how a mother is very explicit in her facial expressions when communicating with her child. We suggest that gestures follow speech, that is, when the interlocutor is perceived as being suited for baby talk the gestures will also become more explicit and engaging.

This hypothesis leads us to believe that our two different versions of the robot will receive different responses. Since the lo-fi interface is very simple designed with only two eyes and a mouth and since we choose "cute eyes" we believe that this look will be perceived as holding a lower cognitive level than a human interlocutor. Whereas the hi-fi robot will most likely be interpreted as holding the same cognitive level as the human interlocutor. We hypothesize this, because as humans we are very used to computers being very skilled and able to handle specific tasks much better than humans.

However, when a robot is designed to appear as having a lower cognitive level, such as the robot Leonardo (figure 6), the videos of interactions between Leonardo and a human interlocutor reveal that the human will also start talking baby talk. This seems to not have been studied any further since the interpretation of Leonardo is not the focus of the Leonardo project.

An initial experiment

As a prerequisite for this study we have conducted an experiment to figure out which gestures are produced when seeking attention and which gestures are perceived as attention seeking. In the literature on gestures not much has been written about the initial attention seeking gestures, that is, the gestures a subject uses when the interlocutor is not yet paying attention and ready for engaging in a conversation. Instead, research has mainly focused on attention during the interaction; how much attention do we pay to the gestures of the speaker (Gullberg & Holmqvist 2006), how does a deaf mother get the attention of her hearing child (Swisher 1999), and how parents get and maintain the attention of their child (Estigarribia & Clark 2007).

The design of our gestural experiment was intuitive and experimental. Although the experiment took place in a lab environment, we designed it to resemble a natural situation as far as possible. Our primary goal was to observe behaviour, not to modify or change behaviour. We were also aware of the extra linguistic cues the experiment leader accidently might give the subject if perceiving the gestures performed by two other persons in the room. To avoid the observer expectations effects of "Clever Hans" (Ladewig 2007), our chosen gestures were performed outside the experiment leader's field of vision.

Figure 7: Experiment condition for initial experiment

The gestures we chose for two of the experiment participants, gesturer 1 (G1) and gesturer 2 (G2), to perform were part of the intuitive design, although they were chosen to represent a wide scope of gestures.

In the experiment the subject (S) enters the room where an experiment leader (EL) is seated, occupied (presumably) listening to music while attending a paper. S is forced to use gestures in attempt to attract attention. After an appropriate amount of time EL responds to S and the two gesturers, G1 and G2, enter the room (figure 7). Then a repeating exercises begins while G1 and G2 starts gesturing following a specific protocol applied to every trial, and may or may not overlap each other in the execution. Figure 8 shows which gestures the subjects attended to, either by looking at the gesturer or by losing focus on the sentence he was to remember. Two linguists watched the recordings separately and noted whether a gesture was attended or not. If they did not agree on a gesture, the reaction of the subject was played repeatedly until agreement was reached.

Results for the attention getting gestures showed that some gestures were more likely to obtain the attention of the subjects than others. We define these gestures as *attention seeking gestures*, and these are shown in table 1.

However, the gestures performed might not be universal, because the appearance of the robot might also affect the gestures performed, perhaps both in shape as well as size. This is what we would like to examine in this extended experiment.

For our present experiment we examine two hypotheses:

H1: A difference in perceived cognitive level (lower for lo-fi looking; higher for hi-fi looking) in the robot causes a difference in attention seeking gestures from the interlocutor.

H2: Interlocutors to the lo-fi looking robot will rate their interaction more pleasant than interlocutors to the hi-fi looking robot, because according to the uncanny valley graph, a face will increase the experience of familiarity, and because the lo-fi look matches the function of the robot better.

This project is interesting for three main reasons. First of all, because not much research has been conducted on attention seeking behaviour; second, because directing attention is an important feature of a robot to be able to determine which speaker it should focus on in order to create successful interactions; and third, because an appropriate appearance could help enhance interaction, as well as an inappropriate appearance could cause the interaction to break down. Our vision has been to create a robot that is able to share its attention to multiple persons and that this attention will be shared in a most natural, human-like manner.

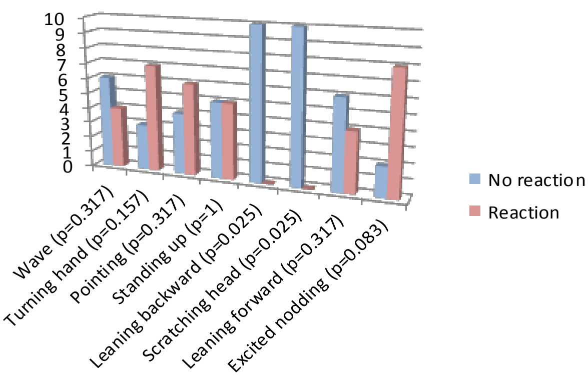
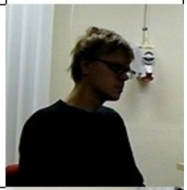
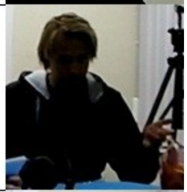
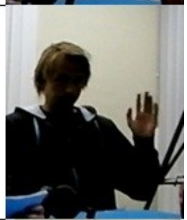
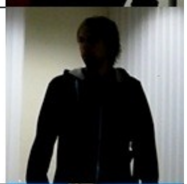


Figure 8: Attention getting gestures

Table 1: Attention seeking gestures

Gesture	Description	Significance level	Picture
Excited Nodding	Intense movement up and down with the head while looking at the interlocutor.	0.083	
Turning Hand	Circular movement with right or left hand ending with palm facing up.	0.157	
Waving	Vertical movement with right or left hand or fingers of right or left hand moving up and down. Both at face level.	0.317	
Pointing	Extended index finger directed towards an object or person supported by gaze in same direction.	0.317	
Leaning Forward	Upper body moving closer to the interlocutor decreasing the distance between the interlocutors.	0.317	
Standing Up	Change of posture from sitting down to upright position.	1.0	
Eye Contact	Continously looking into the eyes of the interlocutor.	No data	

THE ROBOT

The results of the gestural experiment showed that gestures performed by subjects trying to attract attention were scarce. Only one distinct gesture were observed and consisted of a waving hand in the visual field of the experiment leader to seek visual attention since auditive interaction was not accessible. This would suggest that visual attention and search for eye contact is important in seeking attention. The results from the prerequisite experiment also showed that excited nodding and standing up might be interpreted as attention seeking gestures, but we have decided not to implement these into the

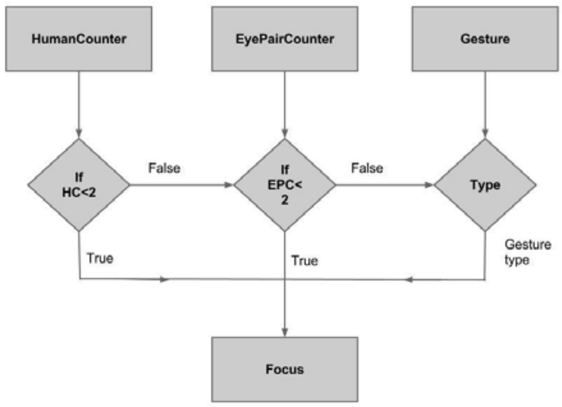


Figure 9: Conceptual design

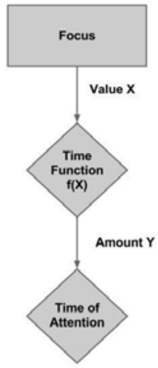


Figure 10: The Focus Node

robot, since the following experiment will only focus on subjects attempting to get the attention of the robot, which is why we expect that the subjects will perform turn taking gestures. Turn taking gestures are mostly represented as hand movements and by leaning forward.

The conclusions of the gestural experiment has been used in designing parameters for a robot to follow. Our goal is to program a robot to be able to register an attention seeking gesture and then turn its attention towards the person performing the gesture. The previous mentioned media equation allows us to implement our results from the experiment to the robot, since possible interlocutors to the robot will be using the same gestures trying to get attention from the robot as they use trying to get the attention of another human. Below we present in what order gestures are attended to more than others. This is the order in which the robot needs to prioritize which gesture to attend to initially as well as during an interaction with a human. Eye contact must at all times be present before reacting to a gesture.

Order for gesture reaction, most important first:

Turning hand / waving / pointing

Leaning forward

Conceptual design

The main idea is that the robot will read its environment by input from sensors and make a decision based on experimental data and logic. The decision will be made by working through a logical schematic. This process will result in a value, describing how much attention the robot is going to assign to a specific person present in its environment. Once an amount of time has been assigned, the process will be repeated. The main input will always be through the video streams materialized by the Kinect, although this data will be interpreted in several ways.

Figure 9 shows the conceptual design of the robot. At the top we get three inputs. *HumanCounter* keeps track of the number of persons present in the robots view, *EyePairCounter* is how many the number of eye pairs are present, *Gesture* will give the type of gesture that is currently being accounted for. These get processed in three decision boxes and becomes input for the *Focus* node at the bottom.

Figure 10 shows the logical schematic inside the *Focus* node. Based on the input to this node we will get a specific time of attention calculated by the *Time Function*.

Robot design

The robot was programmed mainly in C++ using the infrastructure known as the Ikaros-project (Balkenius et al. 2009). Ikaros is a project for creating an open platform library for system-level cognitive modeling of the brain. With this open

Figure 11: Web-gui part of Ikaros

source project it is possible to simulate some different parts of the human brain and these parts functionality, this is needed for setting up an experiment and simulate different functions and reactions of and towards human behaviour.

The system is built upon the idea that the human mind can be simulated with a lot of small interconnected modules, where each module has its own purpose and functionality. The system links all these different modules together and depending on which modules are used in the model the resulting code will behave accordingly. Other functionality built into the Ikaros project is other than modules simulating different parts of the human brain is modules for computer vision and modules for controlling external motors where we are going to use the Dynamixel AX-12+ a small yet strong robot servo.

All modules are written in C++ and for interconnection of modules, Ikaros uses a version of XML. The connections between different modules are all made in an .ikc file for the main program, listing all present connections needed for the data to flow between the modules currently in use. Each module also has its own .ikc file listing which inputs and outputs it needs to run its current code.

Figure 11 shows the web gui-part of Ikaros, the view which you can see here can quite easily be configured to show any desired feedback information from the robot. Here it can be seen how we currently use it for testing and debugging purposes. In this view we display the current states of the servo motors, both their current position, their desired position, temperature, and load information making sure we do not overload the motors, and a picture of the position of the servo motors inside the robot so that the information can easily be understood. We display also the video stream from our Kinect with a corsair and ring system for displaying faces. Eyes are extracted from the picture and displayed, a cropped versions of the face is also shown. The distorted looking image to the right is the depth stream that the Kinect also outputs. This we use to discard a large portion of the erroneous faces that the MPI Face Detector module gives.

Modules in use

Figure 12 shows the modules described below, and their connection.

- Kinect, to get visual input from the Kinect.
- MPIFaceDetector, to detect faces and eyes in the video stream from the Kinect (a wrapper for MPIEyeFinder and MPISearch).
- MarkerTracker, to follow fiducial markers with a BCH coded binary pattern in debugging purpose.
- Attention, our own module, for handling the different inputs and sending the correct output.
- Dynamixel, to control the three dynamixel AX-12+ servo motors in our robot.
- WebUI, only used during development for testing purposes.

Programming the robot

Our robot will use the modules of Ikaros described above as inputs and outputs to our own module. Our module describes how and why to add possible interlocutors to a list for later attention giving. The module's attention giving algorithm used by the robot is partly based on the results from our own experiment. Finally, the robot will decide which interlocutor to direct its attention to by pointing it's "head" towards the attended interlocutor. If the robot detects sudden differences in between the two depth matrixes concurrently as the color matrixes has the same values as an empty room at a coordinate where its possible for a hand to be located, the robot will detect this as a hand movement or other attention seeking gesture and yields attention towards this coordinate.

Inputs used to calculate whether anything important has happened in the view of the Kinect.

- Kinect depth and depth delayed by 2 ticks
- Kinect RED, GREEN and BLUE
- MarkerTracker markers - for debugging purpouse
- MPIFaceDetector - takes in "Faces" and their corresponding "size"

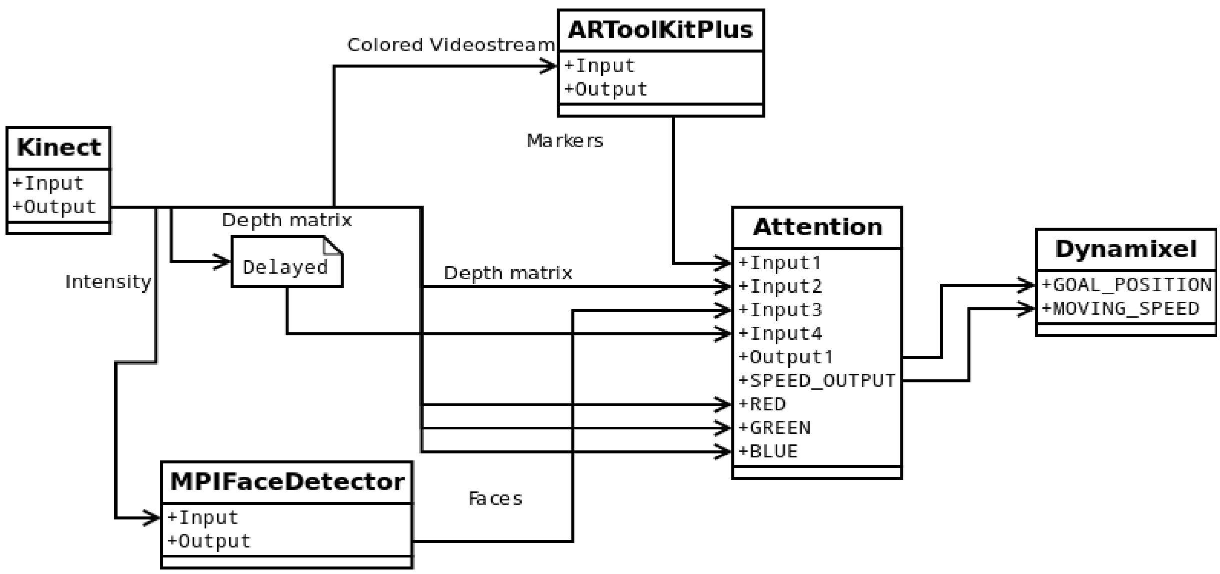


Figure 12: Module flow chart

Outputs for handling robot

- Attention OUTPUT1 - output_array the goal positions for the Dynamixel
- Attention OUTPUT2 - output_markers markers for debugging hand placement in WEBGUI
- Attention SPEED_OUTPUT - speed_array a float describing how fast the robot moves from curr_position to goal_position

They way that we recognize human interaction with the robot is by doing quite general calculations:

1. First loading in the background depth and RGB values of the room that the Kinect is facing.
2. Then we store newly discovered faces and their size for later handling, unless the discovered face already was present or was close to an already discovered face.
    - If we have faces present in storage then we write the position of the first discovered face to the dynamixel, so that the robot will point towards this interlocutor. Then the interlocutor is given a specific attention time from the robot.
    - If somebody in the room moves around in a manner that leads to the robot discovering it, the robot will break it’s current attention towards the current interlocutor switching to the newer interlocutor.
    - If there is no face present in the view of the Kinect the robot will lean slightly forward in a manner similar to a person reading a book thus not giving attention towards the room.
3. Now a person is assigned a specific amount of time and then the program goes back to step number 2.

EXPERIMENT

By conducting this experiment, we aim to examine the attitude toward a robot depending on appearance. This section describes the design, procedure and results of our experiment.

Experiment design

To examine the behaviour on different appearances we design the experiment to consist of two separate condition. In the first condition the robot is manifested by a lo-fi looking appearance, and in the second condition the robot is manifested in its original hi-fi looking appearance (figure 13).

Each condition was carried out in three trials. A total of 12 subjects was recruited; three sessions with two subjects for each condition.

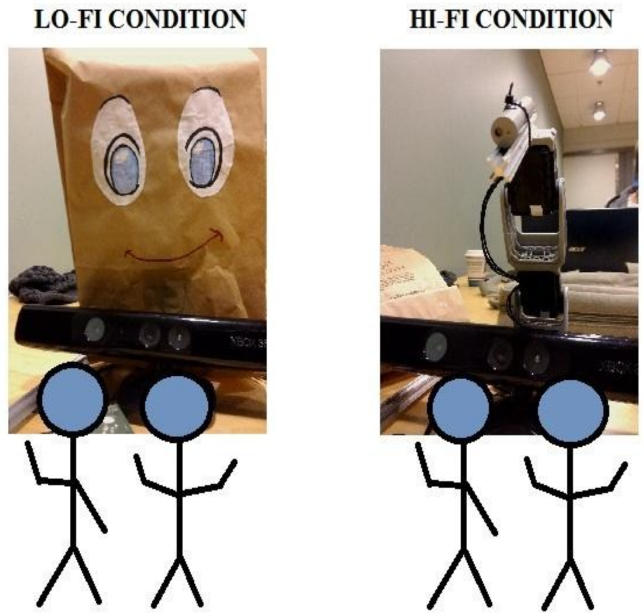


Figure 13: Lo-fi vs– Hi-fi conditions

Procedure

The procedure for the two conditions is identical; the two subjects were told to compete to get the attention of the robot. They were, however, not told what the robot would be attending to. Independent of condition (appearance) the robot reacts the same way; on faces (due to technical difficulties and time pressure, we unfortunately did not manage to make the robot respond to gestures). The trial was set to a limit of two minutes of interaction. After completed trial the subjects were asked to independently rate how they experienced interacting with the robot. Five rating questions consisting of five adjective pairs were rated on a scale of 1 to 7 (table 2). The subjects were also showed the alternative condition and asked which interface they would prefer to interact with. Subjects only participated in one of the two conditions.

The interactions in the two conditions were video recorded which, together with the questionnaire, makes it possible to examine how the subjects were behaving and at the same time how they were perceiving the two different robots.

RESULTS

The results of the experiment will be reviewed in two parts; the first part concerns the behaviour of the 12 subjects towards the two different interfaces of the robot. The second part involves the subjects’ rating of the robot after the sessions were completed.

Table 2: The scales of the five features

English: negative	Swedish: negative	English: positive	Swedish: positive
Inaccessible	Otillgänglig	Accessible	Tillgänglig
Mechanical	Mekanisk	Humanlike	Mänsklig
Unintelligent	Ointelligent	Intelligent	Intelligent
Simple	Enkel	Complex	Komplex
Unpleasant	Otrevlig	Pleasant	Trevlig

Figure 14: Lo-fi condition

Figure 15: Hi-fi condition

Competing for attention

During the experiment it quite quickly became clear that the most preferred strategy for obtaining the robot's attention - regardless of interface - was by waving. The wave differed both in size, speed, and distance, but was by far the most frequent attention seeking gestures, which is interesting when compared to our previous research on attention seeking behaviour towards a human. This will be debated in the Discussion section.

When looking at the two different condition, it did not seem to make any difference gesture wise. As mentioned, waving was the most frequent attention seeking gesture, but it is important to mention that no gestures were performed without looking directly at the robot. Therefore, we can again conclude that eye contact is crucial in attention seeking behaviour.

Regardless of interface it seems that the four most frequent behaviours towards the robot is eye contact, wave, leaning forward, and vocal cues. Figure 14 shows the most common behaviours in the lo-fi condition and figure 15 shows the most frequent behaviour in the hi-fi condition.

The most noticeable difference occurred in the vocal cues in the two conditions. In picture 8 subject3 (the girl in the grey shirt) were kissing at the robot, whereas subject8 (the man in the grey sweater) in figure 15 were giving the robot orders (e.g. "hello", "up here", "kom hit" trans.: 'come here'). Though this might seem to relate to perceiving a different cognitive level of the two robot interfaces, it is not only in the hi-fi condition that our subjects give orders to the robot. Subject12 orders the lo-fi looking robot to look at him (slams hand in table, plus "här" trans.: 'here'), and whistles are produced by Subject4 (lo-fi condition), and Subject6 and Subject7 (hi-fi condition).

Figure 16: Game gesture

It is interesting to see how some movements are based on the previous experience of the subjects with interactive games such as Kinect-based games and Wii-based games. In figure 16 Subject2 (man in white shirt) is producing a hitting gesture which cannot be perceived as an ordinary attention seeking gesture, rather it is clearly inspired by e.g. a tennis game.

It is also very noticeable that the two subjects in a sessions seem to be mimicking each other and therefore producing the same gestures.

Figure 17 shows four examples where Subject9 and Subject10 are mimicking each other's gestures by clapping, drumming on the table, waving directly in front of the Kinect, and by performing large, slow movements while leaned back. This could be a result of a competitive behaviour ("*if it works for you, it must also work for me*") or it could simply be a result of ordinary human interactions which often involves mimicking of one's interlocutor. If the latter is the case, we can argue that the subjects are behaving quite natural in this setting and therefore can use the data to draw conclusions from. However, the only conclusion to be drawn is that none of the six sessions showed any interface specific behaviour.

Figure 17: Mimicking gestures

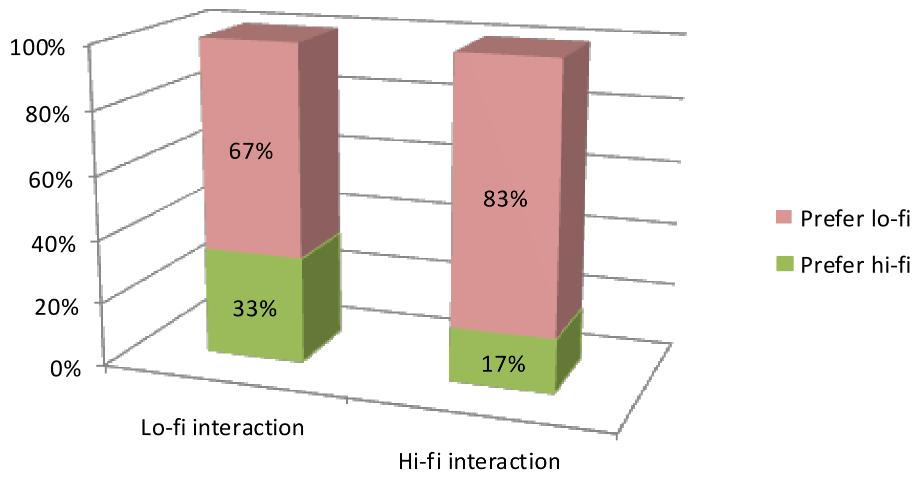


Figure 18: Preferred interface

Rating robots

The experiment showed that the robot most preferred is the lo-fi looking robot. Independent of condition the majority of subjects said they preferred the lo-fi, even more often in the hi-fi condition than in the lo-fi. Figure 18 shows the division between the two conditions.

For examining the perception of the robot we used five parameters (at our personal choice) previously used to rate human to robot interactions (Kanda et al. 2005). Parameters were shown on a scale rating from 1 (negative adjective to the left) to 7 (positive adjective to the right) and involved accessibility, humanness, intellect, complexity, and pleasantness.

Figure 19 mainly shows two things: the average rating of the robot in the interactions *independent* of preference, and secondly average rating of the robot *dependent* on preference. From this we can see that the lo-fi seems to be the most accessible as well as most humanlike. Rating the perceived intelligence it seems that the hi-fi is rated more intelligent regardless of preference. Perhaps the most interesting findings can be seen in the ratings of "complex" and "pleasant". On average, lo-fi was rated more complex, but when hi-fi is preferred (this group consists of only one subject) it is rated higher than the average in both conditions. Could this be due to a choice determiner? This is an interesting question we get back to in the section for analysis. Lo-fi is clearly the most pleasant robot to interact with, on average it is rated more than twice as pleasant as the hi-fi, three times as high concerning preferences.

Now we turn to look at the preference dependent ratings. Figure 20 confirms what we have already seen; lo-fi seems more accessible, although the difference is not convincing. Lo -fi is rated more humanlike by the group preferring hi-fi, and regardless of the condition the robot is rated more intelligent when not preferred. Complexity is rated highest for the those preferring hi-fi in both conditions, whereas lo-fi is rated more pleasant, regardless of preference.

ANALYSIS

In analyzing the results we examine whether our two hypotheses are supported or not. The analysis shows that only the second hypothesis seems to be supported.

Testing hypothesis 1

Our first hypothesis concerns whether a difference in perceived cognitive level in the robot causes a difference in attention seeking gestures from the interlocutor. We expected that the lo-fi looking robot would be perceived as having a lower cognitive level based on its humanlike features, whereas we expected that the hi-fi robot would be perceived as having an equal cognitive level with the interlocutor. In both cases we expected that this perception would lead to a distinction

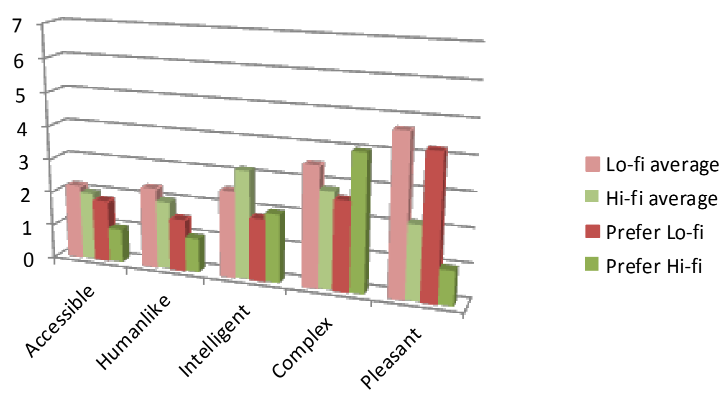


Figure 19: Preference and feature rating average

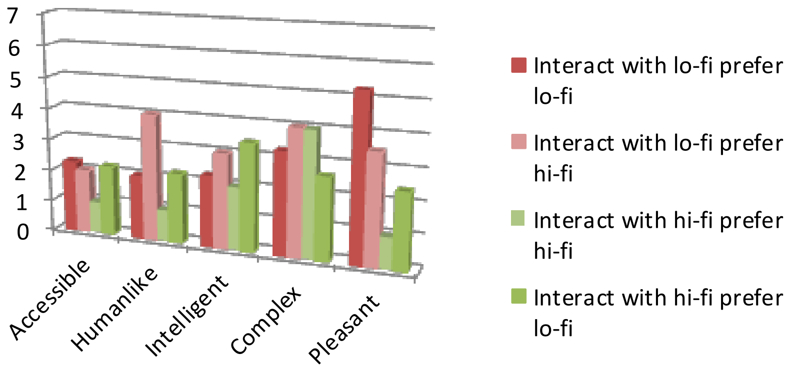


Figure 20: Feature rating based on preference

of the gestures performed towards the robot.

However, this was not what we found. It seems that our subjects did not behave differently towards the different interfaces of the robot. We have described how the most noticeable difference occurred in the vocal cues, but that these differences were more of degree than kind. Therefore, we can conclude that our first hypothesis is not supported.

On the other hand, we experienced some troubles running the robot (see section Evaluating the Robot) which led to inconsistent and sparse feedback from the robot, which made it difficult for the subjects to infer to what the robot might react to. This problem occurred in both conditions and therefore might have led the subjects to respond to both interfaces as if it were having a lower cognitive level. Further studies with two well-functioning robots will have to be conducted before a final conclusion on the matter can be reached.

Testing hypothesis 2

The second hypothesis examines whether interlocutors rate their interaction with the lo-fi looking robot more pleasant than interlocutors rate their interaction with the hi-fi-looking robot. This hypothesis seems to be supported according to our data. Regardless of preference, people tend to rate the lo-fi looking robot more pleasant than the hi-fi robot. Conducting a t-test, we find that this is an almost significant result (table 3).

Not only is the lo-fi looking robot rated highest in being pleasant, it is also the most preferred appearance in both conditions (75% or 9 out of 12 subjects). In figure 19 we could see that the average for lo-fi as well as preference for lo-fi was about the same rate concerning pleasantness, outperforming the other qualities. Based on this we can conclude that being pleasant seems to be the most important trait (of the five) in preferring a (at least a lo-fi-looking) robot.

EVALUATING THE ROBOT

Throughout this project we have been forced to make many decisions of what path to take in the development of the robot. These decision are all results of a realisation of our own limitations and/or the limitations of the technique used for this project.

We started out focusing on programming the robot to successfully be able to register when a possible interlocutor were looking at it. We felt that this task was the most important, since the findings in our experiment showed us that no attention seeking behaviour would occur without eye contact. However, programming the robot to detect when a face is looking at it (simply by tracking faces) proved to be rather difficult since the existing face detector of Ikaros was not accurate enough. This forced us to spend a great amount of time designing a module which could out filter the erroneous faces.

What we have now is a robot that can fairly well identify and count faces in a stressed environment, with many “face like” shapes and colors. The robot still has problems with incorrect face recognition, which occurs when MPIFaceDetector gives an erroneous face due to glare somewhere coupled with dark patterns or a checkered shirt, or by any other particular face resembling patterns in the environment. This erroneous face’s position can then later be accepted as an valid face position by the current use of the Kinect’s depth and RGB stream by differentiating the object interacting with the robot from the surrounding environment when analyzing the streams. What we also have is the quite unstable functionality of the Robot noticing waving gestures and focus attention accordingly.

Technical analysis

Our initial plan was to implement the system following the conceptual design very closely. However this was proven difficult in practice. The face detector module is pretty exact counting faces against a plain background. However when the Kinect is placed in a room with more depth behind the body, the module has problems telling a face from a glare or a field of white. We solved this by filtering out all the faces that has

Table 3: Average rating

	Accessible	Humanlike	Intelligent	Complex	Pleasant
Lo-fi	2,167	2,333	2,500	3,500	4,667
Hi-fi	2,000	2,000	3,167	2,833	2,167
P	0,383	0,232	0,089	0,229	0,006

an almost constant depth. This worked well and we got a number of faces that corresponded better to the actual value. When wearing clothes with a complex pattern the system still has problems telling a face from an area of a shirt with white squares. This problem does not get solved with the filter since the clothes a person is wearing follows when the person is moving and thereby the depth is changing mitigating our filter solution.

The input data to Attention from MPIFaceDetector and the depth sensor was imprecise due to the short distance to the subjects. When doing fast and vivid gestures the Kinect cannot keep up and the result will be that the data contains too many errors in order to assign attention in the way intended.

The Kinect inputs which we use in our Attention to execute to logic relies on exact data. The input is a 640x480 matrix with "pixel elements" that each has a float value that represents intensity, depth or color. When the test subjects continuously moves forward, backward while waving a few centimeters from the camera and depth sensor, the resulting input matrix will contain a 640x480 pixel elements which in this context will be a total disoriented mess. Hence no logic in the code however good it might be will work as intended.

We understand from this experiment that we cannot always rely on 100% perfect input data. We will have to rework the code and make assumptions based on limited information.

When the robot turns straight towards a face and a person, the dynamixel module gets a correct input from Attention. The input is computed by multiplying a float x and y axis value from MPIFaceDetector with a suitable integer value that added with an offset gives an integer value that can cover the whole Kinect´s field of sight. Since the offset value is constant the precision of which the robot can turn towards a person varies with the distance from the Kinect, in our case the robot works best at around 60 cm.

In our experiment all subjects tended to lean forward, thus decreasing the distance to the Kinect. This took down the robots precision with the offset a bit. In addition to this the position of a potential face registered in the MPIFaceDetector was most likely almost random given the excited and active behavior of the subjects. There were cases when the robot responded to the subject. Although as mentioned earlier, it only occurred when the subjects were relatively calm and held their distance to the camera. In that case there were images good enough to analyze and possible to make somewhat accurate readings.

Would another experiment be conducted again some measures could be taken to improve the results. The distance from the Kinect to the test subjects should be fixed to the optimal and desired distance, to prevent any unnecessary disturbances. The computational code analyzing the input from MPIFaceDetector could be improved to more easily recognize number of simultaneous faces, faces in non-optimal environments. The for analyzing and recognizing gestures could be improved in ways of precision and an increasing number of simultaneous number of gestures recognized. If the implemented code is not desirable for use in a test scenario, then another way of controlling the robot should be used as it more than often confuses the test subjects. The final solution to control the robot, if no implemented code works as intended, manually in a wizard of oz-technique.

Final product - A functioning robot in a busy environment

Our vision is that when the robot is finished it will be able to share its attention to multiple persons and that this attention will be shared in a most natural manner. When the robot decides which interlocutor to pay attention to and how long the attention span should last for this interlocutor, it will take many variables into consideration leading to the robot sharing its attention in a human-like manner. If the robot is situated in a crowded room with, for the robot, unknown persons leaving and entering, the robot should not encounter any problems keeping on with its task, that of distributing attention in a human-like manner. Our focus lies in that the persons that are present in the robot's environment have to perform a gesture in order to obtain the attention of the robot. So, if no one is present in the view of the robot and therefore no attempts of getting the robot's attention exist, the robot should appear to focus elsewhere. We program the robot to lower its "head" slightly down in a seemingly contemplating manner.

DISCUSSION

Appearance seems to matter. Although it did not seem to affect the gestures performed by the subjects in our two different conditions. In this section we discuss our findings further, compared to theories presented in the introduction.

When we examined attention seeking behaviours towards humans it seemed that the most common strategy was to simply wait for the desired interlocutor to pay attention. However, when interaction with the robot the interlocutors were much more eager and insisting in seeking attention. This, of course, might be due to the competition we asked the subjects to participate in, that is, get the robot's attention before the other interlocutor. We strongly felt that it was necessary to conduct the experiment as a competition in order to maintain the interest towards the robot in the subjects, though this might have an effect on the behaviours.

We also believe that the confusing and unfinished feedback from the robot were a very confusing and frustrating factor for the subjects, and that this might have caused a perceived lower cognitive level in both conditions which resulted in no great difference in attention seeking gestures performed.

During the sessions it was also detected how the subjects tended to mimic each other, sometimes even cooperating, perhaps because they in some cases seemed to be very good friends. If we had mixed the subjects so that they were competing with someone they did not know, we might have been able to control the cooperating, and perhaps even the eagerness, in the interactions.

Turning now to the perception of the robot, we have concluded that if the robot is perceived as pleasant it will more frequently be preferred to e.g. a more complex looking robot in these kinds of social interactions. From our data we have also found other interesting results concerning how a robot is perceived in an interaction.

In figure 20 we saw that hi-fi was rated higher in intelligence even when the subjects preferred the lo-fi. At first glance this causes us to conclude that the subjects prefer a less intelligent robot. Perhaps this is the case, or perhaps a more plausible explanation is that this is an example of the subjects searching for a match between appearance and performance. It could be that the hi-fi robot looks more intelligent than it turned out to behave like, and therefore the subjects would prefer a lo-fi looking robot.

Only one subject who interacted with the hi-fi also preferred this appearance, which is clearly a group too small to analyse, but it still poses an interesting question as mentioned by figure 18. It might be that this particular subject has an expectation of robots being complex, and when experiencing the robot to be sufficiently complex, then this is the robot of preferred. If so, this is an interesting matter on how expectations determine all humans' choices.

The lo-fi looking robot is also rated higher on humanness, however by subjects preferring the hi-fi (see figure 20). This might be interpreted as the lo-fi looking is too humanlike compared to its behavior. It might also be that - though just painted on paper - a pair of eyes and a mouth gives the interlocutor expectations of more human behaviour than our lo-fi looking robot could manage. This could be a source of disappointment and a reason not to prefer this appearance. As concerning the ratings on intelligence above, this could relate to previous research that a match between expectations and actual behaviour is more important than just a hi-fi or very humanlike appearance.

When asked which appearance he preferred, one subject commented the role of context. If the situation was another than this, say at a hospital where he was a patient who was showed the robot which were to perform his surgery, he would prefer the hi-fi looking robot over the lo-fi. This comment is in line with the human tendency to want the robot to look as complex or simple as it is competent. This is a plausible explanation to why a majority would prefer the lo-fi robot. In our experiment the activity is sitting down, trying to engage in a social interaction, and for this humans want two eyes and a mouth more than a mechanical construction. The role of context and task is something we have not examined, but it has become clear to us that is an important factor to investigate before a further development of any social robot.

Regarding the programming of the robot, the experiment clearly showed that in order to engage in a social interaction using a Kinect many factors will need to be controlled, such as the subjects' distance to the camera of the Kinect, and it might be preferable to place the Kinect behind some kind of barrier so that the subject will not lean in towards the Kinect, obscuring it's view of the room.

We can conclude that programming a human like robot is a time consuming task. To use the Kinect and the different modules was in the beginning a quick way to start and to get something useful fast. However when getting further into the project and adding more functionality to the robot, the complexity of the problem skyrockets. Many problems are fixable with plain coding and iteration with different solutions, but for more dynamical and stable solutions advanced mathematical models and image processing are probably needed.

If we would redo the whole project from the start, several things would be done differently. The problem would be researched more thoroughly in the beginning. This way our expectations would have been more reality based and we could have aimed for more functionality of the robot with non dynamical solutions since this has proven to be very difficult. With more functionality in the robot the test results from our second experiment would probably be more valuable, improving the project as a whole with less of a gap between the technical and theoretical parts.

REFERENCES

Balkenius, C, Morén, J, Johansson, B, & Johnsson M (2009): Ikaros: Building cognitive models for robots. In *Advanced Engineering Informatics*, Volume 24, Issue 1, January 2010, pages 40-48.

Caporael, L. R., M. P. Lukaszewski, & G. H. Culbertson (1983): Secondary Baby Talk: Judgments by Institutionalized Elderly and Their Caregivers. In *Journal of Personality and Social Psychology*, Vol. 44, No. 4, 746-754.

Estigarribia, B. & E.V. Clark (2007): Getting and maintaining attention in talk to young children. In *J. Child Lang.*34 (2007), 799–814. Cambridge University Press, United Kingdom.

Goetz, J., S. Kiesler, & A. Powers (2003): Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation, Proceedings of the 2003 *IEEE international Workshop on Robot and Human Interactive Communication*, Millbrae. Caiibmia. USA, Oct. 31 - Nov. 2.

Guizzo, E. (2010): Who's Afraid of the Uncanny Valley? Blog on http://spectrum.ieee.org, posted 04.02.2010, seen 11.05.2012.

Gullberg, M. & K. Holmqvist (2006): What speakers do and what addressees look at. In *Pragmatics & Cognition* 14:1, 53–82. John Benjamins Publishing Company.

Han, B. S., H. Y. W. Alvin, Y. K. Tan, & H. Li (2010): Using Design Methodology to Enhance Interaction for a Robotic Receptionist, Institute for Infocomm Research, 1 Fusionopolis Way, #21-01, Connexis, Singapore 138632, 19th *IEEE International Symposium on Robot and Human Interactive Communication*, Principe di Piemonte - Viareggio, Italy, Sept. 12-15.

Johnson, D., J. Gardner & J. Wiles (2004): Experience as a Moderator of the Media Equation: The Impact of Flattery and Praise. In *International Journal of Human-Computer studies* 61(3), 237-258.

Kanda, T., T. Miyashita, T. Osada, Y. Haikawa, & H. Ishiguro (2005): Analysis of humanoid appearances in human-robot interaction, *Intelligent Robots and Systems*, 2005. (IROS 2005).

Khan, Z. (1998) Attitudes towards intelligent service robots, *Technical Report* No. TRITA-NA-P9821, NADA, KTH, Stockholm, Sweden.

Ladewig, J. (2007): Clever Hans is still whinnying with us. In *Behavioural Processes*, Volume 76, Issue 1, September 2007, pages 20–21.

Mitchell, R. W. (2001): Americans' Talk to Dogs: Similarities and Differences With Talk to Infants. In *Research on Language & Social Interaction*, 34:2, 183-210

Nomura, T., T. Kanda, T. Suzuki, & K. Kato (2005): People's Assumptions about Robots:

Investigation of Their Relationships with Attitudes and Emotions toward Robots, *IEEE International Workshop on Robots and Human Interactive Communication*.

O'Connor, B. P., & H. Rigby (1996): *P*erceptions of Baby Talk, Frequency of Receiving Baby Talk, and Self-Esteem Among Community and Nursing Home Residents. In *Psychology and Aging*, Voril, No. 1,147- 154

Robins, B., K. Dautenhahn, R. te Boekhorst, & A. Billard (2004): Robots as Assistive Technology - Does Appearance Matter?. Proceedings of the 2004 *IEEE International Workshop on Robot and Human Interactive Communication* Kurashiki, Okayama Japan.

Swisher, M.V. (1999): Learning to converse: How deaf mothers support the development of attention and conversational skills in their young deaf children. In *The Deaf Child in the Family and at School: Essays in Honor of Kathryn P. Meadow-Orlans.* Taylor & Francis.

Syrdal, D. S., K. Dautenhahn, S., Woods, M. Walters, & K. L. Koay (2007): Looking Good? Appearance Preferences and Robot Personality Inferences at Zero Acquaintance. *Multidisciplinary Collaboration for Socially Assistive Robotics: Papers from the AAAI Spring Symposium -Technical Report* SS-07-07: 86-92.

Thieberger Ben-Haim, A., G. Thieberger, T. Thieberger, & S. Thieberger (2009): Infant photo to improve infant-directed speech recordings. *Provisional application* No. 61/035.027. Kiryat Tivon, US.

Walters, M. L., K. Dautenhahn, R. te Boekhorst, & K. L. Koay (2007): Exploring the design space of robot appearance and behavior in an attention-seeking 'living room' scenario for a robot companion. In *Proceedings of IEEE artificial life* (Alife 07) (pp. 341–347), Honolulu, Hawaii, USA.

Appendix

ID ____________ Insp __________ Lo-fi Hi-fi

Interaktion mellan människa och robot

Fråga 1: Hur upplevde du roboten som du nyss interagerade med?

Svara genom att ringa in en siffra 1-7 för varje motsatspar.

OTILLGÄNGLIG TILLGÄNGLIG

1 2 3 4 5 6 7

MEKANISK MÄNSKLIG

1 2 3 4 5 6 7

OINTELLIGENT INTELLIGENT

1 2 3 4 5 6 7

ENKEL KOMPLEX

1 2 3 4 5 6 7

OTREVLIG TREVLIG

1 2 3 4 5 6 7

Fråga 2: Hade du hellre interagerat med den här roboten?

Interagerat med Lo-fi blir visad Hi-fi:

JA NEJ

Interagerat med Hi-fi blir visad Lo-fi:

JA NEJ

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Automatic Generous/Defence Alternator

Carl Bjerggard Tarik Hadzovic Johan Hallberg Olle Klang

Grunden i projektet med rubriken ”Imitation av målinriktade rörelser” är att konstruera ett system som läser av målet med en persons rörelser med en Kinectkamera och sedan utföra samma rörelse med hjälp av en robotarm, t.ex. flytta en kloss. Vi har utgått från grundidén men vidareutvecklat den rejält. Istället för att robotarmen härmar en rörelse vill vi istället att den ska skilja på två rörelser och agera på ett visst sätt, baserat på vilken av dessa två rörelser personen framför utför. Fokus har lagt på två rörelser; en hotfull och en vänlig rörelse. Vår tes är att hotfulla rörelser är hastiga medan vänliga rörelser är långsamma. I detta projekt har vi studerat skillnaden mellan dessa två rörelser i ett användartest. Denna skillnad har vi sedan implementerat i vårt system. Vårt system kallar vi AGDA. AGDA är en höna som känner igen om personen framför är hotfull eller vänlig och agerar utifrån det. Vi har med andra ord skapat ett system för igenkänning av gester/kroppsrörelser.

1 Introduction

I kursen MAMN10 har samtliga studenter tilldelats ett projekt inom ämnet Neuromodellering- kognitiv robotik och agenter. Projektgrupperna består generellt av både teknologer och kognitionsvetare men vår grupp utgörs enbart av teknologer. Detta anser vi har gett oss ett visst tekniskt försprång under projektet medan nackdelen är att den kognitiva biten i projektet kommit i andra hand.

Idéerna kring vad det är vi vill uppnå med projektet har varit många och vår tekniska kunskap har gjort att utvecklingen kunnat fortgå enligt den tidsplan vi tidigt lagt fram i projektet. Efter att ha studerat uppgiften noga har vi tillslut landat i en idé som vi anser är intressant med avseende på både den tekniska och kognitiva biten, vilken beskrivs mer ingående nedan.

I fortsättningskursen MAMN15 var tanken att vi skulle fortsätta utveckla vår lösning. Det huvudsakliga arbetet i denna kurs har handlat om att undersöka våra två rörelser och bygga en teori kring dessa. Vi har utfört ett enklare användartest där vi undersökte hur hotfulla och vänliga rörelser skiljer sig åt.

1.1 Bakgrund

Projektet vi tilldelats är ”Imitation av målinriktade rörelser”. Grunden i projektet är att konstruera ett system som läser av målet med en persons rörelser med en Kinectkamera och sedan utföra samma rörelse med hjälp av en robotarm, t.ex. flytta en kloss [1].

Vi har utgått från grundidén men vidareutvecklat den rejält. Istället för att robotarmen härmar en rörelse vill vi istället att den ska skilja på två rörelser och agera på ett visst sätt, baserat på vilken av dessa två rörelser personen framför utför. De två rörelserna och vår slutgiltiga idé beskrivs senare i detalj.

Det intressanta med vår approach är att undersöka hur mycket av en verklig situation man kan simulera med hjälp av en robot och ett Kinect men vår lösning har ett väldigt mycket större användningsområde. Vi har med hjälp av den information man får från Kinecten lyckats mäta hastigheten med vilken personen framför Kinecten rör sig samt avståndet till samma person. Detta anser vi kan användas inom flera områden där det är av intresse att läsa av och tolka målet med en persons kroppsrörelse/kroppsspråk. Man kan tänka sig att systemet kan användas i avancerade övervakningssystem där man vill att ett system ska känna igen, tolka och agera utifrån en persons kroppsrörelser. Detta användningsområde är tämligen nytt och behandlas i ett examensarbete vi tagit del av [2].

Under ett av de första mötena vi hade i gruppen studerade vi ett tredjepartsprogram som heter FAAST och som utvecklats av University of Southern California [3]. Med FAAST kan man identifiera gester (kroppsrörelser) och koppla dessa till tangentbordsinmatningar. Vår lösning är med andra ord väldigt lik FAAST men inte lika avancerad.

2 Metod

Tidigt i projektet har mycket tid ägnats åt research och vi har satt oss väldigt noga in i IKAROS mjukvara och gjort en del litteraturstudier. Att sätta sig in i mjukvaran har tagit en hel del tid då vi tidigt velat försäkra oss om att vi har ett bra grepp om mjukvarans struktur för att på så sätt kunna implementera en mer avancerad lösning.

Vi har haft väldigt lite att gå på i början och då dokumentationen av IKAROS är knapphändig har vi utforskat mjukvaran genom att implementera mindre avancerade lösningar, vilka senare utgjort en bas för den mer avancerade lösningen vi implementerat. På så sätt har vår slutgiltiga lösning kommit att bli en iteration av de första enklare lösningarna.

Ett första steg har därför varit att skriva en enkel modul som tillåter användare att styra roboten höger och vänster med hjälp av en kloss. Klossen har flera koder som läses in och identifieras av Kinecten och två olika koder används där den ena motsvarar en högerrörelse och den andra en vänsterrörelse. Denna tidiga lösning har gett oss en djupare förståelse för hur vi med hjälp av klossar kan få roboten att utföra vissa enkla rörelser.

Nästa iteration har gått ut på att känna igen klossen och låta roboten följa efter den samma. Med andra ord har vi förflyttat klossen fritt i rummet – höger, vänster, upp och ner – och fått roboten till att noga följa denna rörelse. Efter att denna lösning implementerats bestämde vi oss vad vi ville uppnå med vårt projekt.

Vår slutgiltiga idé kräver att man är väldigt insatt i den information Kinectsensorn hämtar in. Nästa steg har varit att utforska den modul som styr Kinecten och att försöka bearbeta den data denna genererar [4][5]. Detta arbete har varit väldigt tidsödande men i slutändan väldigt givande. Data som är av intresse är djuphetsinformation men även ansiktstracking/igenkänning. När vi väl satt oss in i hur detta fungerar har utvecklingen kunnat fortgå utan att vi stött på några större dead-ends. Detta innebär att vi kunnat implementera en tredje enklare lösning där roboten känner igen ett ansikte och följer efter det. Med hjälp av djuphetsinformationen har vi kunnat göra denna rörelse väldigt precis och på så sätt fått fram avstånd till användaren och positionen för var användarens ansikte befinner sig i förhållande till Kinecten (och därmed roboten).

Parallellt med den tekniska utvecklingen har vi ägnat oss åt att bygga teorin kring de två rörelser som är av intresse för oss och vår robot. Framför allt har stora delar av fortsättningskursen ägnats åt detta. De två rörelserna består av en vänlig rörelse riktad mot roboten samt en hotfull rörelse som får roboten att skydda sig själv och dess tillhörigheter. Eftersom rörelserna utgår ifrån våra egna tankar och preferenser har vi inte kunnat hitta någon teori som styrker våra påståenden. I ett användartest har vi studerat dels hastigheten med vilken rörelserna utförs, dels om det finns fler parametrar, förutom avstånd och hastighet, som man bör ta hänsyn till. Användartestet och resultatet från det samma beskrivs i detalj nedan.

Den sista iterationen i vårt tekniska arbete kallar vi AGDA v2.0 vilken föreställer det vi arbetat med i fortsättningskursen. I slutet av första kursen fastslog vi att det vore bättre om vår robot kunde hålla reda på flera personer. I AGDA v2.0 finns detta implementerat. AGDA kan nu hålla reda på upp till tio personer. Den som är av intresse, dvs. den person som står närmast AGDA och/eller har störst aktivitet/hastighet, är den roboten kommer följa med blicken.

2.1 AGDA: Automatic Generous/Defence Alternator

Samtliga iterationer i vårt arbete och det slutgiltiga målet med vårt projekt har kommit att kallas AGDA [6]. AGDA ska föreställa en höna med ett ägg som den är väldigt mån om att beskydda.

Hönan AGDA (roboten) följer den person som ställer sig framför henne med blicken och framför sig har hon sitt ägg. Roboten ska reagera på hotfulla rörelser (personen vill ha tag på hennes ägg) genom att dra tillbaka ägget och skydda det. Genom att vara vänlig kan man övertyga AGDA om att istället räcka fram ägget.

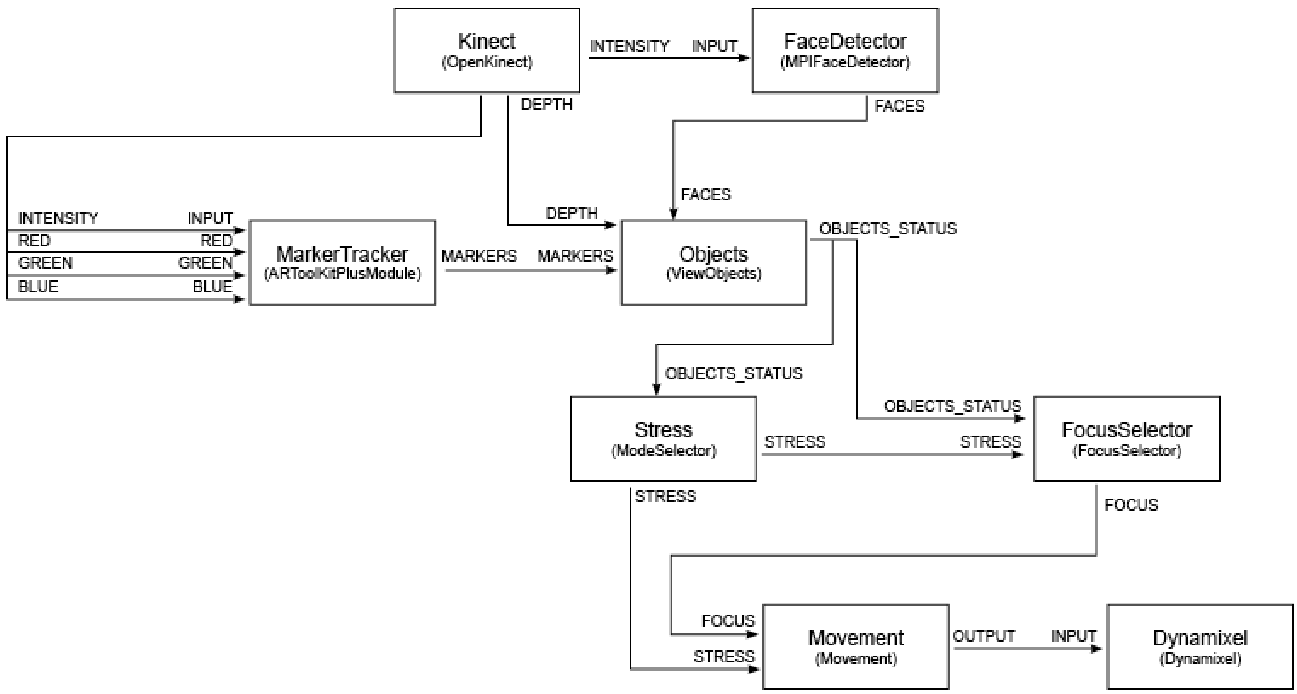
Det framgår ovan vilka två rörelser som utgör grunden för den interaktion som vi tänkt implementera. Vårt antagande är att den största skillnaden mellan dessa två rörelser är hastigheten med vilken de utförs. En hotfull rörelse kännetecknas av en hastig rörelse, där en person så snabbt som möjligt försöker tillskansa sig något som inte tillhör denne. En vänlig rörelse är i sin tur sådan att personen ifråga istället rör sig försiktigt och sakta framåt utan några som helst hastiga rörelser.

3 Resultat

För att skapa AGDA och implementera de rörelser hon ska utföra har det varit nödvändigt att skriva ett flertal moduler. Dessa moduler är sammankopplade genom olika inputs och outputs, dels med varandra men också med moduler som finns fördefinierade i IKAROS.

Modulen Dynamixel är den modul som gör att robotens olika servos ställer sig i olika lägen/positioner [7] och våra moduler utför en hel del beräkningar innan de skickar den valda positionen roboten ska befinna sig i till Dynamixel.

En annan av de fördefinierade modulerna som är viktiga för AGDA är ”FaceDetector”. Med hjälp av FaceDetector tar vi reda på om det finns ett ansikte framför roboten (Kinectkameran) och var i bilden det befinner sig. Denna information skickas till våra moduler där beräkningar görs.



Figur 1: Modulkarta

Nedan följer en detaljerad beskrivning av de fördefinierade moduler vi har ändrat i samt de moduler som vi själva skapat. Modulkartan i figur 1 föreställer hur vår implementation såg ut i slutet av kursen MAMN10. I stycket 3.7 AGDA v2.0 beskrivs vilka tekniska finesser som tillkommit i fortsättningskursen MAMN15.

Figur 1 föreställer en modulkarta där kopplingen mellan våra moduler samt FaceDetector och Dynamixel åskådliggörs. En fullständig modulkarta finns i Appendix A.

3.1 Ändringar i fördefinierade moduler

I modulen OpenKinect har vi ändrat så att outputmatrisen DEPTH innehåller det verkliga avståndet till ett objekt i cm för alla pixlar i kamerabilden. Då vi utifrån information på nätet [8][9][10] kom fram till att djuphetsmatrisen i själva verket innehåller dispariteten mellan kamerans IR projektor och IR sensor beräknar vi det verkliga avståndet till objektet i bilden med formeln

$$z = (b * f) / d,$$

där z är djupet i cm, b är avståndet mellan projektor och sensor i cm, f är den gemensamma fokallängden i pixlar och d är den normaliserade dispariteten i pixlar. Dispariteten normaliseras med formeln

$$d = 1/8 * (doffset - kd),$$

där doffset är ett specifikt offsetvärde för Kinectenheten, kd är dispariteten från Kinectenheten och 1/8 är en skalfaktor pga. att kd anges i enheten 1/8 pixlar.

Den slutliga formeln blir enligt ovanstående källor alltså

$$z = (7.5 * 580) / ((1/8) * (1091.5 - kd)).$$

3.2 ObjectModule

ObjectModule är den modul som med hjälp av ansiktenas x, y och z position beräknar deras förflyttning i x, y och z led. Detta används sedan för att beräkna hastigheterna i samma led samt aktivitetsindex för de aktuella objekten.

Input Input till ObjectModule är den korrigerade djuphetsmatrisen FACE_DISTANCE från modulen KINECT samt matrisen FACES från modulen FaceDetector innehållande ansiktenas positioner i kamerabilden samt deras storlek i procent av bildstorleken.

Output Output från ObjectModule är matrisen OBJECTS som innehåller de aktuella ansiktenas x och y positioner i bilden, deras avstånd till kameran i cm i z led, ansiktenas aktivitetsindex samt deras hastigheter i z led.

Strukturer ObjectModule innehåller de två strukturerna buffer och object_list. Strukturen buffer är en cirkulär buffert som innehåller de 20 senaste samplade x, y och z positionerna för varje aktuellt ansikte tillsammans med tidsstämpeln för varje sampling i mikrosekunder för varje ansikte.

Den andra strukturen object_list är en matris innehållande hastigheterna i x, y och z led för alla samplade ansikten tillsammans med det beräknade aktivitetsindexet för varje ansikte.

Funktioner

void tick()

Vid varje tick beräknas och sparas de verkliga positionerna för x och y, dvs. x_cm och y_cm, på platsen [object_index][current_node[object_index]] i bufferten. Dessutom sparas det verkliga avståndet till ansiktet z_cm ihop med tidsstämpeln t1 på samma plats i bufferten. Därefter anropas metoderna calculateVelocity() och calculateActivity(). Slutligen sparas ansiktenas x och y positioner i bilden, avstånd till dem i cm i z led, ansiktenas aktivitetsindex samt deras hastigheter i z led i output_object_matrix.

void calculateVelocity()

calculateVelocity() beräknar hastigheterna i x, y och z led genom att jämföra de aktuella x, y och z positionerna med de samplade positionerna två platser tillbaka i bufferten (buddy_node). Hastigheterna sparas sedan i x_velo, y_velo och z_velo i object_list.

void calculateActivity()

calculateActivity() beräknar ett aktivitetsindex för det aktuella ansiktet genom att ta absolutloppet för de olika hastigheterna multiplicerat med en skalfaktor, som för tillfället är 1 för alla hastigheterna.

void valid_velocity()

valid_velocity() kontrollerar om hastigheten för det aktuella ansiktet är giltig genom att beräkna absolutbeloppet för hastigheterna i x, y och z led och jämföra med maxhastigheterna i repsektive led. Om någon av hastigheterna överstiger dessa returnerar funktionen false, annars true.

3.3 ModeSelector

ModeSelector är den modul som med hjälp av parametrarna från ObjectModule avgör vilket mode, dvs. tillstånd, som AGDA ska befinna sig i. En beskrivning av AGDA:s fyra olika modes följer senare i rapporten.

Input Input till ModeSelector är matrisen OBJECTS från ObjectModule som innehåller de aktuella ansiktenas x och y positioner i bilden, deras avstånd till kameran i cm i z led, ansiktenas aktivitetsindex samt deras hastigheter i z led.

Output Output från ModeSelector är vektorn STRESS som innehåller det valda mode som AGDA ska anta härnäst.

Funktioner

void tick()

Vid varje tick beräknas först ett genomsnitt för alla ansikten med hjälp av funktionen CalculateFaceAverage(). Detta genomsnitt innehåller ett genomsnittligt avstånd till kameran i cm i z led, ett genomsnittligt aktivitetsindex samt en genomsnittlig hastighet i z led. Därefter valideras avståndet till det genomsnittliga ansiktet för att avgöra vilket mode som ska väljas. Om avståndet till ansiktet är mellan 130 och 350 cm väljs INTEREST_MODE och om det är mindre än 130 cm väljs GIVE_EGG_MODE istället. Om avståndet istället är större än 350 cm väljs IDLE_MODE. Det valda läget sparas i en temporär vektor för att sedan valideras ytterligare. För att jämna ut rörelserna och bytena mellan olika modes håller olika räknare reda på hur många ggr ett visst mode har valts. När räknaren för ett visst mode har ökats till 10 nollställs räknaren och det valda modet sparas ner i output vektorn.

void CalculateFaceAverage()

CalculateFaceAverage() kontrollerar först om det aktuella ansiktet är tillräckligt långt ifrån kameran. Om det är för nära förkastas ansiktets värden och annars kontrolleras om hastigheten är så pass hög att AGDA ska bli rädd. Om så är fallet multipliceras denna hastighet med 10 för att ge den ett högre inflytande än de lägre hastigheterna.

3.4 FocusSelector

FocusSelector är den modul som med hjälp av parametrarna från ObjectModule avgör om och i så fall vilket objekt som AGDA ska fokusera på.

Input Input till FocusSelector är matrisen OBJECTS från ObjectModule som innehåller de aktuella ansiktenas x och y positioner i bilden, deras avstånd till kameran i cm i z led, ansiktenas aktivitetsindex samt deras hastigheter i z led.

Output Output från FocusSelector är vektorn FOCUS som innehåller det valda ansiktets x och y position i bilden samt avståndet till ansiktet i cm i z led.

Funktioner

void tick()

Vid varje tick valideras aktivitetsindexet från ObjectModule för att avgöra om AGDA ska fokusera på objektet eller inte. Aktivitetsindexet skalas om genom att divideras med ansiktets avstånd till kameran för att på så sätt ge ansikten närmre kameran ett högre index. Det ansikte som har högst index sparas som det mest aktiva ansiktet. Därefter kontrolleras om det valda ansiktets z hastighet överstiger 0 och isåfall sparas dess x och y position i bilden samt avståndet till objektet i cm i z led i vektorn FOCUS.

3.5 Movement

Modulen Movement är den modul som avgör när och på vilket sätt AGDA ska anta ett annat mode, med hjälp av input från FocusSelector och ModeSelector.

Input Input till Movement är vektorn STRESS från ModeSelector innehållande det valda modet samt vektorn FOCUS från FocusSelector innehållande ansiktets x och y position i bilden samt avståndet till objektet i cm i z led.

Output Output från Movement är vektorn OUTPUT som innehåller de framräknade vinklarna för AGDA:s tre motorer samt vektorn SPEED som innehåller vilka hastigheter de olika motorerna ska ha i skalan 0-1.

Funktioner

void tick()

Vid varje tick kontrolleras det valda modet, vilken position AGDA har befinner sig i för tillfället samt om hon har ägget eller redan har lämnat ifrån sig det. Utifrån detta avgör Movement sedan vilken ny position AGDA ska anta samt om hon ska lämna ifrån sig ägget, ta tillbaka ägget, skydda det, eller bara picka i marken.

void calc_degrees()

calc_degrees() beräknar vinkeln i x och y led mellan AGDA och det valda ansiktet.

3.6 AGDA:s modes

Hönan AGDA känner nu igen de rörelser personen framför utför och agerar utifrån dessa. AGDA kan befinna sig i fyra olika lägen, eller som vi kallar det modes. Dessa modes är:

Idle mode: Är det läge då det inte finns en person framför AGDA eller då personen som står framför har en aktivitetsnivå som är under ett visst tröskelvärde [11].

Intrest mode: Då personen framför istället har en aktivitetsnivå som överstiger ett visst värde kommer AGDA att bli intresserad av denna person och följa den samme med blicken [12].

Defence mode: Om personen framför har en hög aktivitetsnivå och om denne även befinner sig inom ett visst avstånd från AGDA kommer hon uppfatta situationen som hotfull och dra tillbaka sitt ägg. AGDA kommer också ställa sig i ett beskyddande läge och följa personen med blicken [13].

Give-egg mode: Då personen framför har en låg men intressant aktivitetsnivå som inte uppfattas som hotfull kommer AGDA istället att skjuta fram ägget och på så sätt uppmana personen att ta ägget [14].

3.7 AGDA V.2.0

Den största tekniska uppgraderingen i AGDA v2.0 jämfört med AGDA v1.0 är förmågan att upptäcka, utvärdera samt välja det mest aktiva ansiktet bland upp till tio olika ansikten. Förutom det är nu AGDA:s rörelser anpassade för situationen; t.ex. så rör hon sig snabbare när hon ska skydda sitt ägg (defense mode) än när hon letar efter mat (idle mode).

För att kunna använda flera olika ansikten används nu FACES från FaceDetector som input till ObjectModule istället för FACE_POSITION som användes innan. Ytterliggare har buffertarna i ObjectModule utökats från vektorer till matriser, dvs. en buffert per ansikte.
Tidigare har inte FocusSelector haft någon egentlig funktion, då endast ett ansikte hanterats och skickats vidare till Movement, men nu letar FocusSelector upp det ansikte med högst aktivitetsvärde och skickar endast vidare detta till Movement.

ModeSelector är den modul som varit svårast att uppdatera: inte för att koden blivit speciellt komplex utan snarare för att det varit svårt att besluta hur vi ska välja ett mode när vi har flera ansikten. Antingen kan ModeSelector utgå från det ansikte med högst aktivitetsvärde, dvs. samma ansikte som AGDA tittar på, och då totalt ignorera resten, eller så kan ModeSelector utvärdera alla ansikten och fatta ett beslut utifrån det. Det första alternativet hade onekligen varit lättare att implementera men vi har ändå valt det andra alternativet eftersom det, enligt oss, bättre reflekterar hur en höna uppfattar en liknande situation.

För att kunna hantera input från tio olika ansikten har vi valt att skapa ett “virtuellt ansikte” som i viss mån kan sägas vara ett snitt av de tio ansiktena. Om det finns tre ansikten som tillsammans har 30 i aktivitetsvärde så innebär det att AGDA upplever det som om det totala aktivitetsvärde är 10, och agerar utifrån detta. Detta innebär dock att AGDA lätt får en skev verklighetsuppfattning och om t.ex. om 7 personer står still och en rör sig snabbt mot AGDA kommer hon inte att bli rädd, eftersom den totala hastigheten är väldigt låg. För att undvika detta har vi infört vissa skalfaktorer så att när någon rör sig snabbt mot AGDA väger detta tyngre än när någon står still, vilket resulterar i att AGDA blir rädd då den totala hastigheten upplevs hög.

I slutet av förra kursen hade vi skapat en egen lösning för att ändra motorernas hastighet under runtime men i och med att nya versionen av IKAROS och Dynamixel har en input för att ändra hastigheten, används denna istället. Detta gör det väldigt lätt att ge AGDA olika hastigheter beroende på hennes olika rörelser. Det enda som har behövts ändras som en följd av detta är hur lång tid varje rörelse tar i anspråk.

3.7.1 Begränsningar i AGDA V.2.0

Om vi förutsätter att Kinectkameran och ansiktsigenkänningen fungerar så har AGDA följande begränsningar:

- För det första är hon slö. Även om vi sätter att varje tick ska ta 50 millisekunder körs ändå max 7 ticks/sekund. Detta resulterar i att det tar upp till en halv sekund, om hon har en dålig dag, innan AGDA reagerar på t.ex. en hotfull rörelse.
- Ytterliggare så är trösklarna för hur AGDA skapar det “virtuella ansiktet”, som nämns i sektion 3.7, ganska dåliga. Detta på grund av att hanteringen av flera olika ansikten implementerades väldigt sent i kursen och förfining inte har hunnits med. Två personer går ganska bra, men fler än det har inte testats.
- Förutom dessa begränsningar är hon svår att starta och av någon anledning så krävs det att man startar IKAROS tre till fyra gånger innan motorerna aktiveras. Det är inte heller ovanligt att man måste dra strömmen för AGDA mellan två körningar.
- Slutligen så finns det ingen garanti för att AGDA inte blandar ihop två ansikten med varandra men om ansiktena är långt ifrån varandra borde felet sållas bort på grund av att de har en orealistiskt hög hastighet. Om ansiktena å andra sidan är nära varandra kan felet mycket väl uppstå.

3.8 AGDA:utseende

Eftersom vi vill att omvärlden ska interagera med AGDA är det viktigt att utseendet lockar till detta. För att inte hamna i "uncanny valley" [15], det tillstånd då robotar uppfattas obehagliga, har vi valt att ge AGDA ett utseende som påminner om ett mjukisdjur eller en muppet. Vidare vill vi att AGDA:s utseende även ska spegla robotens "personlighet". Eftersom AGDA är en nyfiken men rädd liten höna har vi valt att ge roboten en uppspärrad och lite skrämd blick. Detta har även förstärkts med ett par ögonbryn.

3.9 Slutsats

Enligt ovanstående har vi lyckats implementera de features vi planerat, helt enligt den tidsplan vi lagt upp. Projektet presenterades först i den första kursen. Där fick vi många nyttiga kommentarer som vi tagit med då vi fortsatt utveckla vårt system. På grund av yttre omständigheter, framför allt en mjukvarouppdatering av IKAROS, har vi varit tvungna att begränsa oss när det gäller den tekniska utvecklingen. Hela fokuset på den tekniska biten har lagts på att implementera en lösning för att AGDA ska kunna hålla reda på flera personer. Detta har vi, trots stor tidsbrist, lyckats implementera. Projektet i sin helhet presenterades inför publik den 3/12 och togs emot med bra respons från handledare och kollegor [16].

Eftersom vi vill att omvärlden ska interagera med AGDA är det viktigt att utseendet lockar till detta. För att inte hamna i "uncanny valley" [15], det tillstånd då robotar uppfattas obehagliga, har vi valt att ge AGDA ett utseende som påminner om ett mjukisdjur eller en muppet. Vidare vill vi att AGDA:s utseende även ska spegla robotens "personlighet". Eftersom AGDA är en nyfiken men rädd liten höna har vi valt att ge roboten en uppspärrad och lite skrämd blick. Detta har även förstärkts med ett par ögonbryn.

4 Användartest

4.1 Inledning

För att AGDA:s beräkningar av vad som är en hotfull respektive vänlig rörelse ska stämma överens med verkligheten har vi varit tvungna att definiera vad det är som utmärker dessa två rörelser. Befintliga studier av skillnaden mellan dessa två rörelser lyser med sin frånvaro. Vi har därför varit tvungna att ta fram en tes att arbeta efter. Vår huvudtes har varit att den huvudsakliga skillnaden mellan vänliga och hotfulla rörelser är hastigheten man utför dem med. Detta känns naturligt och likt andra studier med sociala robotar [17] har vi tagit för givet att så är fallet. Vi har dock till skillnad från dessa studier utfört ett kognitivt användartest för att få belägg för vår tes.

Förutom att undersöka och resonera kring vår huvudtes, försöker vi med vårt användartest få fram flera parametrar än hastighet som skulle kunna vara av intresse då man definierar skillnaden mellan en hotfull och vänlig rörelse.

4.2 Metodik och testdesign

Testet utfördes i ett användbarhetslabb på Ingvar Kamprad Design Center. Testlabbet är uppbyggt av två rum med enkelriktat spegelglas emellan. Detta möjliggjorde att vi kunde observera testpersonen utan att påverka testet. Testen dokumenterades med hjälp av video och ljudupptagning för att möjliggöra analys i efterhand.

Ett enklare pilottest utfördes veckan innan det riktiga testet. Med vårt pilottest ville vi säkerställa att testets utformning gjorde att vi kunde samla in relevant data. Pilottestet visade att den ursprungliga planen för testet behövde ändras i vissa avseenden.

Vid utformandet av testet tog vi hänsyn till att vi ville kunna mäta tidsskillnaden mellan hotfulla och vänliga rörelser. Dessutom var vi intresserade av att testa om testpersonerna upplevde någon skillnad på snabba respektive långsamma rörelser och vad den skillnaden i så fall bestod i. Dessa två vinklingar gjorde att testet delades in i två delar. Testets slutgiltiga utformning beskrivs nedan.

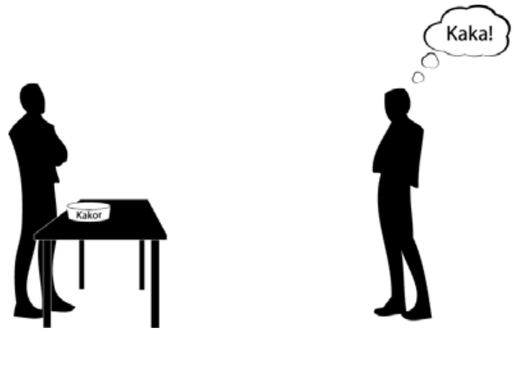
4.2.1 Deltest 1

I den första delen fick testpersonen lösa en och samma uppgift två gånger. Vi ville här mäta skillnaden i tid mellan de två fallen och eventuellt observera andra skillnader. Inför de två testomgångarna gavs testpersonen olika förutsättningar utav testledaren. Vid det ena tillfället skulle testpersonen leva sig in i rollen som en hotfull person och vid det andra tillfället anta rollen som en vänlig person. Det finns alltid en risk att ordningen testpersonerna utför uppgifterna på kan påverka testresultaten [18]. För att motverka denna ”Transfer of Learning Effect” roterade vi ordningen så att hälften av testpersonerna fick börja med att agera hotfulla och den andra hälften fick börja med att agera vänliga. Målet med uppgiften var i båda fallen att de skulle gå in i testrummet och ta en kaka ur en kakburk, som vaktades av en väktare.

Testpersonen släpptes in i rummet i ena kortändan av rummet så att kakburken och väktaren skulle vara det första testpersonen såg, se Figur 2 . Väktaren som satt bakom ett bord och vaktade kakburken på bordet, i ett för övrigt tomt rum, hade i uppgift att agera likadant i båda testfallen. När väktaren fick syn på testpersonen reste han sig upp och drog kakburken mot sig. Denna enkla gest var avsedd att få testpersonen att agera efter sin givna roll.

4.2.2 Deltest 2

Den andra delen av testet var avsett att testa om testpersonerna upplevde någon skillnad på snabba respektive långsamma rörelser. Efter att ha utfört första delen av testet ombads testpersonerna att sitta ned vid bordet och invänta testledaren för att medverka i en kort intervju. Väktaren lämnade rummet. Kort därefter släpptes en funktionär in som agerade servitör och kom in med en bricka med två kaffekoppar. Servitören ställde brickan på bordet och en av kopparna framför testpersonen och den andra koppen på motstående sida av bordet. Därefter tog servitören brickan med sig och lämnade rummet utan att ha haft ögonkontakt eller pratat med testpersonen. Vid hälften av fallen utförde servitören sina rörelser långsamt och i den andra hälften utfördes de hastigt. Efter att servitören gått ut ur rummet gick testledaren in och utförde en kort intervju enligt Appendix B. Efter intervjun ombads testpersonen svara på ett antal frågor i ett formulär utefter en semantisk differentialskala med bi-polära adjektiv, se Appendix B [19]. Syftet med formuläret var att kartlägga hur försökspersonerna uppfattat servitören, för att kunna jämföra resultaten mellan de som träffat en långsam och de som träffat den snabba servitören.



Figur 2: Testupplägg, deltest 1

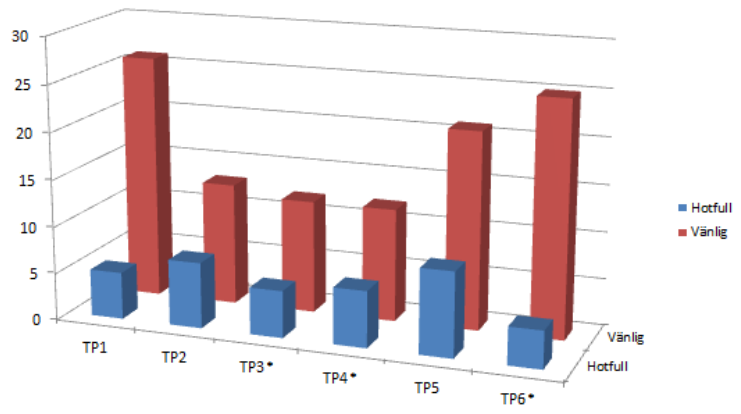


Diagram 1a: Total tidsåtgång (s) att genomföra uppgiften

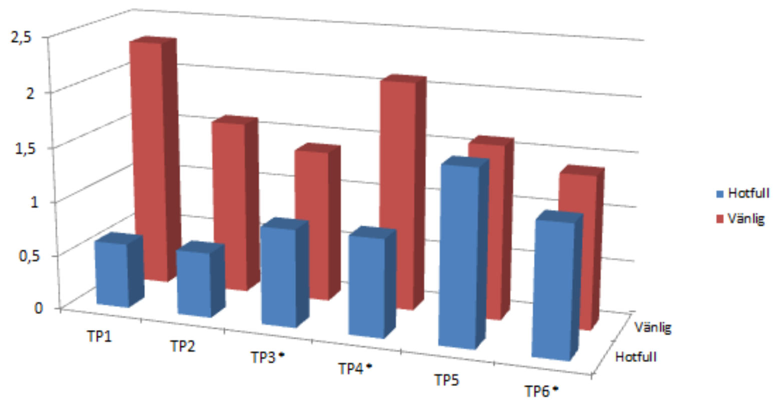


Diagram 1b: Total tidsåtgång (s) att genomföra uppgiften

4.2.3 Testpersoner

Vår målsättning var att utföra testet på åtta personer men pga. sjukdom fick vi nöja oss med sex testpersoner. Det viktiga vid urvalet av testpersoner var att de inte kände personen som agerade servitör på ett personligt plan, då resultatet i så fall hade kunnat påverkas.

Testpersonerna var samtliga i tjugoårsåldern med hälften män och hälften kvinnor. Sett till bakgrund och yrke var personerna ganska lika. Vi ansåg inte att gruppens homogenitet skulle påverka resultatet nämnvärt. Därför la vi inte heller vikt vid att få en större spridning avseende ålder, bakgrund eller yrke.

4.3 Resultat

Vi mätte tidsåtgången för samtliga testomgångar av första delen av testet, från det att de gick in i rummet tills de hade greppat en kaka. Testdatat visar att samtliga testpersoner tog längre tid på sig i fallet då de agerade vänliga än i fallet då de agerade hotfulla, se Diagram 1a. Vi mätte även tiden för den avslutande armrörelsen från det att testpersonen började röra på armen för att sträcka sig mot kakburken, till det att handen greppat en kaka, se Diagram 1b.

Diagram 2 och 3 visar en sammanställning av hur testpersonerna uppfattat servitören i deltest 2. Vi använder oss i dessa fall av en semantisk differentialskala för att på så sätt mäta testpersonernas inställning till servitören.

4.4 Analys

En av testpersonerna hade misstolkat instruktionerna under deltest två. Detta är anledningen till att endast fem värden är införda i diagrammen 2 och 3.

Våra tester visar att det finns en klar skillnad i tidsåtgång då en person agerar hotfullt, jämfört med vänligt. Vi visar med vårt användartest att hotfulla rörelser är hastigare än vänliga rörelser. Dessutom visar testet att hastiga rörelser uppfattas mer hotfulla än långsamma rörelser. Vi ser tydliga med andra ord tydliga tendenser. Den hastiga servitören uppfattas, enligt diagram 2, mer

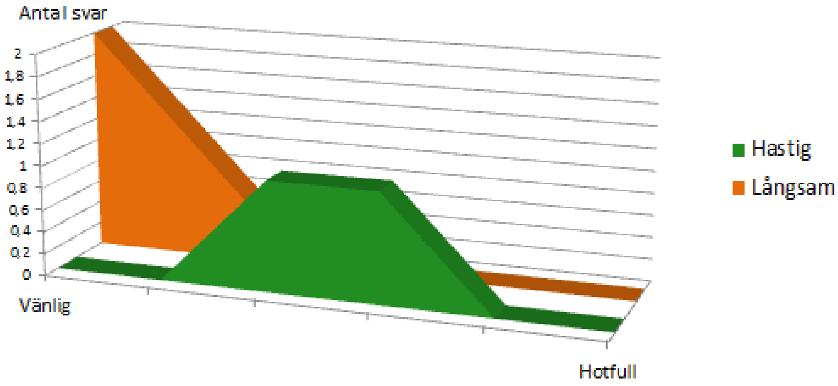


Diagram 2: Testpersonernas inställning till servitören

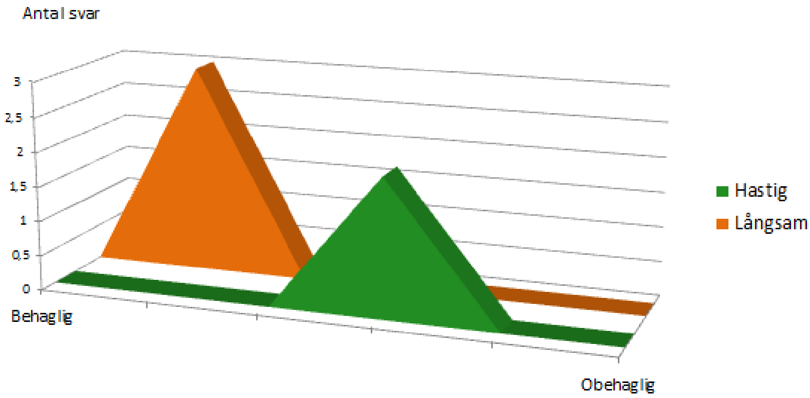


Diagram 3: Testpersonernas inställning till servitören

åt det hotfulla hållet och vi ser med andra ord tydliga tendenser. Vidare ser vi en stor skillnad i hur behaglig servitören uppfattades i de två fallen. Hastiga rörelser verkade testpersonen uppfatta som obehagliga.

Vi finner därmed fog för vår huvudtes och hävdar att AGDA:s beslut baseras på liknande parametrar som människor använder för att skilja hotfulla rörelser från vänliga.

Andra parametrar av intresse kunde vi inte hitta. Hastighet verkar onekligen vara den viktigaste parametern. Den enda tydliga skillnaden vi kunde se var att testpersonerna tenderade att komma närmare personen som vaktade kakorna, då de agerade hotfulla. Detta svarade redan mot vår implmentation i AGDA. En annan skillnad var att testpersonerna var betydligt mer verbala då de var vänliga, med hälsningsfraser och dylikt.

5 Diskussion och vidareutveckling

Av beskrivningen ovan går att utläsa att mycket tid lagts ner på research, framförallt utforskandet av mjukvaran. Detta anser vi har varit en bra approach då vi, när vi samlats för att implementera de olika delarna av vår lösning, kunnat jobba effektivt och lösa eventuella problem snabbt. I båda kurserna har vi hållit oss till vår tidsplanering. Dock var vi i andra kursen tvungna att dela upp arbetet på ett tydligare sätt än i första kursen. I första kursen försökte vi göra så att alla gruppmedlemmar var delaktiga i allt. Nu fick vi istället göra så att två gruppmedlemmar arbetade med den tekniska utvecklingen medan de två andra arbetade med vårt användartest. Mycket av fokus i fortsättningskursen låg i just användartestet, då detta utgjorde den teoretiska bakgrunden till vårt projekt. Vi fick avgränsa oss en hel del när det gäller den tekniska utvecklingen. Mycket tid gick åt tillatt anpassa vår lösning för den nya IKAROS-versionen vilket även gjorde att vi saknade tid åt att implementera flera nya tekniska lösningar.

Gruppdynamiken har fungerat bra. Samtliga gruppmedlemmar har antagit en viss roll och rollfördelningen i sin tur har varit bra. En tydligare rollfördelning antogs i fortsättningskursen. Samtliga beslut har emellertid tagits i samförstånd och i demokratisk anda, vilket gör att alla gruppmedlemmar varit delaktiga i samtliga delar av projektet – allt från planering, till implementation, presentation och dokumentation/rapportskrivning. I fortsättningskursen var vi tvungna att fokusera på två saker samtidigt. Därför delade vi upp oss i grupper om två, där en grupp fick ansvara för projektets kognitiva bit och den andra för den tekniska implementationen.

Vi ansåg tidigt att vi hade ett tekniskt försprång då vår grupp utgjordes av enbart teknologer. Den tekniska biten har verkligen hamnat i fokus i vårt projekt. I slutändan kan vi känna att det vore bra att ha någon i gruppen med en bredare infallsvinkel, eller åtminstone en djupare kognitionsvetenskaplig förståelse. Våra två rörelser har vi definierat utifrån våra tankar och preferenser och satt upp en tes. Det finns knappt någon teori att läsa. Faktum är att det inte finns så mycket forskning eller andra rapporter som behandlar distinktionen mellan en vänlig och en hotfull rörelse. Detta gav oss i sin tur stora möjligheter att genomföra en undersökning med kognitionvetenskaplig utgångspunkt.

Vi har i vårt användartest studerat hotfulla och vänliga rörelser och jämfört detta med vår tes. Testet visar i slutändan på en ganska tydlig tendens när det gäller skillnaden i hastighet mellan de två rörelserna, men även på vilket sätt man uppfattar en hastig och långsam rörelse. Resultatet, som beskrivs ingående i tidigare avsnitt, går hand i hand med vår tes och därmed även vår implementation. I vårt test upptäckte vi inga andra parametrar än hastighet och avstånd (som redan finns implementerat i vår lösning). Som vi nämner ovan var testpersonerna betydligt mer verbala då de var vänliga. Man kan tänka sig att nästa iteration av AGDA skulle kunna innehålla röstigenkänning där hälsningsfraser skulle hjälpa roboten att avgöra om personen framför är vänlig eller hotfull.

Vi kan dock inte på något sätt påstå att resultatet är statistiskt säkerställt. Det vore givetvis bättre om vi kunnat genomföra ett större test på fler än 50 testpersoner för att på så sätt få ett mer statistiskt säkerställt resultat. På grund av tidsbrist och den allmänna bristen på testpersoner har vi istället nöjt oss med att genomföra ett mindre test, för att på så sätt visa på en tendens – snarare än att bevisa vår tes. Detta anser vi att vi gjort bra och att vi nu, i och med testet, fått en större teoretisk (kognitionsvetenskaplig) grund i vårt arbete.

Överlag har utvecklingen flutit på bra. Vi har inte stött på några större dead-ends och eventuella problem har vi kunnat lösa. Det största problemet vi haft i projektet anser vi är avsaknaden av dokumentation. Det tog p.g.a. detta ganska lång tid innan vi på riktigt kunde börja implementera vår lösning. Vi ser gärna att det i framtiden finns en enkel komma-igång-guide samt diverse steg-för-steg guider som gör att man får en djupare förståelse för uppbyggnaden av IKAROS. Eftersom Kinecten är en central del av projektet ser vi gärna en mer utförlig dokumentation av OpenKinect. Mellan första kursen och fortsättningskursen genomfördes även en uppdatering av IKAROS. Denna har bara orsakat problem för oss. I framtiden bör det inte ske några större mjukvarouppdateringar mellan de två kurserna. Detta kommer leda fram till mera utarbetade tekniska lösningar.

I det stora hela är vi nöjda med vårt arbete i de två kurserna. Så här i efterhand tycker vi att den tekniska utvecklingen varit väldigt bra och är också nöjda med det faktum att vi med hjälp av vårt användartest byggt en bredare teoretisk grund för vårt arbete.

Referenser

[1] http://www.eat.lth.se/fileadmin/eat/MAMN10/ProjektbeskrivningarJuly1.pdf [Tillgänglig 12-12-07]

[2] ”Human gesture recognition”, master-thesis by Mohamed Becha – University of Nice-Sophia Antipolis, s. 26-27 [http://www-sop.inria.fr/members/Mohammed-Becha.Kaaniche/medbecha-thesis.pdf, Tillgänglig 12-12-07] [3] http://projects.ict.usc.edu/mxr/faast/ [Tillgänglig 12-10-27]

[4] http://stackoverflow.com/questions/9191994/ using-kinect-to-detect-objects-on-the-floor [Tillgänglig 12- 10-27]

[5] http://openkinect.org/wiki/Imaging_Information [Tillgänglig 12-10-27]

[6] ”AGDA-Beskrivning” [Inlämnad 12-09-20; Elektronisk tillgänglig: http://dl.dropbox.com/u/32717155/AGDA-Beskrivning.pdf]

[7] http://www.ikaros-project.org/module/Dynamixel [Tillgänglig 12-10-27]

[8] http://mathnathan.com/2011/02/depthvsdistance/ [Tillgänglig 12-10-28]

[9] http://tinyurl.com/8hmcf6t [Tillgänglig 12-10-28]

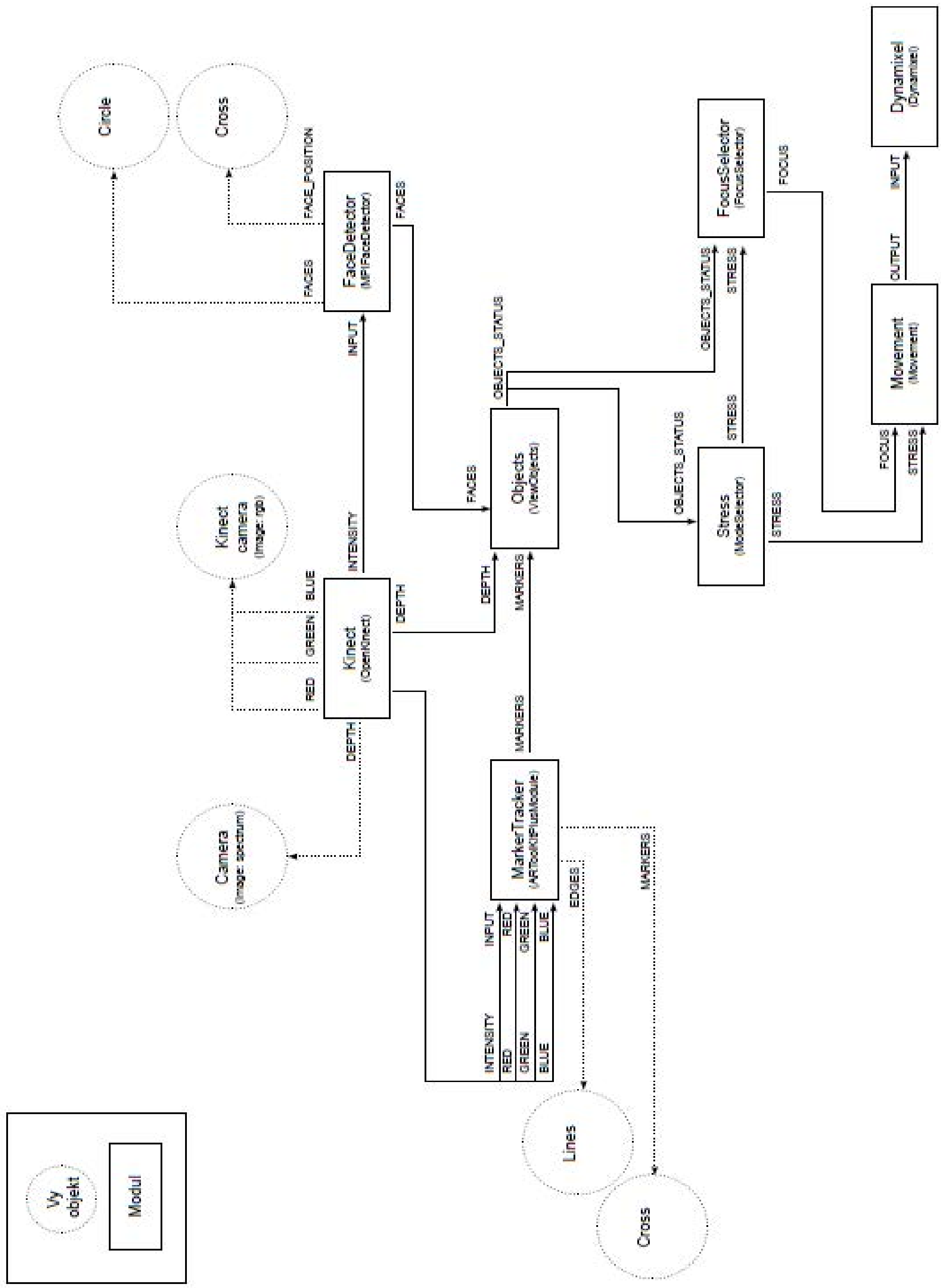
[10] http://www.ros.org/wiki/kinect_calibration/technical [Tillgänglig 12-10-28]

[11] Idle-mode, video-demo: [http://dl.dropbox.com/u/32717155/20121018_104926.mp4, Tillgänglig 12-10- 28]

[12] Intrest-mode, video-demo: [http://dl.dropbox.com/u/32717155/20121018_110051.mp4, Tillgänglig 12- 10-28

[13] Defence-mode, video-demo: [http://dl.dropbox.com/u/32717155/20121018_110403.mp4 , Tillgänglig 12- 10-28
[14] Give-egg mode, video-demo: [http://dl.dropbox.com/u/32717155/20121018_105447.mp4 , Tillgänglig 12- 10-28]
[15] Mori, M. (1970). Bukimi no tani [the uncannyvalley]. Energy, 7:33–35.
[16] http://youtu.be/jZYLrQSpYp0 [Tillgänglig 12-12-07]
[17] Breazeal, Cynthia, A. Edsinger, P. Fitzpatrick and B. Scassellati "Active Vision for Sociable Robots", Socially Intelligent Agents - The Human in the Loop, Special Issue IEEE Transactions on Man, Cybernetics, and Systems, Part A: Systems and Humans. Kerstin Dautenhahn (ed.), Volume 31, number 5, pp. 443-453, September 2001.
[18] Jeffrey Rubin & Dana Chisnell: "Handbook of Usability Testing, Second Edition" (2008)
[19] http://www.psykologiguiden.se/www/pages/ ?Lookup=semantisk+differential [Tillgänglig 12-12-07]

APPENDIX A: Modulkarta



APPENDIX B: Testmaterial

(A) Intervjufrågor

Intervjufrågor

- Tyckte du det var roligt?
- Vad tyckte du var enklast att spela, hotfull eller vänlig?
- **Hur uppfattar du själv att dina rörelser skiljde sig mellan de olika utgångslägena?**
- Vill du ha lite kaffe?
- **Hur tror du att kakburksväktaren uppfattade dina rörelser i de olika utgångslägena?**
- Vilket av de två utgångslägena stämmer bäst överens på din personlighet?
- Vilken av de två approacherna tror du fungerar bäst för att uppnå sina mål?
- **Om du skulle fått göra om uppgifterna, är det något du skulle ha gjort annorlunda så här**
- **i efterhand?**
- Tack för att du ställde upp. Nu återstår bara att fylla i ett kort formulär.

(B) Frågeformulär

Svarsenkät:

Hur uppfattade du intervjusituationen?

vänlig	☐	☐	☐	☐	☐	hotfull
trevlig	☐	☐	☐	☐	☐	otrevlig
behaglig	☐	☐	☐	☐	☐	obehaglig

Hur uppfattade du förhörspersonen?

vänlig	☐	☐	☐	☐	☐	hotfull
glad	☐	☐	☐	☐	☐	arg
trevlig	☐	☐	☐	☐	☐	otrevlig
behaglig	☐	☐	☐	☐	☐	obehaglig

Hur uppfattade du servitören?

vänlig	☐	☐	☐	☐	☐	hotfull
glad	☐	☐	☐	☐	☐	arg
trevlig	☐	☐	☐	☐	☐	otrevlig
behaglig	☐	☐	☐	☐	☐	obehaglig

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Goal Emulation in a Drummer Robot

Hossein Motallebipour, *Member, IEEE,* Björn Andersson, Emil Gunarsson, Filip Larsson, and Jim Persson

Abstract—Learning from demonstration represents a powerful type of learning and a promising alternative to programming robots. However, even after years of research in the field the current level of robot imitation skills can be compared to a 2-year old child. Most of the presented works up to now have been exact imitation of trajectories without any particular understanding about the rudimentary goals of the movements. This paper presents our approach to developing a robot that imitates a human drummer, where the robot is able not to perform a scaled copy of the movements of a human demonstrator but to emulate the goal of the instructor's actions through observation of action and goal extraction based on tests with human subjects. The intention is effectively reproduced and adapted to the robot's situation and capabilities.

Index Terms—Cognitive robotics, Cognitive science, Emulation, Learning from Demonstration

1 Introduction

Employing robots has provided a possible solution for homes, offices, and societies of the future. Creating robots capable of performing tasks that can save human effort and even be deployed in hazardous environments is promising as much as it is difficult to undertake (Motallebipour, 2010). Besides the challenge of producing the hardware, the necessity for developing a reliable software is obvious. However, when a robot enters the dynamic world of humans, it will be in need of the same flexibility they generally possess. That puts extraordinary requirements on the software structure of a robot and calls for other solutions than regular traditional control system algorithms usually provide. In search for such algorithms the matter of robot programming has shifted toward other paradigms inspired by biological systems. The idea here is to make the platform self-learning, self-developing, and teachable, and also facilitating for the end-user of the robotic system by reducing the complexity of search spaces (Billard & Mataric, 2001; Delson & West, 1996; Billard, Calinon, Dillmann, & Schaal, 2008). Self-learning and self-developing are prerequisites for teachability which disentangles the robot from running on predetermined, "hard-coded" parameters, leaving the robot with the choice of the appropriate set of values in each particular situation.

Newborn human infants face the same problem as current robots. At the beginning they possess the capabilities of making but a few, often poorly controlled, single movements. But contrary to robots, they experience processes that help them in developing the necessary sensorimotor skills, entirely based on intrinsic mechanisms, to undertake simple physical manipulation tasks in the beginning of their lives. Experiences from using those skills together with other processes emerge eventually into complex physical actions and later result in high level behaviors as reasoning, and exhibiting the considerably discussed theory of mind—"the ability to reflect upon one's own and other people's inner worlds" (Gärdenfors, 2006; Brooks, Breazeal, Marjanovic, Scassellati, & Williamson, 1999).

For a robot to learn an action there are two approaches to what is called learning from demonstration (LfD). In the older approach, learning occurs by copying the exact movements through extraction of the joint properties in each moment, called *embodiment mapping*, using modalities, *shadowing*, or by directly reading the joint values, *teleoperation* (Argall, Chernova, Veloso, & Browning, 2009). The introduction of cognitive science as a cross disciplinary subject having overlaps with computer science gave for instance rise to a new approach within robotics where learning occurs by extracting the goal of the task to undertake, *goal-directed* imitation. In the former type of learning, the robot motion is constrained to the parameters extracted by the visual and/or haptic modalities, whereas in the latter approach the robot will have to extract the goals and then define a new trajectory between each pair of them to complete the task.

Goal extraction is a persistent issue in the field of imitation, making the current robots at best comparable to toddlers (Breazeal & Scassellati, 2002). This report presents the work carried out within the field of cognitive robotics on a physical 3 degrees of freedom (DoF) robot arm that is to play music on 3 drums without copying the movements, after having seen a human player perform a sequence of drumbeats on 3 equally-sized drums. The idea is to play the sequence without trying to follow exactly the same trajectories between each pair of the drums, although the time intervals between the drumbeats should be maintained; a practical example of learning by imitation.

Since the robot contains neither a particular visual modality, nor processing and/or control units, its hardware is connected to a Kinect™ (*Kinect™ for Windows®*, 2012) motion sensing input device and its software is embedded in Ikaros. Ikaros is a software project for development of an "open infrastructure for system-level modeling", comprised of modules simulating cognitive processes in a living individual's brain-nervous system (Balkenius, Morén, Johansson, & Johnsson, 2010). Kinect™ is a game controller with an infrared sensor, a color camera, and a microphone that enable a player to control the game through bodily activities instead of joysticks or other handheld devices. For the current project, only the proximity sensor is used.

Our contribution to the field of Human Robot Interaction, HRI, is to provide an example of robot learning from demonstration where the interaction occurs without any physical contact between the robot and the instructor and where the robot successfully emulates the instructor's goals based solely on human gestures.

In the following section a compact theoretical background of imitation as a neurocognitive phenomenon will be presented; section 1.1. This theoretical background is rather to be seen as an orientation about imitation and goal emulation, where these are separated as different concepts, out of which the latter is directly connected to the aim of the present work. Both concepts are then considered in the light of opposing theories about their occurrence at the neuronal level.

Section 1.2 will present an overview of some of the previous works

Correspondence Address: hosseinm@ieee.org
E. Gunnarsson and H. Motallebipour are master's students with Lund University Cognitive Science, Lund, Sweden
B. Andersson, F. Larsson, and J. Persson are master's students with Lund Institute of Technology, Lund, Sweden

on imitation as an approach to robot learning, in a chronological order. In section 2, the practical part of the current project is explained together with tests that have been conducted both before and after the implementation was carried out. The report will then proceed with a discussion in section 3, on why the present robot is proclaimed to have a learning comparable to a 2 years old toddler and how the current work differs from the previous attempts on implementing robots that learn by imitation. Also a brief discussion on what is cognitive learning and whether the current robots can be asserted to learn cognitively. end by a discussion on the pros and cons of the current solution.

The report will finally end up with a conclusion, section 4, and the future work, section 5.

1.1 The Neurocognitive Background

Imitation may be used as means for alignment of a discussion (Garrod & Pickering, 2004), or in a teacher-learner relationship. Nevertheless, imitation generally defined as "performing an act after perceiving it" (Bekkering, Wohlschläger, & Gattis, 2000) is a cutting edge research topic within robotics. Children, as the best model for developing "intelligent" artifacts from regular "non-intelligent" robots, have usually been subject for psychological experiments to find out when and how this characteristic arises and what intrinsic features it emanates from.

Human infants were shown to be able of imitating facial and manual gestures as early as 12 to 21 days after birth (Meltzoff & Moore, 1977). Still, in experiments on imitation of manual manipulation and tool-use the chosen ages are 3 and above (McCall et al., 1977; Horner & Whiten, 2004; Dawson & Adams, 1984), because the imitation behavior is not well developed before the ages $1\frac{1}{2}$ to 4 years (Want & Harris, 2002). To find out the reason behind the choice, one will need to study the developmental psychology of children.

Based on Jean Piaget's research in genetic psychology[1] a child's cognitive development comprises four hierarchical stages and to reach each of the stages the child has to fulfill the criteria for the previous ones (Mascolo & Fischer, 2005). In this hierarchy the first stage covers the birth through 2 years.

Piaget and Inhelder (2000) conclude that their studies have shown that during the first stage the movements are impetuous and the coordination occurs on a sensori-motor level without any reasoning or assessment. Piaget holds that learning new reflexes in this scheme occurs nevertheless through *assimilation*, meaning that stimuli perceived from the interaction with the outside world is modified and filtered by the init structure of the child's reflexes, or *action-schemes*.

During the second half of this stage, the child starts combining different intrinsic reflexes into an intricate network of sensors and motor responses, or *acquired habits*, which are eventually progressed into what is called intelligence. In this phase, the *pre-operational*, the intelligence emerges through the operation of semiotic functions, where representations for events and objects are formed (Piaget & Inhelder, 2000). This phase is extended over to the second stage of the child development, 2 through 6 years, when it starts handling objects instinctively. Knowledge representation may be involved in object and image manipulation but is still not well-structured (Mascolo & Fischer, 2005). The absence of theory of mind is yet perspicuous (Gärdenfors, 2006).

There is however empirical studies that prove the opposite to Piaget's proposition about the seemingly purposeless, impetuous arm movements in newborn babies (Meer, Weel, & Lee, 1995). According to the new model, the purpose with the arm movements is to gather information about their effects and ranges giving the infant to start reaching and grasping after 4 to 5 months. For deliberately coordinated movements, the baby will need a point of reference for acquiring information about the position of different body parts through hand eye coordination. Another well-known example of empirical evidence refuting the first stage of Piget's framework is the 12-21 days old babies' ability to imitate facial and manual gestures (Meltzoff & Moore, 1977).

Imitation is intuitively understood as mimicking an individual's action by another individual, without any preferences given to the details of the action, or the quality of the attributes of the mimic. Qualities as good, bad, or exact imitation are the only classifications made in everyday terms. Tomasello was the pioneer who differentiated between the terms imitation and emulation for the first time in 1987 with a stress on the dependency of objects in a causal relationship rather than the mere existence of objects (Tomasello, 1990; Heyes, 2001; Horner & Whiten, 2004). On the other hand Tomasello, Kruger, and Ratner (1993) defined the imitation, in opposition to emulation, as following the exact actions without understanding what function each action is meant to fulfill.

Inspired by the idea, Whiten and Ham (1992) defined the emulation in the context of social learning as achieving a goal without copying a particular course of action—as one of the eight possible types of mimetic processes. Hauser (2000) applied these definitions on research on animal intelligence to classify learning in thirteen different types, by borrowing from other researchers in diverse fields of science. These definitions were then classified in two groups of similarities, where the second group—similarity via social mimetic processes—was divided into *Social Influence* and *Social Learning*. The latter subgroup comprised in turn of five types of learning: *Stimulus/Local Enhancement*, *Observational Conditioning*, *Goal Emulation*, *Imitation*, and *Teaching*. (Hauser, 2000; Breazeal & Scassellati, 2002). He defined the imitation as

> *"A learns some aspect of the form of a behavior performed by B that is novel to A's repertoire, and is capable of expressing the behavior in the absence of B",*

and the emulation as

> *"A produces the same end product as B following observation of B's act, but the form of A's behavior differs from B's",*

both resembling the definitions from the work by Tomasello, Savage-Rumbaugh, and Kruger (1993) on children and chimpanzees.

Within life sciences, some researchers believe that emulation is a human ability to extract goals and intentions as an approach to social learning through interaction (Kilner & Frith, 2008). Whether this ability is unique is still under debate since studies in comparative cognition have been trying—at least for the past 20 years—to prove the opposite (Shettleworth, 2010). Erlhagen, Mukovskiy, and Bicho (2006) state that if somebody presses a button with his right index finger, the baby would also press a button but with his left palm instead of the same finger of the same hand. The baby clearly understands that the goal of this movement is not just to extend the forefinger of the right hand but to press the button by any possible means. Typically when an arm movement is performed, the human observer, regardless of the age, concentrates mostly on the end-point of the trajectory to see what the intention by that particular gesture is (Mataric & Pomplun, 1998). Experiments conducted by Bekkering et al. (2000) also provide evidence for human imitation being goal-directed in that children show a tendency to perform less accurate when arm movements are directed to a physical goal compared to when they are not. Their conclusion from the experiments is that goals or intentions motivate all actions regardless of whether they are performed voluntarily.

Another experiment with children between 3 and 6 years by Rohrer and Hulet (2006) has shown that the imitation behavior is a motor pattern comprised of components from a *pattern library*, "guided by an interpretation of the motor pattern as a goal-directed behavior" (Bekkering et al., 2000). The goals are subdivided into different levels of priority and employed for reconstruction of a motor pattern in accordance with the complexity of the movement and resources to perform the action.

1. Part of the human psychology that discusses the effects of and the causes behind the development of a child and refers primarily to its process of cognitive development, for instance intelligence, perception, and such (Piaget & Inhelder, 2000).

Factors affecting the choice of an action as a goal may be presence of an object, an agent, a movement path, or a salient feature. Hence, the mapping is primarily concentrated on the manipulated object and the manipulator (Hamilton & Grafton, 2006).

By evidence obtained through search for the neurophysiological grounds, many researchers have advocated for the existence of mirror neurons in humans as units for interpretation of observed action (Rizzolatti, Fogassi, & Gallese, 2001). These neurons were discovered for the first time in the ventral premotor area, F5, and inferior parietal lobule, 7B in the left hemisphere of the macaque monkeys' brain (Banich & Compton, 2011). Neural firing have been observed in mirror neurons when physical activity is performed without visual feedback or when previously observed activities are sensed without any physical feedback (Gillmeister, Catmur, Liepelt, Brass, & Heyes, 2008). These neurons are known to be involved in matching visual cues to motor responses, also referred to as "simulation", "resonance behavior" and "shared representations" (Blakemore & Decety, 2001). In humans, the homologous entities are postulated to reside in Broca's area in the inferior frontal gyrus and parietal lobule (Banich & Compton, 2011). The former region is believed to be involved in imitation rather than observation or task execution.

At least theoretically, the existence of the mirror neurons has come to bridge the gap between the perceived action and the motor control of the body in imitation of observed action. The common ground for most of the researchers in cognitive science and related fields is hence that observing an action performed by a conspecific activates those areas in the brain that are in charge of the motor control of the same body parts of the observer; one of the most intricate cases of perceptual-motor coordination. (Meltzoff, 1993; Fadiga, Fogassi, Pavessi, & Rizzolatti, 1995; Vogt, 1995, 1996; Massen & Prinz, 2009; Viviani & Stucchi, 1992, to be considered for an early, extended discussion on the matter).

The imitation abilities mentioned previously are regarded as the effect of a direct mapping of the instructor's movements perceived by the visual system and the observer's motor reactions; what has been called Active Intermodal Mapping theory or AIM by Meltzoff (1993)—and also Meltzoff and Moore (1977)—advocating the existence of a region in the brain standing for a higher level modality that stores both the perceptive and the proprioceptive information. When trying to imitate an action, this region is active until both patterns match.

It has further been suggested that there are particular regions in the brain, including the mirror neurons, that are responsible for mechanisms that enable understanding of long term goals of other's actions (Kilner & Blakemore, 2007). However, despite the strong evidence for pattern matching there are observations that show failure to generate a flawless imitation. This has been shown for instance in imitating ipsilateral versus contralateral arm movements, with and without a goal. Also in patients with Broca's aphasia (Head, 1920, p.118) and autistic children (Heyes, 2001) the inability to generate the desired trajectory has been observed. These studies suggest that pre-school children and aphasic patients exhibit a tendency to generate ipsilateral responses regardless of the type of laterality of the instructors' movements (Schofield, 1976, page 572).

Bekkering et al. (2000) performed the same experiment as Head with the difference that the bimanual movements were not considered in Head's experiment. They suggest that even children of higher age show significantly the same inadequacy. Head postulates also that the failures in lateral bifurcation are not a result of neurophysiological constraints. Even though the same phenomenon is not conspicuous in adults' reactions it can be observed as a latency in response to contralateral movements.

Turella, Pierno, Tubaldi, and Castiello (2009) reject the arguments for the role of the same neurons in imitation and understanding the intention of others while watching their actions (Iacoboni et al., 2005), gestural communication (Rizzolatti & Craighero, 2004), language development (Rizzolatti & Arbib, 1998), emphathy (Leslie, Johnson-Frey, & Grafton, 2004), and many other functions considered as high level cognitive abilities.

Their primary explanation is that there are too few experiments made using *positron emission tomography (PET)* and *functional magnetic resonance imaging (fMRI)* showing the relation between all these functions and the mirror areas where they actually operate as mirror (Turella et al., 2009). The reason may be the result of the inability to perform direct cell-probation for observing activities in the human brain, in particular those areas that are called the mirror neuron system, *MNS*. Even though they never reject the plausibility that action observation, imitation, and execution may have some shared areas for carrying out those tasks, there is still no support, they mean, for a centered mirror system in the brain for understanding intentions (Hamilton & Grafton, 2006). Their argument is corroborated by studies on mirror neurons showing that observation of actions that are novel, do not match their context, or whose intentions are uncertain never result in excitation of mirror neurons (Brass, Schmitt, Spengler, & Gergely, 2007). Instead, some form of reasoning must be lying behind the process of "cultural learning" (Tomasello, Kruger, & Ratner, 1993).

Nevertheless, efforts have been made to explain the role of mirror neurons in emulation. For this reason perception of action has been divided into 4 stages, *kinematic level*, *signal level*, *short-term goal*, and *intention* (Kilner & Frith, 2008). Control signals are sent directly by the premotor cortex through the thalamus and are used to move the muscles during a process for which a space-time plan of action is described at the kinematic level. Short-term goals are the usually conspicuous and extractable results of single parts of an action, based on their salient features, which together comprise the intention as the final long-term goal of the whole procedure. Since the intention of a procedure is not always discernible, the observer will have to make inferences based on what is perceived at the kinematic level and what the short-term goals have been.

This is what the mirror neurons are supposedly responsible for: to interpret the goal level into information about the kinematic and signal level that for example is reflected in the intraparietal sulcus for imitation of hand movements. According to *predictive coding framework*, the observed kinematics provide an *a priori* estimation of the goals or the intention which in turn yield a prediction of the motor commands. How then the real motor commands are reckoned is a matter of recursively reducing the error between the observed kinematics and the kinematics of the observer based on the estimated motor commands. In the same manner, an estimation of the goals and intentions are deduced (Brass et al., 2007). Hence, within this framework, even if the MNS does not provide the intention directly, it is suggested to be involved in the estimation of the intention.

1.2 Previous Work

Delson and West (1996) suggested an approach to an imitating robotic arm that using a range of directly measured joint torques and robot forces could copy a motion trajectory from multiple human demonstrations. It could further omit inconsistencies in the trajectories using a stochastic automaton. The result was a simulation of a robot that could perform assembly tasks in 3 dimensions.

Kaiser and Dillmann (1996) presented an approach to a sensor-based Puma 260 manipulator with the ability to acquire peg-insertion skill from a human demonstrator. As with the previous group, the action was constrained to reading joint forces—torques—from simple physical manipulations and processing them to comply with the robot's physical attributes.

The hybrid dynamic system of Eshed Scorbot 5 DOF manipulator implemented by Hovland, Sikka, and McCarragher (1996) and a six-axis Polhemus position sensor used Baum-Welch re-estimation method for finding the *Hidden Markov Model (HMM)* parameters based on the training data. There were still not enough flexibility in the robot's actions as the system could only handle planar tasks. Since it was based on training sets it was also limited as such, suffering the same disadvantages as ANN's or any other similar system.

Atkeson and Schaal (1997) used an anthropomorphic arm with 7 DoF to perform a pendulum swing up task as an example; implemented redundant inverse kinematics as well as real time inverse dynamics to allow the robot to follow desired hand motions.

In later attempts researchers motivated by results of neurological studies, cognitive science, and psychology, suggested theoretical frameworks for biologically inspired models. Eventually, robots or simulations that produced motor skill imitation based on those models were introduced. Usually, a biologically inspired model implements a model of a very limited part of the brain and/or spinal cord as well as the animal sensor system to propose a solution to task specific functions, even though the recent research have led to complex high-level representations in robot models. Billard and Mataric (2000) presented a connectionist architecture to control avatars for learning 3 types of movement sequences.

As part of the research on mirror neurons, Agam, Bullock, and Sekuler (2005) have made experiments on humans imitating several movements along adjacent straight lines of "just-seen" sequences. Their intention was—in contrary to many other groups, where imitation of a single movement, such as grabbing or moving a hand from one point to another, is studied—to test whether and how well subjects could follow series of movements. Tasks of this kind would necessitate the recognition of each component in the series, i.e. when it starts, and how it finishes. They expected the outcome of the experiment to follow the computationally interesting architecture presented by Bullock and Rhodes (2002) based on competitive queuing (CQ) model established by Hartley and Houghton (1996). The connectionist model developed for the short-term memory was based on two general assumptions. First that more than one plan can be activated in a layer of the architecture at the same time. The other one assumes that after the most active plan in a layer is chosen, it is removed from the queue and the next most activated plan is looked for by running the procedure iteratively. Eventually the plan representations are converted into a serial performance (Grossberg, 1978). They found that leaving out segments occur more frequently for longer sequences compared to introduction of new segments. Also as a parallel to "verbal serial recall", repeating one segment could lower the imitation quality.

Billard and her collaborators attempted to create a framework for imitation of manipulation and gestures (Billard, Epars, Calinon, Schaal, & Cheng, 2004). They based their approach to imitation, on extraction of joint values by utilizing statistical methods—as can be realized from their subsequent works—such as HMM, Gaussian Mixture Regression (GMR), and Expectation Maximization (EM). The joint statistics are then used for extracting the most salient features that the robot will try to imitate (Calinon, D'halluin, Caldwell, & Billard, 2009), which they also used in recognition of human gestures (Calinon & Billard, 2004) and haptic communication between a human and a robot that collaborated for lifting objects (Calinon, Evrard, Gribovskaya, Billard, & Kheddar, 2009). Billard, Calinon, and Guenter (2006) presented humanoid robot that learns through kinesthetic, how to manipulate simple objects. The robot needs to be shown the action several times, with variations of actions but consistency in salient features of the task, to be able to make an approximation of the trajectories, paths and positions using HMM, *Support Vector Machines (SVM)* and EM. Calinon and his collaborators have continued, up until recently, using these algorithms for estimation of joint trajectories. Only now, they have introduced researching on other modalities than vision and haptic modalities, such as proximity sensors only for planar tasks, in contrary to the present project (De Tommaso, Calinon, & Caldwell, 2012).

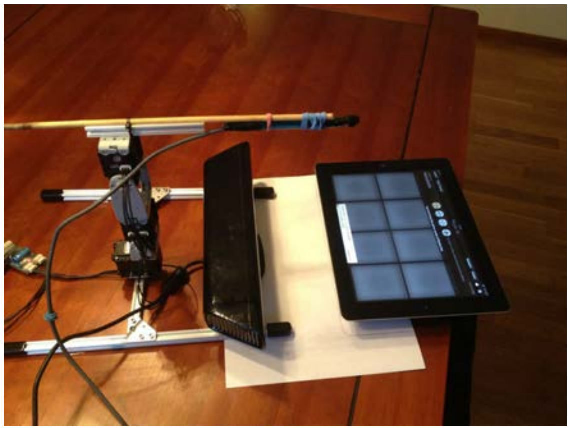
2 METHODS

Since a robot has, at best, an ability of imitation roughly corresponding to a 2 years old child's (Breazeal & Scassellati, 2002), the preparatory test, implementation, and evaluation test—all explained presently—were concentrated on a model of a child in this age.

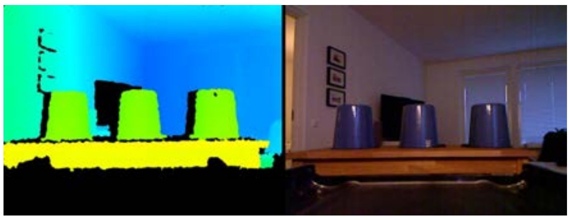
A preparatory test of imitation was planned with one goal: to study movement and behavior of a user, teaching a 2 years old child. The intention was to extract all possible information about poses, gestures, and movements occurring in a one-way instructor-child situation. This information would later be used as a basis for implementing the communication between an instructor and the robot.

The implementation was then based on the same procedure as a non-reasoning child, to observe, emulate critical moments, save in a short-term memory STM, and perform. The feedback from the robot was later implemented as exhibition of head saccades and playback of sentences vocalized by a text-to-speech engine.

The work was finalized with an evaluation experiment on subjects with results that will be presented in the last subsection.



(a) The Robot Arm



(b) A Screen Shot

Figure 1. (a) Shows the 3 DoF robot arm with the stick. The head of the stick was covered with a streamed rubber to enable the robot to play music on an iPad™ touch screen. (b) Left window shows the result of the infrared depth sensor, where the color interval red through purple indicate the proximity "close" through "distant". Right window is the image the color camera is receiving.

2.1 Preparatory Test

Four persons were chosen as subjects for this instructing-test. Each subject was asked to imagine a 2-years old, non-Swedish speaking child facing the subject, each one provided with a set of 3 drums in front. With this setup in mind, the subject was asked to try instructing the child how to play a sequence of 5 drumbeats, without using vocals or touching the child. The instructor was also informed that the child was supposed to start imitating the sequence *after* it was terminated. The whole test was videotaped for further analysis and performed with one test leader being present while the 2-years old child was fictional.

2.2 Result and Discussion

The outcome of the experiment was intuitively plausible. Hence, the test was only performed on 4 subjects to gain tangible evidence to use for the analysis. Would the result differ from the expected, we would consider testing more users. From the instructing test we have made the following conclusions:

- As expected, the subjects were more likely to both play slower and use more excessive arm movements than normal. This gives a minimum speed requirement for the vision module.
- None of the subjects beat the drums more than once. Therefore, for the same drum the robot will need to interpret one drumbeat as one and two as two etc.
- The majority of the subjects tried to make eye contact with the child before starting the sequence, as if they tried to ensure that the child was ready and watching, although the child was imaginary. The essence here is for the robot to show that it is ready to be learnt a sequence. We do meet this requirement as the robot is following the arm movements when standing in front of the camera; more on this issue in the coming sections.

As we had expected, all subjects felt troubled completing the sequence in the correct order in the absence of one of the two modalities. That reproducing a sequence of visually sensed movements would be difficult has already been shown in a more extensive work by Agam et al. (2005). Whether the same would apply to the human auditory system is intuitively plausible but no evidence is known of. As a last point for the current work this information may be irrelevant and consequently unnecessary to continue with.

2.3 Implementation

As previously mentioned, the robot software was implemented in Ikaros. The Ikaros modules used, are the Kinect™ proximity sensor module, the drummer module and the *"Dynamixel"* module. The sensor module outputs a matrix with depth information taken from the proximity sensor to the drummer module. The depth matrix generated by the proximity sensor is used by the drummer module for a variety of functions explained shortly. The drummer module outputs data to the Dynamixel module, which is used for controlling the angles and speeds for the 3 robot motors. What follows is a description of the practical procedure for goal emulation by the robot presented in this report.

After starting up the robot, fig. 1, the instructor is asked, through the text-to-speech engine, to present one of the three drums close to the camera. The size of the drum is measured with the Kinect™ depth sensor and, after asking the person to return the drum to its original position, the measured size is used to locate the $3D$-coordinates of all the three drums. The person is then asked to wave at the robot for it to start learning the forthcoming drumbeats. When the waving is recognized by the robot, it starts to follow the closest point above the drums, the instructor, and asks the instructor to start playing the drums with one hand. Feedback is provided to the instructor, every time a drum is hit, by playing different sounds, depending on what drum is hit. When the instructor is finished playing a sequence, the robot is pointed at by the instructor, with the pointing hand reaching closer to the robot than the drums' positions. The robot then imitates the playing sequence in the correct order and with the same tempo.

The robot is implemented as a state machine, presented in fig. 2, and the basic states are *watch*, *learn*, and *imitate*. In the state *watch*, the robot has already identified all three drums and is now waiting for the user to wave. The waving is identified by looking at the closest point in the space above the drums and recording extreme left and right positions. If four extreme positions in a row meet fixed requirements, such as being a certain horizontal distance apart and occurring at reasonable time intervals, a waving motion is said to have been identified. This has turned out to be a very stable solution.

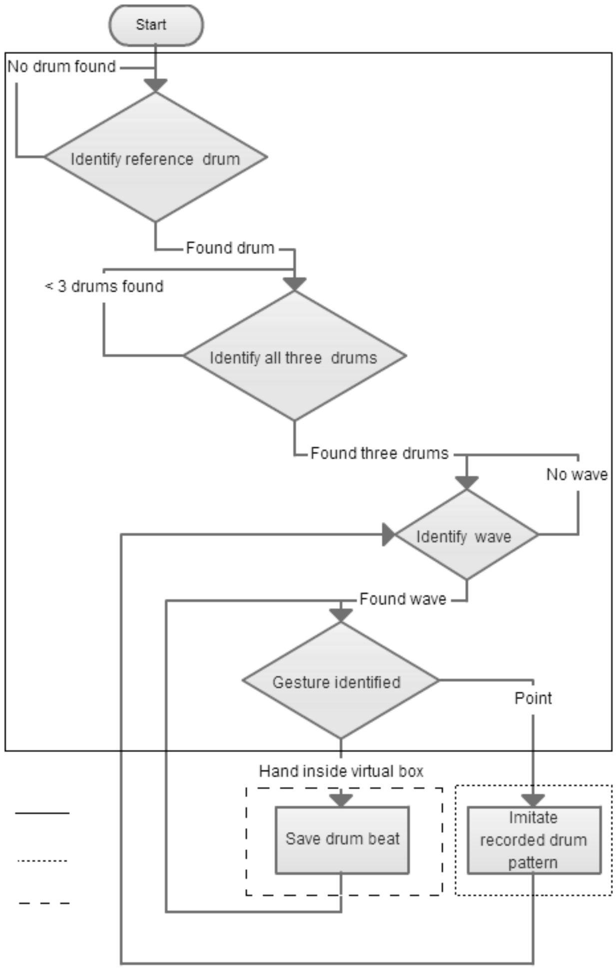


Figure 2. Flowchart for the drummer robot: —— Gesture Recognition - - - - Short-Term Memory (STM) .... Goal Emulation

After identifying the hand waving, the robot switches to the state *learn*. When an object—presumably the human drummer's hand—is detected within a "virtual box" above each drum, a hit is detected and the time and the drummer ID is saved into a list. A drum sound is also played for user feedback. When an object in detected on a certain distance in front of the drums, the robot switched to the state imitate.

In *imitate* the robot goes through the list of beats while beating at its own drums. The robot's three motors are controlled with different speeds in order to make the drumming look realistic.

The drums used by the robot were originally located using markers but in order to make the whole system independent of light conditions and the Kinect™ RGB-camera, the positions of the these drums were statically coded.

Two additional states implemented before entering the state watch, are *identify* and *finddrums*. In *identify*, the robot looks for an object close to the camera and measures its size. This is simply done by looking for remarkable sudden increases of distance in the depth matrix of the proximity sensor where the drum's edges are found. In state *finddrums*, the depth matrix is scanned for objects with the same size as the drum found in state identify. When a drum is found, that part of the depth matrix is no longer searched through. The search ends when three drums are found.

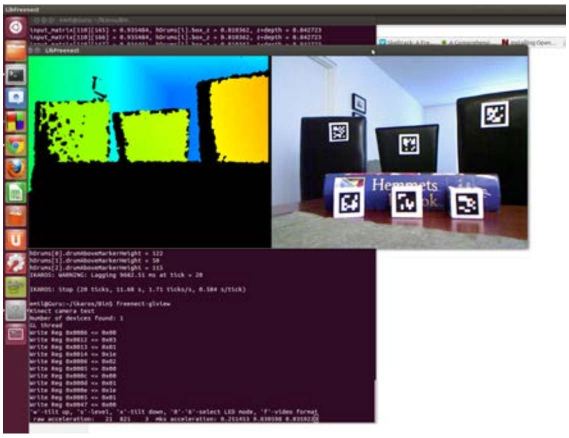


Figure 3. The three parts of the control panel interface: upper left window shows the result of the infrared proximity sensor-feedback, where the color interval red through purple indicate the distances 'close' through 'distant'. Upper right window shows the first setup with markers for locating both the robot's and the instructor's drums in the RGB image. The lower window is the terminal and the feedback text from Ikaros.

2.4 Verification Test

A verification test was carried out with 6 participants to verify the usability for any randomly chosen instructor. During the test, subjects who had never seen or interacted with the prototype before, were asked to follow the vocal instructions given by the robot and have it imitate their drum sequences.

2.5 Result and Discussion

The test showed that the 6 subjects testing the robot, didn't have but minor problems interacting with it. The choice of the drum size did not affect the result in a negative way, as long as all three were equi-sized. There were neither any preferences on the position of the drums under the condition that the distances to the robot did not vary remarkably and all three were fully "visible" by the robot. The test yielded valuable input on two elements during the evaluation test that needed correction:

2.5.1 "Please show me the drum"

To ask the instructor to present the drum, the robot utters the phrase *"Pick up one of the drums and show it to me"*. Since it employs the proximity sensor to identify all movements and drums, it needs the user to hold any of the three drums close to the robot, ahead of the other two. All subjects did however only pick up the drum and placed it above the other two, but never closer to the robot than it was at the beginning. A simple correction made to solve this problem was to have the robot ask the subject further: *"closer to the camera"*.

2.5.2 The gesture to indicate the end of the drum sequence

The instructor was asked in advance, through the robot's speech engine, to indicate the end of a sequence by making a gesture meaning "it is your turn". Given that inactivity of the instructor within a reasonable time interval would trigger the beginning of the imitation procedure automatically, the robot was programmed to only respond to the gesture where the instructor lowered his hands with open palms facing the robot. However, only half of the subjects made this gesture, while the other half pointed at the robot with one hand.

To solve the problem, the "pointing"-gesture could be included as an alternative to the "two-hands"-gesture. On the other hand, using several alternatives could trigger an unexpected reaction before the end of the sequence. Hence, the final decision was to use only one gesture as the standard and notify the subject in advance what the robot expects as the sign of the end of a sequence. Since pointing can be described in fewer words and easier to understand the subject is asked to "point at me when you are done!".

3 DISCUSSION

Based on the findings about mirror neurons and the developmental psychology, the authors postulate that the performance of the implementation presented in this report illustrates roughly the cognitive ability of a 2 year old child. This explains also why the subjects in the preparatory test were asked to imagine a 2 year old child as the subject for learning the drumbeat sequences.

Comparing the presented platform to a 3 to 4 year old child, one can easily conclude that its cognitive difference is that the learning ability of this robot is not epigenetic. In other words, what is learnt is an ad hoc, task specific, solution that cannot be generalized to other forms of imitation. The reason is, first of all, that the robot has no Long-Term Memory (LTM) to save the result for later use and hence lacks the mechanisms to provide the robot with the necessary ''experience". Given that such experience would be present, the change of task domain may involve change of parameters and functions; what the robot is not capable of carrying out, neither physically nor cognitively. Apart from the physical limitations, to be able to extend the task of imitation to other forms and situations, the individual needs an understanding of what imitation is. That knowledge comes in the later stages of a child's development; what is generally non-existent in the current robots.

Nevertheless, the robot is indeed imitating, because it ''notices" the intention and it, like a toddler, follows the movements and captures the moments that are crucial for reaching the intended goal. This is made possible by investigating the particular gestures a human instructor would make—when showing a child how to play the instruments—and providing the results in the form of predetermined positions to the robot, before the procedure of imitation begins; what was carried out in the 2.1. From this perspective, the information embedded as intrinsic knowledge in the robot, is comparable to the information the robots presented in section 1.1 obtained through statistical methods, as HMM and GMR, or kinesthetic learning methods.

There are several recurring tasks in robotics that the experimental, research projects try to solve, to show their robots' task performance skills. In the peg-insertion task, the intention of the action is the same, regardless of the trajectory. The paths or trajectories, may still vary, but since they only approximate their moves around the same statistical joint data they are provided, a robot cannot choose but what it has been ''taught".

In playing table tennis, the robot will have choices in both the goal and the trajectories, but of the same reason as above, the current solutions yield the same result as for the peg-insertion task.

In balancing a pendulum, the uniqueness of the end-point, or the intended point of equilibrium, forces the robot to follow exactly the same strategy to succeed with the task, leaving the robot with no choices but to follow exact trajectories, estimated usually through learning strategies as RL.

In the case of the robot in the current work, the task is similar to a pick-and-place robotic task—as a simple case of peg-insertion. The similarity between the two is that the end-point of the manipulator moves between predetermined positions, but the trajectories are free to decide by the robot. The difference between the two, is that in the former, contrary to the pick-and-place task, the robot has to ''remember" the order of the displacements.

Using HMM or other approximating methods may create some flexibility in the joint values or trajectories but there is still no free choice for the robot to decide its path. This way, the emphasis is on other factors than the goal of the movement. The current work, on the contrary, paves the way for creating a platform, a robot, that makes a general assumption that the movement itself is not essential and hence the robot will have the freedom to choose the most appropriate path with regard to its own physical properties, assumed that it is equipped with a learning method and even better if a long term memory is supplied to this robot system.

Fig. 2 shows that in the current solution, the robot pays attention to the movements of the instructor and emulates their short-term goals, by memorizing the tasks to carry on, and by performing them in the correct order. The robot is thus able to learn tasks from demonstration, only if it can optimize its own parameters (Mataric, 2007). Extracting the trajectories might give us the possibility to emulate the goals but following these trajectories may not be the best policy for performing a task. Hence, a mixture of our model and Calinon's, resulting in extracting goals and understanding the intention by statistical models, and then enhancing the trajectories to create new possibilities by employing learning algorithms as RL can be the optimal solution for imitation.

Through utilization of a well-developed vision system and a sound module, extracting information for recognition of the instructor and the drums, the robot would not be in need of being provided the position of its own and the instructor's drums. Presence of sound would remove the constraint that drums should be on a particular position and that those positions should be provided to the robot in advance. Furthermore, since the robot cannot ''hear" when it has been successful in striking on a drum, there is a risk that it misses a correct hit in the drumbeat sequence without noticing it. Besides, in case of real drums, the hits could be adjusted so as to acquire the same sound effect from each drum.

For the same reason, in absence of a sound module, the order of the drums are also constrained, in that the robot cannot find the correct sound in the sequence by trial and error, a problem that limits the robot learning to a great extent.

Generally, the shortfall of most of the proposed models known to the authors, including the present work, is that the robot is learning positions. If object recognition could be feasible to a great extent, the issue of searching the physical space would reduce to moving to or manipulating a particular object. But this is not the only hinder. Possessing the ability to recognizing objects cannot be limited to certain basic cues as size or color. Rather other properties as how they can be manipulated and would be necessary for the robot to be aware of.

4 Conclusion

In this report a robotic solution has been presented which successfully emulates goal-directed movements of a person playing on a set of 3 equi-sized drums. Although merely the goals of the movements, the drumbeats, are extracted and the hand trajectories are not followed, the rhythm is replicated and the melody is preserved, as expected. The drums and the position of the instructor are detected and the arm movements made by the person are effectively extracted and used for understanding their respective goals, the moments of drumbeats, and other human gestures necessary for a mutually advantageous communication with the instructor.

Preparatory tests with subjects helped in gathering information about possible movements that the robot should be aware of and what salient features the robot needed to extract for a successful goal emulation.

The evaluation tests has also confirmed that the system is robust to changes in the setup of the drums, their sizes, speed of the hand of the instructor, both in waving and playing the drums and that the verbal communication with the subjects creates an effective means for even better result.

5 Future Work

No need to say that the presented work can be further developed towards a more flexible and interactive platform. Combining the information from the color camera and the proximity sensor on top of more advanced algorithms together with speech recognition to promote the ability to recognize drums of different sizes and to recognize gestures and instructor commands are but a few development steps that can be taken. In addition to the issue of communication, creating a module for streaming music into Ikaros and analyzing the sounds of drumbeats will give the robot the ability to follow them by listening and to imitate the sequences by trying to find the sequences via trial and error, utilizing well-known learning algorithms.

6 Acknowledgements

The authors thank Prof. A. Gulz for coordinating the highly valued course, for which this project work was conducted. Deepest appreciations go to Prof. C. Balkenius and Dr. B. Johansson for invaluable discussions about and technical support of this project. Gratitudes are also expressed for contribution made by Prof. M. J. Mataric to the theoretical part of the report.

References

Agam, Y., Bullock, D., & Sekuler, R. (2005). Imitating unfamiliar sequences of connected linear motions. *Journal of Neurophysiol*(94), 2832–2843.

Argall, B. D., Chernova, S., Veloso, M., & Browning, B. (2009, May). A survey of robot learning from demonstration. *Robot. Auton. Syst.*, *57*(5), 469–483. Available from http://dx.doi.org/10.1016/j.robot.2008.10.024

Atkeson, C. G., & Schaal, S. (1997, April). Learning tasks from a single demonstration. *Proceedings of the 1997 DEEE*, 1706-1712.

Balkenius, C., Morén, J., Johansson, B., & Johnsson, M. (2010). Ikaros: Building cognitive models for robots. *Advanced Engineering Informatics*, *24*(1), 40 - 48. Available from http://www.sciencedirect.com/science/article/pii/S1474034609000494

Banich, M. T., & Compton, R. J. (2011). *Cognitive neuroscience*. Cambridge, Massachusetts: Wadsworth, Gengage Learning.

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *53*(1), 153-164.

Billard, A., Calinon, S., Dillmann, R., & Schaal, S. (2008). Survey: Robot programming by demonstration. In B. Siciliano & O. Khatib (Eds.), *Handbook of robotics* (chap. 59). Berlin Heidelberg: Springer-Verlag.

Billard, A., Calinon, S., & Guenter, F. (2006). Discriminative and adaptive imitation in uni-manual and bi-manual tasks. *Robotics and Autonomous Systems*, *54*(5), 370–384.

Billard, A., Epars, Y., Calinon, S., Schaal, S., & Cheng, G. (2004). Discovering optimal imitation strategies. *Robotics and Autonomous Systems*, *47*, 69-77.

Billard, A., & Mataric, M. J. (2000, June). A biologically inspired robotic model for learning by imitation. *Proceedings of the fourth international conference on Autonomous agents*, 373-380.

Billard, A., & Mataric, M. J. (2001, April). *Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture* (Vol. 37). Robotics and Autonomous Systems. Minneapolis, Minnesota: University of California Press.

Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561–567.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007, December 18). Investigating action understanding: inferential processes versus action simulation. *Current Biology*, 2117-21.

Breazeal, C., & Scassellati, B. (2002, November). Robots that imitate humans. *TRENDS in Cognitive Sciences*, *6*(11), 481-487.

Brooks, R. A., Breazeal, C., Marjanovic, M., Scassellati, B., & Williamson, M. M. (1999). The cog project: Building a humanoid robot. In *Lecture notes in computer science* (pp. 52–87). Heidelberg Berlin: Springer-Verlag.

Bullock, D., & Rhodes, B. (2002). *Competitive queuing for planning and serial performance* (Tech. Rep.). Boston, MA: Boston University for Adaptive Systems and Department of Cognitive & Neural Systems.

Calinon, S., & Billard, A. (2004, September). Stochastic gesture production and recognition model for a humanoid robot. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 2769–2774). Lausanne, Switzerland: Ecole Polytechnique Fédérale de Lausanne (EPFL).

Calinon, S., D'halluin, F., Caldwell, D., & Billard, A. (2009). *Handling of multiple constraints and motion alternatives in a robot programming by demonstration framework.* Proceedings of 2009 IEEE International Conference on Humanoid Robots, 2009, p. 582 - 588.

Calinon, S., Evrard, P., Gribovskaya, E., Billard, A., & Kheddar, A. (2009, June). Learning collaborative manipulation tasks by demonstration using a haptic interface. In *Proc. intl conf. on advanced robotics (ICAR)* (pp. 1–6). Lausanne, Switzerland: Ecole Polytechnique Fédérale de Lausanne (EPFL).

Dawson, G., & Adams, A. (1984). Imitation and social responsiveness in autistic children. *Journal of Abnormal ChiM Psychology*, *12*(2), 209-226.

Delson, N., & West, H. (1996, April). *Robot programming by human demonstration: Adaptation and inconsistency in constrained motion.* International Conference on Robotics and Automation. Minneapolis, Minnesota: University of California Press.

De Tommaso, D., Calinon, S., & Caldwell, D. G. (2012). A tangible interface for transferring skills using perception and projection capabilities in human-robot collaboration tasks. *Intl Journal of Social Robotics, Special Issue on Learning from Demonstration*.

Erlhagen, W., Mukovskiy, A., & Bicho, E. (2006, January 31). A dynamic model for action understanding and goal-directed imitation. *Brain Research*, *1083*, 174-188.

Fadiga, L., Fogassi, L., Pavessi, G., & Rizzolatti, G. (1995, June). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608-2611.

Garrod, S., & Pickering, M. J. (2004, January). Why is conversation so easy? *TRENDS in Cognitive Sciences*, *8*(1).

Gillmeister, H., Catmur, C., Liepelt, R., Brass, M., & Heyes, C. (2008). Experience-based priming of body parts: A study of action imitation. *Brain Research*, *1217*, 157-170.

Grossberg, S. (1978). A theory of human memory: Self organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), (Vol. 5, p. 233-374). New York: Academic Press.

Gärdenfors, P. (2006). *How homo became sapiens*. New York: Oxford University Press.

Hamilton, A. F. d. C., & Grafton, S. T. (2006, January 25). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, *26*, 1133-1137.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, *35*(0001), 1-31.

Hauser, M. D. (2000). *The evolution of communication*. Cambridge, Massachusetts: The MIT Press.

Head, H. (1920). Aphasia and kindred disorders of speech. *Brain*, *43*(2), 87-165.

Heyes, C. (2001, June). Causes and consequences of imitation. *TRENDS in Cognitive Sciences*, *5*(6), 253-261.

Horner, V., & Whiten, A. (2004, November). Causal knowledge and imitation/emulation switching in chimpanzees (pan troglodytes) and children (homo sapiens). *Animal Cognition*, *8*, 164-181.

Hovland, G. E., Sikka, P., & McCarragher, B. J. (1996, April). Skill acquisition from human demonstration using a hidden markov model. *Proceedings of the 1996 IEEE*, 2706-2711.

Iacoboni, M., Molnar-Szakacs, M., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005, March). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*(3), 529-535.

Kaiser, M., & Dillmann, R. (1996, April). Building elementary robot skills from human demonstration. *Proceedings of the 1996 IEEE*, 2700-2705.

Kilner, J. M., & Blakemore, S.-J. (2007). How does the mirror neuron system change during development? *Journal of Neuroscience*, 524-526.

Kilner, J. M., & Frith, C. D. (2008). Action observation: Inferring intentions without mirror neurons. *Current Biology*, *18*(1).

Kinect™ for windows®. (2012, November 18). Available from http://www.microsoft.com/en-us/kinectforwindows/

Leslie, K. R., Johnson-Frey, S. H., & Grafton, S. T. (2004). Functional imaging of face and hand imitation: towards a motor theory of empathy. *NeuroImage*, *21*(2), 601 - 607. Available from http://www.sciencedirect.com/science/article/pii/S1053811903005858

Mascolo, M. F., & Fischer, K. W. (2005). The cambridge encyclopedia of child development. In B. Hopkins (Ed.), *The cambridge encyclopedia of child development.* Cambridge, UK: Cambridge University Press.

Massen, C., & Prinz, W. (2009). Movements, actions and tool-use actions: an ideomotor approach to imitation. *Philosophical Transactions*, *364*, 2349–2358.

Mataric, M. J. (2007). *The robotics primer.* Cambridge, Massachusetts: The MIT Press.

Mataric, M. J., & Pomplun, M. (1998, June). Fixation behavior in observation and imitation of human movement. *Cognitive Brain Research*, *7*, 191-202.

McCall, R. B., Parke, R. D., Kavanaugh, R. D., Engstrom, R., Russell, J., & Wycoff, E. (1977). Imitation of live and televised models by children one to three years of age. *Monographs of the Society for Research in Child Development*, *42*(5), 1-94.

Meer, A. L. H. van der, Weel, F. R. van der, & Lee, D. N. (1995, February 3). The functional significance of arm movements in neonates. *Science*, *267*.

Meltzoff, A. N. (1993). The centrality of motor coordination adn proprioception in social and cognitive development from shared actions to shared minds. *The Development of Coordination in Infancy*, 463-496.

Meltzoff, A. N., & Moore, M. K. (1977, October 7). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 75-78.

Motallebipour, H. (2010). *An approach to a complete recognition module for legged robots in the robocup domain*. Master's thesis, Department of Computer Science, Faculty of Engineering, LTH, Lund University, Lund, Sweden.

Piaget, J., & Inhelder, B. (2000). *The psychology of the child* (H. Weaver, Trans.). New York: Basic Books.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188 - 194. Available from http://www.sciencedirect.com/science/article/pii/S0166223698012600

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*(1), 169 - C-4. Available from http://search.ebscohost.com.ludwig.lub.lu.se/login.aspx?direct=true&db=a9h&AN=13985119&site=ehost-live

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*, 661-670.

Rohrer, B., & Hulet, S. (2006). *A learning and control approach based on the human neuromotor system.*

Schofield, W. N. (1976). Do children find movements which cross the body midline difficult? *Quarterly Journal of Experimental Psychology*, *28*, 571-582.

Shettleworth, S. J. (2010). *Cognition, evolution, and behavior*. Oxford, NY: Oxford University Press.

Tomasello, M. (1990). Cultural transmission in the tool use and communicatory signaling of chimpanzees? In S. Taylor Parker & K. R. Gibson (Eds.), *'language' and intelligence in monkeys and apes: Comparative developmental perspectives* (p. 274-311). University of Texas: Cambridge University Press. Available from http://dx.doi.org/10.1017/CBO9780511665486

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993, December). Cultural learning. *Behavioral and Brain Science*, *16*, 495-552.

Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. C. (1993, December). Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Development*, *64*(6), 1688-1705.

Turella, L., Pierno, A. C., Tubaldi, F., & Castiello, U. (2009). Mirror neurons in humans: Consisting or confounding evidence? *Brain & Language*, *108*, 10-21.

Viviani, P., & Stucchi, N. (1992). Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 603-623.

Vogt, S. (1995). On relations between perceiving, imaging and performing in the learning of cyclical movement sequences. *British Journal of Psychology*, *86*, 191-216.

Vogt, S. (1996). Imagery and perception-action mediation in imitative actions. *Cognitive Brain Research*, *3*, 79-86.

Want, C. S., & Harris, P. L. (2002). How do children ape? applying concepts from the study of non-human primates to the developmental study of 'imitation' in children. *Developmental Science*, *5*(1), 1-41.

Whiten, A., & Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. In P. J. B. Slater, J. S. Rosenblatt, C. Beer, & M. Milinski (Eds.), *Advances in the study of behavior* (Vol. 21, pp. 239–283). New York: Academic Press.

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Ankungen, ankmamman och krokodilen - en jämförelse mellan konsekvensfeedback och rätt/fel-feedback i ett lärspel för förskolebarn

Ann-Louise Andersson   Caroline Alvtegen   Dan Thaning   Emil Nikander

1 Inledning

Projektet går ut på att vi ska designa ett spel för barn där de ska lära sig att förstå talsystemet. I detta projekt ligger tyngdpunkt på den feedback som användaren erhåller vid spelande. Vår uppgift är att jämföra olika typer av feedback och undersöka om olika feedback har olika påverkan på användaren.

Frågeställningar som vi jobbar utifrån är följande: Har rätt/fel-feedback respektive konsekvensfeedback olika påverkan på barnets lärande och hur gestaltar sig i så fall denna påverkan? Hur kan man på ett bra sätt implementera rätt/fel-feedback respektive konsekvensfeedback i ett lärospel för småbarn?

Målet med projektet är att på bästa möjliga sätt kunna hjälpa barns lärande utifrån att ge bra feedback. Barns matematiska samt kognitiva förmåga är olika utvecklad om man bland annat ser till ålder. Det finns andra aspekter som spelar roll, t.ex. sociala aspekter men avgränsningar har behövts göra på grund av tidsbrist.

Arbetet har fortlöpt över två kurser, som legat efter varandra under hösten 2012. Under första delkursen skapades en spelidé, en design och två olika prototyper utvecklades, en LoFi och en HiFi. Under andra delkursen utvecklades HiFi-prototypen fullständigt, användartester utfördes och en film skapades.

2 Bakgrund

Inom det här ämnet har det inte forskats särskilt mycket eller utvecklats bra lärospel. Det har därför varit svårt att få en stabil fakta bakgrund att utveckla utifrån. Vi har däremot valt att testa vår prototyp och utifrån resultatet dra slutsatser om hur olika feedback kan hjälpa barn att lära sig siffror och talsystemet bättre.

Det man vet är att barn ofta har lätt att lära sig saker vid tidig ålder, då hjärnan utvecklas [1]. Forskare har visat att vid fyraårsåldern kan barn börja jämföra storlek och börja med grundläggande sifferräkning. Vid denna ålder kan inte barn se en koppling mellan storlek och siffror. Först vid fem/sex års ålder kan barn börja koppla samman storlek med siffror, vilket är revolutionerande. För att en fyraåring ska kunna koppla samman storlekar med siffror kan man använda sig av fysiska objekt.

Tyngdpunkten i vårt spel, ligger som nämnt, på feedback till användaren. Att använda sig av feedback som lärandetyngdpunkt är även det någon nytt. Många spel på marknaden har använt sig av feedback till användaren för att användaren ska förstå hur man ska fortsätta spela spelet eller att förstå vad som händer i spelet. Feedback används dagligen i lärandeprocesser i det verkliga livet.

Att utveckla spel med tyngdpunkt på feedback kan göra att samhället kan dra slutsatser av olika feedbacks-typer man kan använda sig av för att bättre stödja barns lärande.

2.1 Feedback

Feedback är en process där information om det som varit eller det som är påverkar samma fenomen i nutid eller framtid. Man kan se feedback som en kedja av orsaker och verkan [2]. Feedback i tekniska sammanhang kan ses som en input till användaren, där användaren får information om en händelse utifrån ett val som användaren gjort. I alla tekniska system är det viktigt att användaren får feedback för att förstå att något hänt, vad som händer och hur man ska kunna gå vidare i systemet. Detta är för att användaren ska vara med på vad som händer, och ha en förståelse för de val som gjorts.

I vårt fall handlar det om att ge användaren feedback på om man har löst uppgiften eller inte. Det finns flera olika typer av feedback, och kan ges på flertal sätt. I detta projekt är det avgränsat till rätt/fel-feedback och konsekvensfeedback.

2.1.1 Rätt/fel-feedback

Denna typ av feedback ger användaren konkret information om att hen gjort rätt eller fel men användaren får ingen information om varför det är rätt eller fel. Detta kan resultera i att användaren vid nästa tillfälle inte vet hur hen ska göra för att det ska bli rätt, då användaren inte vet vilket konsekvens ett val har utlopp för utan bara om valet var rätt eller fel. Man kan ge användaren feedback om att valet var rätt eller fel på olika sätt, antingen genom en röst som säger fel eller rätt. En text kan komma upp med ordet rätt eller fel. Man kan välja att jobba med färgerna rött och grönt, då de symboliserar antingen fel eller rätt. Men användaren får ingen information om vad som var fel och vilket resultat eller konsekvens valet fick.

En koppling till barns skolgång kan vara när barnet får reda på resultatet på ett prov och enbart får reda på poängsumman och inte vad barnet gjort rätt eller fel. Rätt/fel-feedback är den vanligaste typen av feedback i lärande processer.

2.1.2 Konsekvensfeedback

Konsekvensfeedback betyder att användaren får se, till exempel under spelets gång, hur hennes val påverkar resultatet. Användaren får alltså se resultat eller konsekvenser av det val som utfördes under det givna tidspannet, istället för att bara få reda på om det var rätt eller fel.

Till exempel om man ska hälla upp rätt mängd vatten i ett glas kommer vattnet svämma över om man häller i för mycket. Denna typ av feedback kan hjälpa användaren att förstå vad man gjorde fel och då lättare förstå vad han ska göra för att det ska bli rätt nästa gång.

Konsekvensfeedback är mindre vanligt som lärande metod i skolan, men används mycket inom spelindustrin.

2.2 *Number Sense*

Number sense i matematik undervisning kan ses som en förståelse av tal, storleken av talet, relationer mellan tal och hur de påverkas av beräkningar [3]. När man talar om *number sense* lämnar man det traditionella sättet att undervisa matte, där algoritmer och operatoner är i fokus, utan i *number sense* är förståelsen mellan siffror och storlek i fokus. Lärare kan ställa frågor som, vad är ett tal?, när man pratar om *number sense* istället för att fråga om barnet förstår reglerna som gäller när man ska utföra operationer på siffror. Med *number sense* lär man säg att sätta sifforna och tal i relation till varandra. *Number sense* lär barnen att förstå vilket av talen 9+2 eller 7+3 som är störst, istället för att matematik uträkningen är i centrum. Med att matematiken är i centrum menas att man förstår symboliken för addition och subtraktion för att sedan kunna utföra operationen på talet. Med *number sense* ser barnen på siffrorna och kan tolka storlek.

Number sense är något som utvecklas hos barnet vid en tidig ålder. Forskare har sett att spädbarn kan särskilja ett sett av två objekt från ett sett med tre objekt. *Number sense* hos barn kan dock vara olika, beroende på olika förhållanden. Till exempel om barnet får hjälp hemifrån eller från skolan att utveckla och träna [1]. Forskning har visat att vid fyra års ålder kan barn börja räkna samt att se storlekar. Vid en ålder av fem och sex börjar barnen att kunna slå samman räkning och storleks.

3 Arbetsgång

3.1 Arbetsgång första delkursen

Då vi var en helt ny gruppkonstellation fick vi ta lite tid till att lära känna varandra. Detta för att arbetet skulle kunna flyta på bra under kursens gång. Vi diskuterade vad vi önskade att få ut av kursen, vilka kunskaper vi hade från tidigare och vad vi gärna vill lära oss. Undertiden läste vi på artiklar som berörde området vi skulle komma att jobba med. Resultatet blev att vi kunde arbeta utifrån en gemensam kunskapsgrund, samt att arbetet kunde börja från en union utgångspunkt. När vi kände att vi hade lite kött på benen började vi brainstorming på olika spelidéer som vi kunde jobba utifrån. Brainstormingen gick till så att samtliga grupp medlemmar satt och diskuterade olika förslag som samtidigt antecknades ner. Brainstormingen var fri, vilket resulterade i att den var väldigt kreativ.

Två förslags togs fram, ett som vi gillade mer och ett som backup. Vi diskuterade de två förslagen med våra handledare där vi valde att fortsätta arbetet med huvudförslaget. Den konceptuella designen kunde nu ta fart. Vi valde att rita på papper samtidigt som vi diskuterade olika förslag på lösningar. Vi använde oss av ritningar i denna process för att alla grupp medlemmar skulle ha samma visuella bild i huvudet så att vi arbetade åt samma håll, det är annars lätt att missförstånd uppstår och att man inte diskuterar samma sak, då man har olika bilder i huvudet av vad som menas.

Sedan utformades en LoFi-prototyp i Powerpoint. Powerpoint valdes som verktyg därför att samtliga medlemmar var familjär med verktyget, tiden på att lära sig ett nytt verktyg kunde reduceras, och samtliga funktioner kunde implementeras som vi behövde. Vi gjorde mindre tester och utvärderade prototypen för att sedan kunna förbättra den, dock är det svårt att testa en Powerpoint då den är väldigt begränsad i det den kan göra.

Nästa steg var att utveckla HiFi-prototypen. Vilken utvecklades i JavaScript och HTML5, detta för att basen för mWorld spelet utvecklas i det programmeringsspråket. Prototypen implementeras bara på en grundnivå, vilket betyder att enbart en spelnivå samt ramverk för hur spelet ska vara uppbyggd och grafiken. Anledningen till det valet är att vi hade tidsbrist för att hinna implementera allt denna delkurs, och för att vi ansåg att det inte var en nödvändighet att ha en helt klar prototyp. Nästa delkurs skall den fullständiga versionen implementeras med flera spelnivåer. Prototyp ska även testas utförligt i nästa delkurs, för att kunna jämföra de olika feedbacktyperna på den valda målgruppen. I detta stadium har vi koncentrerat oss på att utveckla de grafiska delarna av HiFi-prototypen.

3.2 Arbetsgång andra delkursen

När del två av kursen drog igång tappade vi tyvärr en gruppmedlem. En konsekvens blev att nya roller delades ut och nya ansvarsområden gavs. Det viktigaste var att vår HiFi prototyp blev klar och att vi kunde börja testa. Vi började med att utvärdera vad som behövdes göra från delkurs 1 och vad vi ville få ut av delkurs 2. Samtidigt som HiFi'n utvecklades började vi fundera ut vilka tester som ska göras och hur vi vill göra dem. Under läsvecka tre var de två olika speltyperna färdigutvecklade och testning kunde ske på en förskola. Därefter analyserades resultatet av testningen och vi drog slutsatser. Vi diskuterade vad vi kunde utveckla med spelet för att få det ännu bättre. Därefter spelades en film in som ska visas under slutredovisningen. Arbetet under denna delkurs flöt även den på väldigt bra och alla gruppmedlemar jobbade på. Då alla tog lika mycket ansvar stötte vi inte på några större problem eller större svårigheter samt att alla gruppmedlemar var involverade i samtliga delar vilket resulterade i att om man stötte på ett problem kunde alla gruppmedlemar diskutera och komma fram till en gemensamlösning.

Vid början av kursen satte vi upp en tidsplan med interna deadlines som följdes mycket bra. De interna deadlines hjälper till att planera arbetet och få en jämn arbetsbörda. Då vi hade olika fokus områden i gruppen blev det så att några i gruppen arbetade mer i början av kursen, och de andra tog mer ansvar i slutet.

V.1	V.2	V.3
Läsa på		Konceptuelldesign
Lära känna varandra	Brainstorming	
V.4	**V.5**	**V.6**
Konceptuelldesign	Mittredovisning	Hifi-prototyping
Lofi- prototyping		
V.7		
Hifi-prototyping		
Slutredovisning		

Figur 1: Tidsplan första delkursen

V.1	V.2	V.3
HiFi		Användartest
Utvärdering från del 1	Fundera ut tester	
V.4	**V.5**	**V.6**
Analys av test	Rapport skrivning	Slutredovisning
Film		

Figur 2: Tidsplan andra delkursen

4 Spelidé och kartläggning av målgrupp

Spelet går ut på att användaren ska lära sig siffror samt kunna koppla samman en siffra med rätt antal föremål. Vi valde att utforma spelet utifrån en berättelse där en anka har trillat ner i ett hål i marken när en ankunge var ute på promenad med sin mamma. För att kunna rädda ankungen måste användaren välja rätt antal hinkar med vatten så att ankungen kan flyta upp till ytan. En hink vatten innehåller storleken ett, motsvarande siffran ett på en tallinje. När användaren har valt rätt antal hinkar med vatten för att ankungen ska kunna åka upp från hålet, kan ankungen komma upp till sin mamma.

Två spel typer har utformats. Det som skiljer spelen åt är vilken typ av feedback användaren får när val har utförts. Anledningen till att det är det enda som skiljer sig, är för att det enbart är vilken feedback som användaren får som ska påverka resultatet. Ena speltypen använder sig av rätt/fel-feedback medan den andra typen ger konsekvensfeedback. Annars är de två speltyperna precis likadan eftersom att vi ska kunna göra en jämförandestudie utan att ta hänsyn till övriga skillnader.

4.1 Kartläggning av användarna

Spelet riktar sig till barn i åldern fyra till sex år. Vid fem års ålder börjar man lära sig och förstå kopplingen siffror till storlek, utan att behöva räkna på föremål. Därför har vi valt att inrikta oss på fyra-åringar, då vi använder oss av föremål som barnen kan räkna för att kunna koppla till rätt siffra på tal-linjen. Om det visar sig vid utförlig testning på målgruppen att det är för svårt för en fyra-åring att utföra kopplingen mellan räkningen av föremål till en siffra, får vi tänka om och gå upp en åldersgrupp. Meningen med spelet är att barnen ska lära sig kopplingen mellan en storlek och en siffra, vilket betyder att användaren inte behöver ha någon kunskap om detta från början.

5 Konceptuell design

När användaren startar spelet hörs en berättarröst som berättar en historia om en ankunge och dess mamma som är ute och promenerar. Ankungen går lite före mamman och råkar trilla ner i ett hål i marken. Berättar rösten kommer att vara med under hela spelet där information behöver ges, då barn vid den givna åldern i allmänhet har svårigheter att läsa.

Rösten förklarar att användaren måste rädda ankungen genom att välja rätt antal hinkar med vatten. Samtidigt som rösten säger detta visas hinkar på skärmen så att barnet förstår att de ska klicka på dem.

På högersidan av skärmen visas mamman som står och väntar på sitt barn. På den vänstra sidan av skärmen finns en krokodil, som ankungen kan åka ner till om för mycket vatten hälls ner i hålet.

Bredvid hålet visas tydliga siffror som beskriver hur många hinkar som behövs för att ankungen ska kunna flyta upp till ytan, genom en tallinje. Spelet fortsätter sedan på olika sätt beroende på om det är rätt/fel eller konsekvensfeedback som används. Spelet bygger på olika nivåer. Nivåerna består av olika höjder som man ska hjälpa ankungen upp från, när användaren har klarat en nivå, kommer användaren till nästa nivå där en ny höjd ska lösas. Vi har även tankar på att det går att utveckla fler nivåer där hinkarna t.ex. kommer byta plats så hinkarna inte är i ordning. Vilket kan göra det svårare för användaren eftersom hen kanske har lärt sig ordningen utantill utan att egentligen förstå den.

5.1 Rätt/fel-feedback

Det som skiljer sig från konsekvensfeedback är enbart den information användaren får när ett val gjorts, resten är exakt likadant. När användaren väljer ett antal hinkar kommer rutan runt hinkarna direkt att markeras med röd eller grön färg, beroende på om det är rätt eller fel val. Beroende på valet som gjorts kommer en ledsen eller glad anka visas på skärmen, med en tillhörande ljudeffekter för att ytterligare förtydliga för användaren att den gjort rätt eller fel val. Har användaren valt rätt kommer man till nästa nivå av spelet, har den däremot valt fel kommer berättarrösten tala om att användaren ska försöka igen, tills användaren klarat den nivån. Färgerna grön och röd används för att färgsymboliken är rätt respektive fel [4].

5.2 Konsekvensfeedback

Konsekvensfeedback bygger på att användaren får se konsekvensen av valen som gjorts. När användaren har valt ett antal hinkar kommer rutan runt hinkarna markeras med en neutral färg bara för att visa vilken som är vald och att användaren får en känsla av att ett val har gjorts.

Om användaren har valt för få antal hinkar kommer vattenytan stiga till den nivå som motsvarar antalet hinkar, ankunge kommer att flaxa med vingarna men inte kunna komma upp till sin mamma. Om användaren däremot väljer för många hinkar kommer vattnet svämma över och ankungen kommer åka ner mot krokodilen.

När rätt antal hinkar väljs kommer vattennivån att stiga lagom mycket, till den högsta nivån och ankungen kan gå hem till sin mamma. Precis som i fallet med rätt/fel-feedback kommer användaren till nästa nivå av spelet om den lyckats rädda ankungen, och har man inte lyckats kommer berättarrösten tala om att användaren ska försöka igen.

6 Prototyper

Två olika prototyper utvecklades under kursens gång, en LoFi samt en HiFi.

6.1 LoFi prototyp

För att utforma en LoFi prototyp finns flera olika verktyg. Vi valde mellan att enbart använda papper och penna, använda onlineverktyget balsamiq eller att jobba i Powerpoint. Vi valde att använda Powerpoint. I Powerpoint fanns alla de funktioner vi önskade använda oss av för att skapa en LoFi prototyp. Andra aspekter som vägde tungt när valet gjordes var att vi redan hade goda kunskaper i hur programmet fungerade och tid behövdes inte läggas ner på att lära sig programmet. Vi valde även att utforma grova utkast med papper och penna innan vi började jobba med Powerpoint. Papper och penna är ett bra verktyg för att samtliga medlemmar i gruppen ska jobba utifrån samma visuella bild och att inte missförstånd ska kunna ske.

Vi valde bort att arbeta med balsamiq mest för att våra behov täcktes av vad en Powerpoint kunde göra och tid krävdes för att lära sig ett nytt verktyg. Balsamiq motsvarade inte heller våra förväntningar på vad vi trodde man skulle kunna utforma.

6.2 HiFi prototyp

Ett huvudmål i projektet för oss var att kunna utföra en studie för att kunna utvärdera konsekvens- kontra rätt/fel-feedback. Redan tidigt i projektet började vi undersöka olika tekniker för en tilltänkt HiFi-prototyp som på ett smidigt sätt som möjligt skulle kunna möjliggöra insamlingen av data. Till en början diskuterade vi Java som plattform som har fördelen att det kan köras på Windows, Mac, Linux osv. Dessutom hade vi befintlig kunskap om Java. Men vad vi sedan fastnade för var HTML5/Javascript som gör själva insamlingen av data mycket smidigare. HTML5/Javascript är en teknik som körs ifrån en webbserver. Sedan kan mobiltelefoner, surfplattor och datorer spela spelet via webbläsaren medan olika speldata lagras på webbservern.

När vi väl hade helt klart för oss i LoFi-prototypen detaljer kring vad som ska hända i spelet började vi titta på hur vi skulle kunna implementera våra spelscenarior. Alternativet att programmera spelet i programkod från grunden föll ganska tidigt bort då vi bedömde att vi inte skulle ha en chans att komma särskilt långt den vägen givet projektets omfattning. Dessutom hade vi ganska begränsad kunskap av just HTML5/Javascript. Vi började istället leta efter olika spelutveckling verktyg som förenklar utvecklings processen.

Parallellt med att möjligheterna för den tekniska implementationen undersöktes började även några i gruppen att ta fram de grafikdelar som skulle behövas för spelet. Detta gjordes med Adobe Photoshop CS5. Efter en del experimenterande med pixelgrafik var det dock tvunget att bytas ut mot slutet av projektet mot vektorbaserad då det blev snyggare och fick jämnare kanter när olika objekt förstorades/manipulerades.

Vidare testades ett par olika intressanta program för spelutveckling i HTML5/Javascript, däribland Game Salad, GameMaker och Construct 2.
Constuct 2 från https://www.scirra.com/ blev det vi fastnade för, eftersom det var gratis och verkade ha en någorlunda bra Community att luta sig mot. Vi var dock hela tiden öppna för att eventuellt byta om någon del av spellogiken skulle vara för krånglig att implementera. Construct 2 verkade dock mer än tillräckligt bra för våra krav, även om det krävdes mycket tid att sätta sig in i ett helt nytt program. Ofta fanns det även inte information om det vi sökte vilket ledde till att vi fick experimentera oss fram. I och med detta märkte vi att när man inte vet hur man ska lösa ett tekniskt problem på bästa sätt så introducerar man flera buggar osv. Vi delade därför upp utvecklingen av spelets i mindre beståndsdelar, moduler, där vi experimenterade fram en fullt fungerande modul, för att sedan sy ihop modulerna till det som till slut blev själva spelet. I varje modul så låg fokus på att få en viss funktionalitet/rörelse osv av simpla bilder (boxar och fyrkanter). Efterhand lades den i Photoshop framtagna grafiken in och experimenterades med.

En stor utmaning var att på ett snyggt sätt få vattnet att svämma över om för många hinkar valts. Olika alternativ testades och till slut fick vi animera översvämningen bild för bild. Vid större förståelse av mekaniken i programmet blev valet att implementera introsekvensen direkt i spelet med hjälp av så kallade "Way-points", riktningar i vilken en figur ska förflytta sig i. Dessa kunde sedan appliceras på ankungen och mamman för att få det rörelseschema som vi önskade.

Vissa problem uppstod vid utformningen av HiFi prototypen vilket medförde att vi inte kunde göra tallinjen dynamisk utan att flytta de flesta övriga objekt och göra nya animationer för översvämningen när ankungen åker ner till krokodilen för varje nivå om vi ville behålla nollans position. Och om vi istället valt att ändrat utgångsnivån för ankan och tallinjen hade vi även i det fallet behövt omarbeta animationen för översvämningen för varje nivå. En nivå i hålet är lika djup som fyra nivåer, vilket inte var tanken, men på grund av begränsad tid fick det bli så ändå vilket man får ha i åtanke i resultatet.

7 Designval

Vi valde att ha designen så pass enkel som möjligt och att inte ha med för mycket onödig information på skärmen så att användaren inte blir distraherad. Vi har endast med det som är relevant för spelet.

7.1 Menyer

Det vi har haft mest funderingar över är hinkarna, tallinjen och hur rätt/fel-feedback ska visas upp på bästa sätt. När det gäller hinkarna hade vi olika idéer om hur hinkarna skulle placeras och grupperas.

En hink motsvarar storleken ett. Hinkarna är även grupperade i rutor samt formaterade i klungor för att användaren ska kunna få en uppfattning vilka hinkar som hänger ihop, samt kunna kopplas till en storlek. Då tre hinkar är grupperade och formaterade som en klunga motsvarar det siffran tre, som i sin tur höjer upp ankungen tresteg i hålet.

För att hinkknapparna ska se klickbara ut har vi valt att skugga dem lite så att det ser ut som att de sticker ut lite från skärmen, och där av ser ut som att man kan klicka på hinkarna. Att vi även har delat in hinkarna i fyrkanter gör att de får utseendet som en knapp. Lagen om likhet [5], samt lagen om närhet gör att användaren förstår att hinkarna hänger ihop och formaterar en meny att välja från. Valet av att ha hinkarna i övre delen av skärmen är för att barnet ska få en inre bild av att hinkarna hälls ner i hålet.

Andra förslag som diskuterades gällande hinkarna var att bara ha dem en och en, där användaren skulle få klicka på rätt antal hinkar som behövdes för att rädda ankungen, till exempel om det behövdes tre hinkar för att rädda ankungen upp till sin mamma klickar man på tre stycken av de enskilda hinkarna. Vi valde bort det alternativet bland annat för att då skulle kopplingen mellan storlek och siffra försvinna eftersom användaren inte får en tydlig bild över antalet när de är placerade en och en istället för att vara tydligt grupperade i antal.

Kopplingen mellan tallinjen och hinkarna är en annan del i designen som analyserades. Tallinjen går från 0 till ett högre nummer, nedifrån och upp, den växer alltså. Tanken med det är att användaren ska kunna få en känsla av att storleken på siffrorna växer uppåt. Det kommer alltid att vara en storlek större till nästa siffra på tal-linjen, 1 2 3 4 osv. Detta för att det blir för svårt för användaren att kunna koppla ihop siffror med storlekar om information saknas i tallinjen. Vi tror att användaren ska ha nytta av att hinkarna är sorterade i växande ordning liksom att tallinjen är växande.

7.2 Figurer

Vid användning av figurer i spel ska man ha i åtanke att barnen använder sig av socialkognitiva strategier, förhållningssätt och tolkningar barnen har inför andra människor på figurerna i spelet. Ankungen är vänd åt mammans håll när den befinner sig i hålet, det är för att användaren ska få en känsla att ankungen vill upp till sin mamma, och där av bli mer motiverad att hjälpa ankungen. Att vi har valt att ha en krokodil nere i högra hörnet är för att användaren ska få en känsla av fara om ett för högt antal hinkar väljs, vilket även det ska fungera som en motivation för användaren att välja rätt antal så att ankungen kan räddas. Om rätt annat hinkar väljs kommer ankungen upp på rätt nivå och kan ta sig till sin mamma, ankungen hoppar runt mamma och blir glad för att förstärka den positiva känslan av att rätt valt har gjorts.

Vi har valt att använda oss av figurer som är enkla och tydliga. Det har vi gjort för att användaren ska kunna fokusera på att lära sig hur talsystemet fungerar och inte ha för många distraktioner. För att få en figur att vara enkel är det bra med bredare och tydligare linjer runt figurerna samt att det inte visas mer än det väsentliga. Att användaren ser att hinkarna är fyllda med vatten gör att användaren kan koppla att vattnet i hinkarna kan höja vatten nivån på vattnet i hålet. Vi valde att använda oss av ankor och krokodiler är för att det är djur som barn känner till och är vanligt förekommande i böcker.

Vid design av figurer i spel finns det olika aspekter att ta hänsyn till, bland annat genus. Vi har här valt att ha en manlig ankunge och en mamma som vuxen karaktär. Vi har i spelet valt att använda oss utav könsrollen och samhälletnormer om att det är mamman som större delen av tiden tar hand om barnet. Krokodilen i spelet är valt som en könsneutral figur. Vi har dock varit tvungna att bortse från hur detta påverkar barnet i spelandet.

7.2.1 Rätt/fel-feedback

De figurer som dyker upp när rätt respektive fel val har gjorts, glad eller ledsen anka. Detta är för att förtydliga om användaren gjort rätt eller fel val. Vi har använt oss av känslor på ankungen för att användaren ska kunna relatera till om något är bra, ankungen blir glad för att den får komma till sin mamma, eller något är fel, ankungen blir ledsen för att den inte får komma till sin mamma. Barn kan tidigt känna känslor och om något är bra eller inte, barn i vår åldersgrupp har dock inte utvecklat ett bra läsande, och därav är text ett olämpligt val som feedback. Ljudeffekter, som ges, är till som en förstärkning av ankungens känsla för att förtydliga om valet var rätt eller fel. Att använda sig av renodlad rätt/fel-feedback är svårt då det blir lätt att man använder sig av förmycket känslor vilket kan göra att man tar in lite konsekvensfeedback och det inte blir renodlad rätt/fel-feedback, detta har vi även haft i åtanke.

7.2.2 Konsekvensfeedback

Användaren kan se att vattnet höjs när hinkar har valts genom att vatten hälls ner i hålet och vattnet i hålet då naturligt stiger. Det blir en naturlig feedback för användaren [5]. Det som även skiljer speltypen med konsekvens från rätt/fel-feedback är att om användaren väljer för få antal hinkar, ska ankungen börja flaxa. Det är för att användaren ska få se att ankungen försöker ta sig upp till sin mamma, men att det inte går och mer vatten behövs alltså hällas i. Att använda sig av flaxande som kroppsspråk hos ankungen när han inte når hela vägen upp till sin mamma vid fel val, kan ses som en naturlig mappning genom att man använder sig av ett kroppspråk som en anka skulle använda sig av in det verkliga livet när ankan försöker ta sig till en högra höjd.

7.3 Färger

Färgvalen som har gjorts i spelet är utifrån hur verkligheten ser ut. Ankorna som är med i spelet liknar till exempel en bad-anka, som i sin tur oftast är gul. Vi har valt att ha färgerna verklighetskopplade för att användaren inte ska behöva fundera på vad alla symboler och figurer betyder för något, utan att en koppling redan finns där. Att hinkarna är rosa är för att inget annat i spelet är rosa, rosa är i sin tur en färg som sticker ut lite. Hinkarna är en central del av spelet och ska tydligt synas.

7.3.1 Rätt/fel-feedback

I rätt/fel-feedback ändras färgerna på bakgrunden bakom hinkarna när ett val har gjorts. Vi valde färgen grön för ett korrekt val. Grön symboliserar att något är rätt eller att man kan fortsätta [4]. När fel val görs ändras färgen bakom hinkarna till rött. Rött symboliserar bland annat att något är fel eller farligt. Färgändringen ger även feedback på att ett val har gjorts.

7.3.2 Konsekvensfeedback

När ett val av hinkarna görs i konsekvensfeedback har vi valt att hinkarna ska ändras till en neutralfärg, som liknar hur det såg ut från början, när val inte gjorts. Det ska enbart symbolisera att ett val har utförts, feedback på att en knapp är ned tryckt. Vi valde att göra bakgrunden till mörkare grå, för att användaren ska få en känsla av att knappen har tryckts in.

7.4 Ljud

Under spelet kommer användaren att höra en bakgrundsmusik. Bakgrundsmusiken är enbart instrumental och drar ingen större uppmärksamhet till sig. Bakgrundsmusiken är till för att användaren ska fokusera mer på spelet. Under spelets gång är det även en berättarröst som informerar användaren om vad som händer. När spelet börjar berättar rösten en historia som ska engagera användaren. Vi har valt att göra spelet till en historia för att barn lyssnar gärna på historier och för att det ska bli lättare att leva sig in i vad som händer på skärmen. Resultat blir att användaren blir mer engagerad i spelet.

När ankungen har trillatner i hålet berättar rösten att användaren ska välja rätt antal hinkar för att rädda ankungen genom att hjälpa den upp till sin mamma. Rösten fungerar som en förstärkande faktor för att hjälpa barnet med att förstå uppgiften. Rösten berättar även för användaren om fel val gjorts att man ska testa igen, eller om rätt val gjorts, att användaren har nått en ny nivå.

Ljud är ett bra komplement att använda istället för att ha text som barnen ska läsa. Barn i fyra-årsåldern är oftast inte väldigt bra på att läsa.

8 Användartest

Testning utfördes på en förskola i norra Lund med barn mellan åldrarna 3-5 år. Testgruppen bestod av 10 barn, varav 5 personer provade rätt/fel spelet och 5 personer provade konsekvens spelet. Det behövs generellt minst 10 test personer för att kunna utskilja indikationer och kunna dra slutsattser [8].

8.1 Syfte

Syftet med testet var att jämföra rätt/fel-feedback och konsekvensfeedback genom att undersöka om våra två speltyper kommer påverka barn på olika sätt.

8.2 Problemformuleringar

- Har de olika feedbacktyperna olika påverkan på barnen?
- Hur gestaltar sig i så fall denna påverkan?

8.3 Testmiljö och utrustning

Testerna ägde rum på barnens skola för att det ska bli så tryggt och smidigt som möjligt för dem. Vi hade två stationer i skilda rum, en för vardera speltyp. Barnen fick en och en komma in för att testa spelet, en lärare fick närvara om så önskades. Testet utfördes på en laptop med tillhörande mus.

8.4 Upplägg vid genomförande

När vi först anlände till förskolan presenterade vi oss och var med på samlingen så att barnen fick se vilka vi var och lära känna oss lite, vilket var bra för att barnen inte skulle tycka det var så läskigt att sedan testa vårt spel.

När testsessionen startade fick barnen först en kort beskrivning av vad som skulle hända och vad de ska göra. Sedan startade spelet och de fick spela fyra nivåer, varje nivå spelades om tills att rätt val hade gjorts. När spelet var slut fick barnen svara på några enkla frågor om vad de tyckte om spelet.

8.5 Rollfördelning

Vid varje testsession medverkade endast två personer eftersom barnen inte skulle känna sig uthängda. En person interagerade med barnen, förklarade vad de skulle göra om de inte förstod och var till hands vid behov. Den andra personen satt en bit bort och förde protokoll, skrev ned observationer och data som skulle samlas in.

8.6 Insamling av data

Följande data samlades in under testsessionen:

- Antal fel innan rätt val
- Reaktioner
- Förståelse
- Intresse/uppmärksamhet

Vi hade förberett ett dokument för att sammanställa den testdata som skulle samlas in. Se Appendix C - Testprotokoll. Vilket fylldes i under testsessionen med hjälp av att iakttagelser gjordes.

8.7 Resultat

Tanken med testningen var inte att vi skulle få ut några konkreta slutsatser, utan mer att kunna se lite trender och se vad det fanns för samband och skillnader mellan speltyperna. Detta delvis för att testgruppen var väldigt liten men också på grund av tidsbrist. För att få ut mer av testningen hade det behövts större testgrupp, och från olika delar av landet.

Utifrån de testerna som gjordes kunde vi se att flertal av barnen testade sig fram för att se vad som blev rätt. Vi kunde dock se att barnen som var fyra år eller äldre lättare kunde se samband mellan siffror och antal hinkar. Större delen av testgruppen bestod av flickor vilket kan komma att påverka resultatet.

8.7.1 Data

	FP1	FP2	FP3	FP4	FP5
Nivå 1	4	1	0	1	2
Nivå 2	1	3	4	1	1
Nivå 3	2	0	2	1	3
Nivå 4	0	0	4	0	0

Tabell 1: Tabellen visar antal fel försökspersonerna gjorde på varje nivå när konsekvensfeedback testades.

	FP1	FP2	FP3	FP4	FP5
Nivå 1	1	0	3	1	1
Nivå 2	1	2	1	3	2
Nivå 3	0	2	1	0	1
Nivå 4	0	2	1	2	2

Tabell 2: Tabellen visar antal fel försökspersonerna gjorde på varje nivå när rätt/fel-feedback testades.

8.7.2 Rätt/fel-feedback

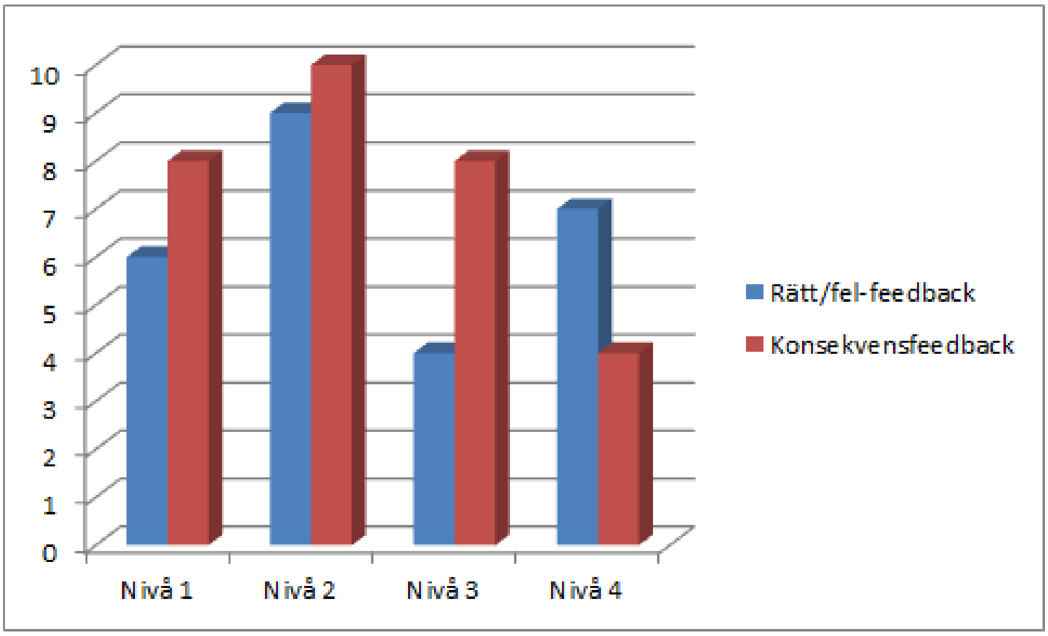
Utifrån resultatet av testerna anser vi att speltypen med rätt/fel-feedback gjorde att barnen blev mindre distraherade av saker som hände runt omkring och att de lättare kunde fokusera på spelet. Spelet gick också snabbare att spela vilket också medförde att barnen lättare kunde koncentrera sig på spelet. Vi kunde se att barnen över fyra år förstod spelet bättre och bättre efter ett par nivåer vad som krävdes av dem. Det verka även finnas en tydlig effekt på rätt/fel-feedbacken att barnen förstod vad som just var rätt eller fel men det kändes dock inte som om att de anpassade sina senare val efter detta. Figur 3 ovan visar att rätt/fel-feedback gav färre antal fel per nivå än vad konsekvensfeedback gjorde vilket kan bero just på att barnen blev mindre distraherade och att spelet gick snabbare att spela vilket kan ha lett till att de fokuserade mer på att lösa uppgiften.

8.7.3 Konsekvensfeedback

Speltypen med konsekvensfeedback visade att barnen blev känslomässigt engagerade och tyckte det var väldigt roligt att spela. Vi märkte att barnen ville hjälpa ankungen upp till sin mamma och ville inte att den skulle åka ner till krokodilen, ett av barnen sa "Det var kul när ankungen kom till mamman". Mellan nivåerna tog det lite för lång tid vilket kan ha medfört resultatet att barnen glömt bort de val som de tidigare har gjort på nivån och att de då väljer samma val flera gånger eftersom de inte kommer ihåg. Vi kunde se ett mönster att barnen valde att mindre/större anternativ efter att de hade fått se konsekvensen av dess tidigare val. Vi antar därför att barnen förstod med hjälp av konsekvensfeedbacken om de valt för mycket eller för lite vatten men de kunde inte riktigt förstå hur mycket mer eller mindre som behövdes. Vi tror detta kan bero på att tallinjen inte är dynamisk, en nivå i hålet är lika djup som fyra nivåer, vilket inte ger barnen en direkt koppling mellan hinkarna och vattnet. Som beskrivit i det förgående stycket visar Figur 3 att konsekvensfeedback ger totalt fler antal fel per nivå än vad rätt/fel-feedback gör. Detta kan bero på att speltypen med konsekvensfeedback tar för lång tid att spela mellan nivåerna och barnen glömmer bort deras tidigare val och väljer samma flera gånger på samma nivå. Diagrammet visar också att antalet fel för konsekvensfeedback minskar efterhand som spelet går, vilket kan betyda att barnen lär sig under spelets gång.

9 Diskussion

Det som vi har funnit mest intressant är att man antagligen kan utveckla barns lärande utifrån vilken typ av feedback som ges. Att jobba med två olika prototyper, som liknar varandra, har varit spännande därför att man då kan se vad olika designval kan resultera i. Våra två speltyper är väldigt lika varandra, men samtidigt väldigt olika. De är olika eftersom den viktigaste delen i spelet, feedbacken, är helt olika och det är feedbacken som vårt spel bygger på. En fundering som uppkommer är om spelarna har olika upplevelser utifrån vilken typ av spel de spelar, eller är det så att det inte spelar någon roll.



Figur 3: Diagrammet visar det totala antalet fel som gjordes på varje nivå i de olika feedbacktyperna.

Då vi har jobbat med ett arbetssätt där många punkter har flutit in i varandra, t.ex. att man har varit tvungen att söka fakta i samtliga steg, har vi hela tiden varit tvungna att ifrågasätta vårt val av design. Detta har lett till att vi har kunnat reducera att vi har målat in oss i hörnen när problem har uppstått. Det har hela tiden varit enkelt att gå tillbaka och diskutera val och möjligheter. De största svårigheterna som vi har haft är att hitta fakta om hur barn i åldersgruppen 4-6 år fungerar och vad de har för kunskaper. Andra svårigheter vi har haft är hur man ska kunna koppla antalet hinkar till tallinjen. Vår lösning var att först gruppera hinkarna så att barnen kan koppla vilka hinkar som hänger ihop. För att sedan kunna koppla antalet hinkar till talsystemet presenteras dessa samtidigt vilket bör göra det lättare att se ett samband mellan dessa.

I första delkursen fokusera vi mycket på utforma och implementera ett spel så vi hade en grund att stå på med endast en enkel nivå, vilket vi ansåg var en bra svårighets nivå att börja med. Här var givetvis också mycket fokus på den konceptuella designen. I delkurs två kunde vi utöka spelet och göra det mer givande och lärorikt genom att implementera fler nivåer med förhoppningen att man kan få se en lärande process genom nivåerna. I och med att vi hade en bra grund från delkurs ett kunde vi lättare lägga upp en tidsplan till delkurs två. Det positiva med detta var att vi kunde lägga tyngdpunkt på att testa prototypen och få mer tid till att analysera den data som samlades in.

För att svara på våra frågeställningar som vi satte upp i början av projektet kunde vi se skillnad på påverkan på barnen beroende på vilken feedbacktyp som barnen testade. Vi kunde se ett större engagemang från barnen som spelade med konsekvensfeedback, vilket i sin tur kan leda till att de ville spela mer och resultatet av det kan bli att de lär sig mer. En anledning till att barnen blev känslomässigt engagerade är att vi använde oss av figurer som barnen kände till från tidigare, och barnen kunde utifrån berättelsen få en anknytning till historien. Att barnen fick se konsekvensen av utförandet gjorde också att de på ett mer engagerat sätt verkligen vill rädda ankungen. Barnen förstod även med hjälp av konsekvensfeedback om de valt för mycket eller för lite vatten men de hade problem med att veta hur mycket mer eller mindre som behövdes. Barnen som spelade med rätt/fel-feedback verkade bli mindre distraherade av saker som händer runt omkring och kunde på ett bättre sätt fokusera på att göra rätt val. Vi såg även att barnen tydligt förstod om de hade gjort rätt eller fel val, men verkade inte direkt kunna anpassa sina val efter feedbacken.

Även om våra tester får anses som en möjlighet till att upptäcka trender och tendenser hos barnen så gav de likväl väldigt intressant information. Det verkar finnas ett slags lärande när barnen passar igenom nivåerna då vi märkte på en del barn att de svarade fel i början men efter ett par nivåer gick det helt plötsligt mycket bättre. Detta på båda typerna av feedback. Vår kanske största upptäckt förblir dock att det verkar som att barnen lyckas anpassa sina val att välja mindre eller mer beroende på konsekvensfeedbacken, där rätt/fel-feedbacken endast verkar ge förståelsen om de just gjort rätt eller fel.

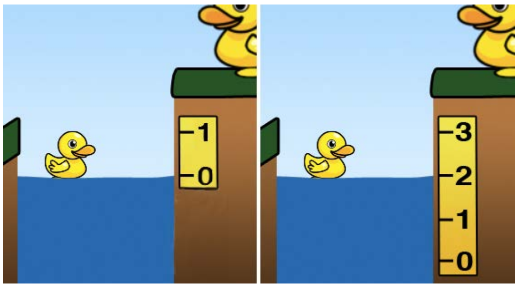
Det finns dock ett par osäkerheter kring detta ”lärande”. Det kan exempelvis finnas ett lärande att man alltid ska koppla rätt svar till siffran som ligger överst på tallinjen. På så vis kan man bara kolla vilken siffra som är där och sen trycka på rätt antal hinkar, även om dessa hinkar givetvis måste räknas. Denna trend lyckades vi se lite på ett av de äldre barnen som lyckades svara rätt förvånansvärt fort. Huruvida det är så här eller inte, så finns det absolut rum för ett antal förbättringar som vi tror kan vara ytterst avgörande för spelets potential.

Det första är som sagts man skulle kunna ändra ordningen på hinkarna. På så vis kommer man undan det faktum att barnet lär sig att åt höger finns det fler hinkar och att vänster finns det färre. Nu måste barnet noga räkna hinkarna innan ett val görs.

Det andra (och troligtvis det viktigaste) är dynamiken i vattnet. Om vattnet startar på olika nivåer uppstår en naturlig koppling till verkligen och en slags mängdlära introduceras vilket troligtvis kan ha en stor påverkan på hur barnet tänker; ser det ut att saknas mycket vatten i hålet borde en större mängd vatten fyllas på, och motsatsen att om det ser ut att bara behövas lite vatten borde man inte behöva så många hinkar vatten. Detta gör också att tallinjen får naturliga längdmått (nivåerna behöver inte sträckas ut) och berättigar därför sin riktiga funktion.

Det tredje är att ytterligare förtydliga att knapptryckandet på hinkarna bidrar till att vattennivån fylls. Genom uppmärksammade trender var denna koppling inte helt självklar för barnen. Man kan förtydliga detta samband genom att ha animerade hinkar som hälls ner eller att barnet får dra ner hinkarna (alla helst med en touch skärm). Denna tydlighet som nu tillsammans med vattnets dynamik blir ännu mer närvarande kan vara väldigt avgörande för att verkligen få fram spelets potential.

En annan punkt (som är en punkt lite vid sidan om, än dock ytterst intressant) är att man nu helt plötsligt får mer möjligheter att experimentera med tallinjen, när dynamiken är naturlig. En idé som skulle kunna vara mycket intressant att prova (som dock blir riktad åt aningen äldre barn) är att göra tallinjen med dessa nivåer ”fasta”. Som det är tänkt med den dynamiska vatten nivån nu, så börjar vattennivån alltid på 0, så är det 3 hinkar man ska välja så börjar vatten nivån helt enkelt i mitten och tallinjen består av 3 nivåer. Om man istället alltid gör nivåerna fasta skulle man exempelvis inte behövt att ändra siffrorna på tallinjen utan alltid ha en bestämd mängd del mått. På så vis om man t.ex. har en tallinje på 5 och den dynamiska vattennivån börjar på nivå 3 så måste barnet själv lista ut att det är 2 hinkar den ska välja, dvs. 3+2 = 5 hinkar, se figur 4 nedan. Detta har två funktioner.



Figur 4: Visar startpostion på det dynmaiska vattnet åt vänster (hur det är tänkt det ska se ut nu) jämfört med startposition på hur det skulle kunna vara med en “fast” tallinje. Båda bilderna har givetvis svaret “1 hink” som rät svar.

Det ena är att barnet inte längre kan ta för givet (om den lyckas lära sig ett mönster) att det är översta siffran som är den korrekta, vilket som sagt en femåring lyckades verka kunna lista ut. Den andra funktionen är att vi faktiskt skapar en slags förståelse för subtraktion nu också. Även om barnet inte nödvändigt förstår att det är just subtraktion så övas detta ”tänk” likväl. Detta hade givetvis krävt en egen variant på spelet då en ny nivå inte skulle kunna plötsligt introducera detta genom att ändra tallinjens funktionalitet. Detta är något som då ska bestämmas från första början. Men det är en intressant utbyggnad/variant på spelet.

Som synes och som upptäckts efterhand har vi endast öppnat dörrarna på vad möjligheter det finns för detta spel att verkligen nå sin fulla potential. Vi tror absolut på att nu när digitala hjälpmedel finns tillgängliga är det inte längre en fråga om att bara förvandla det lärande tänk som funnits innan till digital form då på så vis förändras egentligen ingenting av lärande processen mer än att det är digitalt. Att vi nu har möjlighet att introducera ny slags feedback i form av konsekvens känns det som att både intresset och resultaten av lärandet absolut kan tas till nya nivåer. Med vårt experiment nådde vi en stabil grund att stå på med ett spel som verkligen känns lämpad att prova just konsekvens-feedback på och se vad det kan ha för skillnader och likheter mot just ett lek och lär - spel som är baserat på klassisk rätt/fel-feedback.

Slutligen så även om vi endast uppfattat trender om att konsekvens-feedback har en inverkan på hur barnen utför sina senare val, så har vi i åtanke på att tydligheten fortfarande är långtifrån bra i spelet och om tid hade funnits och detta (vatten-dynamiken m.m. som nyss nämnts) hade implementerats, vem vet hur pass stor inverkan detta egentligen hade kunnat ha på det grundläggande matematik-inlärandet för ett barn?

Referenser

[1] *Teaching number sense.* Griffin, Sharon. USA, Alexandria (2004)

[2] *Wikipedia.* Feedback. http://en.wikipedia.org/wiki/Feedback [2012-10-12]

[3] *Wikipedia.* Number sense. http://en.wikipedia.org/wiki/Number_sense [2012-11-26]

[4] *Wikipedia.* Färgsymbolik. http://sv.wikipedia.org/wiki/Färgsymbolik [2012-09-31]

[5] *The Design of Everyday Things.* Norman, Donald A. USA, New York (1990)

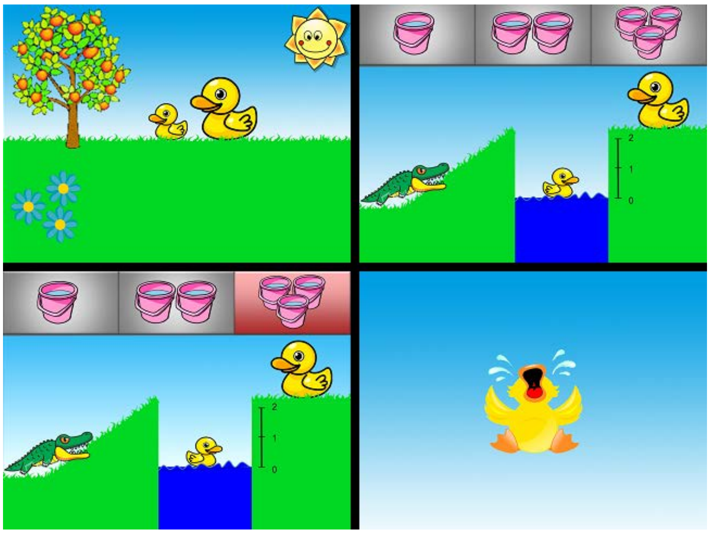
[6] *Building number sense with Number Worlds: a mathematics program for young children.* Griffin, Sharon. USA, Orlando (2004)

[7] *Using Teachable Agent Feedback to Support Effective Learning-by-Teaching.* Roscoe, Rod D. Wagster, John. Biswas, Gautam. USA, Washington, DC (2008)

[8] *Chapter 5. Develop the Test Plan.* Eriksson, Joakim. http://wwwold.eat.lth.se/kurs/material/mam120/Lecture_notes/HUT_kap5.pdf sid. 3. [2012-10-03]

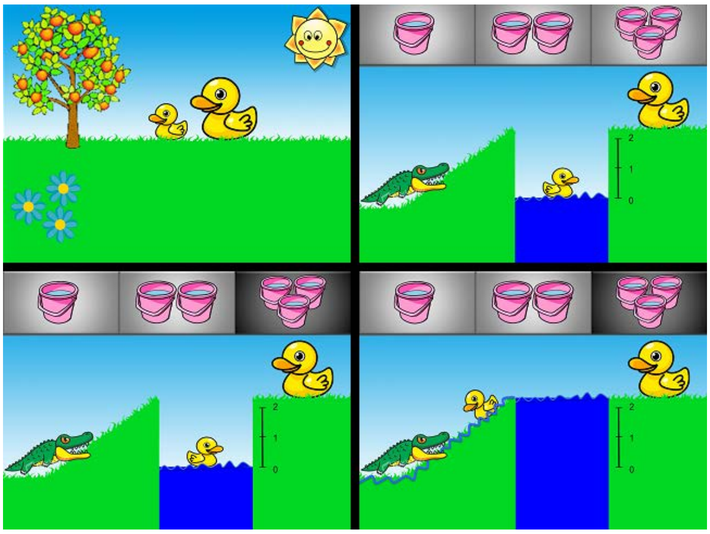
A LoFi prototyp

A.1 Rätt/fel-feedback



Figur 5: Bilden visar vår LoFi prototyp med rätt/fel-feedback.

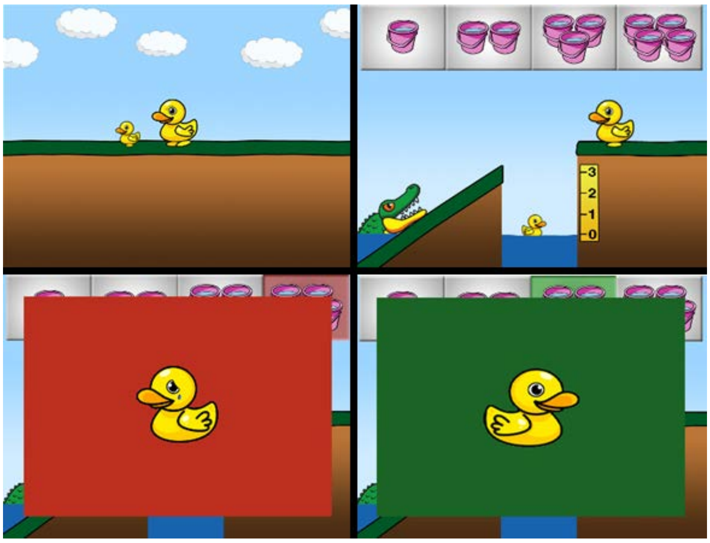
A.2 Konsekvensfeedback



Figur 6: Bilden visar vår LoFi prototyp med konsekvensfeedback.

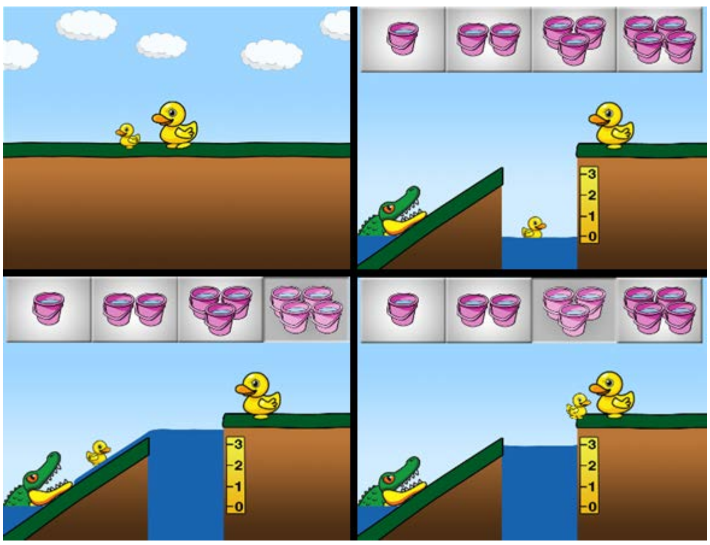
B HiFi prototyp

B.1 Rätt/fel-feedback



Figur 7: Bilden visar vår HiFi prototyp med rätt/fel-feedback.

B.2 Konsekvensfeedback



Figur 8: Bilden visar vår HiFi prototyp med konsekvensfeedback.

C Testprotokoll

Fylls i för varje nivå

Antal fel innan rätt val:	
Förståelse:	
Intresse/uppmärksamhet:	
Reaktioner:	
Övrigt:	

Debriefing

Var det kul att spela?	
Tyckte du att det var lätt eller svårt?	
Vad var bäst med spelet?	
Vad var sämst med spelet?	
Hade du velat spela mer?	

D Bilaga om HiFi-prototypen

Verktyget som användes var Construct 2 från https://www.scirra.com/ ett semi-grafiskt verktyg för att skapa HTML5-spel. Utvecklingsverktyget använder sitt eget filformat kallat capx för att spara projekt. Det innehåller också en export-metod för att publicera spelen i autogenerat HTML5-format. Gratis versionen av spelet är begränsat till 100 så kallade events, en begränsning som nåddes tidigt i den andra delkursen. Då verktyget fortfarande är i beta så har det en del buggar, den som ställde till mest problem var att de inte gick att spela ljud direkt i början på nivåerna utan det resulterade i att alla instanser av det ljudet spelades upp samtidigt, vilket gav en en massa reverb och echo effekter som gjorde ljudet olyssningsbart. Ett annat problem med ljudet visade sig vid själva spelandet av spelet, ibland när man byter flik i webbläsaren eller återställer den från minimerat läge så nollställs nivåerna på ljuden till den nivå de spelades in med, vilket resulterade i väldigt hög bakgrundsmusik i vårt spel.

Filerna för spelet bifogas dels i programmets egna capx format och dels som zip-filer innehållande kataloger med allt HTML5 materialet.

Länkar för att spela spelet
Rätt/Fel-spelet: https://dl.dropbox.com/u/71887276/RattFel/index.html
Konsekvens-spelet: https://dl.dropbox.com/u/71887276/Konsekvens/index.html

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Freds Bondgård: Ett av de första mWorldspelen för förskolebarn

Emelie Brynolf, Axel Hammarlund, Marcus Malmberg, Caroline Rödén & Magnus Walter

Matematik är ett av våra kärnämnen och en viktig beståndsdel i vårt dagliga liv. Barn brukar lära sig siffrornas betydelse i tidig ålder och börjar då att räkna och använda sig av matematiken. Det finns en mängd pedagogiska sätt att förmedla den på. I och med teknikens utveckling finns det nya resurser att tillgå, vi har därför tagit fram och anpassat ett digitalt lärospel för att barn redan i fyraårsålder, för att de ska kunna utveckla sin taluppfattning och få kunskap på ett nytt och roligt sätt.

Vår spelidé baseras på en Bondgård där spelaren hjälper bonden och hans barn med olika sysslor på gården. Spelet ska på så vis ge barnen kunskaper i matematik och förstärka detta genom att lära ut till bondgårdens egna Teachable Agent, Kim. Vi har arbetat från och med brainstorming genom processen med alla dess steg, där bland annat implementering och slutligen testning ingick.

När vi kände oss färdiga med implementeringen, testade vi spelet på ett antal barn i en förskola i Göteborg. Detta för att kunna förstå hur målgruppen uppfattade spelet och för oss att därefter fortsätta utvecklingen av det. Genom spionprogram och observationer följde vi användarnas mönster och kommentarer kring spelet. Testningen visade att svårighetsgraden var aningen för lätt, animeringarna något för långsamma och de förstod inte vår pedagogiska agent till fullo. Däremot uppskattades belöningssystemet av barnen och de kände ett begär för att spela mer. I sin helhet var spelet uppskattat både av barnen och läraren.

1. Bakgrund

Projektet handlar om att utveckla ett spel för barn i fyraårsålder där de ska lära sig grunderna av talsystemet och förbättra sitt “Number Sense”, en intuitiv förståelse av tals storlek. Anledningen till att man vill förbättra så pass små barns “Number Sense” är för att det finns teorier om att det ger en stor fördel när barnen börjar med matematik i grundskolan. Att använda digitala lärospel för att förbättra barns “Number Sense” för den här åldersgruppen är unikt.

Det övergripande problemet vi vill lösa är att utforma ett digitalt läromedel som kan hjälpa till att utveckla barns taluppfattning i ett tidigt stadium. Genom att möjliggöra ett alternativt lärosätt hoppas vi kunna aktivera, engagera och utveckla eleven till att prestera bättre på ett roligare sätt.

I projektet använder vi oss utav "Teachable Agents", vilket går ut på att barnen är lärare till en digital elev. "Teachable Agents" bygger på metoden att "lära genom att lära ut". Intressanta frågor som vi vill få svar på är:

- Hur barnen beter sig när de interagerar med ett digitalt spel.
- På vilket sätt barnen interagerar med en "Teachable Agent".
- Vad barnen kan och vad som är en bra svårighetsgrad för spelet.
- Om man kan se att barnen verkligen förbättrar sitt ”Number Sense” genom “Teachable Agents”.

2. Målgrupp

Målgruppen vi har inriktat oss mot är barn på fyra år med olika bakgrund och kunskapsnivåer inom matematik. Valet av åldersgruppen beror på att ett ”Teachable Agent”-spel aldrig tidigare har utformats för så små barn. Barnen befinner sig även i en ålder där de börjar utveckla sitt så kallade ”Number Sense” vilket kan vara intressant ur en matematisk aspekt.

2.1 Vad kan en fyraåring

Med anledning av att alla barn har olika förmågor, gener och förutsättningar är det svårt att ta reda på exakt vad en fyraåring kan. Vi vet däremot att barn i fyraårsåldern ofta är energiska, initiativrika och har stor fantasi. De lär sig tänka mer abstrakt genom att till exempel fantisera om att ett föremål är någonting annat än vad det egentligen är. Det är i fyraårsåldern som barn också börjar förstå innebörden av begrepp som; störst, längst, lika, mer, och så vidare. Ordningsföljder, färger, siffror och bokstäver börjar också falla på plats. Med andra ord börjar ”Number Sense” utvecklas i denna åldersgrupp. Generellt kan man säga att de kunskaper som kännetecknar en fyraåring är att barnet kan räkna till tio, skilja på olika färger och har börjat få grepp om tidsuppfattning.

3. Teori

3.1 Teachable Agent

En Teachable Agent, är en animerad pedagogisk agent som ska underlätta inlärning och utveckling genom “Learning by Teaching”. Swartz & Blair beskriver fyra funktioner för Teachable Agents som gör det möjligt att förstärka “Learning by Teaching” effekten, det vill säga att eleven lär sig genom att lära en agent det eleven själv lärt sig[1]. De fyra funktionerna är att:

- Tydligt visa Teachable Agentens tänkande och resonemang
- Studenternas lärande ska ge feedback genom att Teachable Agenten presterar bättre eller sämre.
- Agenternas förmåga att förtydliga det som den behöver lära sig
- Placera agenten i miljöer som stöder undervisningen

3.1.1 Protégé Effect

Det har visat sig att elever engagerar sig och lägger ner mer tid i sitt lärande då det finns en Teachable Agent. Den nya tiden läggs inte ner i lärandet för elevens egen skull utan för att kunna lära ut bättre till Teachable Agenten.[2]

3.1.2 The Media Equation

The Media Equation är en allmän kommunikationsteori som hävdar att människor tenderar att behandla datorer och andra medier som om det vore riktiga människor eller platser. Eftersom barnen kan skapa en relation till en agent möjliggör det användandet av Teachable Agents i spel.[3]

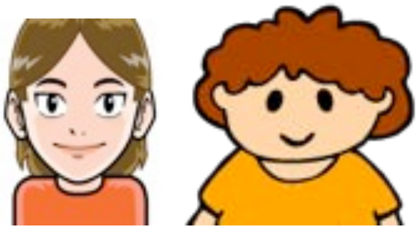
Det är viktigt att ta hänsyn till genus när man konstruerar en virtuell agent. Forskning har visat att folk behandlar en agent olika beroende på genus och därför är det ett smart drag att designa den androgynt, se figur 3.1.

3.1.3 Ego-Protective Buffer

En Teachable Agent har visat sig skydda elevers egon. Detta på grund av att det inte längre är eleven som gör fel utan Teachable Agenten. Då det inte är eleven som gör fel kommer inte eleven i samma utsträckning uppfatta sitt nederlag. Istället har det visat sig att eleven känner motivation för att ta nya tag i upplärningen av sin Teachable Agent. Eleven ser alltså inte problem i sin kunskap, utan i sättet som lärdes ut.[4]

3.1.4 Theory of Mind

I detta sammanhang åsyftas Theory of Mind, eller inlevelseförmåga, att förstå att andra kan uppfatta saker på ett annat sätt eftersom de kan ha andra övertygelser, önskningar och avsikter. För att en elev ska kunna undervisa en annan människa något, eller en Teachable Agent, måste elevens kunskap på något sätt komma fram på ett förståligt sätt.[5]



Figur 3.1 - Den androgyna Teachable Agenten Kim. Första versionen längst till vänster och nya versionen längst till höger.

[1] Blair Kristen, Schwartz Daniel L., Pedagogical Agents for Learning by Teaching: Teachable Agents

[2] Chase, C., Chin, D.B., Oppezzo, M. & Schwartz, D. (2009) Teachable Agents and the Protege Effect: Increasing the Effort towards Learning. Journal of Science Education and Technology

[3] Johnson, D., Gardner, J. & Wiles, J. (2004) Experience as a Moderator of the Media Equation: The Impact of Flattery and Praise. International Journal of Human-Computer Studies 61(3), 237-258.

[4] Chase, C., Chin, D.B., Oppezzo, M. & Schwartz, D. (2009) Teachable Agents and the Protege Effect: Increasing the Effort towards Learning. Journal of Science Education and Technology

[5] http://en.wikipedia.org/wiki/Theory_of_mind, 2012-10-23

3.2 Number Sense

Number Sense, dvs. taluppfattning, innebär en intuitiv förståelse av tals storlek, relationer och hur de påverkas av matematiska operationer. En god taluppfattning ger en känsla för hur tal fungerar genom att förstå sambandet mellan nummersymbolen och själva mängden. Detta underlättar för förståelsen av hur noggrann en beräkning bör vara, gör det lättare att upptäcka räknefel vid uppskattning och bidrar till sunt förnuft vid användning av tal.[6]

En elev med god Number Sense:

- Tittar på ett problems helhet innan han/hon går in i detaljer.
- Söker efter samband mellan tal och operationer.
- Hittar på metoder som stämmer överens med den egna uppfattningen av tal och söker efter den bästa representationen av ett problem.
- Använder hållpunkter för att bedöma ett tals storlek.
- Känner igen orimliga svar vid beräkningar.

4. Bondgården

4.1 Storyn

Den övergripande storyn är att spelaren befinner sig på en bondgård där han ska hjälpa bonden, Fred, med olika sysslor (se figur 4.1). Spelaren hjälper till med dessa sysslor genom att spela olika delspel. Varje delspel är kopplat till ett av de fyra djuren vi använder oss av. Djuren är följande: ko, höna, häst och kanin.

När ett delspel startas ger bonden en syssla som ska utföras och det är spelarens ansvar att hjälpa bondens barn, Kim, som är en Teachable Agent, att utföra uppgiften. Som belöning för att spelaren har klarat ett delspel fås en belöning som sparas vid varje djurbur. Genom detta ser spelaren enkelt hur mycket ett visst spel har spelats.

4.2 Delspelen

På bondgården kommer det finnas fyra olika spel. Spelen är kaninspelet, som vi kommer utveckla mer i detalj, ett hästspel, ett hönsspel samt ett kospel. Tanken med delspelen är att få barnen att stanna kvar och spela spelet längre. Om de inte skulle uppskatta ett av delspelen kan de fokusera på ett av de andra.

Figur 4.1 - Start menyn där spelaren ska välja spel

4.2.1 Kaninspelet

Det spel vi bestämt oss för att lägga fokus på är kaninspelet. På bondgården finns det många hungriga kaniner som behöver matas. Spelarens uppgift är att mata kaninerna med rätt antal morötter. Då spelaren genomfört uppgiften korrekt tre gånger får den en morot till bondgården som belöning.

Om spelaren väljer för få morötter ges feedback om detta genom att de kaniner som inte fått morötter blir ledsna, Teachable Agenten blir ledsen, samt att det kommer en kommentar från bonden. Liknande gäller om spelaren väljer för många morötter, men nu blir inte kaninerna ledsna utan de överflödiga morötterna läggs i en hög bredvid kaninerna. Även här kommer bonden in och påpekar det felaktiga, att spelaren valt för många morötter. Detta hjälper även barnet att inse att de borde ta ett mindre antal morötter nästa gång.

Kaninspelet är indelat i tre olika faser. Under första fasen spelar barnet helt själv medan Teachable Agenten syns uppe i ena hörnet. Denna fas finns för att låta barnet lära sig spelet innan Teachable Agenten kommer in i bilden. I nästa del av spelet kommer Teachable Agenten ner och ber att spelaren ska visa hur spelet fungerar. Det är först nu Teachable Agenten börjar lära sig av spelarens rätt och fel. Efter denna fas är genomförd blir det Teachable Agentens tur att försöka. Spelarens uppgift i detta läge är att rätta eventuella fel som Teachable Agenten gör. Allt eftersom spelaren gör rätt val kommer Teachable Agenten utvecklas och svara fel mer sällan.

Under de olika iterationerna under spelutvecklingsfasen fanns det olika idéer om hur för få, respektive för många, morötter skulle hanteras. Hur för få morötter skulle hanteras växte fram snabbt och kändes naturligt, medan för många

[6] Barbara J. Reys & Robert E. Reys (1995). Perspektiv på Number sense och taluppfattning

morötter inte kändes lika logiskt. Det fanns tankar om att kaninerna skulle få ont i magen, att de skulle explodera samt att en råtta skulle komma in och äta upp de överflödiga morötterna. Dessa idéer skrotades dock då vår aktuella idé är lättförstålig, ger tydlig feedback samt passar in bra i spelidén.

4.2.2 Hästspelet
Hästen Harald ska hoppa över hinder, men han behöver hjälp med hur många gånger han ska hoppa för att klara banan i fråga. Barnen ska här välja rätt antal hopp för att hjälpa hästen. I detta spel samlar barnen på sig medaljer.

4.2.3 Hönsspelet
Det finns två idéer till detta spel. Den första handlar om att bonden behöver hjälp med att samla hönornas ägg i korgar. Det får plats olika antal ägg i de olika korgarna så barnets uppgift blir att välja rätt antal ägg till rätt korg. Den andra idén är att leta rätt på de gömda hönorna på bondgården. Hönorna har nämligen sprungit och gömt sig för den farliga räven Rune. Då barnet hittar en höna räknar en röst upp för att tydliggöra händelsen. I båda spelen får spelaren ägg som belöning.

4.2.4 Kospelet
Kospelet råder det delad mening om. Vi har tagit fram flera olika förslag, men inte bearbetat dem till ett konkret spel. Det finns förslag om att bygga staket, placera rätt antal kor i olika stora hagar, mjölka korna eller paketera mjölk. Även i detta spel kommer barnet att få en belöning som syns vid kons staket på bondgården. Vad det är kommer att bero på hur spelet blir utformat.

4.3 Svårighetsgrad

Att ha en statisk svårighetsgrad i spelet kan göra att smartare barn tycker att det är för lätt medan barn som utvecklas långsammare uppfattar det som för svårt; vilket kan leda till att de tappar intresset och får ut än mindre av spelet.

Kaninspelet kan enkelt implementeras med olika svårighetsgrader. Som det ser ut nu använder vi oss av högar som går från en morot till sex morötter. Grupperingen av morötterna sker i fingerform. Om svårighetsgraden ska öka kan grupperingen av morötterna ske i högar istället för att de är ordnade som fingrar. Våra högar är för tillfället även ordnade efter tallinjen, från minst till störst, men detta kan ändras för att öka svårighetsgraden. Högarna med morötter kan istället ligga i en slumpmässig följd. Utöver detta kan även antalet högar ändras för att öka svårighetsgraden.

Figur 5.1 - Morötter upplagda för att påminna om händer och fingrar, tallrikarna förtydligar grupperingen av morötterna

Representationen av morötter kan även bytas ut. Istället kan spelaren få välja mellan exempelvis verkliga siffror samt representation via tärningar. Tanken här är att även om spelet inte använder sig av morotshögar kommer morötter delas ut. Detta för att fortsätta att använda denna feedback samt att spelaren hela tiden får se hur morötterna delas ut längs med tallinjen.

5. Design av interaktion

5.1 Gruppering

Det är viktigt att få eleverna att förstå att de kan trycka på de olika morotsgrupperna och förstå vilka morötter som hör tillsammans. Därför har vi ett tydligt avstånd mellan varje morotsgrupp, och dessutom en tydlig inre gruppering där morötter som hör tillsammans ligger tätt ihop. Dessutom har vi lagt till tallrikar under varje morotsgrupp som gör grupperingen ännu tydligare. Det blir också ännu tydligare när vi lägger ut morötterna på ett sätt som representerar hur barn räknar på fingrarna, se figur 5.1. Morötterna är även mappade enligt tallinjen där minsta gruppen ligger längst till vänster och den största längst till höger.

5.2 Navigering

Barn i åldern två till fem år navigerar till stor del med hjälp av bilder och ljud men inte med text. Därför använder vi bilder och ljud som hjälper dem att förstå vad de ska göra. Ett barn som aldrig har använt en dator behöver först hjälp av en vuxen för att komma igång, men generellt sett lär sig barnen snabbt hur de ska använda datormusen. Med lite övning kan de även lära sig att dubbelklicka. Konceptet med drag and drop är inte fullt lika utvecklat hos ett barn och därför sker all navigering med datormus i systemet. Då barn i vår målgrupp kan navigera med en datormus kommer de även kunna använda sig av en pekplatta. Detta ger oss möjlighet att utveckla en hi-fi prototyp som blir mångsidig genom att den, navigeringsmässigt, kommer kunna användas både på en dator och en pekplatta.[7]

[7] Medierådet (2010). Småungar & Medier

6. Grafisk design och ljud

6.1 Layout och grid

En viktig punkt med layouten är att vi inte vill ha för många valmöjligheter för eleven. Det kan leda till spatial förvirring och även till att de tappar intresset.

6.2 Färg och ikoner

När eleven ska rätta Teachable Agentens val görs detta med hjälp av två knappar, en grön med glad gubbe som representerar rätt, och en röd knapp med ledsen gubbe som betyder fel.

Vi har valt en grön färg på grund av att det representerar säkerhet, till exempel vid trafikljus. På samma sätt representerar röd färg varning. För att ytterligare förtydliga ger den glada och ledsna gubben feedback på vad knappen symboliserar. Detta förtydligar inte bara knapparna utan gör det möjligt för färgblinda barn att spela spelet.

6.3 Ljud

Då vår målgrupp är så pass ung kan det inte förutsättas att de kan läsa. Detta gör det mycket viktigt med en berättarröst som guidar barnen genom spelet. Tonläget på denna röst är också viktig för att skapa intresse hos barnet. En glad och entusiastisk röst väcker intresse och engagemang hos många barn medan en lite mer nedtonad röst inte gör det.

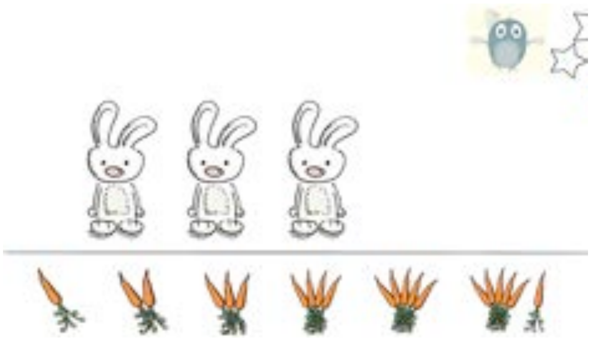
7. Prototyper

7.1 Mid-fi-prototypen

För att underlätta framtagningen av en teknisk prototyp hjälper det om man börjar skissa på storyboards. Genom att noga välja vilka specifika scenarion som skissas upp kan det bidra med mycket information utan att det ger för många bilder. Vi valde därför att fokusera på en story där spelaren är mitt i spelet, ska mata kaniner och väljer fel antal morötter.

Som man kan se i figur 7.1 är kaninerna tydligt uppställda på en rad för att enkelt ge barnen en chans att räkna dem, samtidigt som grupperna med morötter är sorterade från minst till störst antal: vilket är för att framhäva effekten av tallinjen. Morötterna är samtidigt upplagda i grupper för att påminna om händer, vilket ska påminna spelaren om att räkna på fingrarna.

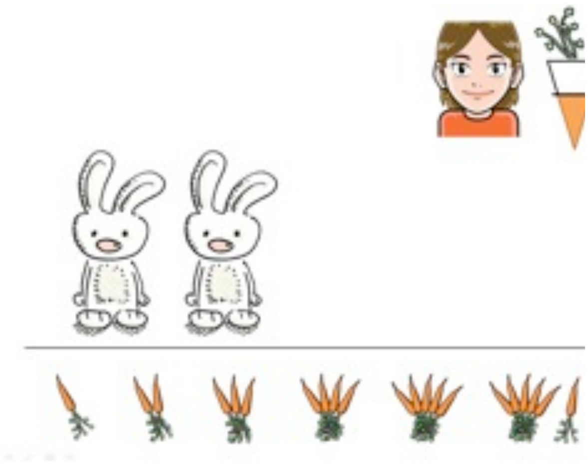
Figur 7.1 - Story board när spelaren väljer en hög med för få morötter



Figur 7.2 - En första implementation av kaninspelet i Balsamiq

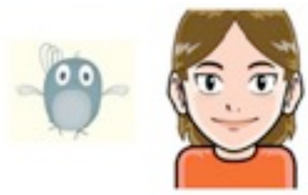
En mid-fi prototyp började ganska tidigt utvecklas i verktyget Balsamiq. Vi kom snabbt framåt med detta verktyg och hade utan större problem implementerat stora delar av vår spelidé. I figur 7.2 kan man se hur en första digital prototyp såg ut. När vi kommit en bit insåg vi att vi behövde animationer i prototypen för att förstärka känslor och vi valde därför att överge Balsamiq och övergå till PowerPoint, vilket alla i gruppen sedan tidigare har erfarenhet av.

I första delkursen valde vi att stanna vid en enkel mid-fi prototyp i PowerPoint, se figur 7.3, för att ge oss utrymme för att finslipa funktionalitet och koncept. Prototypen är grundligt gjord för att få en stabil grund att stå på när vi ska börja utveckla hi-fi. På så sätt kan vi snabbt komma igång att utveckla hi-fi prototypen utan att behöva fastna på konceptuella detaljer, kan fokusera på att ge en bra upplevelse för spelaren, och på så sätt kunna få ut så mycket som möjligt av våra tester. I Appendix A återfinns en manual för de olika scenarierna i PowerPoint-prototypen.



Figur 7.3 - En första implementation av kaninspelet i Balsamiq

Figur 7.4 - De två olika belöningssystemen



Figur 7.5 - De två olika Teachable Agenterna

7.1.1 Förbättringar

När Balsamiqprototypen byttes ut mot PowerPointprototypen skedde en del förändringar utseendemässigt. Till att börja med byttes belöningssystemet ut, se figur 7.1. Istället för att fylla tre stjärnor ska nu spelaren fylla en morot som sedan kommer återfinnas på bondgården. Anledningen till att vi gjorde detta är för att moroten har bättre koppling till bondgården än vad stjärnorna hade. I mid-fi-prototypen krävs det bara två korrekta svar för att fylla moroten, men i hi-fi-prototypen kommer detta ändras till tre korrekta svar. Detta känns som en bra balans, för vi vill inte att nivåerna nås för snabbt, för det leder till för många avbrott. Samtidigt vill vi att barnet ska få belöningarna tillräckligt ofta, så att de känner sig motiverade att fortsätta spela. För att förtydliga antalet rätta svar som krävs av barnet är stegen markerade i moroten.

Förutom att belöningssystemet byttes ut, byttes även Teachable Agenten ut. I Balsamiqversionen användes en blå fågel som Teachable Agent. Denna byttes sedan ut mot bondens androgyna barn Kim, se figur 7.5. Detta genomfördes på grund av att den nya agenten passar bättre in i bondgårdsstoryn.

7.2 Hi-fi-prototypen

Mid-fi prototypen gav oss en bra känsla för konceptet och hur spelet fungerar. Nästa steg var att utveckla en spelbar prototyp som kunde användas för testning på målgruppen. Vi valde att implementera ett spel som kunde spelas direkt i webbläsaren, utan att behöva installera ett program. Det är väldigt strategiskt då det kan spelas på alla plattformar direkt, det vill säga Mac, Windows, och Linux etc. En ytterligare fördel med att spela i webbläsaren är att det går att spela på datorer utan att användaren har administrationsrättigheter, vilket är väldigt vanligt på till exempel skolor.

Det finns flera olika tekniker man kan använda för att implementera ett spel som spelas i webbläsaren. Vi valde att använda HTML5 och JavaScript för att det är ett växande område, så det är enkelt att hitta information och få hjälp vid behov, och det kräver inget utav användaren, förutom att de använder en modern webbläsare. För att vara effektiva och undvika att uppfinna hjulet en gång till använde vi oss av JavaScript-biblioteket KineticJS[8] som sköter alla animationer och övrig grafik. Ljuduppspelningen sker med HTML5-elementet "audio".

För att flera personer i gruppen skulle kunna programmera samtidigt på ett effektivt sätt använde vi oss av versionshanteringssystemet Git[9], i kombination med GitHub[10].

7.2.1 Förbättringar

När vi började utvecklingen av hi-fi-prototypen så gavs det mer möjligheter än när vi använde oss utav PowerPoint. Mid-fi-prototypen är väldigt statisk och man var tvungen att följa ett visst mönster. Men med hjälp av JavaScript och HTML5 fick vi större möjlighet att utveckla en prototyp som låg väl-

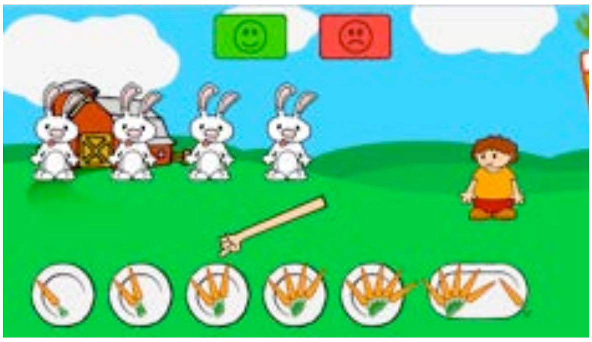
[8] http://kineticjs.com/

[9] http://git-scm.com/

[10] https://github.com/

digt nära en färdig produkt. Detta gjorde det mer möjligt för oss att utföra ordentliga tester på vår idé.

I början av utvecklingen av läget då Teachable Agenten Kim spelar själv så stod Kim stilla på samma plats, medan en hand pekade på den morotsgrupp som Kim valde, se figur 7.6. Denna lösningen lämnade vi senare på grund av risken att barnen inte skulle förstå att handen och Teachable Agenten hänger ihop. Vår slutgiltiga lösning blev istället att Kim förflyttar sig från ursprungsplatsen och pekar på morotsgruppen som Kim valde, se figur 7.7.



Figur 7.6 - Arm som pekar och en stillastående Teachable Agent.

Figur 7.7 – Teachable Agent som går fram och ställer sig vid tallriken.

Figur 7.8 - Layout för morötterna i mid-fi-prototypen

Figur 7.9 - Layout för morötterna i hi-fi-prototypen

Ett problem var om morötterna var tillräckligt grupperade. Det är väldigt viktigt att barnen ser vilka morötter som hör ihop och förstår var de ska klicka, se figur 7.8. Detta löste vi genom att placera ut tallrikar under varje morotsgrupp, se figur 7.9. Det gav också den positiva effekten att de verkligen förstått att man ska mata kaninerna.

I mid-fi-prototypen använder vi oss av en röst för både Teachable Agenten Kim och Bonden Fred. Men i hi-fi prototypen separerade vi dem, så att barnen lättare skulle förstå vem som pratar.

När man startar spelet så visas startskärmen. I vår mid-fi-prototyp hade vi en vit bakgrund som visade de olika delspelen, samt Teachable Agenten och bonden Fred, se figur 7.10. I hi-fi-prototypen gjorde vi nya bilder som både är snyggare och dessutom följer samma stil, se figur 7.11 . Vi la också till en bakgrundsbild som ska göra spelet mer intressant för barnen.

Figur 7.10 - Bondgården med två klarade nivåer på kaninspelet i mid-fi-versionen

Figur 7.11 - Bondgården med två klarade nivåer på kaninspelet i hi-fi-versionen

8. Testning

För att få svar på hur vårt spel kan komma att fungera, samt för att fastställa designproblem, planerade vi in testning i samband med att spelet ansågs tillräckligt körbart. Buggar och andra problem gjorde att testning utfördes senare än beräknat, men ändå i god tid för även hinna med analys. När väl kodningen var klar kunde resten av arbetet i projektet utföras parallellt.

Vi vill att vårt spel skall fungera för alla barn. Därför vore en förskola i Lund direkt olämplig eftersom andelen med akademisk bakgrund i staden inte representerar landet i övrigt. Vi sökte oss därför utanför staden med hjälp av vårt kontaktnät, närmare bestämt Göteborg och Förskolan Lars Kaggsgatan 35[11].

En testplan gjordes, vilken återfinns i Appendix B, och användes som grund för testutförandet. Det visade sig dock att verktyget, Camtasia Studio[12], medförde att spelets hastighet drogs ner. Därför togs beslut att enbart observera och ta anteckningar.

Testet utfördes på totalt åtta deltagare, lika andel pojkar som flickor, i åldern 4-5 år.

8.1 Analys

8.1.1 Huvudmenyn

Det märktes tydligt att barnen hellre provar sig fram än att invänta och följa de instruktioner vi byggt in ljud till. Istället klickades det frenetiskt för att testa vad som hände.

Det var inte tillräckligt tydligt vilka delar av huvudmenyn som faktiskt ledde in till ett spel. Barnen klickade på ladan, bonden, Teachable Agenten, hönan, hästen, kon och kaninen i olika ordning. Speciellt efter att enbart kaninspelet fungerade försökte de klicka på annat, trots instruktioner om att de inte var implementerade. Bättre affordance skulle behövas i huvudmenyn, samtidigt som ljud och animationer skulle behövas även för bonden, Teachable Agenten och ladan.

8.1.2 Första spelfasen

Även här var barnen ivriga att komma igång, och klickade gärna innan instruktioner och animationer var klara. Enbart en deltagare förstod aldrig hur man skulle gå tillväga för att få rätt resultat. Resterande hade olika tillvägagångssätt. En del provade sig fram första omgången, för att sedan klara utmaningen direkt nästa delmoment. Andra började med att räkna kaninerna och sedan räkna morötterna på tallrikarna. Vissa gjorde fel med vilje för att se vad som hände.

Barnen upplevde att animationerna gick för långsamt. Kanske borde man även lägga till fler moment innan man avslutar spelomgången i kombination med att dra upp tempot. Vid en eller två morötter utbrast barnen "lätt".

När första spelmomentet var utfört visades stor uppskattning på belöningssystemet från samtliga deltagare. De var dock ivriga att antingen klicka in sig i spelet igen, eller försöka spela ett annat spel, varför första momentet i spelomgång 2 började med att de av misstag klickade på den tallrik som fanns just där de nyss klickat. Snabbare animation på belöningen samt constraints emot att klicka innan kaninerna radats upp löser förhoppningsvis problemet.

8.1.3 Andra spelfasen, Teachable Agenten tittar på

Barnen lyssnade inte speciellt bra på förklaringen till att man skulle visa Teachable Agenten. Man bör kanske ha en skärm där Teachable Agenten kommer fram och förklarar att den också vill lära sig. Genom att visa en lugnare bild, utan klickbara alternativ, tvingar barnet att ta del av den information vi vill förmedla.

I övrigt gick även denna del som tåget, bortsett från problemet med att klick i huvudmenyn följde med in i spelomgången.

8.1.4 Tredje spelfasen, Teachable Agenten spelar

Enbart en testdeltagare förstod hur man skulle göra, resterande behövde instruktioner. De försökte spela vidare som de gjort innan, klickade frenetiskt på rätt tallrik och undrade varför det inte hände något. Efter att de klickat på rätt ansikte återgick de till att klicka på tallrikarna ett par gånger innan de åter kom på vart de skulle klicka.

En lösning skulle vara att plocka bort alla tallrikar utom den som Teachable Agenten väljer, för att tydligare belysa vad som efterfrågas. Risken med det är att konceptet med tallinjen försvinner, så en annan lösning kan vara att dimma ner de andra tallrikarna. Därefter bör de nya knapparna göras tydligare, i nuläget är de helt osynliga för barnen. Ändrad storlek, animering för att belysa att något nytt skett, eventuellt även flytta placeringen.

8.1.5 Teachable Agenten

Ingen deltagare förstod meningen med vår Teachable Agent. En del ryckte på ögonbrynen när Teachable Agenten pratade med tjejröst. De flesta barnen såg Teachable Agenten som en pojke, Teachable Agenten bör därför göras mer androgyn. Kanske bör både bonde och Teachable Agenten presenteras i

[11] http://goteborg.se/wps/portal/enheter/forskola/forskolan-lars-kaggsg

[12] http://www.camtasiasoftware.com/camtasia/index-camtasia-studio-eu.htm

en filmsekvens innan spelet drar igång. Det var inte tillräckligt tydligt vilken röst som tillhörde vem.

8.1.6 Diskussion med lärare

Främst diskuterades svårighetsgraden. Gemensamma uppfattningen var att barnen tyckte spelet var för lätt. Förmodligen har vi underskattat kunnandet för målgruppen. Samtidigt kan detta balanseras med högre svårighetsgrader, samtidigt som även yngre personer kan använda spelet.

Läraren föreslog att direkt använda siffror, och att även byta ut glad och ledsen gubbe mot ”Rätt” och ”Fel”, eftersom de är i ett stadie där de lär sig känna igen ord och bokstäver.

Barnen samarbetar gärna, varför det vore positivt om man tydligt visar vilken nivå man klarat av. Om ett barn fått tio morötter, medan ett annat bara klarat få fem, kommer det första barnet visa hur man får fler. På så vis får man även en analog Teachable Agent stundvis.

Barnen bör själva få välja en Teachable Agent. Detta kan även användas som inloggning och för att spara information om hur långt man kommit i spelet. Det bör även leda till bättre uppskattning för sin Teachable Agent.

Att få ett barn att spela i 15 minuter är ingen konst om spelet är bra. Förskolelärarna begränsar barnens speltid för att de inte ska sitta för länge klistrade framför skärmen.

9. Diskussion

9.1 Arbetsprocessen

Under inledningsfasen diskuterades mycket om hur vi i stora drag ville lägga upp vårt projekt. Vi beslutade att koncentrera första halvan av projektet till att ge oss en stadig bas att stå på och andra delkursen till att finslipa och utveckla spelet. Detta passade väldigt bra eftersom samtliga gruppmedlemmar tänkte läsa båda delkurserna.

Vi tog beslut rörande intern dokumenthantering, vilket resulterade i Google Documents, samt för intern kommunikation. För att enkelt kommunicera inom gruppen startade vi en grupp i WhatsApp-applikationen, vilket gör att man enkelt kan skicka meddelanden till varandras mobiltelefoner.

Med hjälp av brainstorming tog vi fram idéer till olika delspel. När vi beslutat om en huvudidé jobbade vi utifrån denna och diskuterade fram beslut rörande påbyggnad. Från handledarmötena fick vi mycket värdefull feedback om hur vi kunde förbättra vissa saker, och hur vi gjort bra beslut inom vissa områden. Prototypen togs fram med hänsyn till den teori och kunskap inom Teachable Agents och Number Sense vi tillägnat oss. Vår första prototyp byggdes i Balsamiq, men vi övergick senare till PowerPoint för att få tillgång till animationer för att kunna presentera hela vårt koncept.

Efter att ha byggt upp en stadig mid-fi prototyp kontaktade vi en utomstående expert på området kring spel för barn, där vi ventilerade frågor och fick givande information rörande hur vi bör fortsätta utvecklingen med vårt spel.

Inför utvecklandet av hi-fi prototypen sattes ett versionshanteringssystem upp och det första steget efter detta var att förstå hur implementationen av ett spel i HTML5 med JavaScript skulle ske. Eftersom ingen i gruppen hade tidigare erfarenheter av KineticJS så utgick vi från ett exempel-projekt för att förstå upplägget. Efter att vi förstått hur grundläggande bildinläggning och animering fungerade så implementerade vi första spelfasen, där spelaren spelar själv, med den grafik som användes i mid-fi prototypen.

Vi fortsatte implementera logiken i spelet med att skapa en startskärm där man kan välja spel samt spelfas två och tre, där Teachable Agenten först kollar på när spelaren spelar och sen spelar själv. I läget där Teachable Agenten spelar själv hade vi ingen klar bild över hur det skulle fungera och vi var tvungna att diskutera fram ett handlingsätt för Teachable Agenten skulle fungera i praktiken. Då grunden i spelet var implementerad byttes grafiken ut mot nya bilder som höll en konsekvent stil. Dessa nya bilder är skapade av Magnus Haake utifrån våra idéer och synpunkter.

Det sista steget i utvecklingen av hi-fi-prototypen var att implementera stöd för ljud och att justera alla fördröjningar och animationshastigheter.

Då spelet befann sig i ett testbart stadium kontaktades förskolan Lars Kaggsgatan 35 i Göteborg. Där utfördes sedan den mycket intressanta testningen som resulterade i matnyttig information. Eftersom vi stirrat oss blinda på vissa problem i spelutvecklingen, framförallt angående hur spelaren ska interagera med Teachable Agenten, var det välkomnande med nya synpunkter från målgruppen samt förskolepedagoger som dagligen arbetar med målgruppen. Det bekräftades att saker som tidigare diskuterats behövde förändras. Detta var bland annat svårighetsgraden i spelet. Att istället för det fixa systemet, som användes i hi-fi prototypen, använda sig av en ökande svårighetsgrad är att föredra. För att dra det hela ännu längre kan spelet känna av en passande svårighetsgrad för spelaren, men detta ligger utanför kursens ramar. Den andra inputen från testningen handlade om utformningen, samt hur interaktionen skedde med Kim, Teachable Agenten. Många av synpunkterna här var intressanta och borde inte vara alltför krävande att implementera i spelet. Slutligen gav testningen oss mycket att tänka på samt mycket nyttig information att ta upp i ett senare utvecklingsskede.

9.2 Testresultat

9.2.1 Barns beteende i samband med Freds Bondgård

Spelet Freds Bondgård uppskattades av testgruppen. Barnen tyckte spelet var roligt och var intresserade av att spela de andra delspelen som inte hunnit bli implementerade. Majoriteten av barnen ville klicka sig igenom spelet snabbt vilket resulterade i att en del buggar visade sig. Barnen hade svårt att sammankoppla röster med karaktärerna i spelet. Detta kan lösas genom att skapa animationer i form av munrörelser eller någon annan form av animation.

9.2.2 Hur barn interagerar med en "Teachable Agent"

När vi utförde våra test hade barnen svårt att förstå sig på Teachable Agenten. De flesta hade svårt att se vilket syfte Teachable Agenten hade i spelet. En viktig punkt är att barnen känner en koppling till Teachable Agenten, så dem verkligen vill lära den att spela. Designen av Teachable Agenten som vi använde tyckte barnen var tråkig och de blev inte intresserade av den. Om barnen skulle fått välja en Teachable Agent bland en grupp figurer till exempel djur, så hade förmodligen barnen blivit mer intresserade av vad Teachable Agenten gör och känt en större koppling.

En annan viktig del är att man gör en tydlig skillnad mellan när barn respektive Teachable Agenten spelar. Många av barnen fortsatte att spela som innan fast de egentligen skulle hjälpa Teachable Agenten. Detta skulle man kunna lösa genom att ha en tydligare mellanfas i de olika delspelen. Man kan också göra större designskillnader i spelläget så barnen förstår att det har skett en förändring. Detta kan vara att förtydliga att morötterna inte är klickbara eller, göra “Ja” och “Nej” knapparna mer synliga eller att det visas en kort animerad film om den nya fasen i spelet innan barnet får börja spela igen.

9.2.3 Barns kunskapsnivå och vad som är en lämplig svårighetsgrad för spelet

Då testet tyvärr inte enbart innehöll testpersoner i den aktuella målgruppen kan våra resultat vara missvisande. Barnens ålder varierade mellan fyra och fem år och ytterligare variation om barnen är födda tidigt eller sent på året. Resultatet angående svårighetsgraden blev att den var för enkel. Detta verkade även vara fallet hos de fyraåriga barnen. Dock testades spelet på en liten grupp barn och framförallt bara på tre fyraåringar. Detta gör att resultatet angående svårighetsgraden blir svår att analysera i det stora hela. Sättet svårighetsgradsproblemet bör lösas på är genom att alla barn börjar på den lägsta svårighetsgraden. Allt eftersom spelet fångar upp en förbättring av spelarens resultat ökas svårighetsgraden. Högre svårighetsgrader kan tänkas vara tärningsprickar och siffror istället för morötterna som finns i dagens version av spelet. Utöver detta behöver inte tallinjen börja på ett utan på exempelvis fem.

9.2.4 Number Sense och Teachable Agents i förhållande till målgruppen

Det är svårt att dra slutsatser om huruvida barns utveckling av Number Sense drar nytta av Teachable Agents utifrån det korta test vi genomförde. För att få tillförlitliga siffror hade vi behövt införa fler testpersoner, där en viss andel av personerna spelade en variant av spelet som inte använder sig av en Teachable Agent. Testerna hade även behövt köras över en längre tidsperiod än bara ett kort spelpass, för att vara säkra på att barnen faktiskt har blivit påverkade av Teachable Agenten.

9.3 Framtidsaspekter

Med den goda feedback bondgårdsspelet fått i olika sammanhang känns det som en självklarhet att på något sätt utveckla våra idéer ännu mer. Då det finns ett helt spelkoncept uppritat är det möjligt att en extern part går in och implementerar klart de tre sista delspelen. Detta har diskuterats då vi själva inte har tid på grund av framtida kurser, examensarbeten och liknande. Magnus Haake har själv uttryckt intresse av att ta över utvecklingsarbetet efter kursens slute i mitten av december. Det känns som att spelet därför kommer lämnas över i goda händer samt att vi i gruppen kommer hållas uppdaterade med framsteg och om eventuella saker och beslut som kommer ske runt spelet. Något som skulle vara mycket roligt är att någon gång se att bondgårdspelet används ute på förskolor och att det blir ett nytt och spännande sätt för fyraåringar att lära sig matematik på.

Referenser

[1] Blair Kristen, Schwartz Daniel L., Pedagogical Agents for Learning by Teaching: Teachable Agents

[2] Chase, C., Chin, D.B., Oppezzo, M. & Schwartz, D. (2009) Teachable Agents and the Protege Effect: Increasing the Effort towards Learning. Journal of Science Education and Technology

[3] Johnson, D., Gardner, J. & Wiles, J. (2004) Experience as a Moderator of the Media Equation: The Impact of Flattery and Praise. International Journal of Human-Computer Studies 61(3), 237-258.

[4] Chase, C., Chin, D.B., Oppezzo, M. & Schwartz, D. (2009) Teachable Agents and the Protege Effect: Increasing the Effort towards Learning. Journal of Science Education and Technology

[5] http://en.wikipedia.org/wiki/Theory_of_mind, 2012-10-23

[6] Barbara J. Reys & Robert E. Reys (1995). Perspektiv på Number sense och taluppfattning

[7] Medierådet (2010). Småungar & Medier

[8] http://kineticjs.com/, 2012-12-03

[9] http://git-scm.com/, 2012-12-03

[10] https://github.com/, 2012-12-03

[11] http://goteborg.se/wps/portal/enheter/forskola/ forskolan-lars-kaggsg, 2012-12-06

[12] http://www.camtasiasoftware.com/camtasia/index-camtasia-studio-eu.htm, 2012-12-06

Appendix A

Manual för mid-fi-prototypen

För att ta sig igenom mid-fi-prototypen på bästa sätt, följ denna manual:

1. På första sliden väljs kaninen för att komma till kaninspelet.
2. Nu visas tre kaniner. Här finns två alternativ. Antingen väljs tre morötter för att gå vidare eller så väljs två morötter för att se vad som händer då det valts för få morötter.
3. Om två morötter valts väljs sedan rätt antal, tre morötter. Då tre morötter valdes i förra steget hoppa över detta steg.
4. Nu visas två kaniner. Här väljs fyra morötter för att se vad som inträffar om för många morötter väljs.
5. När det nya försöket med två kaniner kommer välj två morötter för att komma vidare.
6. Nu har spelaren tilldelats en morot och det är dags för Teachable Agenten att komma in i bilden. I detta läge väljs fyra morötter, lika många som kaniner på skärmen.
7. Nu är det dags för Teachable Agenten att välja och spelaren att göra ett aktivt val om Teachable Agenten gjort rätt eller fel. Det som är implementerat i detta läge är att godkänna det som Teachable Agenten valt.
8. Nu visas den sista sliden av prototypen. Det är möjligt att börja om spelet här, men allting nollställs och itereras om som förklarat tidigare.

Appendix B - Testplan

1 Syfte
Utföra testning av hifi-prototyp av spelet Bondgården på målgrupp för att undersöka hur väl det lämpar sig för målgrupp.

2 Målgrupp
Barn i åldrarna 4-5 år.

3 Det som ska testas
- Hur barnen beter sig när de interagerar med ett digitalt spel.
- På vilket sätt barnen interagerar med en "Teachable Agent".
- Vad barnen kan och vad som är en bra svårighetsgrad för spelet.

4 Verktyg
- Laptop
- Camtasia Studio för registrering av prototypnavigering
- Tape-a-Talk Voice Recorder för intervjuer

5 Utförande
1. Inför varje test ställs fråga till varje enskild individ om tidigare erfarenhet av datorspel och datoranvändning.
2. Testobjektet får sedan spela igenom prototypen, varpå alla handlingar loggas för senare analys.
3. Efter genomgång av prototyp ställs fråga om hur testobjektet upplevde spelet, åsikt om Teachable Agent Kim och åsikt om själva spelidén.
4. Testet upprepas för nästa individ.
5. Då alla testbara individer i målgruppen genomfört testet ställs frågor till lärare angående syn på lärande datorspel, dess eventuella och förväntade effekter samt inställning till teori bakom vårt spel.

6 Analys
All data från testning sammanställs och analyseras för att ge svar på:
- Mappning - Hur påverkar vår mappning navigering i prototypen?
- Affordance - Har vår prototyp tillräcklig affordance, eller försvinner tid på att hitta rätt i navigeringen?
- Constraints - Är våra constraints tillräckliga eller klickas även ej klickbara alternativ?
- Spelglädje - Uppskattar målgrupp vår spelidé?
- Lärarsynpunkter - Hur ställer sig erfarna lärare till vår spelidé och lärande datorspel i undervisningssyfte?

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Lekplatsen – Ett lek-och-lärspel där förskolebarn lär sig grundläggande matematik genom att lära ut till någon annan

Simone Andersson Sofie Eliasson Mattias Götling Erik Jönsson Jonas Svensson

Sammanfattning

Vi presenterar ett lärospel i matematik för åldrarna 3-5 år. Spelet är tänkt att utveckla barnets intuitiva känsla för tal, såsom känslan för vad en siffra symboliserar, vad det innebär att ett tal är större än ett annat och förmågan att inse om ett svar är rimligt, så kallad *number sense* . Rapporten följer utvecklingen av två stycken delspel, från idé till färdig implementation, med särskilt fokus kring de designval som gjorts under utvecklingen och vilka resultat dessa designval fick genom testning på en förskola. Med hjälp av en digital karaktär, så kallad *Teachable Agent*, som barnet lär upp genom att visa hur man gör och komma med feedback när agenten gör olika val, ska barnen lära sig matematik genom *learning by teaching.* Målet är att barnet på ett roligt sätt ska utveckla sina kunskaper tillsammans med agenten. Genom testning på den färdiga produkten vid en förskola i Lund kom vi fram till följande:

- Agenten lockar till interaktion med spelet och skapar glädje.
- Olika former och utföranden på interaktions-verktygen såsom prickar och streck återkopplar barnen till en siffra.
- Implementation till en ipad och navigeringen som den lockar till passar barnen bra. Med hjälp av rörelser som drag-and-drop lyckas barnen interagera med spelet.

1 Introduktion

I kursen MAMN10 och MAMN15 har gruppen fått i uppgift att utveckla ett matematikspel för barn i förskoleåldern. Spelets huvudsakliga uppgift är att utveckla barnets så kallade number sense. Detta innebär att användaren ska få en känsla för vad tal innebär, vad det betyder att ett tal är större än ett annat och hur tal kan representeras på olika sätt. Unikt för spelet är också att det ska innehålla en så kallad Teachable Agent (TA), dvs en slags digital karaktär som barnet är lärare till. Teachable Agents har använts tidigare med högre åldrar och har visat sig ha en positiv effekt på inlärningen, men det är ännu otestat hur barn i denna åldersgrupp reagerar på den.

1.1 Bakgrund

Rapporten beskriver utvecklingen av ett delspel till den tilltänkta plattformen Mworld. Mworld är ett samarbete mellan School of Education, Stanford University och Educational Technology Group, Lunds universitet och Linköpings universitet. Spelet är ett projekt där man ska uppmuntra små barn att lära sig grunderna till talsystemet på ett roligt och lättsamt vis. Spelet använder sig även av en TA. Det är ett pedagogiskt verktyg som innebär att barnet, istället för att enbart själv spela, lär upp en agent som sedan får fatta egna beslut utifrån vad den har lärt sig. Tanken är att detta ska skapa ett mer interaktiv spelmiljö än att bara ha rätt eller fel, samt att barnet får följa med agenten under dess utveckling. Genom att barnet får förklara för en annan varför något är fel får den en djupare förståelse och blir mer engagerad. [1] Användandet av digitala karaktärer och lärkompanjoner av olika slag har också visat sig ha en positiv effekt för unga flickors intresse för matematikundervisning. [4]

1.2 Syfte och mål

Syftet med rapporten är att man ska få följa projektet från en idé till en färdig lösning samt fördjupa sina kunskaper inom de teorier som ligger i grund för utvecklingen av spelet till Mworld. Målet är att med den information och det material som getts i kursen tillsammans med information och material som tagits fram på egen hand skapa en färdig prototyp som är redo för testning inför delkurs två MAMN15. Prototypen kommer att bestå utav två delspel där det ena kommer att lägga fokus på jämnvikt medan det andra kommer att representera tallinjen och hur man kopplar samman den med hjälp av siffror och grupper av symboler.

1.3 Struktur

Rapporten kommer att leda läsaren hela vägen från det första utkastet till hur gruppen har skalat ner idéerna och genom feedback från handledare och övriga kursdeltagare lyckats skapa en fungerande prototyp. I texten kommer läsaren att få del av de teorier och budskap som ligger bakom utvecklingsmiljön, agentens roll och dess utveckling i spelet samt hur användaren med hjälp av feedback kommer att belönad och inspirerad så att användaren vill fortsätta att spela. Rapporten innehåller också de frågeställningar som har satts upp av projektgruppen under delkurs 1 och som söktes besvaras genom testen i delkurs 2.

2 Teoretisk bakgrund

2.1 Teachable Agent

Ett vanligt fenomen som ofta uppstår i pedagogiska sammanhang är att en student eller elev studerar en matematisk formel eller begrepp men att budskapet sjunker in först när studenten själv har varit tvungen att förklara den för någon annan, så kallad *learning by teaching*. För att utveckla det beteendet har forskarna utvecklat så kallade TA:s där användaren med hjälp av en digital karaktär kan öka den egna förståelsen för ämnet och förbättra problemlösningsförmågan. En TA är besläktad med andra typer av virtuella karaktärer som är vanliga i lärspel, såsom agenter och lärkompanjoner, men har också egenskapen att de kan lära sig saker. En relativt vanlig variant är att barnet och TA:n turas om att spela och svara på frågor. TA:n använder sig av artificiell intelligens för att generera frågor och svar baserat på de problem som ska lösas och hur eleven tidigare har svarat. Baserat på agentens svar så kan användaren ta ställning till informationen som ges och avgöra om den är korrekt eller ej. TA:n blir bättre och bättre genom att den svarar rätt på en större del av frågorna, eller genom att den kan svara på svårare och svårare frågor. På så sätt kan användaren själv vara med och utveckla sin egen kunskap såväl som agentens. En stor skillnad med att använda sig av agenter istället för människor är att agenterna själva inte blir drabbade om studenten svarar fel eller leder agenten till felaktiga val, något som kan få konsekvenser ute i verkliga livet. [10]

Det har gjorts forskning som visar på att undervisning i matematik med hjälp av TA:s har flera fördelar. Bland effekterna som nämns som positiva är dessa: [7]

- Det är lättare att bibehålla intresset.
- Det upplevs som mer meningsfullt när barnet lär ut till någon annan än om det själv skulle göra uppgifterna.
- Det är lättare att anpassa för svaga elever, genom att dessa lär upp en digital karaktär. Misstag som görs läggs på agenten och inte barnet. Detta brukar kallas ego-protective buffer.
- Genom att barnet får förklara varför något är fel ger det en djupare förståelse än om det enbart hade bett den lösa uppgiften

En TA kan utformas på olika sätt. Exempelvis kan det finnas en variabel som håller reda på hur stor procent av frågorna eleven har svarat rätt, och därefter göra så att TA:n svarar rätt motsvarande antal. En mer avancerad metod är att agenten håller reda på vilka sorters frågor eleven svarar rätt på, eller vilka fel den gör, och därefter svarar på frågorna. En Teachable Agent kan också ge feedback till användarens val, komma med kritiska frågor och ge spelet en storyline och därigenom göra spelet mer spännande.

Det är värt att notera att forskning som har gjorts på TA:s har utförts på barn i äldre åldrar, men att det saknas forskning på hur undervisning med TA:s påverkar barn i fyraårsåldern. Det finns alltså inget som stöder att de fördelar som TA:s-tekniken för med sig även gäller för barn i fyraårsåldern.

2.2 Number sense

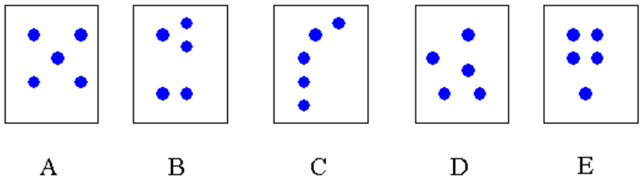
Number sense är ett begrepp som avser förmågan hos en person att förstå och länka samman siffror och samband för att lösa matematiska problem. Det handlar inte i första hand om utantill kunskap, eller att man ska lära sig en speciell algoritm för att lösa ett problem, utan mer om förståelse. Begreppet är delvis svårdefinierat och olika forskare ger olika förklaringar på vad det är, men enligt de flesta definitioner så är number sense: [5, 6]

- Förmågan att förstå vad tal är och vad det symboliserar.
- Förståelse för vad det innebär att ett tal är större än ett annat.
- Grundläggande känsla för vad olika operationer såsom addition och subtraktion gör.
- Förstå saker som ändring i storlek.
- Intuitiv känsla för om ett svar är rimligt.
- Grundläggande förståelse för tallinjen.

Förmågan att observera antalet föremål ur en mängd, utan att direkt räkna antalet, brukar benämnas med det engelska ordet *subitizing*. En person som studerade detta var den schweiziske psykologen och filosofen Jean Piaget. Vid två års ålder kan ett barn upptäcka en, två och tre föremål ur en mängd utan att behöva räkna dem. Vid fyra års ålder kan de identifiera upp till fyra föremål. En vuxen människa kan utan direkt träning eller organisering lyckas identifiera fem föremål. Genom grupperingar blir det lättare för personer att bedöma antalet. En samling av åtta mynt kan exempelvis delas in i två grupper med fyra mynt i varje. Genom att först dela upp det i två högar med fyra i varje, och sedan dela upp dem i två nya högar, kan vi snabbt dra slutsatsen att det är åtta stycken mynt på bordet. Tekniken att organisera prickar i olika formationer och logiska mönster ses på en vanlig tärning. Genom att se två rader med tre prickar i varje kan vi snabbt säga att det är 6 prickar på tärningen. Utan ordning eller uppdelning i mindre grupper måste de flesta börja räkna prickarna för att bestämma antalet.

Genom att observera figur 1 är det möjligt att själv försöka urskilja prickar från en mängd och observera de svårigheter det kan ge. I bild A är den klassiska formen av tärning för symbolen av en femma. I bild B och E går det snabbt genom att analysera grupperingarna, addera ihop grupperna utav prickar och få slutsatsen att antalet är fem. I bild C är prickarna placerade på en linje och direkt blir det svårare att plocka ut antalet och även i bild D är det oreda då prickarna inte tycks följa något konkret mönster. Detta stärker Piagets teori om organisering och sambandet mellan föremål och dess koppling till siffror. [5]

I de lägre åldrarna är det vanligt att barn lär sig tekniker för att räkna, medan förståelse för varför ett tal är större än ett annat ofta kommer senare. Ett barn i 5-års åldern kan till exempel ofta läsa talen 62 och 26, och kan kanske till och med veta att 62 är större, men behöver inte nödvändigtvis förstå varför. För barn i förskoleåldern är det inte ovanligt att barnet kan räkna till 13, utan att förstå att 13 är en grupp av 10 och 3. [2] [3]



Figur 1: Fem olika bilder och mönster som alla innehåller samma antal prickar i figurerna.

2.3 Feedback

2.3.1 Rätt/fel - feedback

Det konventionella sättet att utforma lärspel är att eleven får en uppgift eller fråga, och får sedan veta om svaret var korrekt eller inte. Detta brukar kallas rätt/fel-feedback.

2.3.2 Konsekvensfeedback

Konsekvensfeedback fokuserar istället på vad konsekvenserna av användarens val blir. Om exempelvis barnet bygger en stol, och något av stolsbenen är för kort, kommer stolen att rasa ihop när någon sätter sig på den. Tanken bakom konsekvensfeedback är att barnet ska få en djupare förståelse kring varför ett val leder till en händelse istället för att bara få ett rätt/fel och få veta vad den kan göra för att rätta till sina misstag.

3 Spelidén

Gruppen satte sig tidigt ner för att bolla idéer och redan på ett tidigt stadium beslutades att en lekplats som spelmiljö skulle användas. Anledningen till detta var flera: en lekplats är någonting som alla barn känner till, de förknippar det oftast med något positivt och det finns väldigt stora möjligheter att utveckla flera delspel genom att ha olika stationer på lekplatsen. Det första delspelet blev en gungbräda, då detta är en naturlig miljö för att lära sig jämvikt mellan olika vikter. Det andra delspelet blev en rutschkana, där trappstegen representerar tallinjen. Det diskuterades även om att ha en station med kulkastning där det var tänkt barnet skulle utveckla en förståelse för längder och jämförelser (vilket barn kastar längst osv).

I takt med att TA:n lär sig mer kommer den att växa. Gruppen bestämde sig för att TA:n skulle vara någon form av djur och valde att representera den i form av en tiger. Förhoppningen är att en tiger ska göra att spelet blir lite mer spännande. Genom arbetet av prototyperna har ett flertal olika tigrar använts men i slutändan valdes den som tagits fram i samarbete med en extern grafiker.

Från början var tanken att endast ett fåtal stationer skulle vara öppna. Barnet/TA:n skulle då få möjlighet att tjäna pengar genom att svara rätt på frågor. Pengarna skulle senare användas för att exempelvis köpa en bättre rutschkana och på så sätt låsa upp stationer. Denna idé övergavs dock så småningom eftersom det skulle vara för krångligt för barnet att hålla reda på att både TA:n utvecklas och att de skulle få pengar som senare kunde användas för att låsa upp stationer.

3.1 LoFi prototyp

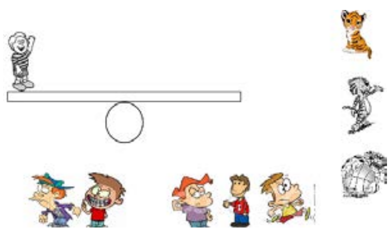
LoFi-prototypen utvecklades i Balsamiq, som är ett prototypverktyg särskilt inriktat på framtagning av användargränssnitt. Verktyget liknar PowerPoint, men med skillnaden att prototyperna har ett mer skissartat utseende. I Balsamiq-modellen användes både egenritade bilder och hämtade clip-art bilder på Internet. Här utvecklades de första prototyperna av delspelen gungbrädan och rutschkanan. Genom att lägga in knappar som tog användare till en annan bild kunde en ganska stor del av funktionaliteten implementeras. Detta gav gruppen en möjlighet att tillsammans utveckla och bygga fram en prototyp och arbeta fram en spelidé som sedan skulle kunna utvecklas till ett riktigt spel.

3.1.1 Gungbräda

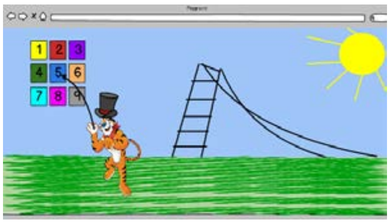
Figur 2 visar en tidig version av gungbrädan. Konceptet bygger på att användaren vid spelets början får tillgång till gungbrädan och att det står ett slumpmässigt antal personer på vänstersidan. Det är sedan upp till användaren att dra över rätt antal barn till högersidan för att på detta vis skapa jämnvikt. Tidigt kom idéen att gungbrädan skulle förflytta sig beroende på brädans belastning precis som i verkliga livet. Diskussion uppstod om det verkligen var lämpligt att använda sig av människoliknande karaktärer. Finns det några risker vid igenkännande och skulle barnen missförstå idéen och istället reta sig på utseendet hos figurerna?

3.1.2 Rutschkana

Figur 3 visar ett tidigt utkast på rutschkanan, spelet går ut på att gissa antalet steg upp till rutschkanan förändras under spelets gång. Användaren ska sedan räkna stegen



Figur 2: Bilden visar vår tidigaste skiss av gungbrädan i Balsamiq. Till höger syns ett förslag till agent och hur de olika stadierna i utvecklingen skulle kunna se ut.



Figur 3: Bilden visar vår tidigaste skiss av rutschkanan i Balsamiq

Figur 4: Bilden visar när agenten spelar gungbrädespelet (HTML5-implemetation).

och med hjälp av siffrorna i boxen till vänster, länka samman detta och få agenten att nå till toppen och kunna åka ner för rutschkanan. Tanken är att på svårare nivåer ska siffrorna kunna bytas ut mot punkter eller streck för att utveckla number sense hos användaren. Ett centralt tema inom number sense är att eleven ska kunna förstå kopplingen mellan exempelvis tre streck och tre prickar och att de har någonting gemensamt, nämligen deras antal.

3.1.3 Agent

Gruppen hade en tidigt en idé om att agenten skulle vara annorlunda, något som skulle locka barnen till att spela och som inte är för töntig. Valet föll snabbt på en tiger. Målet är att den i takt med att barnet blir bättre och klarar uppdragen ska utvecklas och bli större och starkare samt även bättre på matematik. För att detta skulle ske så fanns det idéer om att den efter avslutad omgång skulle få sig en portion av supermat beroende på hur duktig användaren hade varit på den aktuella nivån. Detta väckte dock frågor som ”När ska detta ske?”, ”Ska det ske när användaren har klarat en nivå på vardera delspel eller ska agentens utveckling följa med mellan de olika spelen?”

3.2 HiFi prototyp

Implementeringen i Balsamiq fungerade bra och var utmärkt att använda för att utveckla idéer, dock har Balsamiq sina begränsningar. Framför allt visuell feedback, exempelvis att gungbrädan gungar till lite när det läggs till en karaktär på gungbrädan, trappstegens förändring samt att det var svårt att implementera TA:n och arbeta med dess utveckling. För att komma vidare i projektet började gruppen med att göra en HiFi-implementering i HTML5 med hjälp av JavaScript-biblioteket jQuery. Utvecklingsplattformen som användes är Eclipse. Det är en miljö som gruppen kände sig hemma i eftersom alla i gruppen läst programmeringskurser där detta har använts. Dessutom har den ett bra stöd för sk. *subversion*, dvs. versionshantering och gemensam tillgång av källkoden.

3.2.1 Gungbräda

Med Balsamiq-prototypen färdig gick utvecklingen av spelmiljön snabbt framåt. Utseendemässigt har inte mycket förändrat sig från prototypen, agenten har förflyttat sig från högerkanten ner till vänster och har fått en mer inflytande roll. Den står nu i området där själva spelet äger rum och inte uppe i ett hörn. Utvecklingen med olika nivåer har förbättrats ordentligt. Användaren följs genom sex olika nivåer där användaren först spelar ensam med aporna för att lära känna miljön och interaktionen.

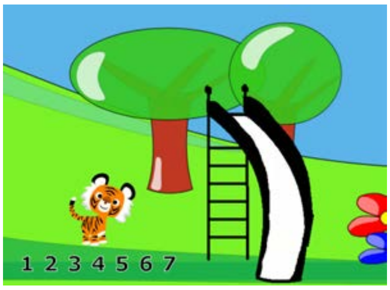
Användaren ska själv flytta upp rätt antal apor på gungbrädan för att klara nivån. På nästa nivå ska användaren sedan förflytta upp grupper utav apor på gungbrädan för att skapa jämnvikt. Här kan användaren flytta upp flera grupper om så önskas. Om det till exempel är tre apor på vänster sida så kan användaren antingen ta upp gruppen med tre apor eller välja att ta upp en grupp med två och en grupp med en apa.

Det har även skapats en nivå där aporna byts ut mot tärningar för att ytterligare utveckla användarens number sense och förmåga att organisera antal. När användaren har spelat klart de första tre nivåerna kommer de sedan att starta om från början dock kommer aporna på vänstra sidan av gungbrädan att bytas ut mot en siffra mellan ett och fyra.

Agenten kan vara med och spela i detta delspel. Efter avslutad nivå tar agenten över och användaren ska då själv avgöra om agenten har tänkt rätt eller fel. Agenten anger antalet apor som den tror ska upp brädan och användaren får själv antingen trycka på den gröna tummen eller den röda tummen för att visa om agenten har valt rätt eller inte, se fig 4. Vid rätt val flyttas aporna till brädan och de kan gunga lugnt en stund och vid fel val kommer brädan att röra sig mot marken på den sida av brädan som blir tyngre. I nuläget är agenten implementerad så att den ger ett felaktigt svar slumpmässigt med en faktor på ett av tre förslag. Anledningen till detta är att spelomgångarna är ganska korta och användaren kommer vidare till nästa nivå efter bara tre korrekta försök. En förbättring av antalet rätt gissningar skulle helt enkelt inte märkas innan man kom vidare till nästa nivå. Agenten kan istället ses som att den ”utvecklas” genom att den kan svara på svårare och svårare frågor.

3.2.2 Rutschkana

Rutschkanan har även den utvecklats. Dock har en del frågetecken uppstått under utvecklingens gång: hur ska rutschkanan och stegen utformas och vad ska räknas som trappsteg? Utseendet på produkten har även förändrats en del och siffrorna från LoFi-prototypen har flyttats ner till nedre kanten av skärmen för att skapa likhet med gungbrädespelet. Siffrorna har också lagts ut på en linje istället för att organiseras som en box, vilket förmodas ge barnet en starkare känsla av tallinjen. Figur 5 visar spelet



Figur 5: Bilden visar rutschkanespelet implementerat i HTML5.

på grundnivån när användaren ska välja vilken siffra som motsvarar antalet trappsteg på rutschkanan. Genom att klicka på siffrorna så förflyttar sig agenten upp och ned på stegen efter valt antal. I och med att siffrorna har lagts ut vågrätt, medan stegen är lodrätt tränar användaren på att tallinjen kan visas både lodrätt och vågrätt, och lär sig därmed att tal kan representera både höjd och längd. Detta kan även leda till förvirring, men lärdomen att samma tal kan representera både höjd och längd är så pass viktig och fundamental för fortsatta geometriska studier att det ändå motiverar designvalet.

Feedback har implementerats till rutschkanan genom att agenten kan klättra upp och ned från stegen. Kommer den till en felaktig nivå så klättrar den antingen ner igen om siffran blev för liten eller upp i ”ingenting” tills den sedan faller ner till marken likt en tecknad serie.

3.2.3 Agent

Efter olika bilder och skisser på tigern så valdes det en bild som var enhetlig för båda spelen. De svåraste frågorna kring agenten under utvecklingsfasen var hur utvecklingen av agenten ska ske och hur tigern naturligt ska kunna klättra upp för stegen på rutschkanan. Gruppen bokade in ett antal möten med en grafiker inför delkurs två där dessa och andra frågor hanterades.

4 Design

4.1 Konceptuell design

4.1.1 Spelets övergripande struktur

Med en målgrupp på fyraåriga barn i åtanke drogs projektetgruppen till att utforma spelmiljön med hjälp av direktmanipulation, en teknik där användaren manövrera sig direkt i tjänsten och inte via en kommandobaserad tjänst. Barnen ska själva kunna starta spelet och komma igång utan någon vuxen till hjälp och för att göra detta krävs en miljö där användaren själv kan manövrera runt med hjälp av ett finger eller med ett hjälpmedel som en datormus. [9] Med dagens teknik är det dumt att begränsa sig till enbart datorer utan det är bra om det funkar på flera plattformer så som surfplattor och mobiltelefoner. Spelet är upplagt så att det ska kunna köras både via datorinkopplad mus, men också via touchscreens och dess egenskaper av drag-and-drop för att kunna interagera med applikationen. När spelet startas så kommer användaren att mötas av en huvudmeny där den kommer att få se lekplatsen och dess redskap. Därifrån kommer användaren att få välja de redskapet som de vill interagera med och sedan kopplas vidare till det valda delspelet.

4.1.2 Synlighet, affordance, mappning och feedback

Med ett spel anpassat för fyraåringar är det viktigt att det känns naturligt att interagera med produkten. Det är viktigt att olika kognitiva lagar och principer såsom lagen om närhet, gruppering och kunskap om människans närminne tas i beaktande under utvecklingen så att produkten blir användbar. I gungbrädespelet har aporna blivit indelade i boxar som är fyllda med olika färger för att öka synligheten vid interaktion. Olika grupper har olika färger och detta för att ge en känsla av att de hör samman. Detta är även kopplat hand i hand med affordance till aporna. Boxarna hjälper aporna att se ”klickbara” ut och ska leda till interaktion sk. *affordance*.

Tanken är att användaren ska vilja fylla tomrummet på ena sidan av brädan med de övriga aporna som visar sig på skärmen. När agenten sedan spelar kommer den att plocka upp apor från grupperingen för att symbolisera sitt val. Agenten visar det antal apor den gissar ska vara på gungbrädan genom att placera dem i en tankebubbla. Detta ger en naturlig mappning till användaren och ger igenkännande. De apor som inte placeras i tankebubblan tas bort för att inte användaren ska tro att de kan flytta på dessa. Detta för att smidigt förklara för användaren vad agenten tänker och vill göra. Feedback ges genom att gungbrädan gungar till vid interaktion precis som i verkliga livet. Den sida av brädan som är tyngre kommer att röra sig ner mot marken. När jämvikt nåtts kommer aporna att gunga en kort stund för att efterlikna hur en gungbräda gungar i verkligheten.

I rutschkanan är siffrorna placerade på samma nivå som aporna i det andra spelet. De har en vit kant för att öka synligheten och skapa markera att de går att interagera med. Mappningen ligger även här i att det används ett redskap som barnen känner igen och det ska vara naturligt att veta vad som ska göras. När en ny omgång startar så kommer stegen till rutschkanan dyka upp en efter en och sedan kommer siffrorna att förflytta sig inåt för att ge både synlighet och mappning åt användaren. I och med att siffrorna och stegen dyker upp vid ungefär samma tidpunkt är förhoppningen att det ska vara enkelt för användaren att förstå att det finns ett samband mellan dem. Vid val av antal steg kommer agenten att förflytta sig valda antal steg. Feedback kommer här att ges till användaren beroende på val. Vid för stort antal steg kommer TA:n att fortsätta att klättra antalet valda steg upp i luften för att sedan falla ner mot marken, vid för litet antal steg kommer TA:n att stanna på stegen efter antalet valda steg och sedan klättra ner till marken igen. Om användaren väljer rätt antal steg kommer målet att uppfyllas och TA:n kommer att åka ner för kanan.[8]

4.1.3 Gruppering och placering

Trots att spelidéerna är olika och har olika sätt att lösa uppgiften, har valet blivit att placera de interaktiva föremålen på samma ställen i delspelen. Då det är naturligt att föra blicken från vänster till höger och det finns naturliga rörelsemönster uppifrån och ner har valet blivit att placera objekten därefter för att få in hela sinnesbilden.

Agenten är placerad till vänster och speciellt när agenten själv spelar krävs det att den är tydlig, så att användaren får upp ögonen för vad den vill säga. Till höger är själva förmålet där spelhändelsen kommer att ske, både rutschkanan och gungbrädan är placerad där och här kommer synintrycken att påverka användaren för att kunna välja några av tjänsterna placerad längst ner i fönstret. I spelet med gungbrädan är aporna placerade både enskilt och i mindre grupper där de går att urskilja dem med hjälpa av olika färger för att visa att de hör till samma grupp. I rutschkanespelet är siffrorna placerade på samma ställe som aporna för att ge en känsla av likhet mellan delspelen. Strecken och prickar kommer att placeras i rutor eller boxar för att visa att de hör ihop. [8]

4.2 Design av interaktion

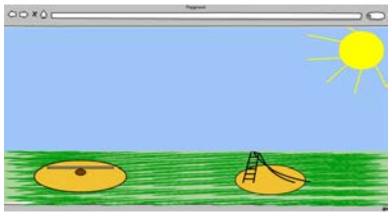
4.2.1 Interaktionsteknik

Eftersom målgruppen är små barn i åldrar runt fyra år så måste produkten vara lättanvänd och den ska inte innehålla onödiga funktioner som kräver text eller förklaring. Det blev tidigt bestämt att produkten ska i huvudsak ska använda sig av direktmanipulation av enkla grafiska objekt så att barnen ska kunna använda den utan en vuxen person närvarande som ska förklara informationen på skärmen. Spelet har dubbelkodning för att hjälpa barnet med inlärning av siffror och i både gungbrädespelet och rutschkanespelet visas hur många apor som är på gungbrädan/antalet steg med hjälp av siffror. Detta även på de lägre nivåerna när de inte används i själva spelet. Detta gör att barnet passivt får in kopplingen mellan ett antal och en siffra.

Ett annat designval är att hålla gränssnittet så rent som möjligt och inte ha en massa knappar och verktyg som inte fyller någon funktion. [8, 9]

4.2.2 Översikt, orientering och navigering

Som tidigare nämnts är huvudtanken bakom spelet en lekplats, med delspelen gungbrädan och rutschkanan. Som huvudsida visas en översiktsbild av lekplatsen, där användaren kan välja mellan olika delspel. Huvudsidan är hittills bara implementerad i LoFi-prototypen och ett exempel syns i figur 6. I arbetet med LoFi-prototypen fanns även en idé om att spelet skulle innehålla pengar som användaren kunde spendera för att "låsa upp" olika banor. Denna idé övergavs senare för att det riskerade att göra spelet och navigeringen för rörigt. Hur användaren ska kunna ta sig tillbaka från de olika delspelen till huvudsidan hade inte bestämts i LoFi-prototypen.



Figur 6: Spelets huvudmeny (gränssnittsskiss från Balsamiq)

4.2.3 Felhantering

Ett viktigt mål vid utformningen av interaktiva produkter är att produkten ska tala för sig själv och möta användarens krav. Designen är därför väldigt sparsam och produkten är utformad mot målgruppen på ett sådant sätt att inga onödiga knappar är synliga eller att grafiken innehåller för många element som inte går att interagera med. De fel som i första hand kan uppstå i spelet är när användaren i ett delspel väljer ett felaktigt alternativ i något av delspelen. Det är i första hand detta som avses med felhantering i avsnittet nedan. Ett felval hos användaren, inte eventuella buggar i spelet.

I spelet med gungbrädan syns felet klart och tydligt genom att gungbrädan lutar åt det håll där brädan väger tyngst. Detta är en form av konsekvensfeedback. Vid jämnvikt kommer den lyftas upp mot mitten för att låta aporna gunga en kort stund innan spelomgången avslutas.

I rutschkananespelet kommer agenten oavsett val av siffra kommer att klättra upp på stegen. Väljer den ett för kort antal steg så kommer den att stanna på motsvarande plats på stegen vid för högt val kommer den att klättra upp i luften för att sedan landa på marken och symbolisera felet på detta vis. Även i detta delspel används konsekvensfeedback.

4.3 Grafisk design

4.3.1 Layout

Målet är att hålla layouten enkel. Den ska inte ha några onödiga detaljer och var sak ska kunna få sin plats på skärmen. Då spelet även ska fungera på olika plattformer innebär det en ökad förståelse för vad som ska finnas med och inte. Med en surfplatta med skärmstorlekar från 8 tum ska den fungera lika väl på den typen av skärm som på en datorskärm med betydligt större storlek och upplösning. Det är också viktigt att knapparna och de interagerbara objektens storlek är anpassad på ett sådant sätt att de går att interagera med för barn med små fingrar.

4.3.2 Färg

Då målgruppen är barn är det positivt om produkten är färgglad och har detaljer och mönster som lockar barnet till att använda produkten. Då delspelen även utspelar sig i en utomhusmiljö har färger och former valts som uttrycker detta. Som tidigare nämnts så är aporna placerade i färgade boxar för att ge dem en grupptillhörighet i gungbrädespelet. Olika färger symboliserar olika tillhörighet och hjälper användaren att lösa uppgiften.

4.3.3 Ikoner och interaktiva element

Spelet använder sig sparsamt utav ikoner och knappar för att leda användaren under spelomgången. De ikoner som finns i spelet är knappar i rutschkanespelet som anger antal steg. Aporna i gungbrädespelet är drag-and-drop-objekt som är tänka att förflyttas från sin position genom att dra dem till gungbrädan. Siffrorna är tänkta att fungera likt en knapp och vid interaktion med denna kommer de att aktivera agenten. Då spelet är tänkt att fungera både till datorer och surfplattor, är ikonerna utformade i en storlek som möjliggör interaktion i båda miljöerna. Uppfattningen var under spelutvecklingen att det är mer intuitivt att dra olika element på skärmen än att klicka på dem, och därför används drag-and-drop i så stor utsträckning som möjligt där det verkar naturligt.

4.4 Ljud

Spelet använder ljud istället för text för att guida användaren genom spelet. Detta eftersom spelet är inriktat mot förskoleåldrarna och de flesta barn i målgruppen inte kan läsa. Alla instruktioner är inspelade med ljud, och även vissa ljudeffekter som jubel. Det spelas även upp ett ljud när TA:n klättrar upp för stegen som räknar upp för varje steg som TA:n tar. Rösterna är inspelade med ett så kallat text-to-speech program, för att ge en klar och tydlig inspelning som samtidigt var neutral. TA:n är implementerad med en manlig röst medan berättarrösten är kvinnlig, men detta går enkelt att ändra på i källkoden eftersom de allra flesta ljuden är inspelade för båda könen. Verktyget som använts är acapella-box. [11]

5 Arbetsprocessen

Gruppen valde att jobba mycket tillsammans under början av projektet och gärna sitta i samma grupprum och jobba. Två fasta tider i veckan sattes upp där vi diskuterar projektet och sedan jobbar sida vid sida. Alla gruppmedlemmar har kunnat delta nästan alla gånger. Mellan dessa fick medlemmarna uppgifter de kunde jobba med på egen hand. En stor del av tiden i början gick åt till att diskutera olika aspekter medan vi utvecklade LoFi-prototypen. Detta ledde troligvis till att alla i gruppen fick en väldigt bra förståelse för spelet, vad gruppen ville göra och teorin inblandad i det.

Gruppmedlemmarna har inte haft så formella grupproller i första delkursen utan alla hjälper till där det behövs. Självklart skapas ändå naturliga indelningar och någon kanske sitter med LoFi-prototypen för gungbrädan, någon annan med rutschkanan, medan någon tredje tittar på testning. Till en början var projektledaren den enda formella rollen. Det är först i och med arbetet med HiFi-prototypen och rapportskrivningen som en riktig arbetsindelning började ta form.

Inför delkurs 2 planerade vi att utveckla koden vidare. Fokus skulle ligga på att koppla ihop spelen, skapa konsekvens mellan dem och få TA:n att bättre komma ihåg vilken nivå den är på. Gruppen fick också kontakt med en extern grafiker, Lisa Wallin, som hjälpte till och förbättrade grafiken utifrån de idéer vi hade i den tidigare delkursen. Några veckor in i den andra delkursen gick gruppen ut och testade spelet på en förskola. Gruppen spelade också in en film som presenterar projektet/spelet och lade till röster på spelet, så att barn som inte kan läsa förstår instruktionerna. Under andra delkursen har arbetet mera delats upp, men med täta samarbeten och långa sessioner där vi sitter tillsammans under dagarna innan och kring testningen. Kontakten med grafikern har skett genom möte på fik, skype-möten och mailkontakt.

6 Testning

6.1 Introduktion

I delkurs två, MAMN15, hade projektet nått en nivå där gruppen ville testa spelen se hur väl det uppskattas av målgruppen. Med hjälp av ipads lät vi barnen testa på spelet i sin naturliga "arbetsmiljö". Genom observationer och samtalen med barnen fick vi en bekräftelse på vad som var positivt och vad som kan tänkas uppdateras.

6.1.1 Problemformulering

Med oss från slutredovisningen i delkurs ett MAMN10 hade vi samlat ihop en del frågor som vi önskade få svar på under kursens gång. Vi hoppades även att frågorna skulle ge oss svar på hur lärorikt och användbart spelet är. Frågorna listas nedan:

- Hur enkelt är det att orientera sig i spelet?
- Hur lång tid tar det för testpersonen att komma igång?
- Hur lång tid tar det för testpersonen att klara en nivå?
- Förstår testpersonen kopplingen mellan redskapen och interaktionsverktygen?
- Känns agenten naturlig i spelet?
- Får skadebeteendet hos avataren vid "misslyckande" testpersonen att vilja fortsätta med detta?
- Känns det naturligt med drag-and-drop för testpersonen?
- I gungbrädespelet, kommer spelaren att se symboliken hos grupperingen utav färgerna?

6.2 Metod

6.2.1 Urval av försökspersoner

Vi besökte Backens förskola, avdelning Stenen. Det är en avdelning för barn i åldrarna fyra till fem. Av dessa barn ville tolv stycken delta i vårt test. Förskolan har i dagsläget ipads i sin undervisning där barnen får spela spel, så de barn som ville testa spelet hade ipad-vana sedan tidigare. Några av dem spelade även spel hemma både på surfplattor och mobiltelefoner med touchskärm.

6.2.2 Genomförande

Inför testet hade vi en kort frågestund med barnen för att få dem att känna sig bekväma men även få tillgång till information om ålder, kön och ipad vana. Barnen testade båda delspelen och inför varje spelomgång förklarade testledaren snabbt hur spelet går till för att sedan låta dem leka fritt med spelet under observation. Varje spelomgång planerades till att ta sju minuter detta för att bibehålla koncentration genom hela testet samt för att få tillräckligt med tid för att samla in testdata. Datan bestod av kvantitativ data i form av tid, missförstånd i användandet av produkten samt för genomförandet av de olika nivåerna i delspelen och kvalitativ data i form av svaren insamlade under frågestunden samt information de delade med sig av under spelomgången. Efter att båda spelen hade testats så avslutades testen med en kort frågestund där frågor ställdes om speluppleveIsen och vilket spel de tyckte var roligast.

6.2.3 Testmiljö och utrustning

Testerna utfördes på Backens förskola i ett hobbyrum där de båda testgrupperna satt i varsitt hörn och spelade spelet. Vi använde oss av två stycken ipad 4 med webbläsaren Google Chrome för att kunna starta upp och köra spelet.

6.2.4 Testansvarigas roller

Testledare Vid testets början förklarade testledaren för barnet att det är spelet om testat och inte de själva samt att om det skulle uppkomma problem under själva testets gång är det inte barnet som gör fel utan att det är verktyget. Testledaren ansvar var att förse försökspersonen med instruktioner innan testningen påbörjades. Under testet hade testledaren en mer avvaktande roll utan att bli för passiv och om det krävdes assisterade barnet i uppgiften utan att ge alltför ledande information om hur man löser den.

Observatör Observatörns roll var att sitta bredvid och ta anteckningar om hur försökspersonen går tillväga och vilka eventuella svårigheter som uppkommer under testet. Detta eftersom det kan vara svårt för testledare att hinna med detta och agerar som ett extra par ögon om svårigheter skulle uppstå.

6.3 Resultat

Under testens gång så användes ett formulär som observatören fyllde med information om hur väl barnen kunde använda produkten samt hur lång tid det tog för dem att lyckas med nivån. I tabell 7 kan man se hur väl barnen lyckades genomföra de båda nivåerna. Bedömningen delades upp i fyra nivåer där varje barn fick resultaten: J = Ja, lyckades lösa uppgiften utan hjälp eller med lättare introduktion; JmH = Ja, lyckades lösa uppgiften med hjälp utav testledare; N = Nej, lyckades inte på egen hand lösa uppgiften.; - = Spelade inte på nivån.

I tabell 8 kan man se hur lång tid det tog för barnen att genomföra nivån. Tiden mättes från att nivån startade till agenten tog över och spelade sin omgång. Efteråt avrundandes tiden till närmsta hel eller halv minut för att ge en tydligare uppfattning om tidåtgången. Vi hade inför testdagen beslutat att varje spelomgång inte skulle behöva ta längre än två minuter att spela och i figur 8 kan man se att resultaten av tiderna hölls väldigt bra.

6.4 Analys

Hur enkelt är det att orientera sig i spelet? Överlag fungerade det bra att orientera sig i appen, vissa utav barnen kunde få en liten tankeställare hur de skulle agera vid byte utav spelmiljö då placeringen av siffrors respektive apors placering skiljde sig mellan spelet. Dock hittade barnen snabbt till aktuell interaktionshörna och kunde påbörja spelet.

Ett problem som vi upptäckte var dock att några av barnen hade svårt att trycka på skärmen och placera fingret rätt för att interagera. Några barn använde fingertopparna och nageln för att styra vilket blev problematiskt och frustrerande för dem då interaktionen inte gick som de ville. Vi upptäckte även att många utav våra knappar var för små eller hade en form som inte var anpassat för ett barnsfinger. Vid knapparna för tumme upp och ned samt några av siffrorna i rutschkanespelet krävdes det ett ”vuxet” finger för att interaktionen skulle underlättas.

Hur lång tid tar det för testpersonen att komma igång? Under testdagen så ville tyvärr inte ljudet fungera med den aktuella hårdvaran men efter en snabb introduktion så kändes barnen direkt redo att ta sig an uppgifterna som kom. Många av dem kände även att de inte behövde någon ny introduktion till de övriga delnivåerna utan spelade på som vanligt och löste det bra.

Hur lång tid tar det för testpersonen att klara en nivå? Vi hade inför testdagen provspelat spelen för att få en uppfattning om hur lång tid det eventuellt skulle kunna ta för ett barn att spela spelet och klara av nivåerna. Vi dubblade vår ”experttid” och kom fram till att tider runt åtta minuter inte kändes helt orelevant, vilket ger ett resultat på ca två-tre minuter per bana och det lyckades barnen med.

Förstår testpersonen kopplingen mellan redskapen och interaktionsverktygen? Överlag var det inga större bekymmer för barnen att genom förklaring förstå att de skulle flytta upp aporna till gungbrädan eller att siffrorna representerade trappstegen. Det som svar svårt och onaturligt för dem var att förstå kopplingen utav tumme upp och tumme ned. Det var inget naturligt interaktionssätt barnen emellan vilket gjorde dem förvirrade. Vid förklaring av fallen att grön uppåtvänd tumme var positiv och röd nedåtvänd tumme var negativ eller genom att enbart referera till färgerna kunde barnen arbeta vidare och lärde sig snabbt bekräftelsen när det var agentens tur att spela.

Några av barnen hade svårt att se när det var fyra apor på gungbrädans vänstra sida då inte alla fick plats på en rad utan fick placeras lite längre ner och därför föll

Person	Kön	Ålder	Gungbräda				Rutschkana		
			Nivå1	Nivå2	Nivå3	Nivå4	Nivå1	Nivå2	Nivå3
1	T	4	J	J	J	J	J	J	J
2	K	4	N	J	-	-	J	J	-
3	T	5	J	J	-	-	J	J	-
4	K	5	J	J	-	-	J	J	J
5	T	4	J	J	-	-	J	J	J
6	K	4	-	-	-	-	J	JmH	J
7	T	4	J	J	J	-	J	J	J
8	T	4	J	J	J	-	J	J	J
9	T	4	J	N	J	-	J	J	-
10	K	4	J	J	J	-	J	J	J
11	K	4	J	J	N	J	J	J	J
12	K	4	J	-	-	-	J	J	J

Figur 7: Tabellen visar hur väl barnen lyckades lösa uppgiften

Person	Kön	Ålder	Gungbräda				Rutschkana		
			Nivå1	Nivå2	Nivå3	Nivå4	Nivå1	Nivå2	Nivå3
1	T	4	1min	1,5min	1min	-	1min	1min	1min
2	K	4	-	-	-	-	-	-	-
3	T	5	of	1min	-	-	1,5min	1min	-
4	K	5	2min	1min	-	-	1,5min	1min	1min
5	T	4	2,5min	1,5min	-	-	1min	1,5min	1min
6	K	4	-	-	-	-	1min	1,5min	1,5min
7	T	4	1min	1,5min	0,5min	-	1,5min	0,5min	-
8	T	4	2min	1,5min	-	-	1,5min	1min	1,5min
9	T	4	1min	2min	1min	-	1min	1min	-
10	K	4	1min	0,5min	0,5min	-	2min	0,5min	0,5min
11	K	4	1min	1min	-	-	1,5min	1min	1min
12	K	4	1min	-	-	-	1min	1min	1,5min

Figur 8: Tabellen visar hur lång tid det tog för barnen lyckades lösa uppgiften utan agent

i glömska. De såg då bara tre apor när man bad dem räkna högt vilket indikerade att den lägre placerade apan smälte in för väl i brädan.

Känns agenten naturlig i spelet? Ja, många utav barnen tyckte att det var roligt med agenten men de tyckte speciellt vid övergången från rutschkana till gungbräda att det var tråkigt att inte tigern fick vara med och leka och ville gärna placera honom på gungbrädan tillsammans med aporna. Många utav barnen ville även bekräfta agentens val genom att trycka på boxen eller siffran som angavs i tankebubblan vid korrekt svar från agenten. I gungbrädespelet ville de även hjälpa agenten genom att dra bort eller lägga till apor antingen i molnet eller på brädan för att hjälpa tigern så att den skulle få rätt. Detta varierade dock och inte alla barn ville detta. Eventuellt skulle man kunna ta tillvara på barnens vilja att vilja interagera med agenten genom att ge dem enn möjlighet att exempelvis dra agenten med pekfingret och flytta den mellan de olika platserna på skärmen. Detta påverkar ju inte själva spelet. Man skulle också kunna ha att tigern pratar när man klickar på den.

Får skadebeteendet hos avataren vid ”misslyckande” testpersonen att vilja fortsätta med detta? Nej, denna oro är något vi mer eller mindre kan utesluta. Många utav barnen reagerade och fnissade till men det var enbart något av barnen som under en kort tid fick agenten att flyga ovanligt mycket. Vad som uppfattades av testledarna var att barnen inte alla gånger kände att konsekvensfeedbacken gav dem tillräckligt med information om det bakomliggande felet då agenten flög upp i luften vid för högt svar och klättrade ner vid för lågt. Detta var något som testledaren fick förklara för barnen när felen uppstod. Dock är det osäkert om informationen hade upplevts som mer hjälpsam om ljuduppspelningen hade fungerat som tänkt. Mer om detta kan läsas under ”Svagheter i undersökningen”.

Känns det naturligt med drag-and-drop för testpersonen? Vid introduktionen och förklaringen att man skulle dra upp aporna så kände många att drag-and-drop var en okej rörelse. Några hade svårt till en början att veta när de skulle släppa och var på skärmen de skulle göra det. Några kom utanför det område som implementerats och därför upplevt en viss förvirring eller förvåning när apan inte ville sätta sig på brädan.

I gungbrädespelet, kommer spelaren att se symboliken hos grupperingen utav färgerna? Många av barnen hade till en början svårt att skilja dem åt. Antingen ville de dra upp rätt antal "apgrupper" för att motsvara antalet på andra sidan. T.ex. två apor på vänster sida, då ville de kanske dra upp grupp ett och två för att tillsammans skapa två. På denna nivå fanns där även ingen direkt gräns på hur många apor man kunde placera på en gång så några av barnen ville att alla apor skulle få gunga, en gräns som fanns i nivån tidigare. Bad man däremot de barnen som upplevde svårigheter att räkna antalet apor av en viss färg så förstod de efter ett tag kopplingen och kunde placera upp rätt grupp till brädan.

6.5 Slutsats

Överlag var både spelen uppskattade utav barnen, några tyckte dock att det var för lätt och kanske kan man även göra ett antagande att barn i åldern tre kan manövrera spelet. Dock får man nog i sådana fall försöka ge en kort introduktion av surfplattor då interaktionen annars kan uppfattas som lite svår. Något som uppmärksammades under spelomgångarna var dock att många utav barnen hade svårigheter att använda sig utav surfplattan. Några utav barnen försökte navigera med fingertopparna vilket gjorde att nageln kom emot och hindrade navigering på spelet.

6.5.1 Svagheter i undersökningen

Testledaren hade i förväg informerat om att internettillgång via wifi krävdes för att genomföra spelet men väl på plats så var det en del problem med nätverket. Då testgrupperna tilldelades gästkonton till nätverket så skedde det en timeout efter en stunds spelande vilket skapade problem i interaktionen. När en timeout skedde så slutade spelet att fungera och knappar och interaktions objekt gick inte att navigera. Detta skapade en del frustration hos barnen men med glatt humör hos testledarna och en ny sökning utav nätet så fick man börja om. Hälften av barnen drabbades hårdare av detta vilket syns i fig 7 där många utav barnen inte hann genomföra lika många nivåer under det uttänkta tidsintervallet.

Något som upptäcktes inför testningen var att ljudfilerna som utvecklats till spelet inte gick att spela upp på de två ipad 4 som vi hade fått tillgång till. De hade till de nya surfplattorna lagt in spärrar på autouppspelade ljud i HTML5, något som inte upptäckts tidigare då vi spelat upp dem på en ipad av tidigare modell eller via android telefoner. Det ledde till att testledarna fick agera berättare och förklara hur spelet fungerade och vad som skulle göras i en viss situation. Det gör det svårt för att veta om agentens egen berättelse kunde hjälpt till att en del missförstånd kunde minskats eller om spelupplevelsen hade blivit annorlunda.

7 Diskussion

7.1 Problemområden

Under utvecklingen har vi stött på vissa saker som kan betraktas som svårigheter och som fortfarande är ämne för diskussion. Dessa är bland annat:

Utveckling av agenten Hur ska det visas att agenten utvecklas/blir bättre? Ska det ske med en animation? Vad ska hända när agenten "backar" i sin utveckling, dvs om barnet svarar fel på för många frågor? Kanske är det för negativt om man går bakåt i utvecklingen och en utvecklad agent kanske ska förbli utvecklad.

Belöning av agenten Hur ska agenten belönas när den svarar rätt? Finns det en narrativ det ska passa in i? Vi har diskuterat detta och har gått från konceptet med pengar till mat. När tigern äter maten kommer den att växa och bli kraftfullare. Men detta kommer inte att ske varje gång agenten svarar rätt eftersom det bara finns ett begränsat antal utvecklingssteg. Hur kan man uppmuntra agenten att svara rätt och ge omedelbar positiv feedback?

Trappstegen Hur ska trappan till rutschkanan visas? Det har visat sig vara svårt att hitta ett perspektiv så att både trappan och själva rutschkanan syns bra. Diskussioner har också uppkommit i arbetet med HiFi-prototypen om det översta trappsteget ska räknas som ett steg eller en ansats till rutschkanan. Vi har tagit upp denna fråga både på möten med handledare och på gruppredovisningarna och alla har olika åsikter om det. Vårt möte med grafikern gav oss en rutschkana som fungerade bättre då den visade rutschkanan från ett annat perspektiv. Under testet med förskolebarnen visade det sig att några ville räkna den översta trappansatsen som ett steg, medan de flesta tryckte på det antal gruppen hade tänkt var rätt. I regel behövdes det bara ett eller två försök för att barnet skulle förstå hur det var upplagt.

Variation Spelen tränar ju upp olika saker, och det bästa vore om man kan få barnet att träna på alla delspel. Vad ska hända om ett barn bara vill träna på rutschkanespelet och inte alls är intresserad av gungbrädan? Ska agenten fortfarande utvecklas? Ska agentens smarthet i ett spel vara beroende av hur den har presterat i det andra spelet?

Felval Kan ett spel som bygger på att agenten är aktiv deltagare leda till att användaren medvetet vill göra fel? Med en tiger som klättrar upp i luften och sen flyger ner till marken lockar det till enbart skratt eller får det användaren att vilja fortsätta med detta? Vi känner personligen att det är ett roligt sätt att låta användaren förstå att siffran den valde var för stor och där med fortsatt klättring ut i luften. Användaren kanske lockas att välja en avvikande siffra för att se vad som händer, men är det ett sätt att interagera på så lär ju användaren sig till slut vilken siffra de måste undvika. På så sätt kommer användaren att lära sig number sense, dock ej på det önskade viset. För att kunna medvetet trycka fel måste ju barnet veta vilken siffra som är rätt. Agenten kommer inte att utvecklas och spelet kommer kanske efter ett tag att bli tråkigt. Vår förhoppning i slutet av delkurs 1 var att att användaren själv väljer att ta ett aktivt val och fortsätta utvecklingen framåt. Denna förhoppning bekräftades i testet då det visade sig att denna risk kan uteslutas. Barnen var mycket målmedvetna och benägna att göra rätt, även om det är en rolig animation då man gör fel.

7.2 Egna roller

Vi har upplevt det väldigt positivt att inte ha några låsta roller under början av projektet, det har ökat den kreativa sidan under den första halvan av kursserien och det har gjort att alla medlemmar har fått sin röst hörd och känt att de har varit delaktiga i prototypen. Med den öppna rollsättning i början och utvecklingen i LoFi-prototypen kändes det dock naturligt att antingen ta en mer teoribaserad roll och fördjupa sig i de texter som gruppen hade blivit tilldelad och börja skissa på rapporten eller ta en teknisk roll och fortsätta utvecklingen vidare mot HiFi-prototyp och läsa på HTML och börja kodandet. Med veckomöten har vi alltid hållit koll vad de andra personerna i gruppen har gjort och det har blivit ett öppet tillfälle att komma med feedback till utvecklarna och ge varandra en klapp på axeln.

7.3 Reflektioner kring testresultaten

Det har varit intressant att se att de flesta barnen var väldigt vana vid att använda ipad. Dialogrutor som dök upp på grund av nätverksfel trycktes enkelt bort av barnen, utan någon som helst tecken på förvirring eller att de skulle behöva hjälp. Ett av barnen gick till och med in på ipad-menyn och började öppna andra spel när det tyckte vårt var för tråkigt. Försökspersonen gjorde det ett flertal gånger så det är inte någon knapp den kom åt av misstag.

Nätverksproblemen var någonting som svårade till testningen och även att ljudet inte fungerade.

Drag-and-drop var något som var problematiskt men vi är nästan säkra på att detta till viss del berodde på att nätverksanslutningen inte fungerade. När programmet inte kommer åt nätverket syns allting som vanligt, men interaktionen fungerar inte. Detsamma gällde knappar. Inget av barnen verkade ha några problem med att förstå själva konceptet med drag-and-drop utan det var mer att det i vissa fall inte gick att utföra.

Det faktum att ljudet inte fungerade under testningen var också något som var negativt. De flesta av instruktionerna var inspelade som ljud och det hade varit bra att testa om detta fungerar. Det som framför allt påverkades av detta var att vi inte kunde testa om de instruktioner som var inspelade var tillräckliga för att ge barnet den information det behöver, samt att det var svårt att testa om barnet förstod konceptet med TA. I spelet är den kvinnliga rösten inspelad som berättarröst, medan en manlig röst representerar TA:n. När vi gjorde testet var vi istället tvungna att låta testledaren stå för både instruktioner och TA. Det gjorde att vi inte riktigt kunde testa om de förstod att det var en annan självgående agent som spelade och att de lärde upp den. De hade i vilket fall som helst inga problem att spela spelet när TA:n spelade.

7.4 Ej implementerade idéer

I arbetet med prototypen och den färdiga produkten är det några saker som vi ursprungligen hade tänkt oss, som inte kom med i den slutgiltiga versionen. Vi skiljer på features som har valts bort på grund av att vi tyckte att de var dåliga, och sådana som inte har kommit med för att vi inte hann:

7.4.1 Nedlagda idéer

Pengar Vi hade en idé i ett tidigt stadium att vi skulle ha pengar som användaren/TA:n kan tjäna ihop. Dessa skulle sedan användas för att låsa upp olika banor. Som beskrivits ovan lades denna idé ner redan på LoFi-stadiet eftersom vi trodde att det skulle bli för krångligt för barnet att man både tjänar ihop pengar och att TA:n utvecklas. Vi var också osäkra på om barn i 4-årsåldern förstod konceptet med pengar.

7.4.2 Ej implementerat

Mat Vi hade tänkt oss att agenten vid rätt svar skulle få en bit mat, som så småningom gör att han växer och blir större. Denna idé implementerades aldrig, på grund av tidsbrist. För att göra detta skulle vi behöva lägga ner en stor del av tiden på animationer. Det ansågs viktigare att utnyttja grafikerns tid på att få animationen med när agenten åker ner för gungbrädan och få nickande/skakande huvud.

Utveckling av agenten Agenten svarar alltid rätt med en sannolikhet av 1/3. Det kan därför anses att den inte "utvecklas". Argumentet som försvarar detta är att den blir bättre och bättre genom att den kan svara på svårare och svårare frågor.

Kulkastning Gruppen hade från början en idé om att vi skulle ha ett delspel med kulkastning. Även om denna idé aldrig lades ner officiellt var det redan på ett tidigt stadie av LoFi-prototypen tydligt att vi skulle fokusera på gungbrädan och rutschkanan och ha kulkastningsbanan som en extra sak att implementera om vi får tid.

Hopkoppling av spelen Delspelen hade kunnat kopplats ihop mer. TA:ns nivå följer inte med från ett delspel till ett annat. Till viss del har spelen ändå gjorts till en helhet eftersom grafiken har gjorts mer konsekvent mellan spelen och siffrornas placering är ungefär samma i de båda delspelen, något som de inte var i slutet av delkurs 1.

Integrering med mWorld Grundtanken var att vi skulle utveckla en spelmodul till mWorld. I slutändan blev det ändå att vi mer utvecklade ett eget spel, dock med inspiration från mWorld.

Referenser

[1] *A Teachable-Agent Arithmetic Game's effects on Mathematics Understanding, Attitude and Self-Efficacy*, Lena Pareto, Tobias Arvemo, Ylva Dahl, Magnus Haake, Agneta Gulz. http://www.lucs.lu.se/wp-content/uploads/2011/10/Paper158_ TeachableAgentGameAchievementEffect.pdf [2012-10-26 19:00]

[2] *Number Sense Series: A Sense of 'ten' and Place Value*, Jenni Way. http://nrich.maths.org/2479 [2012-11-27 19:00]

[3] *Understanding a Child's Development of Number Sense*, National Council of Teachers of Mathematics. http://illuminations.nctm.org/Reflections_preK-2.html [2012-11-27 19:00]

[4] *MathGirls: Toward Developing Girls' Positive Attitude and Self-Efficacy through Pedagogical Agents*, Yanghee Kim, Quan WEI, Beiji XU, Youngah KO, Vessela ILIEVA. http://create.usu.edu/publications/ MathGirls_Kim_etall.pdf [2012-10-26 20:46]

[5] *Number Sense Series: Developing Early Number Sense*, Jenni Way. http://nrich.maths.org/2477/index [2012-10-26 19:30]

[6] *The Number Sense*, Bruce White. http://www.math.twsu.edu/ history/ topics/ num-sys.html#sense [2012-10-26 19:56]

[7] *Föreläsning, Neuromodellering - kognitiv robotik och agenter MAMN10*, Agneta Gulz. http://www.eat.lth.se/fileadmin/eat/TA-system.pdf [2012-10-26 21:09]

[8] *The Design of Everyday Things*, Norman Donald A., Basic Books, 1988.

[9] *Föreläsning, Människa- datorinteraktion MAM061*, Mattias Wallergård. http://www.eat.lth.se/ fileadmin/eat/MAM061/Interaktion_1.pdf

[10] *Pedagogical Agents for Learning by Teaching: Teachable Agents*, Kirsten Blair, Daniel L. Schwartz, Gautam Biswas, Krittaya Leelawong. http://aaalab.stanford.edu/papers/ Teachable_Agent_Lite.pdf [2012-10-27 12:47]

[11] *Acapella box*, ett system utvecklat av Acapella Group. http://www.acapela-box.com/ [2012-12-03 15:00]

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Worlds' first "Teachable Agent"– game outside the mathematics and science domain

Axel Forsberg, Morten Hougen, Linus Håkansson and Johanna Truong

This paper focuses on the development of the worlds' first Teachable agent game outside the mathematics and science domain, namely the development of the first teachable agent game within history.
A teachable agent, TA, is a virtual pedagogical agent that can be taught to perform certain tasks or solve certain problems in a specific environment. By letting a student teach an agent, the students benefits from the effects of learning by teaching. The student teaches the agent and then evaluates the agent's knowledge by asking questions or by letting the agent solve problems. In our game we let the student evaluate the agent's intelligence by supervising him through a problem-solving.

Being the first TA-game in history ever, we were free to create our own framework for such a game. The downside of being first also meant that we had to do a lot of groundbreaking work. To our knowledge this is also the first time someone is building a learning game on the game-platform Unity. With no other games for references all solutions to upcoming problems had to be solved by us. The only mandatory guidelines we received were that the game should include three given historical epochs and be a teachable-agent based learning game. Early in the process, we decided that we wanted to create an interesting and captivating history game. We simply wanted to create a history-game that was interesting and captivating.

As an agent in a secret agency the students' mission is to save a colleague, the TA, who has been trapped in time. To save this unfortunate colleague, the students must travel through time using the company's database, and gather historical information. This gathering is done through visiting various places and interviewing historical persons, and examining these persons' environments. When the student has gathered information he/she returns to the classroom where the teaching of the TA takes place. After the TA has learned all he needs to learn, he will start a test activity. If he has been taught correctly and sufficiently he will be able to pass the test and travel on in time. This will make a new era available for him and the student-teacher.

Introduction

Our group consists of three technology students and one cognitive science student. What brought us together were the interest of a good history game and the idea of learning by teaching.
Studies have shown that the method learning by teaching is a very powerful and motivating educational tool. The task we were given was to create and develop a history game for 10-11 year old students, where we focus on discoveries and inventions during three eras.

In this history game there are three phases. The first phase is called the knowledge acquisition, where the student collects historical information. Next phase is the learning activity, in this phase the student will work with the gathered information. An example of a learning activity is to let the student match a picture of a historical person with facts. The testing activity, which is the last phase, is the activity where the TA performs novel but similar activity as the activity in the second phase.

Roles

During our first group meeting we decided on the various project roles, which we kept throughout the whole interaction course. We appointed Linus to be our project leader and his main tasks were to write the project logbook for each meeting and the weekly agenda. These documents were then sent by him to our supervisor Agneta Gulz every week before our meeting. Axel and Morten were responsible for the conceptual design and Johanna was responsible for testing and evaluation. Because Morten is a cognitive student it was a natural choice to let him share the responsibility of the conceptual design with Axel, who has a lot of knowledge in this topic. The responsibility of testing and evaluation was given to Johanna since she has done this many times before.

Group processing

All groups must go through different levels and stages of developments, each containing specific challenges and dynamics. These processes can be dealt with reactively or proactively. Our group chose the proactive approach. By introducing some small lectures and debates we tried to enhance the group processes and avoid some of the best known pitfalls sometimes occurring in projects like this. The theoretical approach towards teambuilding was based on the framework of Fundamental Interpersonal Relations Orientation, FIRO (Schutz, 1958). The FIRO model claims there are 3 kinds of relations on the level of behavior: inclusion, control, openness. These elements of relations correspond to significance, competence and likability on the level of feelings. Different group members will have different preferences towards these elements, ranging from high to low. A group member with high preference for control will have different needs and expectations towards leadership than a member with a low preference for control.

Ambition

The goal with this project was to develop a history game for students in grade 4-6, i.e. students in ages 10 to 12. Our ambition for the first half of the project was to finish the whole first world, which included four worlds with intractable historical persons, their inventions, the learning and testing activity.

A vital part of the game is the dialogues between the student and the historical persons, and therefore we spent a lot of effort and time in choosing the most important facts. Furthermore we tried to select discoveries and inventions that played a crucial and important role in history as well as being able to attract the attention and interest of the students.

The decision to program the game in a 3D program opened up exciting possibilities graphically as well as in terms of cognitive learning theory. On the other hand, there are many games with stunning graphics on the market which leads to high expectations. This meant that we spent a lot of time working with the design and graphics.

Educational Games

Playing is an important aspect of training for the adulthood, not only for humans but mammals in general. Play is a term employed in psychology and ethology to describe a range of voluntary, intrinsically motivated activities normally associated with recreational pleasure and enjoyment. Intrinsic motivation refers to motivation that is driven by an interest or enjoyment in the task itself, and exists within the individual rather than relying on any external pressure. Intrinsic motivation is based

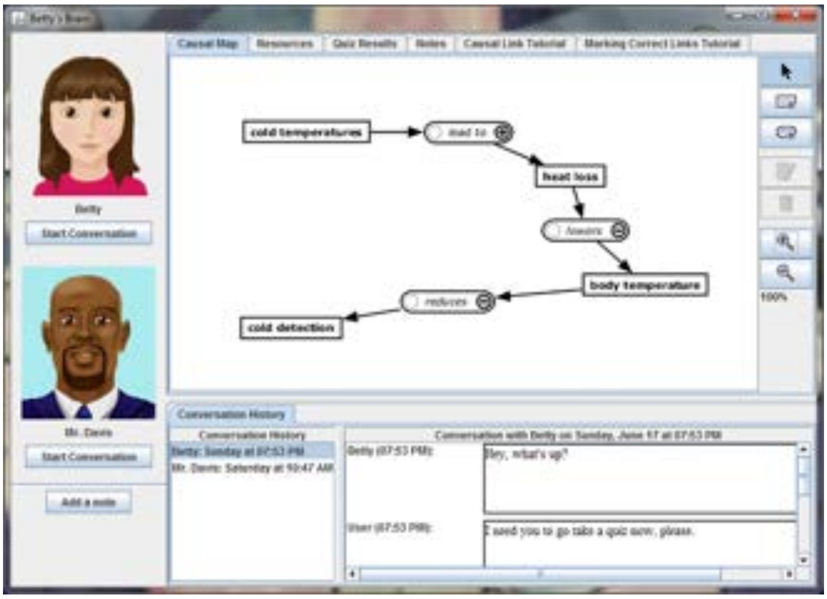


Figure 1 Betty's Brain

on taking pleasure in an activity rather than working towards an external reward. This intrinsic motivation is also found when people are playing games, and are an important aspect of serious gaming. Serious games are the definition of computer-programs and video games intended to not only entertain users, but have additional purposes such as education and training. One subcategory of serious games is educational games produced for and used with an educational purpose.

There are many educational games on the market, and several designed around the theory learning by teaching, but these games are all about mathematics and science. One example of a science game is Betty's Brain, *see figure 1*, in this game the students' task is to teach Betty science through models of a system or process (TA-group Vanderbuilt Univeristy). Because of the lack of educational games in history we were asked to develop one.

Methodology

School curriculum

The first thing we did before we started drafting the game was to look at the school curriculums history section for students in grades 5 to 6. This was because we wanted to get a better understanding of what is expected from the students after these years. The Swedish Riksdag and the Government sets out the goals and guidelines for the preschool and school through e.g. the Education Act and the Curricula. The Curricula for the sixth grade in History states among others that: "through the teaching of History, students shall be given the opportunity to develop their ability to:

- Use a historical reference frame that includes different interpretations of the periods, events, characters, cultural encounters and trends
- Critical review, interpret and evaluate the sources that provide grounds for creating historical knowledge
- Reflect on their own and others' use of the history in different contexts and from different perspectives."

Theory

It is important to recognize that this project bases on a cognitive learning perspective, which differs from the more traditional behaviouristic view on learning. More than just presenting different facts about persons, inventions and historical events, we try to facilitate situations where the student themselves can experience history and draw their own conclusions about the importance of different historical events. From a cognitive science perspective, one interpretation of understanding is to see patterns (Gärdenfors, 2010). Through the game the students will meet and interview different historical persons and witness historical events which will expose them to historical facts. This will present the History in an animated and interactive way, where students can engage in dialog with historical persons based on their own needs and motivation.

It would be to demand too much of the students expecting them to find the historical links and make all the right conclusions themselves. Therefore it is important that the game steers them through the story and provides them with opportunities to extract relevant information when needed. Through the game, history can be presented in a great variety of formats, like motion pictures, texts, dialogs sounds, tests.

The narrative and layout of the game will create a path through knowledge which we think will optimize the learning. It is also possible to steer the students' attention within each picture frame. Ongoing research shows that children between 10-12 years old have underdeveloped oculometer gaze control, which is the ability to direct or inhibit reflexive eye movement (Holmberg, 2012). This poor ability to inhibit reflexive eye movement implies that their eyes unintentionally steer towards elements with high degrees of salience or movements. With this in mind, important features or events can be highlighted and the students' attention can be steered towards these elements. Since there is a strong connection between fixation and attention, we can create situations where the students unconsciously are being led towards important pieces of knowledge.

Research shows that when information is conveyed through several types of media, more senses are activated which in turn improves the chance of learning. "In many cases it is a matter of processing information, not only at different levels but in different ways. Memories that are distinct and well worked through, it will be remembered well" (Helstrup og Kaufmann, 2002. p.131). We believe that mediating historical information through a platform which enables the use of a great variety of different instruments will enhance the students' inner motivation. Furthermore it gives the students a tool for understanding the larger historical context, rather than just enabling them to reproduce a memorized schema of historical events. We hope to provide the students with the ability to not only know what, where and when, but more important also the why and how I know this. If we manage this, which in no sense is an easy task, we will facilitate situations where students learn how to learn. An invaluable skill which they will be able to transfer to other knowledge domains (Gärdenfors, 2010, p. 168). A game as a platform of learning provides excellent opportunities for creating powerful effects like animations, interactions and dialogs.

As a tool for visualizing our learning goals for our game we will use Bloom's Taxonomy (Bloom, 1956). Bloom's Taxonomy is a classification of learning objectives. It refers to a classification of the different objectives that educators set for students (learning objectives). Bloom's Taxonomy divides educational objectives into three "domains": Cognitive, Affective, and Psychomotor. Within the domains, learning at the higher levels is dependent on having attained prerequisite knowledge and skills at lower levels. Applied on the cognitive domain and incorporating the School curriculum, the desired staircase of knowledge for our project, *see figure 2*.

Teachable Agents

Central in this assignment is the use of a Teachable Agent. An agent in this context is defined as: "a software application with an ability to act on its own and to reproduce human behavioural and cognitive capacities by means of artificial intelligence" (Haake, 2009, p. 10). And as the name infer, teaching is a prominent feature in this concept of learning. Teaching is an important part of man's culture, and there are hardly any human that does not teach. Therefore teaching creates a strong motivational drive in all of us. To teach, one has to understand and understanding is to see patterns. If a student does not fully understand a topic, teaching will often be an impressive tool to gain such understanding. To teach, one has to understand what the pupil does not understand, which highlights difficult aspects of the topic. The student could teach each other, but that would demand twice as much time as when they teach an agent. Teaching an agent also removes the risk of error learning between students. The teachable Agent is a virtually created character specifically inserted in this game as an object for student teaching. He never get tired, he is always as motivated as we programmed him to be. He is also programmed to disclose how he think, spurring the students meta-cognition about their own learning. "The TA uses artificial intelligence to learn and reason about what is has been taught. TA is a hybrid; they reflect their owners' knowledge, yet they have a mind of their own." (Chase, in press, p. 4).

Through working with the TA the student will gain knowledge, not only about the subject they are teaching, but they can also gain meta-knowledge about how to acquire knowledge. A TA will also shift the focus from the students own performance to the performance of the TA, or to the protégé. "…the Role of the TA's in creating a distinctly social set of motivations to learn, which are supported by an ego-protective buffer, an increment list approach to learning, and a sense of responsibility" (Ibid). A protégé offers an ego-protective buffer, EPB. The EPB shields the students from forming negative beliefs about themselves, because the blame for failure can reside elsewhere. For instance, when a TA is failing, it can absorb part of the blame (Ibid). This will in particularly benefit students with lower self esteem. TA's also offers a motivational effect, "...the protégé effect: students are willing to work harder to learn for their TA's than for themselves, and this is especially true for low-achieving students."(Ibid, p. 12). In other words,

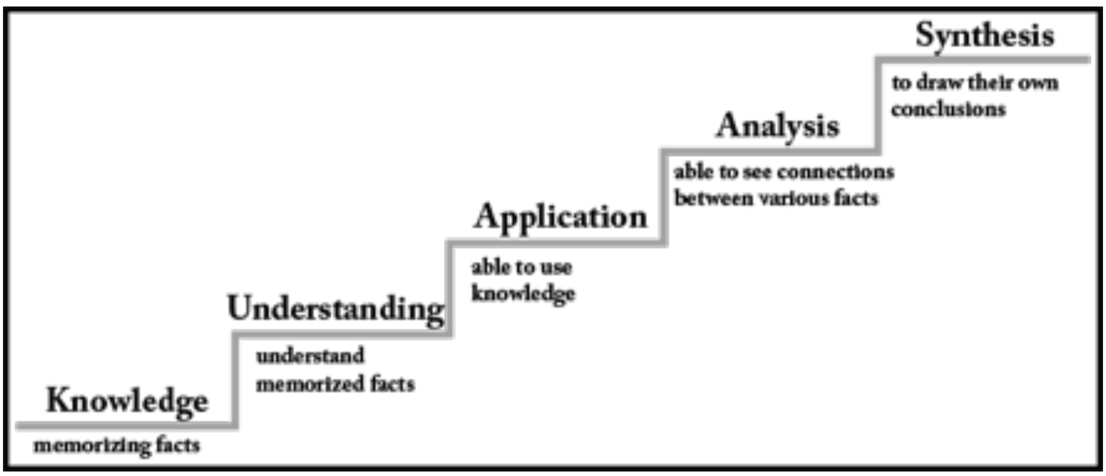


Figure 2.

the student becomes ambitious on the behalf of their TA's.

The game HistorieAgenterna

After many hours of group meetings and brainstorming we finally agreed on a gaming idea. The main concept is to let the student take part as an agent in a secret agency which mission is to ensure that the space time continuity isn't disturbed. Due to a breakdown of the time machine, one of the student's colleagues is trapped in a different epoch than ours. The time machines programming is structured around historical events and fact, and due to different breakdowns, our students must sort out different historical hick-ups in the machine.

We were given three different epochs: the Middle Ages, age of Liberty and the Industrial Revolution. These epochs are divided into three worlds and in each world there will be five historical figures. Hence there will be only one learning activity and one testing activity in each world.

In this game the colleague is the TA. To be able to save the colleague, or TA, the student needs to travel back in time and interview different historical figures. By interviewing these famous historical figures the student learns history. In the game the students are given a pair of virtual reality glasses which enables them to travel within the agencies database, where all historical information is stored. To travel in time the student chooses which city to visit. These cities are pinned on a world map, *see figure* 3, and are based on where the historical person is from.

Layout of the game

The avatars and agents

In the first world we chose to have four historical characters, hence we have four worlds. The historical persons that we chose to introduce were: Johann Gutenberg, Leonardo da Vinci, Galileo Galilei and Robert Hooke.

The head-up-display

In the game there will be a head-up-display which will be visible through the whole game. It can be withdrawn but it will not disappear completely. In the head-up-display there are a few icons, all the icons are placed in a way that makes it easy to click on. For example the notepad, agency boss and the teachable agent icons are placed in corners, the reason for this is that it is easier to "throw" the mouse pointer towards the corners than it is to navigate to a specific area.

Figure 3 Cities pinned on the map

The notepad
For students who like to take notes, we placed a notepad in the top right corner. During the interviews with the historical persons the student can click on the notepad and take notes for later purposes, e.g. to read during the learning activity.

Agency boss
In the bottom left corner there is a picture of the agency boss. When she has been clicked on the environment changes to the tutorial view, here the student can talk to her again.

Colleague
To navigate to the learning activity the students clicks on the picture in the bottom right corner. The environment changes to the learning activity and the teachable agent will explain to the student how the learning activity works.

Globe
In the bottom centre there is a globe, clicking on this globe opens up the map. There are four cities pined on the map, to travel to one of the cities one just clicks on the pin. When the mouse pointer is laid over one city the name of the city will show.

The different worlds

As stated above we have four worlds, one for each historical person. We chose place the characters in their respective birthplace or in the city where they lived. In each world we present an invention and an interactable object is placed beside its inventor.

Johann Gutenberg
We placed Gutenberg in his birthplace, Mainz. The characteristics in this world are the houses and the room where Gutenberg is standing. The houses in this environment are from a project called, Malmö 1692. Although the houses are from Malmö during the 17th century, we thought that characteristics of the houses fitted well into the 15th century Mainz.

Gutenberg is most famous for his printing press and this is the invention that is presented in this world. By inventing movable types he could fasten the printing process and reuse the types. His invention changed the world, thus information and knowledge could spread faster. The printer accelerated the book printing process which resulted in a decreased book price. With the decreased book prices the people of the lower social classes could afford a book

Leonardo da Vinci
At first we decided to place da Vinci in Rome, but when we started to model the world we inserted an ocean and Rome is not near the ocean. Therefore we had to state in the dialogue that da Vinci was a couple of kilometers outside of Rome, by the Tyrrhenian Sea. When one first enters the world an animation of a flying object is started, this was one of da Vinci's inventions. Because da Vinci's invention crashes, he does not want to speak with anyone. Hence one does not speak with da Vinci himself; instead an apprentice presents da Vinci and his work.

When someone speaks the name Leonardo da Vinci one instantly thinks about Mona Lisa. But da Vinci was more than just a painter; he was also an inventor and an inspiration source for his successors. In this world we highlight his work as an inventor; although some of his inventions did not work they were inventions of the future. During his lifetime he made sketches of helicopters, parachutes and weapons. The inventions that are presented in this world are the flying object, i.e. the helicopter, and the parachute. The reason why we decided to present two objects is because we wanted to show how much further in time da Vinci was. If he had the right resources many of his sketches could be modeled.

As a painter da Vinci did not always use the traditional painting oils, instead he used newer painting methods. Because he used newer techniques paintings like The Last Supper already showed signs of cracks during da Vinci's lifetime.

In this world we also made an atelier where one can explore da Vinci's most famous paintings, such as Mona Lisa and The Last Supper.

Galileo Galilei
In this world Galileo is located in Pisa, where he was born and lived. In the background one can see the leaning tower of Pisa; this is to give an additional feedback that one is located in Pisa. We decided to place Galileo during his last years spent in Pisa, i.e. when he was in house arrest. Having him placed in house arrest we could introduce the power of the Catholic Church.

For this world we introduce the telescope that Galileo improved. The reason why we chose the famous telescope was because it is often said the Galileo invented it. But this is not the case, the telescope was invented by a Dutchman and Galileo improved it. With this improved telescope Galileo started to study celestial bodies. His observations assured him that Copernicus conception of the world was correct. In his book "Dialogue Concerning the Two Chief World Systems" he compares the two world pictures and therefore he started a war again the church. Because of his friendship to Pope Urban VIII he was not sentenced to death, he was instead to be in house arrest for the test of his life.

Robert Hooke

Hooke was born in Freshwater but we decided to place him in Oxford where he studied. This world was very simply made thus we did not have the time to hone it like da Vincis and Gutenbergs world. The environment in this world is also taken from the Malmö 1692 project.

Hooke is a person that is often neglected in the history books therefore we wanted to uplift his work and reintroduce him as an important person in history. Because Galileo lived in a different time than Hooke, he did not have the resources to further develop the microscope. Hooke successfully developed a working microscope and with this microscope he opened up a whole new world, the world of biology. He could very precisely draw the structure of a flea which is included in his book Micrographia. In this book Hooke also applied a new word, cell, to describe biological organisms. Furthermore Hooke was also a physicist; he came very near to deducing a law about gravity. His idea was subsequently developed by Newton and therefore we let Hooke be suspicious and think that the student is a spy sent by Newton.

Dialogues

Before we could start writing the actual dialogues between the student and the historical person, we had to look up some facts and sort out what we thought was interesting and important. When we gathered all the information we needed, we started to draw a dialogue tree, *see figure 4*. By drawing a dialogue tree we can easily guide the student through the information and another advantage of a dialogue tree was that it was easier for us, visually, to implement the dialogues.

Knowledge gathering

The concept behind the knowledge gathering is that the students can move around in the world and gather information either from the historical person itself or through other people. When the student approaches a figure, a window will pop up where the figure greets the student. Then by using multiple choice questions the student can gather facts about the historical person and the object invented. More than just presenting cold facts about persons, events and discoveries the game tries to convey information about the impact they had and the role they played in the years to come. Knowledge gathering shall bring the student up the first two steps on the staircase of knowledge, and provide the student with a factual platform on which to build a deeper understanding.

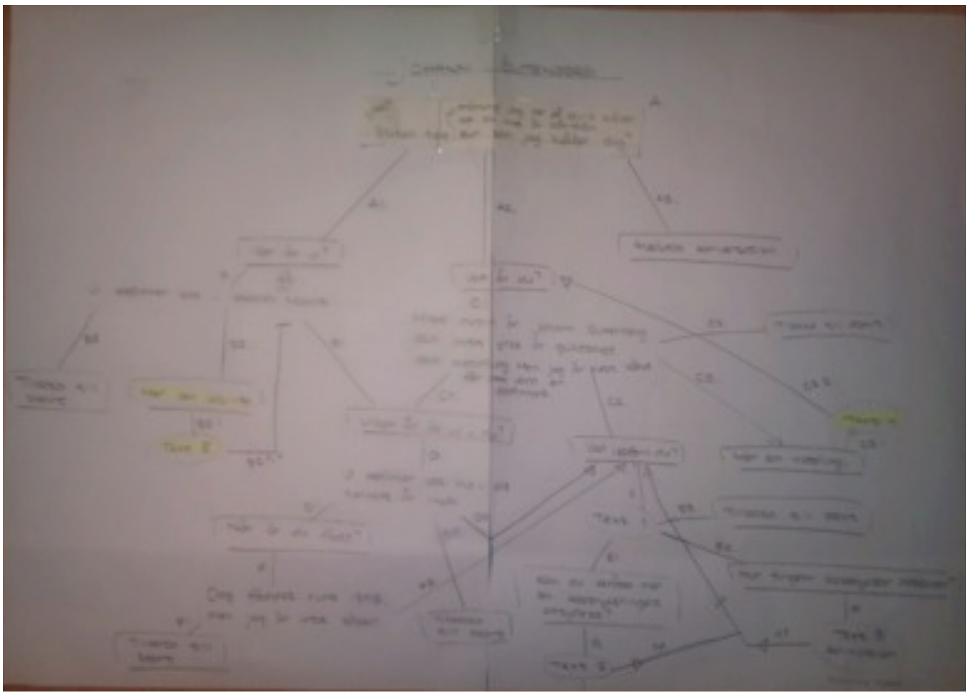


Figure 4 Dialogue tree

Learning activities

After interviewing some or all of the four historical figures the student will have to teach the TA. This is done by correctly placing different blocks of history which have fallen out of the time-machine. The TA will be present under this whole activity and will frequently ask questions or cheer on the student. Through this acquired knowledge the TA will rebuild the programming of the broken time-machine. In this partial course we will only focus on the first epoch, the Medieval, and therefore we only have one learning activity and one testing activity. Because of the time constraints we did not implement our second learning activity. This will be presented in a mid-fi version. Our intentions were to create three different learning activities, with varying levels of difficulty. These levels correspond to different stages on the knowledge staircase. Each new level will provide the students with favorable conditions for developing the ability to synthesis.

Learning activity 1

The student will teach the TA by matching different information to each of the historical persons, *see figure 5*. Each historical person will have a separate row in the matrix and it is up to the student to match the correct brick of information. Note that the historical person's name will be given in the first column, to give the student a bit of a hint. Having more bricks than boxes makes it harder for the student to guess. In this state we have a total of 24 boxes and 29 bricks in the pit. On this level, which is the easiest, one, advancing to the next step require that the student manage to place all boxes correct.

Testing activity

The next step is to let the TA work with the historical fact, the boxes, sorted by the student in the learning activity. This is named testing activity because the TA will have to pass this before he can travel on in time. Since the first level has to be correct before proceeding to the testing activity, some random errors will occur in the TA matrix. The purpose of this is to motivate the student to observe when the TA puts the boxes into the matrix. When the TA is done he will ask the student if this is correct and the student can choose to either answer yes or no. The student has to answer no and then change the order of the boxes. When all the boxes are in their right positions the student can confirm that all is correct and will watch the time machine leave this era. The surrounding in this activity will not be the same as in the teaching activity. The environment will be somewhat like the TA is programming his technical instruments. Therefore we changed the green board to something that looks like a circuit board, *see figure 6*.

Once the student clears this stage a new world will open up, which means that the colleague has successfully travelled in time into a different era, to an era closer to the present era. When this new world opens up the student will get familiar with five new historical per-

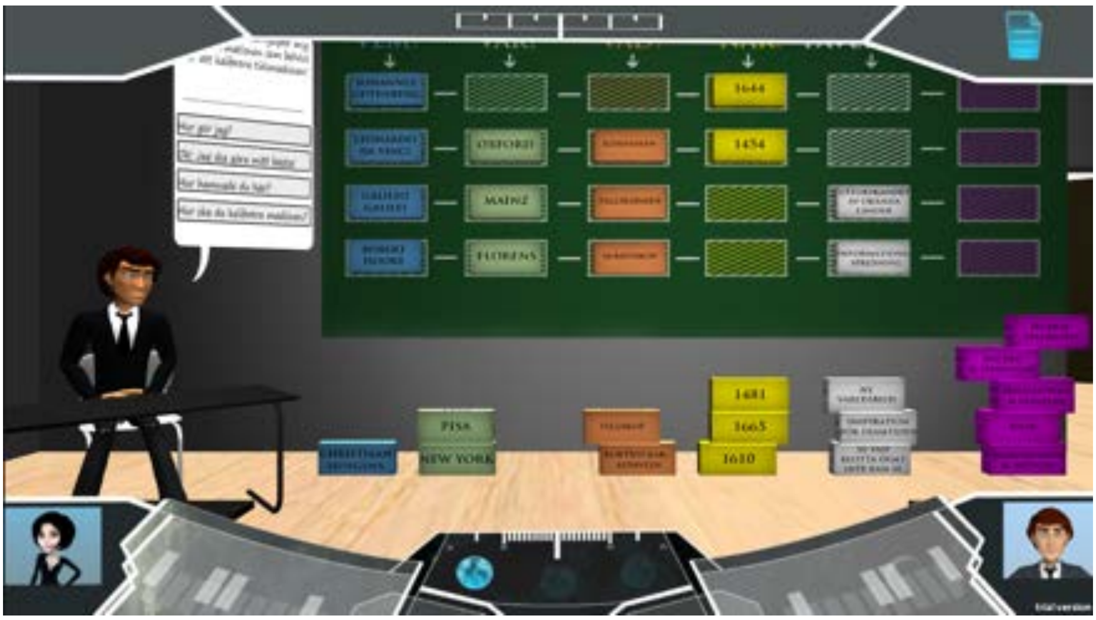


Figure 5 Learning activity

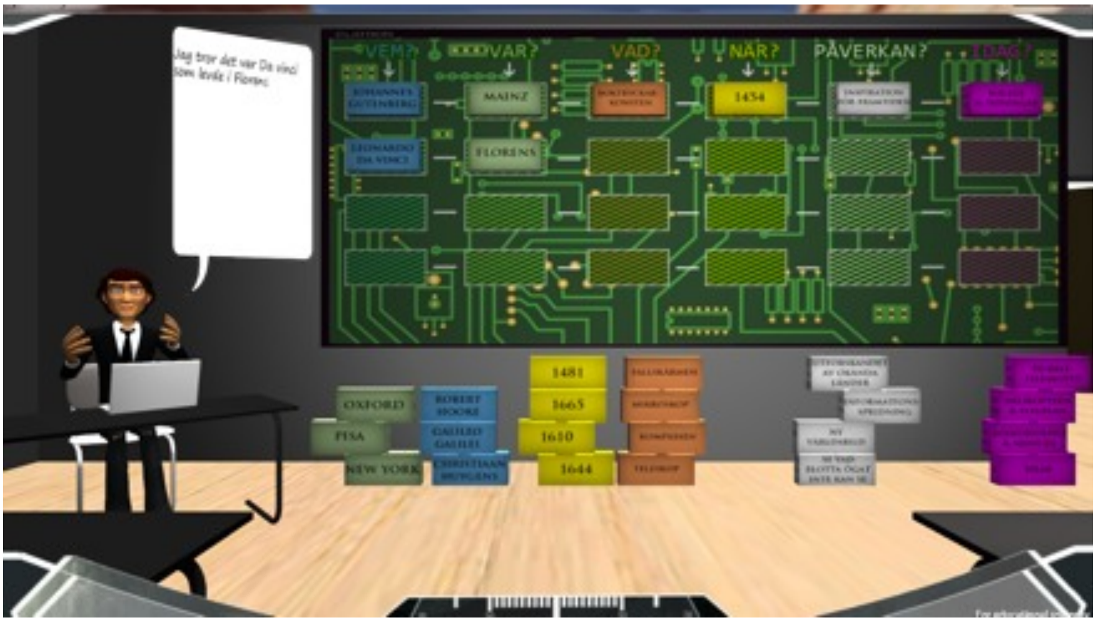


Figure 6 Testing activity

sons, a new learning activity and testing activity will also be introduced., *see figure 7*.

Learning activity 2

The overall objective of the second learning task is to help the student to climb further on the learning taxonomy staircase. We want to enhance the students' abilities to see specific historical events in a larger context. The importance of History lies not in the enumeration of historical events and dates, but in understanding the influence each event had upon the following history.

Through the dialog with the TA we want to spur the student to reflect over the importance of distribution of ideas and the impact this had on the development of our modern society. A main historical event in this process is the development of the Gutenberg press. Until Gutenberg invented the press and the moveable types, the main channel for information-transition was orally. The majority of the population was illiterate. The essence to be extracted here is the revolution in knowledge distribution provided by Gutenberg and his press. It will be unreasonable to expect that all student's possesses the ability to extract such knowledge by themselves. Therefore we want the TA to spur such thinking by asking open questions about connections, contexts and the historical development. One sound motivation for this can be that the time machine in this stage is dependent of the historical continuum. It is important that these questions do not place the TA in the position of being the teacher. He is still to be the students' student.

Learning Activity 2 is still at the draft stage, *see figure 8*. We work on the basis about building a cognitive map that resembles the map of a rail network or a metro station. The idea is that the map will grow as items are placed on the map. For example, when Gutenberg's press positioned as a "station" on the map, a new "publishing track" will grow out of it. On this track, which will cross other tracks, new stations will appear along the way. These new stations will represent other person's publications. Similarly, there can be "tracks" of art, inventions, years etc. The idea behind this is sided. On is to visually display the relationships and connections between different events. If a station or junction lacks the further network does not grow. The other is to created numerous occasions where the students interacts with persons, places, events and so forth, creating multiple opportunities for learning.

Figure 7

Prototyping

The prototyping process was divided into three phases: low fidelity, mid fidelity and high fidelity. The reason why we decided to breakdown the prototyping process into three minor tasks was because of simplicity. It was

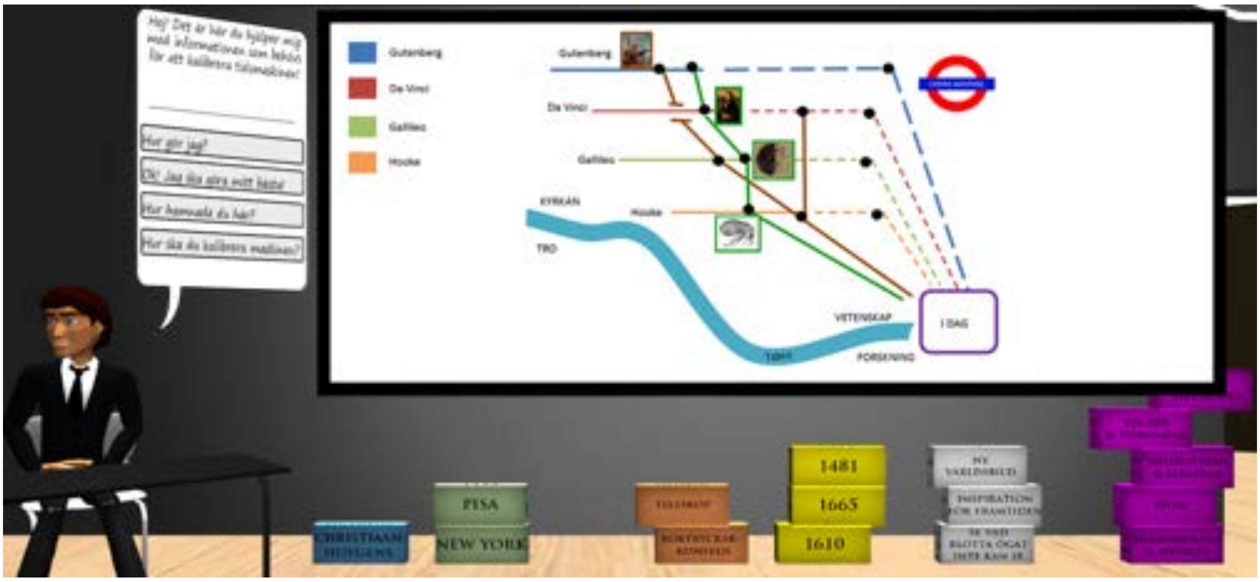


Figure 8 Draft of learning activity

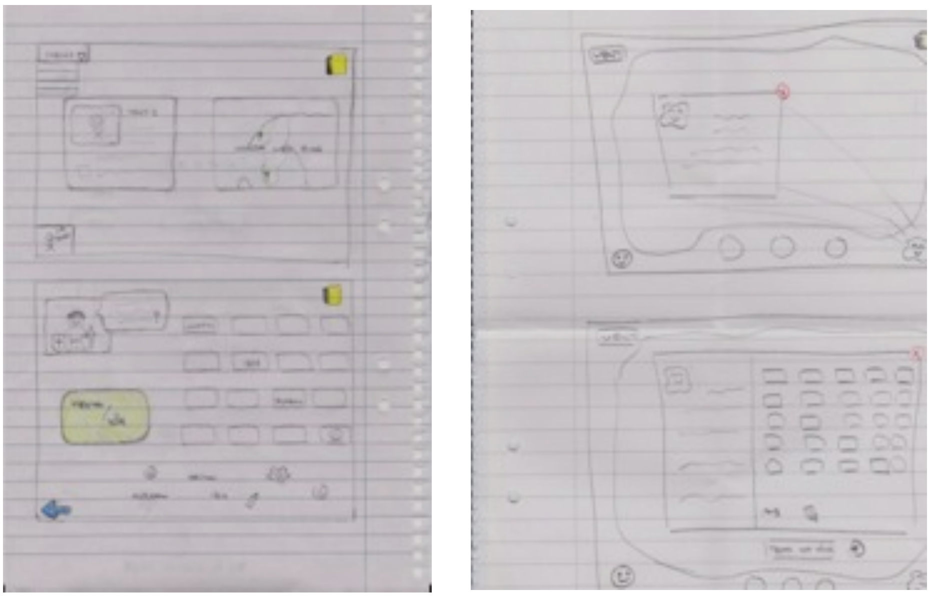


Figure 10 Lo-fi version 2

also easier to get an overview of the whole project and the different tasks that we had to do.

Low Fidelity

We choose to make the first prototypes with pen and paper. Thus this is the easiest way to modify the prototype and quickest way to sketch while brainstorming. In the first version of the prototype we decided to have a control panel, where the student could interact with both the TA and the world map, *see figure 9*. This control panel was later removed as a result of a design choice. The control panel was instead replaced by a head-up-display, *see figure 10*, making it possible for the student to interact with the TA, the agency boss, a menu bar and the three worlds at the same time. A notepad was introduced, making it available for the student to take notes while interviewing the historical person. We made this decision for the students, whom learn more by taking notes. This makes the student reflect over the collected information again.

Figure 11 Mid-fi version 1

Figure 12 Mid-fi version 2

Mid Fidelity

When we had an approximate layout of the user interface, we started to draft on the mid-fi prototype using Balsamiq. This program enabled us to get a better view of the project and more details about the layout could be done. The first version, *see figure* 11, of the mid-fi was based on the first version of the lo-fi. In the second version, *see figure 12*, we added a historical figure and added text boxes. Yet again, this was done to get a greater picture of how the game could look like.

High Fidelity

The implementation of the game was made in Unity and the historical person's faces and the intractable objects were made in 3ds Max. Implementing the game in Unity gave us more space to develop and design the game as we wanted. The choice of making it in 3d was simply because we had the knowledge and technique to do so. Another reason was that students now days play a lot of online games in 3d. During the implementation of the game we made it possible for the students to hide the

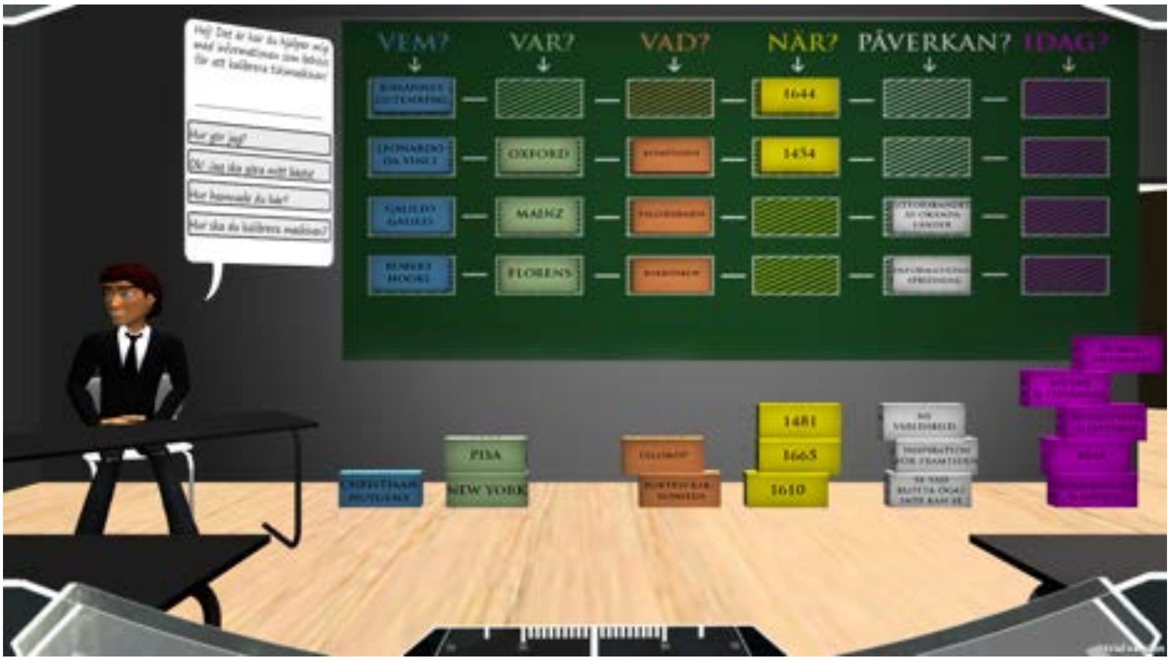


Figure 13 Learning activity with hidden head-up-display

Figure 14 Inside Gutenberg's house

head-up-display, *see figure 13*. If the student wishes to look around in the world he/she must press down the left button of the mouse and then move the mouse toward the wished direction. We made this modification because of the icons in head-up-display. If we did not do so the students would have a hard time pressing the icons thus the view would move as the student moves the mouse.

We also received a block of Malmo city during the 17th century, from a project called Malmö 1692. This block was inserted in Gutenberg's world and some modifications were made in 3DStudio Max and Unity.

Because the houses were hollow we had to add floors and decorate, *see figure 14.* Please read the *Appendix A* section for more information about the historical persons and controls.

Tutorial

The game starts with a tutorial explaining the idea behind the organization and the student's mission. The narrator of this tutorial is the agency boss thus she is the person who welcomes the student to the agency and assigns the mission to the student, *see figure 15*. After explaining the mission, the agency boss will go through the navigation and the different icons and their function. She will also urge the student to click on the icons and walk around to ensure that they have understood. At the end of the tutorial she will recommend the student to first visit the TA before heading out in the worlds. We want the student to interact with the TA first because he will explain a more thoroughly about the learning activity.

Cognitive design

Not only did we choose to take a cognitive stand in the theory of learning, Cognitive design was a very essential part of our development of this game. In this part of the report we highlight aspect as visibility, feedback, sound and text and mapping.

Figure 15 Tutorial with agency boss

Visibility

We chose to have a clean design and as few icons in the head-up-display as possible, the main reason for this were that we did not want the head-up-display to become messy and confusing. All the icons, on the head-up-display, are big and clear which makes it easy to press on them. The different icons have each a distinct symbol making it easy for the students to both recognize the symbols as well as tell the symbols apart. Furthermore the different symbols should provide the gamers with affordance, because affordance provides strong clues to the operating of things (Norman, 1999, p.9). The game should be as intuitive as possible, demanding little or

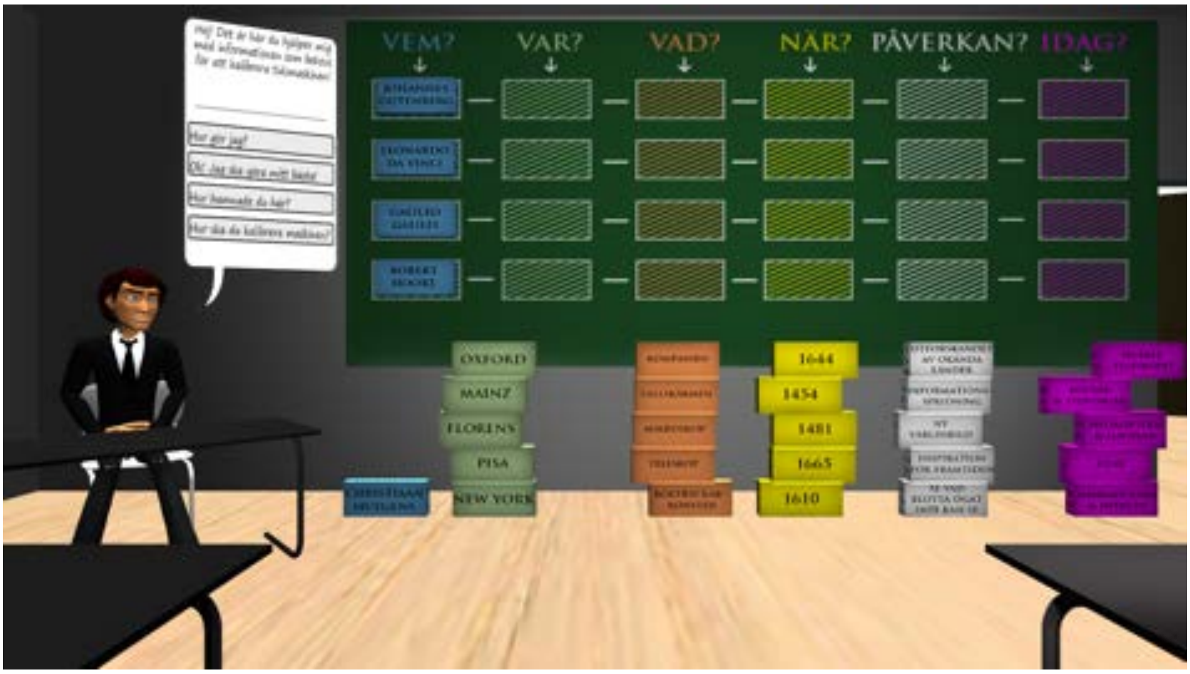


Figure 16 Classroom environment in learning activity 1

none instructions other than what is given through the game itself. As mentioned under the chapter Theory, children in our target group has an underdeveloped oculometer gaze control. Their poor ability to inhibit reflexive eye movements means that their attention is involuntarily drawn to objects of high salience or movements. We have taken advance this phenomena extensively throughout the game. For example by the light-setting and illumination of different objects, we can steer the attention of the student to important features in the game. Models which moves or have moveable parts will attract the attention of the players.

Feedback and Mapping

Feedback and mapping are important features in a game, were an important ambition is to make an easy understandable and intuitive game. Well planned and carefully created feedback and mapping ensures a user interface customized an audience with very different playing experiences. "Sending back to the user information about what action has actually been done, what result has been accomplished - is a well known concept in the science of control and information theory" (Ibid, p.27). The expression mapping is a technical term for the relationship between two things. Natural mapping is to take advantage of physical analogies and cultural standards which leads to immediate understanding (Ibid, p.23).

When the students enters the learning activity the environment will take form of a classroom with board, chairs and school-desks, *see figure 16*. The reason for this was that we wanted to give the student an additional feedback for securing an intuitive understanding that this is the place where you teach the TA. As seen in the figure, we added a green board to the classroom and again this was to strengthen the feedback about the learning activity.

When a student clicks on one of the multiple choice questions, the font color of that question will change from one color to another. This is equivalent to browsing the Internet, the font color of the link will change from blue to purple when the link has been clicked on. Using this well known feedback will ensure mapping and produce a seamless transition from previous experiences on the Internet to the game.

In the game there will be a head-up-display which will be visible through the whole game. It can be withdrawn but it will not disappear completely. Neither will the icons on the display change through different parts of the game. Changing the display or icons can easily arise confusion and we want to avoid that. Therefore we chose to let the icons and display have a fixed position.

We created a distinct environment for the teaching mode. By placing the TA and the student in a classroom and let the teaching take place on what looks like a green chalkboard we try to emphasize that this situation differ from the knowledge gathering. When in teaching mode the transfer of knowledge takes placed through the green board. The student has to combine the right persons with the right year and the right place and so forth. Different topics and boxes are colored in several colors, *see figure 16*. This is a way to map the boxes to their right topic, to give an additional feedback we choose to color the boxes in the matrix as well. It is important to remember that this learning activity only intends to take the student up the two first steps of the taxonomy. Later interactions shall provide the students with tools for a higher grade of learning.

Sound and text

When the student is interviewing the historical persons a text will appear. Because of the possibility that some students are struggling with reading problems we added sound effects, which reads out the appearing answer from the historical person. If some students experiences that the sound is annoying he/she can turn it off.

To separate the questions from the actual answer, we put the answer in the top of the dialogue box. This was because it is more natural to read from the top down. The questions are listed and has box around them, *see figure 17*. To give additional feedback we put icons besides some the questions, as seen in the figure. The calendar icon represents that the information given from the question is taken from present day.

Testing HistorieAgenterna

In order for us to gather some objective feedback on our game, we performed two testing sessions in a middle school by two fifth grade classes. Our main goal with the testing session was not only to receive objective feedback but also to obtain comments about the user interface and the quality of the game. These comments would help us improve the game and see things we missed during the development. Thus we can see the game through a student's eyes and make the necessary changes.

Figure 18 Text above invention

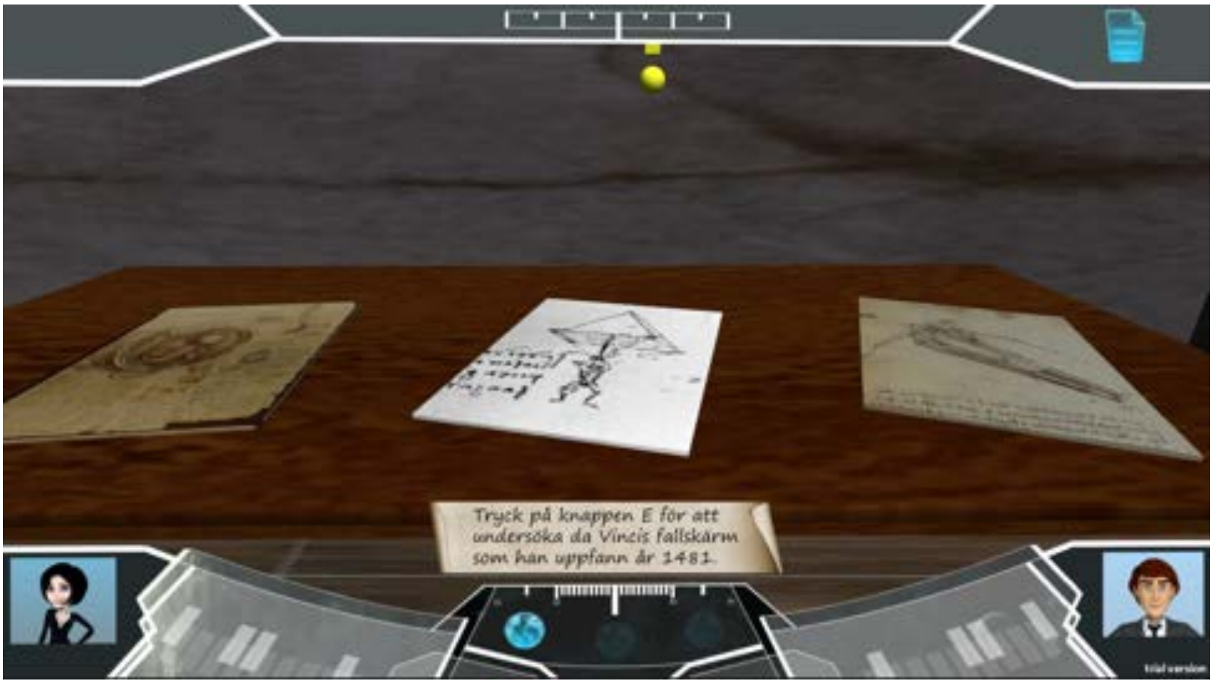


Figure 19 Text underneath invention

Figure 17 An example of a dialogue

Picture 1 Ongoing testing session

The testing was divided into two days, the first day we had 19 students participating and the second day we had 21 students. During both days the students were separated into two groups, with about 10 students per group per session, *see picture 1*. Each testing session lasted for about 40 minutes. Having smaller groups of students enhanced our connection with the students and it was easier to reach out and help them.

Before we let the students' login and start the game we had a brief introduction about the game. Each student got their own unique username; depending on the name the student either got a troublemaker TA or an ordinary TA. We then gave the students approximately 30 minutes to play. After the test we asked the students to fill out an inquiry, *see Appendix B*. We made an electronic inquiry so we could save all information directly to our shared folder in Google Drive. Doing so we could share the gathered information faster and everybody had access to it. The inquiry consisted of 7 questions and the two last questions were asked because we wanted to get additional feedback about how to further develop the game. Questions one to five was graded with the scores one to seven. By having an odd number of score points we forced the students to take a stand about the statements.

Evaluation

After compiling the results we could see that we got fairly good feedback on our game, *see Table 1*. Counting with those who answered with a five or higher, 72% of the students wanted to play our game again and 90% of the students thought that the game was a good way to learn history in. The main goal of our game was to enhance the interest in history and provide an alternate way to teach history. Receiving such high percentage in both of the questions was a huge success.

The main purpose with the testing sessions was to receive feedback about the user interface, the four worlds and the four historical persons. Again counting with those who answered with a score point of five or higher, 72% of the students thought it was fun to talk to the historical figures and 64% liked the different worlds. A reason why we did not receive a higher percentage, in the questions of the different worlds, may be because the

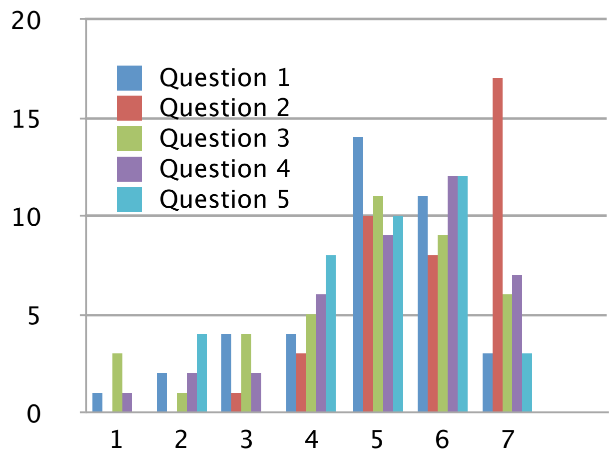


Table 1 Result of inquiry

shortage of time. We believe that if we had more time the worlds would be more detailed, in terms of more interactable objects and furniture. Obtaining such high percentage in the question about the historical persons was another victory. We had indeed put a lot of effort in modeling the historical persons and finding interesting facts about the inventions.

During the first day the students asked us and wrote in the evaluation why they could not write more in the notepad. Thanks to their comments we noticed that we had accidentally put a word limit to 200 words in the notepad, this limitation was removed till the next day.

One thing we noticed during the first session was that the students occasionally leaned over the computer. We realized quite fast that the students had trouble reading the texts in the purple boxes, in the teaching activity. Due to time constrains we did not fix this problem until the second day. Since w did not have this problem, we believe that the problem was due to the color resolution in the schools computers. To prevent this problem from being repeated we will change the background color of these boxes.

Another thing the students struggled with was during the tutorial where they had to mark the notepad and write something in it. The students clicked on the first row of the notepad, which is quite understandable because one usually starts writing from the first line. We had to explain that they had to press in the middle of the notepad to mark it. Because we had gotten used to clicking in the middle of the notepad we did not realize the problem until the students pointed it out.

When the students reached the teaching activity they had problem with the yellow boxes. The common comments were that we had put the wrong year in the box and that they could not find any information about that year. We had to explain that they had to explore the invention again or calculate to acquire the year. When the students explore an invention a short description about the invention will show up above the object, *see figure 18*. We noticed that the students often missed these texts and therefore we decided to change it, *see figure 19*.

What really touched us were the comments from question six and seven, *see Appendix C*. When developing the game we never thought of receiving so many positive comments, which made us think that all the effort we put in the game was worth it. In question seven we asked the students what we could change in the game. Some wrote that the sound could have been better, and we agree to that. We spent one whole day recording the dialogues using our own voices and one thing we struggled with during the recording was the different dialects. But the outcome was fairly acceptable. However the result would have been better if we used voice actors, thus they are trained to read dialogues with more empathy.

After the testing sessions the students asked us if they could play the game at home or in school and if we could make an app out of our game. Once again this implies that our game was a huge success and that the students enjoyed playing it.

Collaboration

During the first interaction course, we initiated cooperation with the History troublemaker team and we continued this collaboration throughout this course as well. In the first course their task was to write the dialogs between the TA and the student and they had the responsibility to develop the TAs and the boss' facial expression. In this second course they continued with this work and they also helped us with the actual implementation. Because we purchased four Unity licenses is was much easier to share the programming code between the groups.

To prepare for the testing sessions we sat down together and came up with the questions for the inquiry. Initially the inquiry was going to be on paper but the troublemaker group made an electronic version. As mentioned in the text above all data was stored into a document immediately after the student pressed on the button "submit".

Because the other group was interested in logging the students time in each world and their actions we made a log file. It was also interesting for us to see what the students were doing in the worlds and how many times they interacted with the objects.

Challenges

Every project will encounter many different obstacles that must be overcome, this project was not different. We had our ups and downs but in the end we always managed to get a mutual agreement.

One challenge we met where when we engaged into the collaborated with the "Troublemaker group". The reason for this was that the group suddenly consisted of eight persons, all with many different ideas. It was also time consuming to sell our idea to the other group. Another kind of problem we had to encounter and solve was the technical problems. Some of our initial ideas had to change because they were hard to implement in the game.

Yet another challenge was to get world one done before the end of this course, unfortunately this did not happen. World one was not finished due to time consuming problem solving. Many of our ideas have never been implemented in a game before and therefore we have to work out our own solutions. Modeling the objects was time consuming but quite straightforward, e.g. it took two hours to complete Gutenberg's pressing machine, *see figure 20*.

Discussion

At the beginning of the course we were recommended to meet once a week for group meetings. Our group met up more than once a week, this was because we had a lot to do and we wanted to be a bit head in time. Even though we did not set up an actual working schedule, we put up a few milestones during the project and we managed to reach almost all of them in time.

As mentioned we decided to have a proactive approach toward our own group process. An important part of our project was therefore to focus on group processing, which was a good way to get to know each other on a whole new level. One of our group members has extensively experience in facilitating processes enhancing group performances as well as team coaching. The group therefore decided that it wanted to keep a focus on group processes as well as on the invention of the game.

Figure 20 Gutenberg's pressing machine

Morten was in charge of lecturing us about group processing. The first time he spoke about this he gave us a task till the next group meeting. This task was to figure out where we believed our own preferences were towards the elements: inclusion, control, openness.

When we later discussed in plenum the task, we got to know a little bit about each other's preferences and ambition towards the group process. Furthermore we could reflect over the distribution of personal roles and responsibilities within the group. The second time Morten spoke about group processing, he read imaginary scenarios which created a common ground for reflection and discussions about our group process. Sharing our experiences from the imaginary scenarios enhanced both personal as well as group knowledge about preferences toward group-work. Through sharing experiences from different exercises and discussions, the group establishes a common ground of knowledge regarding individual preferences related to group working. Shared knowledge enhances the social glue in the group and minimizes the risk of misunderstandings and conflicts. Key factors are openness and self-disclosure.

To use the school curriculum as a starting point was a very good way for us thus we could use it as a guideline through the project. It gave us a hint of what is expected from the students and indirect that is expected from our game.

It took a few hours before we came up with the ultimate gaming idea. The student is portraying an agent whom works for a secret agency, HistorieAgenterna. His/hers first mission is to save unfortunate colleague who has been trapped in time. To save the colleague, the TA, the student need s to interview historical persons and solve a teaching activity test. When this is done the TA will take a similar test.

Dividing the prototyping phase into three made it easier for us to design the game and to get a better view of what was needed to be implemented. While brainstorming we used pen and paper to sketch our ideas, this is the easiest and quickest way to express oneself and to get all the ideas down immediately. Using Balsamiq was a way for us to actually see how our final product could possibly look like. This tool also made it easier for us to explain our thoughts to our supervisor because we could actually show them our idea step by step by the animations. Implementing the game in Unity really gave us a good view of our game. Using Unity gave us more freedom to implement our ideas and to create our ideal game.

Cognitive design is a vital part in designing an interface; therefore we put time and effort in designing the layout of our game. We want our game to be able to reach out the all students, with that in mind the color choices where important especially in the teaching activity. Green and red placed together was never an option and adding sound to the answering text provides the user with more channels for perceptions, which hopefully will motivate students who is struggling with reading. All the icons appear in the same position throughout the game, as we do not want the students to get confused.

Initiating collaboration with the Trouble making group lifted some of the workload from our shoulders but introduced us to new challenges concerning group process. Their task was to develop the TA and the agency boss. Hence we could focus on implementing our game. When we had some implementation questions we could turn to the trouble making group, this way we could quickly eliminate a problem and continue on with the implementation. What differed this course from the first interaction course was that we purchased Unity licenses, doing so we could easily commit new code and share it with the troublemaker group. Another advantage was that the implementation went a bit faster, because we could continue our work from home.

The whole project we had a race against time. We have made quite ambitious goals. There are a lot of things that needed to be implemented and the correct models needed to be found. A great deal of time was put into Photoshop, where we for example modified Gutenberg's face. Maybe this is too pretentious, but we believe it is important that the gaming experience is as authentic as possible. Many teenagers spend a lot of time on gaming and must therefore be considered to be a demanding audience. You could say that we were thorough, which we were, and want everything to be neat before we moved on to the next task. Being neat and thorough is what made the game as it is now. Finishing Gutenberg's world was a huge step since we could copy and reuse many of the elements and models it in the forthcoming worlds.

One of our goals in this course was to implement the testing activity, which we manage to do. Because we wanted the game to have different levels we wanted to implement a second learning activity, which could take the students up in the stairs of knowledge. We did not mange to implement this activity but we have a vision. In the second learning activity the student is it to see patterns and understand that one invention could change the world.

The new conversations and dialogues will be a crucial pivot point for the game. It is through the dialogs we can help the student to discover the relationship between the various historical events and extract the important perspectives. By doing this we will promote meta-cognition and help the student learn how to learn. This is far more important than remember the birthday of Leonardo da Vinci.

In the first course we stated that we wanted to create more historical models and surrounding. In addition to Gutenberg's world we did manage to fully create another world, da Vinci's world. Although Hooke's and Galileo's worlds are not fully done yet, one can still move around in a time typical world, for instance the famous leaning tower in Pisa in Galileo's world and the typical English environment in Hooke's world. As we stated in the first course we wanted to give the student additional feedback about the geographical place with landmarks. Not only did we use landmarks to give a hint about the city, we also gave the historical persons dialects.

The testing sessions resulted in a few bug fixes, e.g. the word limitation in the notepad was removed. The fixes that we made were problems we had missed during the implementation, having test sessions opened our eyes to see the game from a user's perspective. The main goals with the tests were to test the game quality and the user interface. From the inquiry that the students participated in after the test we could see from the diagram that the game was e huge success. In both the questions about the user interface and the quality of the game more than 70% of the students answered with a score of five to seven. We interpret these results as very good and that the game is something that is awaited. Therefore we are working on an online version of the game, having an online version opens up a larger target group.

In the future we will continue to develop the game, but first and foremost we will do some finishing touches in Hooke's and Galileo's world. When this is done we will probably start implementing the second learning activity and later we will hopefully perform another user test.

References

Schutz, William C.1958. FIRO: A three-dimensional theory of interpersonal behavior. Oxford, England: Rinehart. ix 267

The Teachable Agents Group at Vanderbilt University [http://www.teachableagents.org/research/bettysbrain.php]

Gärdenfors, P. 2010. Lusten att förstå, Författaren och Natur & Kultur, Stockholm

Holmberg. N. 22.10.2012, Lecture at Eye-tracking course 2012

Helstrup. T., Kaufmann. G; 2002; Kognitiv Psykologi, Fagbokforlaget, Bergen

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D.R. 1956. Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain New York, Longmans, Green.

Haake M, 2009. Embodied Pedagogical Agents - From Visual Impact to pedagogical Implications, 3-56, Dept. of Design Sciences, LU

Chase, C., Chin, D.B., Opezzo, M., & Schwatz, D. L. (in press): Teachable agents and the protégé effect, Journal of Science Education and technology

Norman, D.A.1990. The Design of Everyday Things. The MIT Press. London, England.

Appendix A

The following commands are executed by pressing the following buttons on the keyboard:

Left – A
Right – D
Walk forward – W
Backwards – S
Enter – E
Jump – Space
Hide/view head-up-display – H
Look around in the world – hold down the right button on the mouse

Appendix B

Om Historie-agenterna

Kryssa i den ruta som passar bäst in på vad du tycker

Jag är

- Tjej
- Kille

Agentnamn:

1. Jag vill gärna spela detta spel igen

	1	2	3	4	5	6	7	
Håller inte med	○	○	○	○	○	○	○	Håller verkligen med

2. Detta är ett bra sätt att lära sig historia på

	1	2	3	4	5	6	7	
Håller inte med	○	○	○	○	○	○	○	Håller verkligen med

3. Jag tyckte om min agentkollega

	1	2	3	4	5	6	7	
Håller inte med	○	○	○	○	○	○	○	Håller verkligen med

4. Jag tyckte det var roligt att prata med de historiska personerna

	1	2	3	4	5	6	7	
Håller inte med	○	○	○	○	○	○	○	Håller verkligen med

5. Jag tyckte om hur de olika världarna såg ut

	1	2	3	4	5	6	7	
Håller inte med	○	○	○	○	○	○	○	Håller verkligen med

6. Detta tycker jag också var bra i spelet

7. Detta tycker jag kunde bli bättre

Appendix C

1	**6. Detta tycker jag också var bra i spelet**
2	Jag tyckte att grafiken var bra. Jag tyckte också att det var bra att man kunde stoppa planet när det höll på att åka in i tunnorna i Rom.
3	Jag tyckte det var väldigt bra och man kan lära sig mycket.
4	allt var bra fast det tog lite tid att lära sig det
5	allt det jag kanske gillade mest var att vara i dem olika värdarna
6	Slutet när man skulle puzzla ihop.
7	att dom sa bra instruktioner
8	Man kunde få fakta som inte fanns på klossarna.
9	Att man fick besöka de olika länderna
10	Att man kunde "bugga" flygplanet hos Davinci
11	att man lär sej om historia och man kan lära sej om hur saker har uppfans som vi använder i dag
12	Det mesta var kul!
13	det var kul
14	Grafiken var bra
15	Man kunde gå omkring istället för att bara stå still och svara på frågor
16	Graficken i vattnet och på husen.
17	Jag vet inte
18	allt var bra
19	man kan hopa upp på monga saker
20	Man lärde sig mycket om historia!
21	Jag tyckte det var bra att jag fick reda på mycket om olika uppfinnare som jag inte visste något om.
22	Fint animerat
23	man fick anteckna så man kommer ihåg
24	detmästa
25	att man shälv kunde gå runt
26	Dom förklarade det väldigt bra så att man förstod
27	Det var välidigt bra speciellt att man inte behövde läsa en bok ! :)
28	att man lärde sig något

1	**7. Detta tycker jag kunde bli bättre**
2	Rösterna till personerna som var med i spelet.
3	Att man skulle kunna skriva mer.
4	Jag tycker man skulle kunna ´skriva mer på anteckningar och skrolla när man skrivit mycket.
5	inget
6	jag tycker att värdarna skulle kuna vara lite störe och mer att uptecka.
7	Grafiken mussen att man inte behöver hålla inne höger knappen. Och jag hadde inget ljud så att det är därför frågerna e på 1-3
8	att man inte kunde gå till baka till staden om man skulle sätta in klossar typ
9	Det kunde vara lite svårt att upptäcka utropsteckena ibland.
10	Ljudet, men annars.. TOPPEEEEEEEEEEEEEEEEEEEEN!!!!!!!!!!!!!!!!
11	ljudet försvann ibland
12	Att man skulle kunna spela hemma och att man skulle kunan gå in i alla dörrar/hus i omgivningen
13	att alla saker på historia glasögonen skulle funka när man klikade med musen på dem.
14	Den var lite trög
15	inget
16	Några glitchar
17	Mer information vid dom olika gubbarna för man fick inte veta vilket år saken gjordes hos alla!
18	Grafiken. Och att man kan klicka på E och undesöka från alla vinklar.
19	det fins inget betre
20	Lite ljusare färger
21	informationen
22	kanske kunde varit lite större ställen
23	lite fler platser och lite intresantare
24	Det var jätte bra
25	Kanske skulle det komme lite fler världar.

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Developing an Educational Game In the subject of history and with a teachable agent

Jesper Funk Johan Lilja Johan Lindblad Sofia Mattsson

This report aims to give a wide overview of the progress of creating an educational game in history with a teachable agent. By first developing a low-fi prototype of the game and thereafter a mid-fi prototype, the plan was to later on perform usability evaluation tests to gain an insight of the game's ability to develop into a good learning material for students. The usability of the mid-fi prototypes was tested on a group consisting of six potential users.

From the work done the conclusion was that the overall experience of the game was satisfying. A compilation of the test results showed that there still were some areas in the game that needed to be improved. Along with the feedback given from the usability evaluation and with the work of the hi-fi prototype it became clear that it is hard to develop these kind of games, especially the part where the player shall teach the TA. Before implementing the hi-fi prototype, rework and improvement of the part "teaching the TA" was made by working on its conceptual design again.

Due to time constraints no usability evaluation was made on the hi-fi prototype. Therefore the prototype was evaluated along with the nine learning mechanisms by Stellan Ohlsson instead. The evaluation focused on the reworked and improved part of the game; the teaching of the TA. The game does cover almost all of the nine learning mechanisms, although they were not used in the beginning when the game was created.

1 Introduction

Educational games with a Teachable Agent (TA) are computer based pedagogical tools, which have been proven to be powerful and motivating. The TA is a pedagogical agent which is taught by a student who asks it questions and helps it to solve problems, with the results that the student can test its own knowledge while teaching. They can also be used as powerful research tools for studying how pupils learn. This chapter will give the reader an introduction of the work within the project, presenting the purpose and goal, scope and definitions needed for the report.

1.1 Brief introduction

All existing TA games target mathematics and natural science, therefore this project aims to develop the first history orientated TA game. The project was performed in two courses, Interaction 1: Neuro modeling, Cognitive Robotics and Agents followed by Interaction 2: Virtuality and Cognitive Modeling. The stakeholders of the project are Educational Technology Group LU/LiU, Stanford School of Education and Natur & Kultur.

1.2 Purpose and Goal

The report's purpose is to give an overview of the progress of creating an educational game in history. By working with the conceptual design in the beginning of the project and thereafter moving on and developing a low-fi prototype, the plan was to quickly gain an insight of the game's structure. A mid-fi prototype was developed later on and tested on the target group at a school in Lund. Before implementation of the hi-fi prototype was started once again, work was made with the conceptual design and with focus on the part of teaching the TA. The goal with this report is to deliver a documentation which should be viewed as a complement to the project, containing clarifications and justifications of the game's design.

1.3 Scope

Although the project is carried out throughout the semester, time is a limiting factor in both courses. Therefore the focus in the first course, Interaction 1: Neuro modeling, Cognitive Robotics and Agents, was on developing a low-fi prototype and a mid-fi prototype. The mid- fi prototype will contain one historical scenario and was tested on the target group in order to receive feedback.

In the second and following course, Interaction 2: Virtuality and Cognitive Modeling, the desirable focus was to be on developing a hi-fi prototype in HTML5 and Javascript. Due to very little time and limited knowledge in HTML5 and Javascript the focus in the second course was changed to rework and improve the part "teaching the TA" along with the feedback given during the usability tests. A smaller hi-fi prototype of the game was developed with help by personnel at the department of Design Sciences.

1.4 Definitions, acronyms and abbreviations

AI – Artificial Intelligence

EA – Embodied Agent, a graphical representation that in different degrees can be humanlike.

Hi-fi prototype – High-Fidelity prototype, a prototype with many details and numerous functions and is close to the final product. Used in order to investigate the usability in detail and thereafter make conclusions about how it will relate to the use of the final product.

IA – Intelligent Agent, a type of virtual character which performs tasks it has been given from users in a digital environment.

Low-fi prototype – Low-Fidelity prototype, a prototype that is incomplete but has the characteristics of the target product. Used in order to quickly produce the prototype and test broad concepts.

Mid-fi prototype – Mid-Fidelity prototype, a prototype with detailed design, which serves as a reference for the functional specifications. It is used in order to evaluate the usability.

TA – Teachable Agent, a type of pedagogical agent which can be used in education.

2 Background

The purpose of this chapter is to give the reader a more in depth description of the area in which the project has its setting. The use of educational games and their purpose is explained. This chapter will also explain what a teachable agent is and how it can be used to teach pupils.

2.1 Educational games

Educational games are used to teach in certain subjects. Games often contain modern design and interaction and hold the attention of the player for hours. Games and simulations can be seen as powerful tools for teaching theoretical concepts, practical and technical skills [1]. According to the media equation theory, people show signs that they react the same way towards computers as towards real people. Studies have shown that people may react to praise and flattery coming from a computer in the same way as if it was uttered from a real person. [2]. These findings are quite relevant for educational games based on a TA, since it shows that even praise by a TA could have the same positive effect on students as praise from a real person would.

Almost all of today's students have grown up with video and computer games, and especially digital technology. Therefore it is a natural step within the modern technology to use games in order to teach students in different subjects. Within games there are features that are optimal for learning. Some of these features are;

Goals – In a game which is designed with a structure that is clear to follow, the learning goals often also get clearer. The player can easily understand why he/she is learning something and how the new knowledge can be used, both in the game and outside, e.g. in the classroom [1].

Continuous monitoring of progress – The feedback in the game is clear and often direct. This can be useful in order to follow the player's development within the game and possibly adjust the difficulty of the learning level [1].

Interaction and encouragement – A game that has high interaction with the player often motivates him/her to continue playing the game and seeking for information he/she needs to learn. High interaction will also catch the player's focus and have him/her engaged to the game [1].

Infinite patience – A teacher can lose its patience and will for example conclude that a student will never understand mathematics. In a game, one valuable feature is that it contains infinite patience, which gives the player an opportunity to try again and again [1].

By using games as an educational environment subjects can be taught in new ways and it also gives an opportunity to think about meta-cognitive learning abilities. An educational game can be more interactive by having one or more agents in the game. Depending on the type of agent used in the game, the player and the agent can interact in different ways.

2.2 Agents

As just mentioned, agents can be used in games in order to make the game more interactive for the player. In educational games, an agent can in general be defined as a software and/or hardware component which is given tasks from its user and acts precisely to finish them [3]. A software agent is a virtual character and there are different types of these, which are described further down in this section.

2.2.1 Intelligent agents

An intelligent agent (IA) behaves autonomously and has some capacity of taking decisions. The IA performs tasks it has been given from users in a digital environment; the IA can be seen as a digital actor who may affect the real world due to its behavior. If an IA is used in an educational game it is often simulated humanlike, depending on the target group [4].

2.2.2 Embodied agents

An embodied agent (EA) is a graphical representation that in different degrees can be humanlike. By using a EA it is easier for the user to interact since it feels more natural. Embodied agents can be divided into four categories [5];

Synthetic agents – these agents are most common in computer games and other virtual environments. Their behavior is very humanlike [5].

Animated agents – a variety of the synthetic agents, often used as guides in order to help the user if he/she has questions. They are also defined as digital characters that moves and they are often designed as cartoons instead of humans [5].

Emotional agents – the user can affect these agents' predefined personalities and emotions [5].

Embodied conversational agents – these agents have a humanlike appearance and try to simulate a human conversation [4].

2.2.3 Pedagogical agents

Embodied agents can also be used in education as pedagogical agents. These agents are a subtype to IA and can be divided into two categories; those with authoritative roles and those with non-authoritative roles [5]. The authoritative roles involve teachers, instructors and mentors and the non-authoritative roles involve study partners, learning companions and teachable agents. A pedagogical agent can be adjusted along each student and their level of learning [4].

The use of virtual agents has increased gradually in the recent past and they can be found in different contexts. The different types of agents can be combined, for example as a virtual intelligent embodied agent. The main reason for their use is that they allow a smoother interaction between a user and the application it uses [4]. An example is the virtual agent that IKEA uses on their website in order to simplify it for customers to ask questions and find items they are looking for.

2.2.4 Teachable agents

A Teachable Agent (TA) is a pedagogical agent. The TA is taught by a student who asks it questions and helps it to solve problems, with the results that the student can test its own knowledge while teaching. The TA's answers to questions and problems are produced by using techniques within Artificial Intelligence (AI) and are based on what it was taught. The knowledge of the TA is revisable depending of its answers, therefore this gives an opportunity for the student to revise both his/her knowledge and the TA's [6].

The TA can also be designed as a troublemaker. It interferes with the student by suggesting solutions that sometimes are correct and sometimes are wrong, which challenge the student's self-confidence in learning [4].

A player of a game with a TA can have what is called a theory of mind. This means that the player can attribute to others and themselves such mental states as knowledge, belief and emotions. It also makes one realize that others can have different thoughts than oneself. Theory of mind is still a theory because the mind is not something you can observe and therefore ascribe similar awareness through an analogy from one's own [7].

One way to test whether someone has a the theory of mind is a so called false belief test, where you have to be able to understand what another person is going to answer on various questions, where the person knows or sees something different than you to yourself: very young children cannot do this. Their mental development of a theory of mind tends to not be fully completed until the ages of nine to eleven [7].

The theory of mind is relevant for games with a TA, because it builds on the concept that the users can interact with the TA as if it possessed a theory of mind. The user has to understand that the TA's actions are built upon the facts that the user has taught it.

2.2.5 Effects

When playing a game with a teachable agent, different effects can occur;

Protégé effect – means that children spend more energy to teach others than they learn for themselves. These others may be younger children or teachable agents. Children who use a teachable agent usually get more points than students who learn alone, but the effect seems to have greater impact on children who underachieve in school than those who already performs good [8].

Proteus effect – means that a person who uses a digital representation changes his/her digital behavior to adapt to the representation [9]. A person's behavior in the virtual can then differ greatly from the reality.

Ego-Protective Buffer – refers to a kind of motivation-based intervention to increase a person's persistence in continuing after failure [10]. An ego-protective buffer shall ensure that there is a steady sense of being competent, by suppressing the damage a failure can have on the psyche. It may be that the blame for the failure to put away from oneself. An example would be a game where the outcome is influenced by skill and luck, which means that you can often blame bad luck if you lose [10], see for example the difference between chess and backgammon.

2.3 Avatars
An avatar is virtual character controlled by the human it [8]. The difference between an avatar and an agent is that the avatar is controlled by a human and the agent is controlled by the system [4]. A TA has a combination of properties from both an agent and an avatar [6].

3 Methodology

Information about educational games and agents were sought in articles, both from the given course compendium and on internet. Scenarios were to be developed during the creation of low-fi prototype and mid-fi prototype. One of the scenarios was tested out on the mid-fi prototype. The result from these tests led to focus on rework and improvement of one part of the game, which formed the basis for the design proposal of the hi-fi prototype for an educational game in history with a TA.

3.1 Literature Review
To gain required knowledge about educational games and agents, literature review was carried out. The search of relevant information was made mostly at the beginning of the course but was also carried out in parallel with the course.
The types of literature chosen were reports, conference papers, articles and books. In addition to the literature review, the course material from the Department of Design Sciences at Lund University was also a source of information. The curriculum for 10-11 years old students was also a source of information in order to get information about what students are expected to learn in history.

3.2 Test persons
The test group consisted of six persons in the age between 10-11 years old, both females and males. The reason why the number of test persons was six is because according to Jeffery Rubin's "Handbook of Usability Testing" 4-8 test persons are enough to expose most of the usability problems. This is not enough participants for a statistically valid result but a good basis for future work and studies. This makes for qualitative results instead of quantitative [11].

3.3 Test Environment and Equipment
The test was set up at a school in Lund. The test was performed in a smaller room next to the classrooms. The room was decorated with a table and some chairs. The test was supervised in the same room where the test was performed.
The test equipment was one laptop with connection to the internet. The mid-fi prototype of the game was shown on the laptop. The test of the mid-fi prototype needed to be as interactive as possible for the participants so they could get as real experience as possible. This was solved with Balsamiq's presentation mode where the participants could navigate through the game as intended. The test's time during all of the tests was about 30 minutes per participant.

4 Identification of target groups and services

We received a project description and a deeper system description from our supervisors to obtain information how the game should work and who it should target.

4.1 Who are the end users?
The intended end users of the game are boys and girls aged 10-12 years. The game will be designed along with the curriculum for students between the ages 10-12. The game will have a fairly limited audience since it will be too difficult for younger students and it will be too easy for older students.

4.2 What requests do the stakeholders have?
By using the curriculum and its content as well as the project description, the following requirements are considered important to fulfill;

• The user should be able to o read
o find information from a given era
o teach the found information to a TA
o end and save every step
o see the TA's knowledge development
o create an avatar
o use different learning activities to take in information
• There should be different levels of difficulty
• The user should quickly be able to understand how the game works
• There should only be a few clicks to continue the game after the first time the user
played the game
• There should be a button to go back in the game
• The user should have access to previously learned facts

• There should be dual coding for color blind

• The TA should be able to take a test while the user observes

5 Conceptual design

The second week of the project we had a meeting where we brainstormed about the conceptual design of the game. Thereafter the conceptual design was partially updated continuously as the project proceeded.

5.1 The game's overall structure and the user´s mental model

When a user starts the game he/she shall be able to choose an existing avatar from previous gaming or to create a new one. If the user chooses an existing avatar, he/she should then continue the game where he/she left it. If the user chooses to create a new avatar, he/she will then start the game from the beginning.

During the first meeting we all agreed that the following design points should be implemented in the game;

• Rewards – when the user and the TA perform well in tests or in other tasks, they should be given rewards for their achievements.

• Progress bar – will make it easier for the user to see his/her progress in the game.

• Levels – users can have different knowledge and the game should adapt to it.

• Tests – both for the user and for the TA.

• A background story – this should be a way of capturing the user interest.

5.2 Grouping of historical events into natural groups

Grouping could be made by having a number of main groups with unique names where possible historical events, discoveries or persons can fit in. The grouping will provide an easily understandable overview of what historical areas the game consists of.

A further development could be to present the different historical areas with icons and related title. This will provide good visibility and the user can easily choose the area he/she wishes to learn more about since pictures is interpreted faster than text.

5.3 Use of language/terminology

The game should be adjustable to the students' different knowledge in history. The game should have an easy and stylish design, as well as it should be easy to understand the texts. It should also be easy to understand how to play the game.

5.4 Metaphors

Metaphors will be a natural part of the solution, especially to simplify the navigation in the game for the user.

5.5 Background story

From an early point on it was decided that the game would have to have a plot that would capture the player's interest and present them with a goal to accomplish.

The story of the game is as following; the player, represented as an avatar visits the Timecastle of Professor Chronos, who is the current guardian of history. Inside the castle the player meets a time gnome called Tidsnisse who tells them the story of the castle and it's ruler Professor Chronos; Since his employment of three thousand years is winding down he is looking for a successor among his helper gnomes. However the one to take over after Chronos must show extensive knowledge of history. To be able to learn historical facts the gnomes have a time machine at their disposal. Unfortunately Tidsnisse can't use the time machine since time travels give him travel sickness. So what he now asks is that the player uses the time machine and teaches him the necessary facts from the different historical periods, so that Tidsnisse can become the next guardian of history.

6 Low-fidelity prototyping

The art of giving the user an intuitive interface to work with is dependent on four important factors; visibility, affordance, mapping and feedback. If the design does not take into account these factors the experience will be lacking from the users point of view and therefore leave the player with much to ask for.

To ensure that the design is up for the task the first step in designing a program or game in this case is to make a low-fidelity prototype. By limiting the work to pencils and paper and going back to the basics of play school the project group was able to deliver a vast amount of ideas and visions without having to compromise too much time or resources. This is the main advantage of the low-fidelity prototype.

6.1 Visibility, affordance, mapping, feedback

Visibility – Rather self-explanatory. Things need to be visible, including the conceptual model of the system, the alternative actions, and the results of actions. But! Less is more, the mind can only focus on a few things at one time so it is therefore important that buttons and such must be "in use", have a purpose, or they shouldn't be visible for the user. Whilst designing our low-fidelity prototype we made sure pretty early that the layout of the game followed a structural design standard throughout the game so that the user could identify certain buttons and such right away even though they were presented with a new step in the game.

Affordance – When used in the sense of affordance of objects, it refers to the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used. It is important that the design is made so that it is easy to determine what actions are possible at any moment for the user. By doing so we don't get any problems with users not knowing how to orientate themselves through the game. In the process of low-fidelity design we made sure that it was clear how a certain part was coded to act when touched.

Mapping – Is the technical term meaning the relationship between two things, in our case between the controls and their moments as well as the results in the game. Natural mapping, taking advantage of physical analogies and cultural standards leads to immediate understanding. For example, a designer can use spatial analogy: to move an object up, moves the control up. This was made as straightforward as possible when designing the low-fidelity prototype. We made notes in the margin explaining the thought process of the current paper or screen.

Feedback – Is a process in which information about the past or the present influences the same phenomenon in the present or future. For a user it's very important to get input and response from the program and/or game on what the current status is and if my latest action was registered.

It's difficult to simulate feedback when using pen and paper; therefore we simply had feedback as a train of thought when working with the design.

6.2 Selection and design of interaction techniques

The game has in its specifications, to be playable in a web browser and has therefore a web browser's limitations and benefits. In the first stages of the low-fidelity design all interaction was thought to be made with mouse clicking action but in a time of tablets we pictured our user using a full touch interface together with our game.

Interactions like animations, sound and force feedback are things that we want and are going to have in our high-fidelity version of this game.

6.3 Overview, orientation and navigation

Having a structured standard of the game is important for the user's ability to orientate throughout the game. This will facilitate for new users to feel "at home" and become more familiar on how the game works. The introduction of the book "Resor" is an example of easing the overview pressure on the user. By that he or she doesn't have to feel that the amount of information given on each time trip is too much to handle. According navigation the design includes a backward arrow giving the user something familiar to use when in need of going to the previous page. For future versions of the game hand gestures will with high probability be introduced for orientation purposes. Especially younger users are well aware of how it works and it would be a mistake not to use this knowledge.

7 Mid-fidelity prototyping

The mid-fi prototype of the game was created in Balsamiq.

7.1 Layout and grid

The layout of the game was developed to be as simple as possible and with icons keeping their positions through the different stages of the game for consistency reasons. The menu, help and book icons were put on the top of the pages. For the historical periods part the main information about the visited time period is presented in the middle, since that is where the main focus is expected to be. The other options, like persons, events, and discoveries, are listed below, so that the player may advance to them after the main text has been read. On the

left side there is an information bar about the facts that have been learned and are ready to be passed on to the TA.

7.2 Concept the design follows

The game was going to start with a part where the player could create a new character, which would act as the avatar of the player through the game. This would also be the part where already existing characters could be chosen. After that new players would be guided through a short introduction to the story and goal of the game were the TA, the Professor and the time machine are introduced for the first time. Playing the game, the player would be able to use the time machine to easily navigate to the different time periods and there choose which subjects are the ones that should be explored. In the time period there should firstly be some general knowledge about the time period in question. Secondly there are the facts from each period to learn more about. When one is picked. There is first some text with information to read followed by some instructive and fun games about the subject. The game should be simple to understand without taking away too much from the exploration face of the game, so that it not bores the player to fast. The teaching of the TA should also try to give the player an understanding of how a learning process could work.

The three learning steps in the "brain" of the TA were an attempt to visualize this; where the first part is supposed to be a sort of dialog with the TA, where a "thought" is created. Theses "thoughts" consist of the facts about a certain subject, such as the discovery of Greenland. Afterwards the "thought" is placed on a timeline and lastly the "thought" can be linked together with other thoughts in a cluster of thoughts, which is supposed to work like a mind map. The testing of the TA will be through a dialog with the Professor, where questions are asked and the TA gives answers according to how the "thoughts" are organized in his brain.

7.3 Animations

The decision was made to use simple cartoonish animations for the game, since that makes it easier to design characters that avoid the uncanny valley [12]. The animations for the game as a whole were made in Photoshop with a Bamboo tablet, which gave the game a simpler look. The motivation for this was also that, since the game was going to be web based, simpler animations would be easier to load and use on different kinds of computers, which might be an advantage since not all schools might be able to afford the newest computers.

7.4 Sizes

The sizes of the different parts of the game were as big as possible, so that it would make it easier to navigate through the game. This was also made possible since the way the animation was designed there wasn't that much detailed during most parts of the game. An attempt was made to keep the text in the icons in a size that was comfortable for 10-12 year olds.

7.5 Color

The colors that were used in the mid-fi prototype were mainly chosen to give the user a clear way to distinguish the different parts of the game. The colors were simple and helped to enhance the cartoonish feel of the game. Different colors were used for the icons for historical persons, events and discoveries. The avatars one can create have a wide variety of skin colors to represent different human skin types, which reflect the diversity in modern day classrooms.

7.6 Icons

The icons used were a menu button that can be used to open, save and exit the game (In mid- fi this is only represented by a non-clickable button, since the game is going to be web based a short command doesn't make as much sense), a small book, which by clicking gives information about the previous information that has been gathered and a help button, where players can find tips if they encounter any problems during the game. In the time machine there is a control switch that the user needs to push in order to travel to the chosen time. During the different time periods the same icons were used for persons, events and discoveries.

7.7 Typography

No emphasis was put on the typography during the mid-fi prototyping. There were only some thoughts spent on how much text should fit into each speech bubble without making the text too small to read easily.

7.8 Sound

In the game there are supposed to be a variety of sounds that help the user understand what is happening, which couldn't be used since Balsamiq doesn't have that function. One of the sounds would be a sound effect for the time travel part and another one for the sound of a pen writing, when new information is added to the travel log.

8 High- fidelity prototyping

Time was a limiting factor during the second part of the project, and due to limited knowledge in HTML5 and Javascript the focus was changed to rework and improve one of the items mentioned in the section 10.3.1.

8.1 Rework and improvement

The part of the game where the player teaches the TA was perceived as difficult to understand when it was tested during the usability evaluation. Therefore the focus in the second course was to rework and improve this part of the game.

Meetings were held for new brainstorming sessions about the conceptual design of teaching the TA. Along with the feedback given during the usability tests and the meetings it became clear that the interaction needed to be expanded and enhanced and the steps in teaching the TA needed to be easier and clearer to understand.

The first step where the player shall teach the TA wasn't in need of much improvement but the second step "arrange thoughts" was changed a lot. First a sketch of the second step, including the changes, was made, shown in figure 1.

By doing the iteration properly from scratch we reworked the conceptual thinking behind the part where the TA was taught the gathered information from the player's time machine trips. Since the game story with all of its appurtenant details has been highly appreciated, these parts weren't in need of that much work, rather than small tweaks.

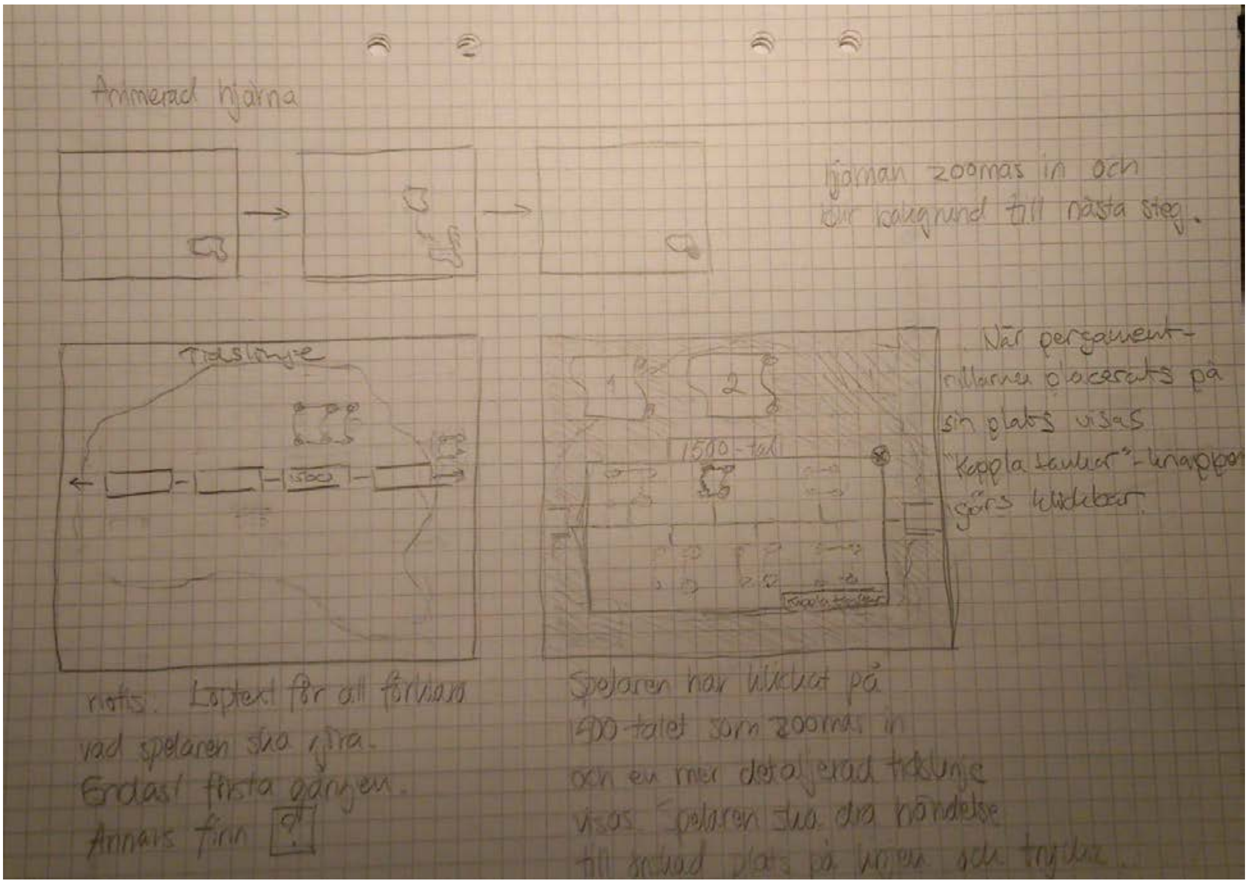


Figure 1. Sketch of the new creation of the TA's thought.

Figure 2. Dialogue between TA and the player.

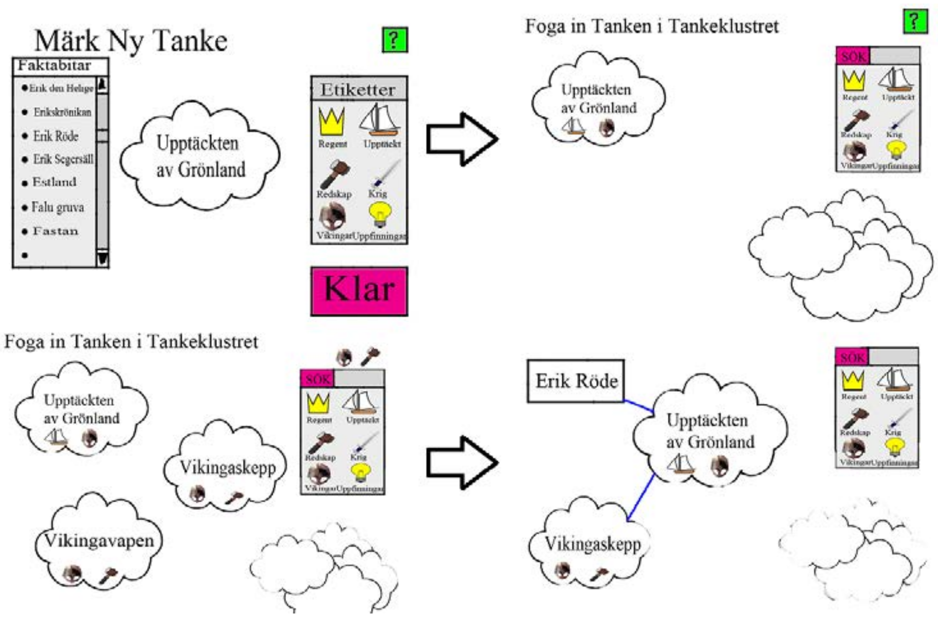


Figure 3. The linking and tagging of thoughts.

When doing the rework of the part "teaching the TA" the initial stage was designed to be in a dialogue mode. When the player returns from his or her historical trip and meets up with the TA they commence to speak to each other. The dialogue is almost in an interviewing manner because the TA is so very interested in where you as a player have been. This also grants a great way of handling the learnability since the player can see and reflect how the TA handles the information he is given, see figure 2.

For every new subject discussed between the two, a parchment scroll is written by the TA. This parchment is in the end of the dialogue absorbed into the TA's brain by an animated loop through the air while his headdress is tipped to the side, as seen in figure 1 above. Whilst this happens an indicator of the TA's brain fills one section to indicate that he has learnt something new. But after three parchments have been absorbed the indicator is full. The TA now has too many new thoughts and must have the player's assistance to arrange them. The player either clicks on the brain or is through dialogue sent into the TA's brain.

Once inside the brain, three steps follows. First, the parchments that the TA just absorbed are now hovering over each respective year where it originates from. The player is to click on the correct year with appurtenant parchments on the timeline that is presented. A zooming effect puts the player at the second step. The player is presented with another timeline within the chosen year. There is now a drag-and-drop situation where the player drags the hovering parchments to a fitting location on the zoomed in timeline, as seen in figure 1 above. When satisfied with their location the player is now presented with the "Koppla tankar" button. This takes the player even deeper within the TA's brain and to the third and last step.

The last step of "creating a thought" where the user shall connect different thoughts to each other also needed much rework. When entering this step the player sees the thoughts floating around with different factual content. The parchment scrolls have changed appearance into "thought bubbles". In this step the player shall link the thoughts bubbles and tag the bubbles with related sorting icons. The selectable icons are predefined, for example a crown could represent kings. This will make it possible to sort the thoughts and to get a better overview. A sketch of the step of creating thoughts and its context is shown in figure 3.

The following phase with the test of the TA's knowledge is also changed into a dialogue mode to keep consistency throughout the game. This makes it easier for the player to get a familiar sense and learn how the game works, see figure 4. This part of the game will also be extended with sound to help the player understand everything that is presented. This phase also holds an important part in the educational purpose linked to the TA understanding and reasoning behind his discussions. In the test you can see how the thoughts are linked and how the TA processes his knowledge into the answers he gives.

Figure 4. Test of the TA.

8.2 The nine learning mechanisms

During one of the lectures given in the second course Stellan Ohlsson's nine learning mechanisms were presented. According to Stellan Ohlsson the nine learning mechanisms are [13]:

1. **Direct, verbal instruction** - it is important to have guidance in an educational game.
2. **Examples that are explanatory** - before the player performs a certain task it is good to have a simulation of "how to perform" the task.
3. **Analogies** - the player uses his/her memory to map his/her current task onto an analogical task.
4. **Reasoning based on declarative knowledge** - the player learns how to reason if he/she already knows one part of the task.
5. **Results of trial and error** - the player learns to try another solution if the first didn't work.
6. **Positive feedback** - the player receives positive feedback when he/she performs a task correctly.
7. **Negative feedback** - the player receives negative feedback when he/she performs a task incorrectly.
8. **Experience and memory of past tasks** - the player will learn how different tasks are performed and remember it for future tasks.
9. **Pattern recognition/ regularities in the environment** - as mentioned in number 8, the player learns to recognize patterns by learning how different tasks are performed.

These nine learning mechanisms constitute "a scientific theory of how skills are learned during practice" according to Stellan Ohlsson. He also states that these nine learning mechanisms are important for educational software. The nine types of information helps education software to be more effective if the types have been in mind when developing the software [13].

8.2.1 Evaluation

Due to time constraints and an incomplete prototype no usability evaluation was made on the hi-fi prototype. Therefore the prototype was evaluated along with the nine learning mechanisms by Stellan Ohlsson. The evaluation was focused on the reworked and improved part of the game, teaching the TA.

Number one of the nine learning mechanisms is achieved in the game. When teaching the TA the player will be given guidance in how to perform the tasks in the different steps. The player will always be able to press on a button with a question mark if he/she needs help to understand a task.

At the moment there are no simulations of how to perform a task, there will instead be instructions explaining what to achieve. Therefore the second learning mechanism can be seen as partially achieved. The player will quickly learn how to drag and drop the parchment scrolls onto the timeline and also how to connect the different thoughts. This will improve the ability of recalling this knowledge at a later moment, which is related to number four, eight and nine of the learning mechanisms.

The player will receive both positive and negative feedback depending on how well the TA performs at Professor Chronos's test. This feedback could help the player to adapt his/her choices in accordance with it and is related to number five, six, seven, eight and nine of the learning mechanisms.

This educational game with a TA does cover almost all of the nine learning mechanisms, although they were not in mind when the game was created.

8.3 Creation of the hi-fi prototype

In dialogue with the supervisor an agreement of getting assistance with the flash implementation was made. This gave an opportunity to make a full scale iteration with the most prioritized parts that had to be changed before implementation of the hi-fi prototype could start.

To provide the supervisor with the material needed to start implementing the high-fi version of the game a folder system was created consisting of Balsamiq slides, deconstructed into the folders with explanatory instructions on how they should look and act in flash. This made it easier for the supervisor to handle the material since the mid-fi with its Balsamiq slides were quite the bunch.

For the hi-fi version we also wanted to add another element to the gaming experience through sound. Although it's not completed, the entire dialogue session from creation of thought has been recorded and will be implemented in later versions of the flash game. By doing this the player gets at least dual coding (text and sound) whilst interacting with the game. In certain situations there will also be pictures enhancing the efficiency of the learning process.

Due to time constraints for the assistant the hi-fi prototype wasn't completely finished when it was supposed to be presented at the oral presentation in the end of the second course. The hi- fi prototype is still under development and will be finished sometime in the future.

9 Usability evaluation

The main focus during this test was to evaluate the overall experience of the game. The participants got a short description about what was going to happen during the test.

The mid-fi prototype was run in Balsamiq's presentation mode during the test, where the participants were able to click themselves through the game. The participants were given a scenario before they started to play the game.

After the test there was a short interview based on some post test questions and a briefing session. The following main areas were discussed;

- The participant's game experience
- The overall thoughts of the game If it was easy to understand how to navigate in the game
- If it was easy to understand how to choose an era to go to inside the time machine
- If it was easy to understand the progress bar and what "added to trips" meant
- If the learning to the TA was easy to understand

9.1 Results

9.1.1 Participant A

Test person A understands how the time machine works. He/she thinks the font size of the text in "Description" is small and should be one or two sizes bigger. He/she finds it difficult to understand that you should scroll in the textbox to be able to press "next" and navigate to the next "page". Instead person A goes back to the time machine.

When it comes to the progress bar, person A understands the purpose of it but doesn't understand "Added to trips" and the connections to the book "trips".

Person A finds it difficult to understand when teaching the TA how the different steps should be added to the TA's brain. He/she is trying to drag the whole box of fact. He/she doesn't find these steps easy to do.

He/she is an experienced game player and thinks the game and its story is very funny. Person A thinks it is hard to understand how to navigate in the game, for example, at the page where the TA is crying.

9.1.2 Participant B

Person B doesn't understand where he/she should press to navigate further when he/she is at the page where the TA is crying. Person B thinks he/she should press on the TA or on the door above the stairs.

He/she finds it is easy to understand how to choose an era inside the time machine.

Person B has some difficulties finding which discovery he/she should choose to learn more about. He/she does understand that he/she should scroll in the textbox of the description to be able to read more of the text, but not how to navigate to next "page". Person B also has problems knowing where to press to finish his learning activity. He/she continues by going back to the time machine instead of pressing "Finished".

The progress bar isn't any problem for person B to understand but he/she doesn't know what "Added to trips" could mean.

Person B finds it difficult to understand where he/she is supposed to click to start teaching the TA. He/she tries to press on the speech bubbles as he/she did on the page where the TA is crying. During the part of teaching the TA person B finds it difficult to understand how the different steps is designed.

Person B is an experienced video game player and likes games with missions and he/she thinks the game has a nice design. He/she feels it is problems knowing how to navigate but overall it is ok.

9.1.3 Participant C

Person C has trouble finding where he/she should click for further navigation on the page where the TA is crying. After a while he/she notices that the mouse icon changes when it hover the clickable speech bubble.

He/she understands how the time machine and how to choose an era works. When leaving the time machine and choosing a discovery person B has difficulties understanding how he/she should choose a discovery.

Person C has no problems understanding the progress bar and the purpose of the book "trips" as well as the connection between the book and "Added to trips". When he/she is finished with his/her learning person C presses on "To time machine" instead of "Finished".

When it comes to teaching the TA, person C finds it difficult to navigate and to understand the different steps.

Person C plays video games now and then and thinks it occasionally is very difficult to understand how he/she should navigate further, especially when teaching the TA.

9.1.4 Participant D

Person D doesn't understands where to click for further navigation on the page where the TA is crying. He/she tests his/her way and finds where to click.

He/she doesn't find it difficult to understand the time machine and choosing an era to travel to. Person D thinks it is a smart and simple way to choose an era.

Person D finds the text in the text box of the description should be in bigger font size or with more air between the paragraphs.

He/she understands the purpose of the progress bar but have difficulties knowing what "Added to trips" means. Person D presses the "go back"-button during this step of the game and he/she sees a lamp under discoveries but doesn't know what it indicates.

When it comes to teaching the TA, person D doesn't have problems with the different steps. He/she thinks it is difficult to understand step 3 but manages to do right.

Person D has an enormous interest of history and mostly plays strategy games. Since he/she has this interest of history, this shows the importance of having different degrees of difficulty in the game.

9.1.5 Participant E

Person E finds it difficult creating a new user. He/she is trying to press on the text "Create new" and on already existing users. After a while he/she understands that it is possible to press on the plus sign to create a new user.

Person E finds it difficult to know where he/she should press for further navigation on the page where the TA is crying.

He/she doesn't have difficulties understanding the time machine and choosing an era to travel to.

He/she doesn't always read the texts carefully and therefore he/she finds it difficult to understand what he/she is supposed to do, especially after his/her own learning activity. Person E has problems understanding the purpose of the progress bar and also what "Added to trips" means.

When it comes to teaching the TA, person E doesn't know what to do in step 2 and 3. In step 2 he/she is trying to drag the whole box of fact to the timeline.

Person E plays free online games now and then. He/she thinks that the game is overall easy to play but sometimes it is hard to understand where you should click to navigate.

9.1.6 Participant F

Person F does quickly understand how to navigate further on the page with the crying TA by noticing that the mouse icon changes when it hover the speech bubble.

When it comes to the time machine, he/she would only use the drop down menu to obtain the era he/she wants to travel to. After a while he/she also says that one alternative would be that clicking on the desired era on the timeline.

Person F doesn't have problems understanding the purpose of the progress bar but finds it difficult to know what "Added to trips" means.

He/she doesn't have problems understanding the step 1-3 when teaching the TA but has difficulties knowing what to do in step 4. Person F is trying to drag step 1-3 to the brain instead of pressing "Add the steps in the time elf's knowledge bank".

Person F doesn't play that much games, occasionally free online games. He/she thinks the game is interesting and that the story is funny. He/she finds the navigation fairly easy to understand.

9.1.7 Merged results
Below is a summary of all participants' opinions;

- 6/6 of the participants are satisfied with the level of the game. It wasn't too hard or too easy.
- 6/6 of the participants like the idea of having missions in the game.
- 6/6 of the participants understand how to choose an era in the time machine.
- 6/6 of the participants have difficulties understanding the different steps when teaching the TA.
- 5/6 of the participants consider that it sometimes is hard to understand how to navigate in the game.
- 5/6 of the participants understand the purposes of the progress bar.
- 2/6 of the participants understand what "added to trips" meant.

10 Discussion

With the objective to create an educational module for history that uses teachable agents as a tool to both challenge and help students, whilst keeping them interested in the task at hand, we as a group had a daunting task.

Games with a TA have proved to be powerful and highly motivational, computer based, pedagogical tools that not only teach the student but also give research material on the topic of how students learn. Our game is constructed in three phases; information gathering and usage, teaching the agent or "creating a thought", and finally the test phase where the TA proves if he or she has learnt enough information.

In the first phase our greatest challenge was to present the given information about history in such a way that the users' interest wasn't compromised. Even though it's a game, it's still a game played for educational purposes and therefore must emit a certain level of seriousness

whilst providing something that's more interesting than an ordinary textbook. In this project we didn't prioritize the handling of the information because of the lack of time. What we did do was to organize and give the structure on how we would like to present the information for the user and how we thought it could be optimized in the given media. It's understood from the tests, and presentations that the story of the game is well received and that the overall structure is on the right track. We also acknowledge that with the proper animations and functionalities in the high-fidelity version the game will be a much better product.

The second and most difficult phase is the "creating of thought" scenario where the user teaches the agent about his trip back in time. By doing that he/she prepares the TA for a test from the professor. Because we knew from the start that this was going to be difficult, we tried to present the "creating of thought" process in such a way that it would be easy for the user to understand. In the first course our primary goal was to create an overall game concept. Unfortunately this had the consequence that we couldn't focus specifically on developing the learning process. This could be seen in the test results where the main problem seemed to be the understanding of "creating a thought". At that stage we hadn't either thought of a way to align the knowledge learnt by the TA, making it possible to connect thoughts to each other. The test results from the first course triggered the ideas of improving the "creating of thought" scenario. During the first week of the second course we summarized the most important deficiencies according to the test results. Several children had problems understanding how to enter the TA's brain and how to navigate inside it. The difficulties of entering the brain have now hopefully been solved by adding a more complete dialogue between the TA and the avatar where the avatar is guided how to proceed. During the dialogue there will be an animation where the TA makes notes to provide continuous feedback to the user that the learning process is continued. Before entering the brain the user is prompted whether he or she wants to continue the learning dialogue or not. Then, if the user chooses not to continue the learning dialogue but to enter the brain there will be an animation where the brain is enlarged representing an entering of the brain. The brain will be fully enlarged and transparent in the background while the user is helping the TA organizing it's thoughts to remind the user that he or she is inside the brain. The development of the learning process during the second course was described in detail in chapter 8, "High- fidelity prototyping".

The third and final phase is the test from the professor. Closely aligned with the second phase this is also something that hasn't come that far development wise. We are aware that this part will consist of statistics and rights and wrongs, presented in such a way that the user can go back to correct certain issues that the agent might have had with a specific question. We also know that it's important for future versions to present the agents "train of thought" in such a way that the user can profit from this knowledge and learn from it, seeing that this certain part of the learning process is a highly efficient one.

During our first presentation of our project, the question emerged if there had been any thought of implementing noise in the TA and the learning process. With noise the TA would randomly give wrong answers during the testing, even if it had been taught everything correctly, which the user could respond to, by stopping the TA and select the right answer instead. This would add another step for the pupils to pay attention and test their knowledge, while bringing the TA to a closer simulation of real world cognition (people very often make mistakes, even when they might have been taught the correct answer). While this is a valid point, especially considering that the user, in our version, would be quite inactive during the TA testing, we consider the implementation of noise in the TA would bring another level to the programming of the game, that would consume more energy and time than it would give benefits. But even though we choose not to consider it for our high-fi modeling, it doesn't mean that one could later add it to the game.

Throughout the project the working process and progress has been well documented and combined with weekly meetings both within the group and with our coordinators. This has given us great feedback on our work, which in return has hindered us from having setbacks. By catching any arising issues at an early stage, we had to spend very little time on "redoing" things we shouldn't have done. Notable is also that the mid-fi design has a lot of things in common with the original lo-fi design according their appearance. Although the mid-fi brings a bit more operations and functions to the table it's still very straightforward.

10.1 General discussion of the result

The results so far indicate that the overall experience of the game for the target group is satisfying. There are however some obstacles that have to be overcome, for example, to get users to understand how the TA should store its learned knowledge and use it.

The game was first visualized by a low-fi prototype and thereafter Balsamiq was used to create a mid-fi prototype of it. A lot of time and effort was put into the design of the mid-fi prototype to obtain a similar game environment as if the test participants were playing a real game. The focus during the usability evaluation was mainly to understand the purpose of the game and its story and thereafter how the flow of navigation in the game was perceived. Using mid-fi prototypes gave us quick and high quality feedback during the tests.

10.2 Methodological considerations

The result from the usability evaluation is based on information, discussions and thoughts from six test persons during usability testing at a school in Lund. As mentioned in section 3.2 the number of participants was selected due to information read in Jeffrey Rubin's "Handbook of Usability Testing" [11]. Jeffrey states that if your time is limited four to five participants is enough to expose the huge majority of usability problems. This is not enough participants for a statistically valid result but it represents a good basis for future work and studies. It was a good decision to use Balsamiq for our mid-fi prototype and for our tests. Balsamiq's presentation mode made the navigation in the mid-fi prototype very similar to the navigation in a real computer game. Therefore this gave the impression that the test participants quickly understood the scenario and how the functionality was intended to work in real life.

10.3 Future work

10.3.1 First course

During the first course, Interaction 1: Neuro modeling, Cognitive Robotics and Agents, the testing sessions and the discussion following the test made it clear that the following issues need to be dealt with:

The navigation – The test participants made it very clear during the tests that it sometimes was hard to understand how to navigate in the game. So for future work the navigation needs to be more obvious for the user. There shouldn't be any confusion for the user when navigating through to the different phases within the game.

The book "trips" – Most of the test participants felt that it was difficult to understand what the book "trips" purpose was. In the mid-fi prototype the book wasn't implemented which may have contributed to confusion for the participants. For future work the book needs to be more visible in the game, throughout the different phases. Also the correlation between "added to trips" and the book "trips" needs to be more obvious for the user.

Teaching the TA – All of the participants felt that the different steps when teaching the TA were hard to understand. Therefore these steps need to be more clarified in the hi-fi prototype and, if needed, a more detailed describing text for each step should be visible. One outcome might also be to replace one or more of the steps to make them more understandable for the user.

10.3.2 Second course

During the second course, Interaction 2: Virtuality and Cognitive Modeling, the work with the hi-fi prototype and the evaluation along with the nine learning mechanisms made it clear that the following issues need to be dealt with:

Dialogues with historical persons – The game would become even more interactive with the player if it is possible for him/her to have "dialogues" with historical persons when traveling with the time machine.

Sound implementation – Adding sound to the game would increase the player's attention. Hopefully this can be done further on.

Flash implementation – The hi-fi prototype is under development and is implemented in Adobe Flash. The game experience will increase significantly for the player with this implementation. Adding sound and animations would be highly appreciated by the target group.

The nine learning mechanisms – When continuing developing this educational game it is favorably to have Stellan Ohlsson's nine learning mechanisms in mind, for example having videos in the game displaying how the player shall perform the different task. These nine types of information will help the game to become as educational as possible for the target group.

The process of teaching the TA – During the development of this game it has become clear that it is technically difficult to implement the part of the game where the player shall teach the TA. If the thoughts are wrongly connected, how should the feedback be given to the player? This is something that the developers need to have in mind in the future when continuing on developing this game.

Expansion packs – The game could in the future be divided into different expansion packs, for example one with European historical events and one with Swedish historical events, or expansion packs with focus on different eras. One expansion pack could contain the possibility for the player to change the design and details in the time machine to make it more personal.

11 Conclusions

The key points of this thesis can be summed up as follows. Today there only exist TA based educational games in mathematics and natural science. The need and the demand of educational games with a TA in other subjects such as history are huge.

The test participants in the usability evaluation thought the game and its story were easy to understand. Even if the usability evaluation was performed on a mid-fi prototype, the participants thought the design and structure of the game were good.

During the testing sessions it became clear that there were some issues and they needed to be fixed for the upcoming hi-fi prototype. Especially the part where the user should teach the TA was hard for the test participants to understand. We also felt it was hard to know for them how this should be implemented. We also came to the conclusion that our test method of using Balsamiq for testing our mid-fi prototype was a good way to provide the participants with the most realistic feel of the game as possible.

When working with the hi-fi prototype, focus was set on the rework and improvement of the deficiencies that came to light during the testing. Since time constraints and limited knowledge in HTML5 and Javascript made it difficult to develop a functioning hi-fi prototype before the deadline, it was thought to be more useful to focus on the development of the underlying concepts. Especially the part of the game "teaching the TA" was in focus and new brainstorming sessions about its conceptual design were held.

During the development of the hi-fi prototype the technically difficult to implement the part of the game where the player shall teach the TA became clear. We found it hard to come up with good solutions on how the linking between the thoughts should be implemented technically to make them represent learned facts in a database like system, whilst still being easy enough for fifth graders to understand and use. There was also a problem about how feedback should be given to players, when they had linked them wrong. To try to test and solve these problems, a hi-fi implementation would have been necessary.

Adding animations and sound to the game will increase the game experience significantly and would be highly appreciated by the target group.

Finally, we came to the conclusion that it is hard to develop an educational game in history, where a TA is supposed to play a central part, since historical facts are not as easily reducible to abstract facts, that can be easily taught to a TA. Trying to overcome this obstacle was what took up a large part of the work for this project.

Bibliography

[1] Federation of American Scientists (2006), "Summit on Educational Games – Harnessing the power of video games for learning", [digital], Available HTTP: http://www.fas.org/gamesummit/ Resources/ Summit%20on%20Educational%20Gam es.pdf [2012-11-25]

[2] Johnson, D., Gardner, J. & Wiles, J. (2004), "Experience as a Moderator of the Media Equation: The Impact of Flattery and Praise", International Journal of

Human-Computer Studies, Volume 61, Number 3, pages 237-258
[3] Ndumu, D., Nwana, H. (1996), "An Introduction to Agent Technology", [digital], Available HTTP: http://webcache. google usercontent.com/search?q=cache:dlcWBAL4pxQJ: www.agen t.ai/doc/upload/ 200302/ nwan96_2.ps +&cd=1&hl=en&ct=clnk&gl=se [2011-11-25]
[4] Haake, M. (2009), "Embodied Pedagogical Agents – From Visual Impact to Pedagogical Implications", Dept. of Design Sciences, LU, pages 3-56
[5] Preece, J., Rogers, Y., Sharp, H. (2011), "Interaction Design: Beyond Human – Computer Interaction", 3rd Edition, Wiley
[6] Blair, K.P., Schwartz, D. L., Biswas, G. and Leelawong, K. (2007), "Pedagogical agents for learning by teaching: Teachable Agents", Educational Technology, Volume 47, Number 1, pages 56-61
[7] Gazzinga, M., S., Ivry, R. B., Mangun, G. R. (2009), "Cognitive Neuroscience: The Biology of the Mind", 3rd Edition, W. W. Norton & Company, New York
[8] Chase, C., Chin, D.B., Oppezzo, M., Schwartz, D. (2009), "Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning", Journal of Science Education and Technology, Volume 18, Number 4, pages 334-352
[9] Yee, N., Bailenson, J. (2007), "The Proteus Effect: The Effect of Transformed Self- Representation on Behavior", Human Communication Research, Volume 33, Issue 3, pages 271-290
[10] Chase, Catherine (2011), "Motivating Persistence in the Face of Failure: The Impact of an Ego-Protective Buffer on Learning Choices and Outcomes in a Computer- Based Educational Game", Stanford University
[11] Rubin, Jeffrey (1994), "Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests", 2nd edition, Wiley, page 128
[12] Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005), "The Uncanny Valley: does it exist?", 19th British HCI Group Annual Conference: workshop on human- animated character interaction
[13] Ohlsson, Stellan (2011), "Deep Learning: How the Mind Overrides Experience", Cambridge University Press, 2nd edition, page 202-204

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Metaphorical gestures, Simulations and Digital characters

Martin Viktorelius[1], David Meyer[2], Lisa Eneroth[3]

[1]Department of Philosophy and Cognitive Science, Lund University
[2,3]Faculty of Engineering, LTH, Lund University

A framework of metaphorical gesture comprehension and interaction is presented. The mechanisms behind the creation, activation and interpersonal exchange of metaphorical conceptualizations are spelled out. An experiment exemplifying a new methodological paradigm for investigating gesture comprehension was conducted showing that metaphorical gesture might have a low communicative value, indicating, based on their role in use, a more hidden functionality.

Introduction

Much has been written about when and why gestures are produced (Krauss, 1998; Kita, S., & Özyürek, A., 2003; McNeill, 1992) and if they play a cognitive role or serve a communicative function that isn't redundant given the co-produced speech (e.g., Krauss et al., 1995). A common theme in the literature is the embodied and situated nature of communication in which both speech and gesture form the relevant unity of analysis (McNeill, 2000). In a recent article, Hostetter & Alibali (2008) present a framework of the spatial-representational origin of gesture production after surveying a large body of evidence grounding gesture in the sensory-motor system. There also exists an extensive literature on how language is understood, in particular on how meaning is grounded in perception and action (Zwaan & Madden, 2005; Glenberg & Kaschak, 2002).

Gestures have been found to very often depict objects and places far away from the communicative setting, sometimes even referring to non-existent or abstract entities (Morford & Goldin-Meadow, 1997).

However, not much has been said about how people go about *understanding* all the frequently co-occurrent gesture during speech. How, if ever, could such bodily movements be interpreted by a viewer as meaning anything at all, without the speaker having to explicitly describe why and what is being gestured.

We propose a model of how this is possible by extending Hostetter & Alibali´s (2008) Gesture-as-simulated-action framework (GSA) to include both the producer and the comprehender, as these have been seen to very often co-structure each other in dialogue (Pickering & Garrod 2004; Garrod & Anderson 1987), and furthermore, by specifying a proposed simulation mechanisms used in multimodal communication based on Barsalou's (1999) theory of the conceptual system. The focus of our analysis is on *metaphorical* gestures as these exemplify a fundamental feature of the human conceptual system, viz., the mapping of content between different epistemic domains (Lakoff & Johnson 1980), and play an important role in learning and teaching (Williams, 2008; Núnez, 2008; Roth & Lawless, 2002).

The paper is divided into three main parts; The first (I) sets the ground for the model of gesture comprehension presented in the second part (II), and the third part (III) describes a valid methodological paradigm to experimentally investigate gesture comprehension and report the results from a pilot study on gestural conceptualization.

Multimodal communication

Do gestures communicate?

Historically there have been a number of types of scientific approaches to gestures. These bodily movements have either been treated simply as involuntary motor leakages and thus not related to language, or as partly conscious and facilitating accompaniments to genuine language processes, or as primitive linguistic attempts, a kind of *lingua franca,* or finally as a proper part of the full repertoire of human communicative resources (Haviland, 2000).

The question of whether gestures communicate is usually divided into two separate questions; Do speakers gesture intentionally?, and, Do listeners benefit from gestures? The questions can be treated independently, and a negative answer often means that gestures don't facilitate the immediate communicative exchange (Hostetter, 2011). This, of course, doesn't mean that gestures, especially if they are frequently occurring, cannot serve more indirect functions, or not have subtle effects.

In this paper we concentrate on metaphorical gestures, which have not been observed to benefit the viewer/listener significantly in communicational comprehension. According to Hostetter (2011), the reason for this might be that metaphorical gesture, as opposed to other kinds, are not that frequent. But, even if that is the case (which seem wrong if you think about how often people express the conduit metaphor by making an open palm gesture when talking, se section 6), the articulated and attention attracting nature of those gestures (Cienki & Müller, 2008), in an otherwise attention limited mind, indicate that some non-obvious functionality is to be found. In this paper, we will try to argue that metaphorical gestures might play an important role in the communal exchange of conceptualizations of the world around us.

It is admittedly partly a matter of definition whether gestures form a substantive part of language use and communication. Neither a too inclusive, nor a too restrictive notion of communication will prove any valuable points. Hence, "whatever has an effect on the listener" would be a bad definition of the function of communication. If the effect is unintentional, i.e. not based on a conscious will, then it might be considered as non-constitutive of a communicative endeavor. However, evidence has shown that people often automatically align to each other in dialog, facilitating the ease with which interlocutors communicate (Pickering & Garrod, 2004). We therefore consider a definition of communication along traditional and broadly Griceian (1989) lines, as an activity of volitionally or intentionally inducing a particular mental state in the recipient that recognizes this intention as such, as a bit too strong.

There is, however, evidence that gestures, in some settings, are produced intentionally, i.e., based on a conscious effort to communicate. Melinger & Levelt (2004) used a picture description task in an experiment with the purpose of determining whether iconic gestures (se section 4. for a definition) produced during speech were made with a communicative purpose. They defined the necessary minimal content of a message and measured if gesture, on occasion, took over parts of the communicative burden. As predicted, participants sometimes omitted expressing linguistic information in favor of gestural expressions. They concluded that people sometimes gesture intentionally.

We therefore think that gestures could be produced and comprehended along a continuum of conscious acts, ranging from fairly automatic and unconscious to quite aware.

Not everyone has been convinced that gestures do communicate, i.e. benefit the viewer. Kraus et al. (1995) devised an experiment in which participants were exposed to subjects that gestured spontaneously in order to see if it enhanced the communicative effectiveness of the message, in comparison to a control condition without gestures, which they found it didn't. However, Krauss and colleagues committed the obvious fallacy of thinking that reference exhausts the semantic realm of communication, by ignoring narration and free conversation, in their argument against the communicative value of gestures in general. The results of their experiments on the semantic function of gestures shows that the ability to pick out the right referent after just hearing a speaker describe it, isn't affected by the occurrence of naturally accompanying gestures. Based on these results they draw the fallacious conclusion that gestures don't serve any communicative function.

But ever since Frege's (1892) work it has been clear that *sense* and *reference* isn't the same thing. A referent can be presented in many different modes each with different cognitive significance. Hence, Hesperus and Phosphorus both refer to Venus, but by means of different expressions. This means that even if observing gestures doesn't change the ability to pick out the correct referent they might changed the mode of presentation, or conceptualization, of the referent. And there actually was a small statistically non-significant effect in one of the gesture conditions in Krauss et al.'s (1995) results suggesting this.

Furthermore, the modality of manual acts, including representational gestures, has the potential to *directly* depict a referent, whereas a single term or description is only arbitrarily, i.e. by convention, hence indirectly related to its referent. Gestures represent things spatially, verbal representations are temporal and sequential (see section 4.). The fairly simple entities communicated in Krauss' et al. (1995) study could have been too easy to describe verbally. The otherwise allegedly disambiguating function of gestures was not necessary for communicational success in their experimental setting with its low ecological validity.

We thus tentatively conclude that gesture might communicate, at least on the level of conceptualization or cognitive significance, and not always as a result of a conscious effort. We also suggest a distinction between research questions regarding comprehension of gestures, specifically metaphorical gestures, and investigations of *if* or *when* gestures have a communicative value, i.e. add some non-redundant information.

Interaction, gesture & alignment

Automatic, distributed and embodied processes in communication

An important distinction to make when analyzing the role of gestural communication is one pointed out by De Jaegher et al. (2010). Although they suggest a very general analysis of the role of interaction in social cognition it can still help us with our purpose of specifying the role of gesture in the understanding of what an interlocutor tries to convey, which is the aim of this section.

Research about social cognition, they argue, should take into account three different degrees in which interaction could be related to a social phenomenon, such as learning. First, interaction can be a *contextual factor* of a social phenomenon, meaning that changes in interaction produces variation in the phenomenon; second, interaction can *enable* a social phenomenon, i.e. be a necessary condition for the social phenomenon to occur; third, interaction can be *constitutive* of a social phenomenon, be part of the process of the phenomenon itself. Translated to our subject matter one should ask whether gesticulation sometimes function as (a) a contextual factor, (b) an enabling condition or (c) a constitutive part of a communicative act. Condition (a) is arguably often satisfied; gestures can change the interpretation of verbal content, as for example when a person says "the fish was quit big" simultaneously holding two hands with the palms facing each other at some particular distance indicating the seize of the fish. A gesture in this situation could certainly direct the interpretation of the speakers fishing luck in either direction depending on the interpreters default understanding of "quite big". So, the question is rather if conditions (b) or (c) are satisfied in some cases. We think that there is evidence that they are.

Goldin-Meadow (1999) describes findings indicating that deaf children have the ability to spontaneously develop a gestural system in order to compensate for their lack of other communicative channels. Toddlers not yet in possession of a language use gesture to communicate their needs and children beginning to speak single words combine those with gestures, creating complex expressions not yet occurring in speech alone. Gestures have also been seen to play a crucial role in many learning settings, both in subjects teaching (Williams, 2008; Núnez, 2008; Roth & Lawless, 2002), and pupils solving problems (e.g. Goldin-Meadow et al., 1993a). In the latter case conceptual mismatches between speech and gesture during the learning phase of children, where the correct answer is expressed in gesture but not in speech, indicate that multiple ideas can simultaneously be entertained and conveyed during task engagement, and importantly that instructors and teachers have the possibility to guide pupils in a transition period towards procedures leading to solutions that they already implicitly understand (Goldin-Meadow et al., 1993b).

This dual level of meaning will later (section 5) be seen to bear a crucial role in the instantiation of metaphoricity in speech and in speech accompanying gestures.

Mental convergence through metaphorical gestures

Interactive alignment is the unconscious joint activity resulting in the paring of representation during communication. It is a distributed (i.e. involving mechanisms outside individual subjects) and dynamic (i.e. self-autonomous) process that facilitates conversation by aligning the interlocutors' individual situation models, or mental representations, Zwaan & Radvansky (1998), that are used to interpret the semantic content of the alternating speaker. This process starts at the level of phonology, making interlocutors pronounce word similarly, and continuous through syntax and semantics to the situation model.

Interactive alignment at the level of conceptualization, or cognitive significance, we propose, concerns the process by which interlocutors strive to establish a common way of understanding or conceptualizing the world. This takes place on a higher conceptual level than reference agreement and relates to a more profound form of intersubjectivity.

We think that metaphorical gestures, being expressions of concepts and abstract processes, do serve an important function in this process.

We propose that *common ground* (Brennan & Clark, 1996) and *alignment* (Pickering & Garrod, 2004) could be created through the use of speech-accompanying gestures. The reason for believing this is that situation models can be constructed by multimodal means; they can be based on pictures, verbal description or written text (Zwann & Radvansky, 1998). Secondly, Kendon has shown that "interactional synchrony" do spontaneously occur during conversation, meaning that the bodily movements of subjects conversing reciprocally respond and are sensitive to each other (1970).

If interactional alignment is correctly characterized as a 'mutual coupling constituting an autonomous self-sustaining organization', to use De Jaegher's definition of interaction, and gesticulation constitutes a valid level in the process of alignment, then gesticulation is an interactional constituent of communication. This means that gesture is not just a peripheral or contextual phenomenon, it *is* social cognition. However, communication, as every other complex cognitive feat, is most probably dependent on individual mechanisms and representations inside the agent as well as interactive and externalized processes.

What is then, a metaphorical gesture. The next section considers some of the most established ways of defining the different types of gestures.

Metaphorical Gestures – A definition and taxonomy

Gestures can be divided into two categories; those that represent or refer to some entity(s) in the producer, such as a thought or emotion, or in the environment, and those that doesn't represent or refer to anything at all. McNeill (1992) identified four types of gesture that can be placed in these categories; *Iconic gestures* are representational and often depict some object refered to in speech, e.g. speaking of hammering while making an as-if hammering movement with one hand. *Metaphorical gestures* are like iconic gestures but represent some abstract entity, like a concept, idea or process. *Deictic gestures* refer to the object pointed or noded at and could be made in conjunction with demonstrative expressions such as 'that one'. *Beats* are non-representational and figure most often as rhythmic pulsations or syntactic articulations.

There is a type of gesture that isn't iconic, like the metaphorical gestures defined above, but still expresses a metaphorical or figurative content; *Abstract deictic gestures* function as pointing gestures, but do not refer to an object in the immediate environment, or any actually existent object at al. These gestures can have a metaphorical content if their imagined referent represents some other entity in a metaphorical manner (se section 5., below for a discussion of metaforicity).

Based on Kendons (1982) initial gradation of gestures, ranging from sign language to spontaneous gesticulations, McNeill (2000) devised a contiuum in four dimensions, forming analytically differing relations and properties inherited by all gestures. According to him gestures differ in their relation to (1) speech, (2) linguistic properties, (3) conventions and (4) semiotic character. The first dimension concerns the co-presence of speech and gesture; whether the gesture could be meaningfully produced without a verbal accompaniment. The second dimension is about systematic constraints on the gesture, stemming from syntactic, semantic or pragmatic rules (e.g. if it can be compositionally combined with other gestures). The third dimension of the contiuum, related to potential pragmatic rules tapped into by the previous dimension, deals with socially established standards in a broader context; such as when a gesture is conventionally accepted. Some gestures might for example be inappropriate in some situations. The final dimension actually concerns two features; whether the meaning of the gesture is *globally* or *locally* determined, that is, if its meaning results from a combination of individually meaningful parts or from a holism of content, with the whole gesture being the only meaning determining level. The other property is attached to *synthesis* contra *analyticity*; a gesture is synthetic if its meaning could be complex without its vehicle (e.g., sentence or actual hand movement) being complex too.

Now when we know how to distinguish metaphorical gestures from other kinds it is crucial to understand how gestures can be metaphorical in the first place. Another question, related to our claim of conceptual alignment caused partly by metaphorical gestures canvassed above, is how this process of metaphoricity is to be understood. The next section is dedicated to these topics.

Metaphors as grounded in a dynamic cognitive activity

Müller (2008) presents a theory in which metaphor is understood as a multimodal phenomenon that is constantly being re-created in thinking, communication and interaction. She denies the traditional distinction between dead and alive metaphors, i.e. the distinction between metaphors that have lost their initial non-figurative meaning and become literal, and those still expressing a figurative-through-non-figurative meaning. The denial is based on a separation of the level of a linguistic and collective conceptual system from the level of individual use. The latter level, as opposed to the former, is always transparent in the sense that a dualism of meaning is being actively perceived by the producer, owning to a dynamic mapping between a source domain (concept, meaning,) and a target domain (concept, meaning) that is taking place through a *seeing-in-terms-of*. Metaphors on the individual level can be more or less conventionalized; both novel and entrenched ways of conceiving the world fit within this dynamic cognitive framework. Müller's emphasis on the structure and organization brought onto language through its *use*, is part of a more general claim that *communication* isn't just the "defective instantiation of whatever system" (Ibid., 14), rather it constitutes a major part of the system.

Metaphoricity ensues, according to Müller, when one thing is seen in terms of another and hence two concepts or epistemic domains are active at the same time, with one concept or domain projecting elements of meaning onto another. This active establishment of metaphor implies a *triadic structure*, as opposed to a merely dualistic mapping of meaning. The structure involves an entity or process A, which relates two entities B and C, such that C is seen in terms of B. Müller cites Lieb's extensive work identifying different concepts of metaphor since Aristotle and concludes that the triadic structure is the common denominator in all the studied accounts. When the active mapping become conventional, or part of a standard way of conceiving a thing, then the metaphor surfaces the system level, and the convention, expressed through e.g. a verbal expression, takes the role of relating the source- with the target domain. Moreover, nothing prevents all sorts of expressions to stand in for the active and in different degrees conventionalized mappings, such as gestures, pictures, lexeme or idiom. This “sedimented product of collective processes” then guides “the individual process of seeing-in-terms-of along a collectively paved path” (Ibid., 114).

The framework of the triadic structure accounts for the process of metaphor creation. It starts within an isolated mind, actively mapping two domains of knowledge, and goes on to the collectively recognized uses of metaphorical constructions. But the framework does also implicate an interaction between the levels (system and use) and is open for the possibility of reciprocity of projection between the source- and target domains. Furthermore, as was described in section 3., in the process of alignment, interlocutors conform to each other in order to converge on a conceptualization. Systematization of metaphor could be considered as a culturally extended alignment.

Müller (Ibid.) focuses, as most linguists interested in the study of verbal metaphors do, on the production instead of the comprehension of metaforicity. But given Müller´s emphasis of the fact that metaphor, on the individual level at least, is a matter of cognitive activity in the speaker *or* listener and from the claim that “…it takes what is *inter*personally salient to be salient *intra*personally”, it follows that a direct focus on the comprehension side is necessary for an proper examination of metaforicity.

Müller speaks primarily of verbal-gestural metaphors, not about gestural metaphors *per se*. But this form of metaphoricity, we suggest, should also be acknowledged (cf. Cienki, 2008). The purely gestural metaphor could be the result of an activation of a triadic structure on a sub-linguistic, i.e. primordial conceptual, level. The verbal context is of course of great importance for the interpretation of the gesture; the content directs the viewer's inner attention into a situation model (se section 3.) in which a gestures metaphorical meaning can be made sense of. A non-metaphorical verbal context, e.g., with the subject matter being cancer, triggers, we propose, mental imagery or simulation of the concept, which then easily fuses with the simulation of the images schema expressed in the gesture.

The empirical evidence put forward by Müller for her claim that metaforicity is an active mapping between two domains that can be realized in different modalities comes primarily from a case study in which a person describes a particular memory of her first boyfriend. Now, in speaking of him as being “clingy”, she simultaneously opens her palms and touches each other repeatedly in a slow manner as if having something clingy in between (2008:100). This and a number of other examples from the case indicate that the subject understands and tries to communicate the character of the relationship in metaphorical terms with the help of gesture.

Now, it could be objected that metaphorical gestures are just a peripheral phenomenon, and that they therefore probably doesn't play such an important role in the communication of abstract content. The next section tries to give some examples of the usual scenarios and frequencies of metaphorical gesture use.

Metaphorical gestures – their use in communication and learning

Here we present empirical findings of how and when metaphorical hand gestures are used. We then try to argue that the reported gestures should be situated within a broader framework of whole bodily enactments grounded in inner mental imagery.

Chui (2011) makes a case for the prevalent use of metaphorical gestures by giving some examples from a transcribed face-to-face conversation on the topic of tea procedures and love relationships. He takes the evidence to show that metaphoricity emerges as a dynamical process and that gestures often complement the co-occurring speech by expressing additional information about “the speaker's focus of attention in real-time multimodal communication”. The types of metaphorical gestures that are reported relate to five different ways of conceptualizing different aspects of a tea procedure and a love relationship. The common feature of these conceptualizing gestures is that they all allegedly express a conceptual mapping between an abstract concept and an experiential component. The following is a short summery of the types of metaphors expressed by gestures reported in Chui's data; OBJECTIFICATION OF NON-PALPABLE ENTITIES such as love relationships or tea procedures: the gesture in question assumes a posture as if holding something concrete in the hand; PERSONIFICATION OF THE PROPERTIES OF AN OBJECT: a gesture of hitting and punching expressing the fact that the object in question endures “abuse”; SPATIAL ORIENTATION AS THE CHARACTER OF A SOCIAL RELATION: the dropping of a hand indicating being “bogged down in a mess”; SPATIAL RELATION AS TIME: pointing to an imaginary timeline indicates a spatial understanding of the concept of time; WEIGHING AS COMPARING ALTERNATIVES: two open palms on each side of the speakers body going up and down is if balancing something.

The existence of bodily (the whole, not just the hands) expressed metaphors explains why seeing good acting in the theater or in the movies can give aesthetic or poetic pleasure; the actor is able to instantiate multimodal metaphors by his or her actions and bodily expressions either through a mapping between the immediate content of his or her speech and the conveyed emotions and experiences found in the bodily expressions, or between a mapping between a larger thematic or situational meaning and the concrete instantiation of bodily shapes and movements. This complexity or multi-channelization of metaphor instantiation makes the aesthetic experience particularly profound.

Another example, from the literature on embodied conceptualization, that serves as a reason for thinking that the whole body expresses ways of understanding and representing knowledge, comes from (Bargh et al., 1996) showing that words priming the elderly stereotype make participants walk slower.

Cienki (2008) points out that metaphorical gestures can vary along a continuum of conventionalization; from culturally sedimented ways of understanding (e.g., the conduit metaphor as expressed by a cupped or flat hand with palm up mapping containers onto ideas) to spontaneous expressions of structural similarities between domains (e.g., understanding degrees of dishonest behavior as different levels in a vertical space expressed by having one hand with the palm facing up move downwards and the other hand facing down moving upwards, Cienki, Ibid., 12). The continuum often goes with degrees of explicit awareness. The culturally sedimented metaphorical gestures are often unwittingly made, whereas the novel gestures require a certain level of conscious thought.

Roth & Lawless (2002) conjecture, based on a reanalysis of a database on student learning, that "gestures have an important scaffolding function in students' development of scientific language…" (Ibid., 302), and that "students use metaphorical gestures to represent abstract concepts and processes" (Ibid., 299). Students from different countries and in different grades explained and described theoretical entities such as *force, electron, energy* and *mechanical systems.* In doing this students often represented the abstract entity talked about by gestures (one boy, for example, made a statically shaped hand gesture that he moved in one direction and talked about the way the force is going. This expressed the idea of changing velocity contra static force in kinetic energy). Furthermore, they argue that these students used their common perceptual environment to situate representations of non-palpable theoretical constructs and that these perceptual and indexical pointers also anchored the metaphorical gestures that were made in order to "free up the verbal modality to focus on constructing new conceptual statements…" (Ibid. 302).

Two related observations made by Roth & Lawless (2002) was that metaphorical gestures, registered over the time course of the class, were often used before the corresponding scientific vocabulary or linguistic conceptualization got established and used, supporting the finding described in section xxx, of a dual conceptual representation during a transitional phase of learning reported in (Goldin-Meadow et al., 1993b). The second observation concerns the fact that the students subsequently established a *common ground* (cf. Clark & Marshall, 1981) in their environment and gesture, representing the concepts discussed and constituting a more tangible conceptual platform from which they could proceed in their elaborations.

Embodied language understanding

After this lengthy review of the function and use of metaphorical gestures we have now come to the point where it can be asked *how* these gestures are comprehended. In this section we allude to some previous accounts of the embodied and embedded nature of language comprehension, in particular the understanding of fictive, abstract and metaphorical expressions. The following section will foreground the later proposed answer to the above question.

Do people undergo mental imagery or simulation during metaphorical language comprehension?

Before answering that question we first need a terminological clarification. It has been suggested that we need to distinguish between different notions of the term 'embodied cognition'.

Wilson (2002) uses the terms *on-line* and *off-line* cognition to point to the fact that the intentional object (the thing thought about) of a mental processes can be either present in the immediate environment or out of reach, perceptively speaking. Very roughly, in on-line embodied cognition, the ongoing mental processes are causally dependent or constitutive (see Shapiro 2008 for an overview of different kinds of cognitive relations between the mind and body) of the informational flow of input and output, trough perception and action. In off-line cognition, on the other hand, the situation does not contain information that could facilitate the mental process in question, e.g. in planning. But this does not prevent the mind to simulate the situation or bodily state in order to carry out a similar computation as that during on-line cognition. Communication, i.e. production and comprehension of language and gesture, incorporates an example of this distinction; the referent of a discourse can either be present, as in demonstrative reference, or absent, as in figurative or metaphorical narratives.

Richardson & Spivey (2007) argue for the embedded nature of reference understanding in conversation; they review their own and others' studies showing that we tend to externalize processes of comprehension by using, i.e. looking and pointing at, entities in our environment in order to index our memory, parse syntax effectively and create a common ground in conversation. An interesting example from their survey, from (Matlock, 2004), is the finding that the reaction time for responding to statements, framed in figurative motion terms ('crossed', 'follows', 'runs'), about stories involving traveling, depended on the distance, rate and terrain described. A long, slow and rugged travel story gave rise to longer latencies. Furthermore, between story types (e.g. fast vs. slow) the differences in reaction time were not significant when the statements were phrased in literal terms, indicating a genuine effect of processing figurative linguistic content. (Matlock, 2004) take this to show that linguistic content indicating fictive motion or action is comprehended through a "dynamic visual-spatial simulation" of that action or motion.

Richardson et al. (2003) demonstrated in two experiments a significant response interference effect in subjects that were comprehending a sentence with either a concrete (e.g. 'pull', 'point', 'smash') or an abstract verb (e.g. 'argue', 'give', 'warn'), and at the same time being presented with visual stimuli with a contradictory image-schematic content. The reaction times were longer in trials where sentences like 'The husband argues with the wife', (in which the verb had been seen to express a horizontal image-scheme, in a previous norming study) were combined with schematic pictures illustrating a vertical relation. This indicates, they argue, a perceptual-spatial representation of verbs, as opposed to an amodal and decontextualized one.

Gibbs (2006) describes studies purported to show that prompted mental imagery connected to verbal metaphors provides evidence of embodied simulation. The hypothesis being "that people's mental images should provide evidence of the embodied character of metaphor understanding, even when the actions referred to are impossible to perform." (ibid, p. 445) In the first experiment participants were asked to imagine performing actions such as grasping a concept and were then asked to answer some questions about their imagery. The results indicated that the participants´ imagery in fact was constrained by the assumed underlying conceptual metaphors, which means that the imagined abstract actions, e.g. grasping a concept, were being understood in terms of embodied knowledge. In a second study participants watched, imitated or imagined performing the actions mentioned in the first study, to see if this affected the understanding of the metaphorical expressions used. As expected, participants did increasingly make reference to aspects of bodily actions in their reported understanding of the abstract concepts and related actions.

Gibbs draws three conclusions from these results; first, that interpretation of metaphorical expressions referring to abstract actions is based on embodied simulation, second, understanding of metaphorical expressions is situated, and third, understanding of metaphorical expressions is multimodal.

However, the obvious problem with the experiments conducted by Gibbs concerns the demand characteristics; that the participants were asked to explicitly imagine performing the abstract actions by directly being exposed to words stating the physical actions thought to underpin the abstract ones. It is therefore very likely that the participants felt compelled to answer the questions in a way that mentioned the physical action, by understanding the assignment as one of elaborating on that tangible theme. However, the crucial predictions concerning the differences between the metaphorical and non-metaphorical conditions, i.e. increases of conceptual descriptions using source-target mappings, as opposed to descriptions of the bodily movements *per se*, were satisfied.

Another set of studies conducted by Gibbs (2006) tested whether performing or imagining doing different actions facilitates metaphor comprehension, i.e. speeds up the reports of language understanding. The results showed that indeed it did; imagining actions congruent with metaphorical content facilitates abstract language comprehension. In a third set of studies Gibbs showed that embodied interpretation of metaphorical narratives even affect the way participants moved their own bodies after hearing different types of stories involving either metaphorical or non-metaphorical content.

Embodied gesture production

The theme has been implicit from the beginning but in this section we will present some accounts and reasons for the embodied nature of gesture production.

Every theory of gesture comprehension or production should consider both the role of gesture in social interaction as well as its role in thought, or 'thinking-for-speaking' (cf. Slobin, 1987). The reason is simply that the neuro-cognitive processes of speaking, or language production, are presumably similar to the processes of language comprehension, although executed in the reversed order. (Gazzaniga et al., 2009:397,420) And therefore, if gestures play an important role in utterance understanding and interaction, then their role should at least be accounted for in speech production as well. Secondly, discursive mechanisms (se section 3. on interactive alignment) imply interdependences between production and comprehension, action and thinking, the private and the public.

A number of theories have been presented as accounts of the cognitive role of gesture production.

In McNeill's (1992; 2000) Growth Point theory and Kita's (2000; 2003) Information Packaging Hypothesis "[g]estures together with language, *help constitute thought*" (Kita 2000: 163. Originally in McNeill 1992: 245; his italics). Kita takes McNeill's theory as his starting point and defends his view against two competing alternatives; the essential difference between these latter theories and Kita's being that they emphasize a unidirectional, whereas Kita argues for a bidirectional relationship between the, what he terms, "spatio-motoric", or imagery-based, and the "analytic", or propositional ways of thinking or representing information. They all claim that cognition that is contingent on perception and action (e.g., mental imagery), and that is expressed by gestures, influence propositional or linguistic representations. Unfortunately it is outside the scope of this paper to argue that Kita's evidence of cross-cultural variation in gesture production in relation to particular linguistic constraints, really supports his theory, although we think it does. However, the take home message from the debate certainly is the link between imagery, gesture and speech. Furthermore, the probability of this link can be strengthened if we consider the nature of the relata, in particular the "analytic thinking"-side. Kita dose not characterize it substantially but claims that it is independent of context and some specific perceptual modality (Ibid; 164). However, the view of conceptual and linguistic representation as *a*modal and *de*contextualized have forcefully been argued against by a number of psychologists, cognitive neurologist and philosophers (e.g. Hesslow, 2011; Barsalou, 1999; 2008; Prinz, 2002; Glenberg; 1998). They argue that conceptual knowledge, and meaning in general, is grounded in perception and action.

Hostetter & Alibali´s (2008) framework for studying the production of gestures is based on three main fields of research. The first source of evidence comes from studies pointing to the intimate coupling between perception and action; the relation between perceiving the environment and acting in it is one of mutual dependence. This dependency can be seen on both a behavioral, as well as on a neurological level. Furthermore, cognition, in particular conceptualization, is thought to be constituted by the sensory-motor system in the brain, underlying the perception-action coupling. The constitution explains, amongst other things, the ability to think about objects that are absent from the immediate environment in which the thinker is placed, without postulating any abstract symbols representing those object. The second source of evidence comes from research showing that language understanding is in fact embodied, i.e. grounded in perceptual experience. People use their perception to index the referents of their communications when the referents are present in the environment, and use perceptual simulations when they are not. This goes both for concrete object, such as chairs, and abstract object, such as love relationships. Action is also involved in language interpretation in as much as it is simulated in the mind, even though not necessarily conscious, when hearing utterances such as "close the drawer"(Glenberg & Kaschak, 2002). Ones again these findings are seen on many levels of cognitive implementation. The last source of evidence comes from studies about mental imagery showing that imagining perceiving things, e.g. imagining transforming them in some way or another, as well as imagining performing some action without actually doing it, is made feasible by a mental simulation of that would-be perception or action.

The authors then point out the close affinity between gestures and mental imagery; the former being a natural by-product or manifestation of the latter and in having the same global and synthetic semiotic properties (se section 4), as well as mnemonic and attention directing functions.

According to the GSA model "gestures emerge from the perceptual and motor simulations that underlie embodied language and mental imagery" (Ibis., 502). The activation of a gesture depends on three factors; (1) the strength of activation of the simulated action; (2) the height of the speaker's current *gesture threshold*; and (3) the simultaneous engagement of the motor system for speaking, i.e. the facilitatory effect of an active motor program for speech.

The GSA model can be improved in order to, (a), fit a communicative framework that,(b), focuses on metaphorical mappings in gesture and speech, and (c) spells out in detail the mechanisms of simulation and integration in metaphorical mappings. (a) is important to account for because the level of dialog brings crucial changes to the account of language comprehension and production (se section 3) , and (b) is important because it prioritizes the alleged central and widespread role of metaphors in the conceptual system (Lakoff & Johnson, 1980; 1999). (c) is important because specificity often translates to testability.

The next section tries to live up with these possible improvements.

A framework of metaphorical gesture comprehension

Verbo-gestural metaphors are, as we saw in section 5, based on a triadic cognitive structure. This structure, in particular, the mapping between domains is, we suggest, implemented in a mental simulation. The simulation can give rise to a gesture depending on the strength of the simulation, the level of the threshold and the engagement of the speech system (se section 8) The interpretation of a metaphorical gesture basically reverses the mechanism.

Metaphorical gestures, we suggest furthermore, are involved in discourse processes such as alignment, but do not necessarily involve mimicry of particular representational gestures. Rather, gestures, through their representational expressivity, have the power to effect an alignment on the level of the situation model.

We would now like to show that Barsalou´s perceptual symbol system theory (1999) readily explicates the mechanisms behind metaphorical gesture production and comprehension, in detail.

Simulation theory and multimodal conceptual metaphors

According to Barsalou (1999) the conceptual system's most basic elements are perceptual symbols, which are modal and analog neurological features extracted by the system from the perceptual states experienced by a subject. A perceptual symbol is a highly schematic record of some property of those perceptual states with that property, e.g. a particular shade of red. These symbols are organized into frames that form clusters of different symbols extracted from entities (mental or environmental) of the same type, encountered previously. Seeing for example cars everyday results in a frame of all the different perceptual features normally connected with experiences of cars. Later on, after a frame of some entity have been created, the system is able to simulate the content of the frame and combine the perceptual symbols in different ways in order to imagine unperceived events or anticipate future behavior. In this way countless different *conceptualizations* of one and the same concept is made possible by the simulations performed in the system.

A simulator, furthermore, is according to the theory constituted by perceptual symbols from different sensory modalities as well as introspection. The concept of e.g. fighting most probably contains perceptual symbols from vision (the swinging of fists, the facial expression of the adversary etc.), from audition (the sound of punches etc.), from proprioception (the action scheme related to fighting such as the bodily posture assumed in situations that require self defense) and introspection (the feeling of anger or fear etc). Perceptual symbols from all these experiential domains bring content to the concept of fighting. It might be objected that people are able to possess a full-blown concept of fighting without having ever been into a proper fight oneself. This fact is accounted for in the theory by the productivity of a conceptual system enabled by the schematic properties of perceptual symbols.

Now, the simulations from different domains can be combined in mental imagery, even if they are not related experientially. You might e.g. fuse experiences from seeing fights on TV with your own memories of anger and sensations of hitting inanimate things. On this account a conceptual metaphor, or a novel metaphorical mapping between domains, can be cashed out as a simulator that combines perceptual symbols from very divergent experiences. An active, i.e. simulating, simulator of the notion of time, for example, which often combines an abstract concept with a concrete perceptual experience of horizontal plane is, on this account, not surprisingly often accompanied by a metaphorical hand gesture that moves from left to right depicting the time line simultaneously referred to in speech (Calbris, 2008, for an interesting analysis of the left to right gesture).

To share your understanding of a concept you don't have to cause a replication of the concept in the other. This would on Barsalou's account be fairly impossible due to the fact that a concept is equalized with a simulator, i.e. a frame of all the extracted perceptual symbols related to the concept and a mechanism for endless combination of those symbols. Instead, by expressing your current simulation, or occasional conceptualization, of the concept through language and gesture you externalize it and, if successful, induce a simulation of that conceptualization in the other.

How frames get integrated during communication

A simulation of the embodied knowledge of a source domain of a conceptual metaphor is not sufficient to generate understanding of a metaphorical expression, be it verbal or gestural. Based on the activation of the source domain, a link has still to be established that activates the target domain in parallel. The link, or fusion, does not necessarily have to be explicit, rather it can made through a fairly automatic additional simulation, the basis of which is another simulator representing the target domain. Both simulators are composed of frames integrating the perceptual symbols extracted during perceptual encounters with instances of the concepts, or experiences. According to Barsalou (1999), frames often integrate multimodal, including introspective and proprioceptive, perceptual symbols into a simulator. This readily explains the possibility of metaphorical projections from one type of embodied experience, e.g. grasping of an object, to another domain such as UNDERSTANDING. If the simulator for this latter concept already integrates perceptual symbols (from vision, proprioception and introspection) extracted from grasping-experiences, then, given a conversational context in which a grasping gesture is executed, the simulation of such an experience is very likely to trigger the simulator for UNDERSTANDING, that subsequently simulates its content. Thus, resulting in comprehension of the metaphorical content of the grasping-gesture.

Frames are roughly equivalent to predicates, which allows them to be productively combined by combinatorial and recursive mechanisms. Frames can stand in different dependency or framing relations; some concepts are necessarily understood against a background of other concepts and cannot be identified independently of them, other change as a result of changes in the network of concepts within which a particular concept is situated, even though it could be identified on its own. Furthermore, every word of a language is associated with a component of some frame, e.g. the auditory and visual surface of the word "car door" is associated, through a frame, with the simulator for CAR, and in particular with the sub-region of the frame corresponding to perceptual symbols of car doors. Because of this, productively construed verbo-gestural utterences can be understood by simulating, i.e. combining and transforming, frames that interpret the content of the utterance.

Interpretation of both novel and entrenched combinations of frames and perceptual symbols could readily be explained by Barsalou's framework. This could also be seen as the cognitive ground for production and interpretation of both novel and entrenched verbal and gestural metaphors. Take for example the utterance of the sentence "our relationship felt as a smelly soggy blanket" together with the iconic gesture of holding something repulsive in one hand, e.g. by making a down-pointing pinch gesture away from ones head direction, as if holding something with a bad odor. The interpretation of this multimodal complex (verbal metaphor and metaphorical gesture) could be understood as being partly constituted by a mapping, in a simulation, of the abstract concept of a love (or lack of) relationship, and the experience of encounters with smelly soggy blankets, wherein which aspects of these two domains are combined to form a simulator of really unhappy relationships. In enacting such an encounter the speaker specifies certain aspects of it that would not explicitly have been there in the communicative message without the gesture; aspect of how you often actually go about when confronted by such an object. This information could not have been communicated without a longer verbal description, but was instead, in this fictional example, conveyed through a global and synthetic (cf. McNeill, 2000) metaphorical gesture. Being guided by this extra-linguistic information, the listener-*cum*-viewer is prompted to simulate a completely novel combination of frames and understands as a result of this a novel way of conceptualizing an entity in his world.

The above framework is consistent with Ferriera et al.'s (2008) proposed model of multimodal integration of language and vision as well as theory of mental representation and memory. In this theoretical model conceptual, spatial, visual and linguistic representations form a network of tightly connected nodes, that can trigger and support each other. The integration between these modalities is substantiated during encoding and re-activates during retrieval.

Our framework is also consistent with Zwaan & Madden's (2005) theory of language understanding and development, in which multimodal experiential traces (perceptual-, sensorial- and motor memories) from linguistic (spoken or written words and sentences) and referential representations (concrete or abstract objects, properties and actions) get integrated during learning in a complex knowledge network according to the principal of 'associative strengthening by co-occurrence'. When not yet encountered, non-existent or abstract referents are introduced in dialogue for the first time, people are able to combine experiential traces (activated by situational features and additional discourse content) in order to create a cognitive ground from which the world and novel uses of language can be understood.

The framework is, furthermore, compatible with Glenberg's (1998) functional theory of memory, claiming that categorization consists in a meshing between projectable and non-projectable properties (similar to top-down and bottom-up processes) and that conceptual creativity consists in a combination of non-projectable properties, or *affordances* (Gibson, 1979). In this theory of Glenberg's everything is in the service of bodily interaction.

In relation to Lakoff and Johnson's (1980) claim that the entire conceptual system is permeated by conceptual metaphors it should be noted that it doesn't account for all the features of metaphoricity and understanding of gestures described here. First, it doesn't explain how conceptual metaphors are created or understood and second, it doesn't allow spontaneous metaphoricity in communication. Our model accounts for both aspects.

Summary
Gestures do form part of human multimodal communicative resources. One type of gesture that is both frequent and cognitively important is the metaphorical gesture. It emerges as a result of a dynamic cognitive activity, or simulation, and can ensue on different levels of conventionalization; from the individual and novel to the collectively entrenched. Metaphoricity in thought, language or gesture depends on the level of activation of the underlying source and target domain and their integration. This activation-process can be understood as a simulation of perceptual and introspective experiences in the sensory-motor system of the brain. Moreover, and this is the papers main thrust, it is claimed that comprehension of metaphoricity in language, thought and gesture likewise depends on simulations.

Experiment

In order to empirically investigate the effect of metaphorical activation in speech and/or gesture on comprehenders, during a short pseudo-conversation, we conducted an explorative experiment (pilot) in which participants watched a film clip of a *digital character* talking about cancer. The purpose of the experiment was to learn about the proposed relation between the simulation of multimodal metaphorical content and conceptualization.

In order to test the assumed relationship we had to create an experiment that probed metaphorical speech and gesture in a systematic way and that didn't depend on arbitrary evaluation. The aim of the experimental manipulation was to engender metaphoricity in the verbal and gestural expressions of the stimuli in order to measure the effect on how the participants later judged statements expressing a certain conceptualization of a concept, 'cancer'. We predicted a positive correlation between the degree of metaphoricity in the independent variable and the degree of metaphorical conceptualization in the dependent variable.

Method
We manipulated with the convergence or simultaneous reinforcement of metaphoricity in speech and gesture. Other parameters that play a role in the degree of metaphoricity is the size and timing of the gestures as well as the attention directed on the gesture indicated by gaze direction. However, we only manipulated the existence of simultaneous reinforcement of metaphoricity.

Gesture research often base its findings in case studies we ,however, were able to implement a proper experiment in which variables were manipulated and causal hypotheses tested. The use of digital characters telling a story and gesticulating creates a new possibility to connect gestures with speech in a natural way; it opens up the possibility to study the importance of speech and gestures independently, by changing only one parameter at the time.

Measure, stimuli & procedure
A norming study with 36 participants was conducted in order to select the items for the questionnaire measuring the metaphorical conceptualization, or in this case, personification (cf. Lakoff & Johnson, 1980), of 'cancer'. The statements included were supposed to generate a higher level of support from the subjects with a higher level of personification of 'cancer'. Therefore, the score of every item were checked for correlation with the sum of all the items on the scale. A high correlation between the answer on one item and the sum of all answers should indicate that the question is in line with the overall trend of the answers and thus is measuring the desired parameter. Three items were clearly more correlated with the sum (table 1). They also created less confusion among the subjects than the other statements. In the pilot study, the subjects were not exposed to any stimuli or asked to think about cancer before answering the questions.

The narrative presented in the experiment either included metaphors such as, "I fought the cancer", or equivalent expressions like "I tried to treat the cancer". These two types of speech were matched together with two kinds of gesticulation. Two conditions contained a lot of metaphorical gestures (hitting and punching), illustrating the same kind of personification of cancer as the metaphorical speech. The other two conditions contained non-metaphorical gestures (mostly beats). This created four different conditions of stimuli (table 2):

The conditions were meant to initiate a metaphorical conceptualization of 'cancer' that responded to the amount of metaphor in the stimuli. After watching the film clip, the subjects answered the questionnaire. Two filler questions were included requiring the participants to think about the film clip just seen. They were also asked to state their age, sex and if they had any personal experience of cancer.

79 participants (22 women, 57 men) were recruited during one day for the experiment (at LTH, Lund University campus). The participants were placed in front of a computer with headphones, watched the short film clip (ca. 1,5 min) and fill out the questionnaire. The procedure took only a couple of minutes.

Results
The average total scores of the three questions for each participant in the experiment were compared between the

Table.1. Items included in final questionnaire

	Correlation with total score excluding other items
"Cancer is a disease with evil intentions."	0.79, $p < 0.01$
"Cancer is a disease that cannot be trusted."	0.63, $p < 0.01$
"Cancer is a violent disease."	0.61, $p < 0.01$

Table. 2. Experimental conditions

(A) Metaphorical speech with metaphorical gestures
(B) Metaphorical speech with non-metaphorical gestures
(C) Non-metaphorical speech with metaphorical gestures
(D) Non-metaphorical speech with non-metaphorical gestures

four conditions. The differences were not found to be significant due to the large variance in the answers (fig. 1). Comparing the scores for each question separately gave similar result, with no significant differences between the conditions.

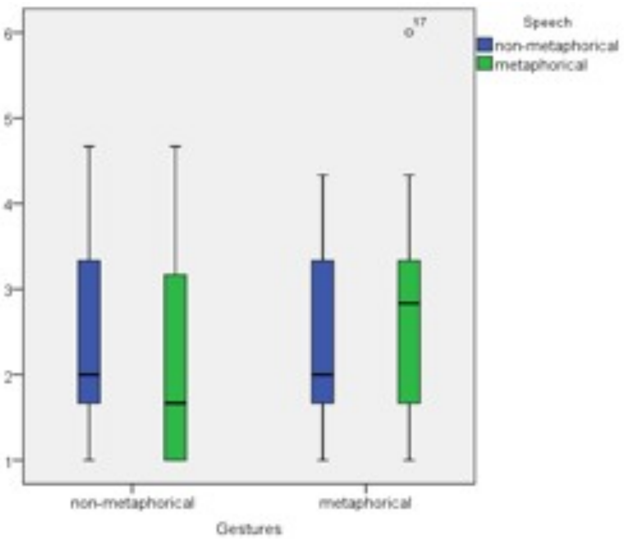


Figure. 1. Y axis: Scale from questionnaire 1 = "I agree", 6 = "Don't agree at all".

Discussion

Part I of this paper tried to answer the question whether people do undergo mental imagery or simulation during metaphorical language comprehension and gesture production. This was, based on the literature, found probable, and a model accounting for gestural comprehension and conceptual integration and alignment was suggested (Part II). The experiment (Part III) was conducted to answer the further question of whether people, watching a film clip and presumably undergoing mental imagery or simulation, change their conceptualization as a result of that simulation. We found that they didn't do that within the time frame of the experiment.

It seems, more precisely, as if our participants in the narrative experiment were not affected in any particular way by watching the metaphorical gestures. Because of the lack of any significant trend in the data, we cannot conclude that gestures increased (or decreased) the level of personification of the concept 'cancer'.

There could be at least two reasons for the absence of difference between the conditions; either the participants didn't understand or listen to the metaphorical content expressed in conditions A, B and C, (table.2), or the metaphorical and literal conceptualizations expressed in the stimuli clips, were not in fact differing in degree of personification. We think that the former alternative isn't probable based on the conversations had with the participants during debriefing. More likely, expressions like "I got attacked by cancer" were metaphorically dead (or sleeping, cf, Müller, 2008) for our subjects, and could not be resurrected by gestures allegedly activating sleeping metaphors (cf. section 5). If the metaphors were sleeping, then the test subjects interpreted the metaphorical stimuli literally and didn't pay attention to the personifications expressed in them.

This might mean that the theory of metaphoricity proposed by Müller (2008, se section 5) and built on in the present framework, should take into account the possibility of a high multimodal activation of an metaphorical mapping between domains on the behavioral level (utterance and gesture), and a low awareness of its figurative content, i.e., entertainment of a sleeping metaphor, on the phenomenological level. This dispersion could be reserved to the speaker/gesturer. But we think that it is equally likely that the listener/viewer that is exposed to multimodal metaphorical expressions could be unaware of the figurative content.

We suggested, in section 2, a distinction between research questions regarding comprehension of metaphorical gestures, and investigations of *if* or *when* gestures have a communicative value, i.e. add non-redundant information (that can be experimentally measured by requiring direct verbal reports) to the conscious and immediately accessible content of the communicated message. Gestures might be understood without leaving an obvious reportable trace. We tried to capture that elusive effect without success. The question now is why we didn't find any effect.

The cognitive basis of metaphorical gestures, i.e., the mapping between, sometimes unconscious image schemas of perceptual knowledge, and abstract ideas, might be hard to access upon hearing questions, in an obviously linguistic form, probing a condensed and conscious semantic essence, rather than the underlying representations.

The participants in the experiment might still have simulated the content of the gestures in exactly the way that the proposed model suggests. Our test subjects could have simulated the metaphorical content of the gestures on an unconscious level. Just as an unconscious simulation of an action, in a gesturer, can cause the execution of an overt gesture, the perceived content of a metaphorical gesture, we suggest based on our negative results, could be unconsciously simulated. This is in line with Barsalous (1999) account of simulation and our model of gesture comprehension.

Perhaps metaphorical conceptualizations drop as they become more explicitly entertained in conscious thought. This does not go against the spirit of Müller's theory of metaforicity as *active cognition*, because *"Activation does not presuppose consciousness"* (Ibid., 12). We suggest, based on our lack of findings, that metaforicity in speech and gesture is likely to go along a fairly unconscious use of those expressions. The continuum of conscious communicative acts were described in section 2. Cienki (2008), as was seen in section 5, also endorse such a continuum.

The framework is, however, probably a better explanation of gestural communication of metaphorical content on a much greater time scale then the one tested in the experiment. The real effect of metaphorical communication might unveil itself first at the cultural or entrenched level of expression.

On this account, metaphors get expressed through gestures and are slowly integrated into the collective mind of conceptual metaphors. Novel gestural metaphors might be understood if the conceptual mapping is structurally isomorphic on the experiential level, but changes in recipient's conceptualization and the conventionalization of metaphor is probably a temporally extended process.

Does the result mean that Krauss et al. (1995) were right when they claimed that gestures do not communicate? Either the gestures didn't communicate or the questionnaire didn't probe the understanding of the content. The questions were formulated with the purpose to measure the temporal change in conceptualization, rather than the understanding *per se*. The questionnaire could have been non-valid. The three items (intention, trust, violence) in it might not have probed the relevant level of conceptualization. Instead, only superficial answers, out of social convenience, might have been elicited. This relates to the risk of imposing demand characteristics onto an experiment, which might be what happened in our case.

We also wanted to test the assumption, explicated in a previous section, of conceptual alignment resulting from metaphorical gestures alone. Our study shows that if that assumption is correct, then the time frame for such a process, is longer than 1,5 min. Which in retrospect isn't that surprising. An actual interactive setting, rather than the pseudo-conversation of our experiment, is probably also a requirement for that kind of alignment to occur. It was a bit too optimistic to think that a positive effect of multimodal metaphorical content presented by a digital character would be found after experimental trials lasting up to 1,5 min.

Moreover, perhaps the participants own simulations would have been more accurately measured if they were required to answer a questionnaire verbally, allowing them to gesture at the same time. Our model suggests that gestures influence a listener/viewer's conceptual situation model directly, if the concept communicated is already formed in the interlocutor. In the case of a child learning a new concept, as was seen in section 6, gestures often express an understanding before the same idea could be expressed linguistically (Goldin-Meadow et al., 1993a). A more accurate proposal, of our model, would therefore probably be that gesture understanding does not always transplant to the linguistic level, even in mature individuals that master the concept.

Conclusion

We have presented a framework for the role of metaphorical gestures in a broader communicative and cognitive context; conceptualization of a common world. It accounts for the mechanisms of the creation, activation and exchange of metaphors as expressed in gesture. The narrow or immediate functionality, or communicative value, of metaphorical gesture is still, however, unclear, as our experiment indicated a lack of effect of gesture on conceptualization.

References

Barsalou, L. (1999) Perceptual Symbol Systems. *Behavioral and Brain Sciences* 22:577-660

Bargh, J. A., Chen, M.,&Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology,71,* 230–244.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645.

Brennan, S. E. & Clark, H. H. (1996) Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 22, No. 6,1482-1493.

De Jaegher, H., Di Paolo, E., Gallagher, S. (2010) Can social interaction constitute social cognition? Trends in Cognitive Sciences, October 2010, Vol. 14, No. 10.

Cienki, A., Müller, C. (2008) *Metaphor and Gesture.* John Benjamins Publishing Company.

Chui, K. (2011). Conceptual metaphors in gesture. Cognitive Linguistics 22(3): 437-458. Garrod, S. & Pickering, M. J. (2004) Why is conversation so easy. *TRENDS in Cognitive Science* Vol. 8., Nr. 1.

Frege's (1892) On Sense and Nominatum. In

Martinich, A. P. (Ed.) *The Philosophy of language*, Oxford University Press.

Ferreira, F. Apel, A., & Henderson, J. M. (2008). Taking a new look at looking at nothing. Trends in Cognitive Science, 12(11), 405-410.

Glenberg, A. M. & Kaschak, M. P. (2002) Grounding language in Action. *Psychological Bulletin & Review* 9(3):558-565.

Goldin-Meadow, S. *et al*. (1993a) Transitions in learning: evidence for simultaneously activated strategies *J. Exp. Psychol.* 19, 92–107

Goldin-Meadow, S., Alibali, M.W. and Church, R.B. (1993b) Transitions in concept acquisition: using the hand to read the mind *Psychol. Rev.* 100, 279–297.

Goldin-Meadow, S. 1999. The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 11 (3), pp. 419-429.

Gibbs JR., R. W. (2006) Metaphor Interpretation as Embodied Simulation. *Mind & Language* 21(3): 434-458.

Grice, P. (1989) *Studies in the way of words.* Harvard University Press.

Gazzaniga, M. S., Ivry, R. B., Mangun, G. R. (2009) *Cognitive Neuroscience: The Biology of mind.* Norton & Company.

Garrod, S. & Anderson, A. (1987), Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27 (1987) 181-218

Glenberg, A.M. (1997). What memory is for. Behavioral and Brain Sciences 20,1-55.

Haviland, J. B. (2000) Gesture. *Journal of Linguistic Anthropology* 9(l-2):88-9.

Hesslow, G. (2011). The current status of the simulation theory of cognition. *Brain Research.*

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15,* 495–514.

Hostetter, A. B. (2011). When Do Gestures Communicate? A Meta-Analysis, *Psychological Bulletin*, Vol. 137, No. 2, 297–315

Kendon, A. (1990) Conducting Interaction: Patterns of Behavior in Focused Encounters, Cambridge University Press.

Krauss, R. M., Dushay, R. A., Chen, Y., Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of experimental psychology* 31, 533-52.

Kendon, A. (1982) The study of gesture: some observations on its history. *Recherches Semiotique/Semiotic Inquiry*, 2 (1) 25-62

Krauss, R.M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science* **7**, 54-73.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory & Language*, **48**, 16-32.

Kita, S. (2000). How representational gestures help speaking. In D. Mc-

Neill (Ed.), *Language and gesture* (pp. 162-185). Cambridge: Cambridge University Press.

Lakoff, G. & Johnson, (1980) *Metaphors we live by*. University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). Philosophy in the flesh: The embodied mind and its challenge to Western t hought.N ew York:B asic Books.

Matlock, T. (2004b). Fictive motion as cognitive simulation. *Memory & Cognition, 32*, 1389–1400.

Melinger, A., Levelt W.J.M., Gesture and the communicative intention of the speaker, *Gesture,* 4(2), 119-141, 2004.

Morford & Goldin-Meadow (1997) From here to there and now to then: the development of displaced reference in homesign and English *Child Dev.* 68, 420–435.

McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-forspeaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141-161).Cambridge: Cambridge University Press.

McNeill, D. (1992) Hand and Mind: What Gestures Reveal about Thought. Chicago: University of Chicago Press.

Müller, C. (2008) *Metaphors Dead and Alive, Sleeping and Walking.* The University of Chicago Press.

Núnez, R. (2008) A fresh look at the foundations of mathematics: Gesture and the psychological reality of conceptual metaphor. In Cienki, A. &

Müller, C. *Metaphor and gesture*. John benjamins Publishing Company

Prinz, J. 2002; *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

Richardson, D. C., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition, 102*(1), 129–138.

Richardson, D. C., Spivey, M. J., McRae, K., & Barsalou, L. W. (2003). Spatial representations activated during

real-time comprehension of verbs. *Cognitive Science, 27*, 767–780.Slobin, 1987).

Richardson, D.C., Dale, R., & Spivey, M. (2007). Eye movements in language and cognition: A brief introduction. In Gonzalez-Marquez, Coulson, Mittelberg, & Spivey (Eds.) *Methods in Cognitive Linguistics*. John Benjamins.

Roth, W. M. & Lawless, D. (2002). Scientific investigations, metaphorical gestures, and the emergence of abstract scientific concepts. *Learning and Instruction* 12: 285–304

Shapiro, L. (2008) *Embodies cognition*. Routledge.

Wilson, M. (2002) Six views of embodied cognition *Psychonomic Bulletin & Review, 9 (4), 625-63*

Williams, R. F. (2008) Gestures as a conceptual mapping tool. In Cienki, A. & Müller, C. *Metaphor and gesture*. John benjamins Publishing Company.

Zwaan, R. A. & Madden, C. J. (2005) Embodied sentence comprehension. In *The grounding of cognition: The role of perception and action in memory, language, and thinking.* Pecher & R.A. Zwaan (Eds.), Cambridge, UK: Cambridge University Press.

Zwaan, R. A., & Radvansky, G. A., (1998) Situation Models in Language Comprehensionand Memory. *Psychological Bulletin*. Vol. 123. Nr. 2., 162-185.

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Size certainly matters – at least if you are a gesticulating digital character: The impact of gesture amplitude on information uptake

Mette Clausen-Bruun, Tobias Ek, Tobias Andersson and Jakob Thomasson*

It is important to investigate what types and forms of gesture might be most effective, so that digital interventions can be designed to take the fullest advantage of the communicative power of gestures. As no previous study has been made on gesture amplitude with virtual agents, we tested information uptake on 120 addressees in conditions with natural gestures vs. amplified gestures and with a humanoid vs. an ”alien” character. Information uptake was significantly better on the amplified gesture condition. Also, comparing a humanoid character with the “alien” counterpart, participants rated the humanoid significantly more natural in its movements and less distracting than the alien, despite both of them being animated with the same motion capture data. We also found a strong trend that the humanoid was more facilitating on information recall than the alien.

Background

Almost fifty years ago Mehrabian and Ferris (1967) claimed that a spoken message consists of 55% nonverbal behaviour, 38% voice modulation and only 7% words. If this is correct, an overwhelming part of the message is conveyed by the body’s motor system and only a minor part by speech. Researchers therefore got interested in the communicative role of nonverbal behaviour and in gestures in particular. In this paper we define gestures as the movements of arms and hands that occur during speech and that are generally considered as intimately linked to communication as speech itself (Kendon, 1980; Kendon, 1986; Kendon, 1994; McNeill, 1985). Much research has been done on the role of gestures in speech *production*, but it is only during the last decade or so that researchers have begun exploring the role of gestures in speech *perception*.

Today, it has been widely accepted that speech perception most definitely involves the motor areas of the addressee’s brain (see for instance Willems and Hagoort’s 2006 review). But why would motor areas be involved in speech perception, one might wonder? Giacomo Rizolatti and his colleagues first discovered what he termed *mirror neurons* in monkeys and later in human beings as well (Rizolatti Craighero, 2004; Rizolatti, 2007). He asked participants in his experiments to watch a movement, to make the same movement or to simply imagine performing the movement. In each case a region of the lateral frontal lobe in the left hemisphere, including Broca’s area"/>, was activated. These neural discharges did not precede the movements, but instead they occurred in synchrony with the movements. As it takes time for a neural message to go from the frontal lobe to i.e. an arm or a hand, controlling the movement would mean that the neurons would have to fire before the movement. But they did not. The neurons must therefore be simulating a movement as it is taking place**.** This implies that people apparently recognize the gestures made by others because the neural patterns produced when they observe those gestures are similar to those produced when they themselves make those same gestures (which is why Rizolatti coined the term “mirror neurons”). Thus, mirror neurons seem to provide a link between the sender and the receiver of communication (in this report referred to as the *speaker* and the *addressee*). Some people go so far as to claim that human speech evolved from manual gestures, with vocal gestures being gradually incorporated into the mirror system in the course of human evolution (Corballis, 2009).

Viewing gestures may also elicit overt mimicry as when we more or less automatically mimic a smile in order to interpret it (Niedenthal, Mermillod, Maringer, Hess, 2010). Mechanisms like those proposed in the *Simulation of Smiles model* of Niedenthal and his colleagues are also involved in interpreting the spontaneous hand gestures that people produce when speaking. That is, overt or covert embodied simulation is also involved in gesture comprehension (Alibali Hostetter, 2010). It has even been shown that areas in addressees’ non-primary auditory cortex show greater activity when speech is accompanied by beat gestures"/> than

"/>[1] An area in the left temporal lobe linked to speech production.

"/>[2] A rhythmic strike that creates emphasis and attracts attention.

when presented alone or with nonsense gestures (Hubbard, Wilson, Callan, Dapretto, 2009).

A massive body of experiments conclude that gestures contribute to spoken utterances from the speaker's point of view (Kendon, 1986; McNeill, 1985), and, as mentioned above, researchers have recently begun investigating the contribution of gestures to the addressee's comprehension of spoken messages as well. Communication is about transferring the speaker's mental representations to the addressee and speakers' gestures have been shown to contribute significantly either to addressees' semantic/pragmatic comprehension, to memory, or to both, as compared with utterances with no gestures (for examples, see McNeill's review, 1994). In accordance with this, Autumn Hostetter (2011) conducted a meta-analysis of the effect size from 63 samples with totally 2396 participants and found that across samples, gestures provided a significant benefit to addressees' understanding of a message as compared with speech alone.

Hubbard and colleagues (2009) studied how the addressee's brain integrates hand gestures with co-occurring speech and found a common neural substrate for processing speech and gesture, which is likely to reflect their joint communicative role in social interactions. Furthermore, gestures have been shown to reinforce deep learning (or rule learning) at the expense of shallow learning (or rote learning), which means that people draw more discourse-based inferences from the information explicitly stated and that they have poorer recollection of the verbatim of the discourse; in short, they understand and synthesise the information instead of merely learning it by heart (Cutica Bucciarelli, 2008). In line with this, Hostetter (2011) concludes that gestures "not only facilitates an immediate message, but also strengthen the memory trace of the message and improve conceptual comprehension and understanding of the message to the point where the knowledge can more effectively be applied in a new situation" (p. 311).

One thing is that gestures *do* contribute; another is *how.* Gullberg and Holmqvist (2006) investigated visual attention to gestures in human interaction by means of eye tracking and found that only a minority of gestures drew fixations. The face dominates as a fixation target (90-95% of the total fixation time), presumably due to cultural paradigms prescribing eye contact during face-to-face communication, (ibid.). As there is considerable evidence that gestures do contribute to addressees' comprehension and memory it is a surprising finding that fixations on the gestures are as rare as 5-10%. Furthermore, in most cases the fixations are cued by the speaker's own gazing at his or hers gesture (ibid.). One should have thought that because gestures illustrate spoken information, they should be targets of gaze direction as part of the addressee's attempt to retrieve information.

Furthermore, it has been shown that gestures that are incongruent with the semantic meaning of the spoken word have hindered addressees' comprehension, to the degree that the discrepancy between the spoken words and the gesture elicits the N400[1] response, which is typical of the brain's reaction to words that are semantically 'out of order' in a given context (Habets, Kita, Shao, Özyurek, Hagoort, 2010). Thus, even if we focus on the speaker's face most of the time, there is little doubt that we do notice and compute the speaker's gestures. As we direct our overt attention to the speaker's face, his or hers gestures seem to be registered by our peripheral vision, the photoreceptors of which are dominantly for motion detection (Gazzaniga, Ivry, Mangun, 2009). This is supported by a later study by Gullberg and Kita (2009), which concludes that there is little evidence of a relationship between addressees' direct fixations of gestures and their uptake.

Gullberg and Holmqvist (2006) compared a live face-to-face setting with two video conditions of the same setting: one shown as a full size video recording and the other shown on a 28" screen. Considering the social eye contact paradigm, it was an unexpected finding that less (though not significantly less) time was spent on the face in the video conditions. Gullberg and Holmqvist suggest that this could be a consequence of the person in the video not being a 'real' person and that social parameters like eye contact therefore are not as dominate in video interactions as in a real life. Furthermore, overall less overt attention was paid on gestures in the video conditions.

Gullberg and Holmqvist suggest that "gestures could compete for attention with the face in face-to-face interaction … because they constitute a source of competing information" (ibid., p. 54). This competition could also exist between our foveal and peripheral vision. If this is the case, one might suspect that gestures with strong amplitude would 'win the competition' – maybe to such a degree that they divert the addressee's attention from the spoken word. In that case, one could speculate that too much attention on the gestures would occupy such a large part of the addressee's cognitive resources, that he or she would not properly encode the simultaneously spoken utterances, resulting

[1] N400 is an Event Related Potential (ERP) in the brain, which is elicited 400 milliseconds after the presentation of a stimulus that is semantically incongruent with the meaning of the sentence in which it occurs.

in impaired comprehension and recall. Furthermore, studies have shown that gestures alone convey little information (Hostetter, 2011), so too much attention on the gestures could result in the addressee ending up with not having registered the message the speaker is trying to convey.

Maybe this is why many consultants and coaches within media training, rhetoric's, and related areas warn that too many and too large gestures are distracting and steals the addressee's attention from the message. For instance, SKTF's[2] web page teaches that "Vill du vara tydlig ska du försöka undvika gester och annat som kan distrahera"[3] (SKTF, n.d.), career coaches CVfabriken (CVfabriken, n.d.) and Kelly Services (Kelly Services, n.d.) both instruct their clients to "Nicka gärna, men var försiktig med för yviga gester. Överdrivna rörelser kan nämligen distrahera intervjuaren"[4] and career coach Rachel Green (Green, n.d.) advises to "Keep movements small".

Neff, Wang, Abbot, Walker (2010) studied how gesture rate and amplitude could be varied to increase or decrease perceived extraversion, a personality trait that is often considered advantageous for people conveying a message. Contrary to the restrictive advice above, Neff and his colleagues conclude that in addition to extraverts being less formal, talking faster and using more positive emotion words, high gesture amplitude or expansiveness have shown to be key indicators of an extravert personality – in short, the bigger the better.

Thus, gestures can be a powerful tool in many different real-world applications, as for instance advertising and education. It is important to investigate what types and forms of gesture that might be most effective, so that digital interventions can be designed to take the fullest advantage of the communicative power of gestures. For instance, do larger gestures result in more benefits for the addressee?

Most gesture studies have been conducted with human speakers, on video or in real life. But as technology continues on to make huge advances and digital hardware is becoming more and more accessible, the use of virtual characters will grow rapidly as well. One reason for this might be that virtual characters have a distinct advantage over human actors: every detail can be controlled and manipulated in a way that is not possible with even the best trained actor, whose performance can never be controlled for confounds like an involuntary blink of the eye or a brief delay in the onset of a gesture.

In order to create trust and rapport, these characters will need to be able to communicate naturally with people using both verbal and nonverbal signals (Ahn, Fox, Bailenson, 2012). One goal in multi-modal virtual character research is to determine how to vary expressive qualities of a character so that the addressee will perceive it in the desired way. As said, trust and believability are important in the relationship between user and the virtual counterpart. Studies have shown that virtual characters with human traits are more likable than a more non-human appearance and behaviour. The character must not be too human-like, though, as this might have the opposite effect, the *Uncanny Valley* effect (Mori, 1970). It is not possible to fool the user a hundred per cent – the animation of a virtual character will sooner or later show lapses and errors that will disclose the underlying computerisation and we then tend to perceive the virtual character as a living dead or a zombie (Draude, 2011; Mori, 1970).

Research Questions

No previous study on the effect of gesture amplitude has been made with virtual agents. Hopefully, our study will contribute to the discourse on how to control the expressivity of animated virtual agents for the benefit of users as well as promoters. As we prominently use technology that has small screens, we have chosen to use 15"-17" screens in our study.

We ask two questions:

1) Will addressees' uptake of a short narrative be affected differently when the gestures are amplified as to compared with normal amplitude?

2) Will the appearance of two very different virtual characters (a humanoid and an 'alien') affect addressees' uptake of a short narrative?

Method

Participants

The study was conducted at campus of the Faculty of Engineering, Lund University. A convenience sample

---

[2] Sveriges Kommunaltjänstemannaförbund, a public organisation for municipal officials.

[3] "If you want your message to be clearly understood, you should try to avoid gestures and other things that could distract" (our translation).

[4] "Feel free to nod, but be careful with too extensive gestures. Exaggerated movements can distract the interviewer" (our translation).

of 120 participants (31 female, 87 male, 2 n/a) was recruited at the premises. As incentives to participate we handed out candy to all participants and every tenth participant received a free cinema ticket.

Material

The digital characters for the study were created by means of motion capture data[5]. The motion capture data featured a woman sitting on a chair, retelling two narratives presented in animated cartoons (Appendix 1). She retold the narratives directly after viewing them, in her own words and nonverbal behaviour, without having been given any directions. We created two characters (Appendix 2), one with a simplified stick-man appearance with blue skin (referred to as the *alien*), and one resembling a female human being (referred to as the *humanoid*). Our intention was to have the characters represent two distinct and diverse poles in the uncanny valley spectrum.

The characters in the normal gesture amplitude condition were manipulated to create a high gesture amplitude condition by manually extending and fine-tuning the movements until a natural but higher amplitude gesture range has been reached. This process was based a case-by-case assessment of each gesture. To synchronize the gestures with speech, we opted to increase the velocity of the gestures at their initial phase as well as in their striking phase, keeping most, if not all, of the gesture delivery and meaning intact, instead of merely reproducing it as a speed-up version of the original gesture. Four videos were created, each with duration 2 minutes and 20 seconds: 1) alien, normal gesture amplitude, 2) alien high gesture amplitude, 3) humanoid, normal gesture amplitude, and 4) humanoid, high gesture amplitude. In all four conditions the voice was the original recording of the woman retelling the cartoon-narratives.

Procedure

After having given their informed consent, participants were seated in a quiet room, four at a time, randomized to one of four 15”-17” laptops, each featuring one of the four video conditions. Participants were facing each other, so that they could not see each other’s screens. The laptops were equipped with headphones to ensure an adequate and similar acoustic environment for all participants.

Data collection was conducted on three consecutive weekdays, during the same hours of the day. Each participant was instructed that he or she was taking part in a digital survey about memory. The survey included instructions how to complete the survey, one of the four video conditions, a distractor text, 22 multiple-choice statements about the story and finally a questionnaire. The distractor text (Appendix 3) served the purpose of ensuring that the narratives had passed from participants’ working memory to their long-term memory before presenting them to the multiple-choice statements about the narratives. 18 blocks of statements were related to the narratives. Each block consisted of four statements: one correct, one almost, but not quite correct, one obviously incorrect, and one both unrelated and incorrect, thus constituting a 25% chance level of a correct answer. The order of the statements was balanced to control random variation. In order to reveal any participants that made no attempt to truthfully fill in the survey, further four blocks were mixed in as fillers that could be related to the narratives but were actually not mentioned or shown.

The final part of the survey was a questionnaire (Appendix 4) asking participants to rate their agreement with the following statements on a 1-6 Likert scale (1= I strongly disagree; 6 = I totally agree):

- I find that the character was moving naturally
- I find the character distracting (reversed scale)
- I find that the character’s movements made it easier for me to remember the story
- I do not find that character’s movement matched the story (reversed scale)
- I like the character.

The questionnaire also inquired about the participant’s age and gender and asked open-ended questions of what they thought the real purpose of the study was and if they had any generic comments.

Results

A two-way between-groups ANOVA was conducted to explore the impact of gesture amplitude and type of digital character on information uptake, as measured by the total score on 18 multiple-choice statements relating to the information given by the digital characters and four control questions to which there were no correct answers. Participants were randomized into four groups according to gesture amplitude and type of character (group 1 (N = 31): normal gesture amplitude, alien character; group 2 (N = 30): high gesture amplitude, alien character; group 3 (N = 30): normal gesture amplitude, humanoid character; group 4 (N = 29): high gesture amplitude, humanoid character).

The interaction effect between gesture amplitude and type of digital character was not statistically sig-

[5] Motion capture is a recording with infrared (IR) cameras filming a person wearing small reflective items distributed across the body. The IR-cameras register the coordinates *(x, y, z)* of all these reflectors, as the person is moving and thus creates a ‘skeleton’ that can be applied to and animate various ‘bodies’.

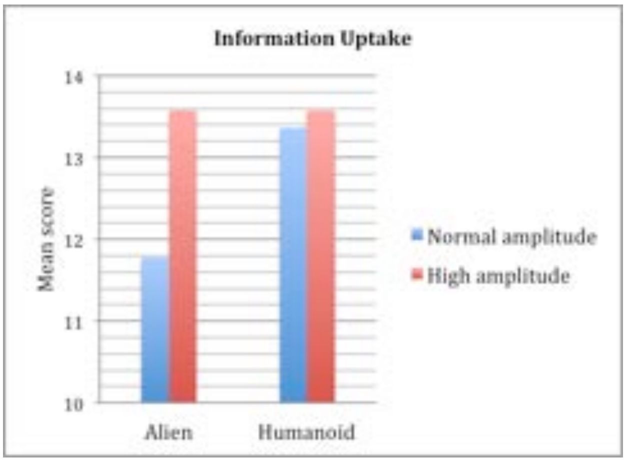


Fig. 1. Participants in the high gesture amplitude conditions had a significantly higher information uptake (score) than those in the normal gesture amplitude con-

nificant, $F(1, 116) = 2.44$, $p = .12$. There was a marginal statistically significant main effect for gesture amplitude, $F(1,116) = 3.85$, $p = 0.05$, indicating that participants in the high gesture amplitude conditions had a higher information uptake than those in the normal gesture amplitude conditions (Fig.1). However, the effect size was small ($\eta^2 = 0.03$). The main effect for type of character, $F(1,116) = 2.39$, $p = .13$, did not reach statistical significance.

Participants rated the gestures of the humanoid character ($M = 3.14$, $SD = 1.26$), significantly more natural than those of the alien ($M = 2.62$, $SD = 1.19$), $F(1, 108) = 4.12$, p = .045, $\eta^2 = .04$ and they found the humanoid ($M = 4.02$, $SD = 1.42$), significantly less distracting than the alien character ($M = 3,41$, $SD = 1.62$), F(1, 108) = 6.21, p = .01, $\eta^{2\,=}$ .05 (Fig. 2). Furthermore, we found a tendency that participants rated the gestures of the humanoid ($M = 3.32$, $SD = 1.39$) more facilitating on information uptake than the gestures of the alien ($M = 2.75$, $SD = 1.56$), $F(1,108) = 3.57$, $p = .06$, $\eta^2 = .03$. Participants expressed no emotional preference for either of the characters $F(1,108) = .16$, $p = .69$, $\eta^2 = .00$ nor did they find any discrepancies between narrative and accompanying gestures $F = 2{,}31$, $p = .13$, $\eta^2 = .02$.

There was no interaction between gender (female N= 31, male N = 87, 2 n/a) and age (min: 18 years, max: 57 years, M = 22.78 years, SD 4.14) of the participants on the ratings on the six point scale which means that none of the genders were overall positive or overall negative in their ratings and that age did not have any influence either.

Discussion

Considering that research on the effect of virtual agents' gestures on addressees' information uptake is virtually in its cradle, we decided to settle for an explorative study. The present study is thus definitely to

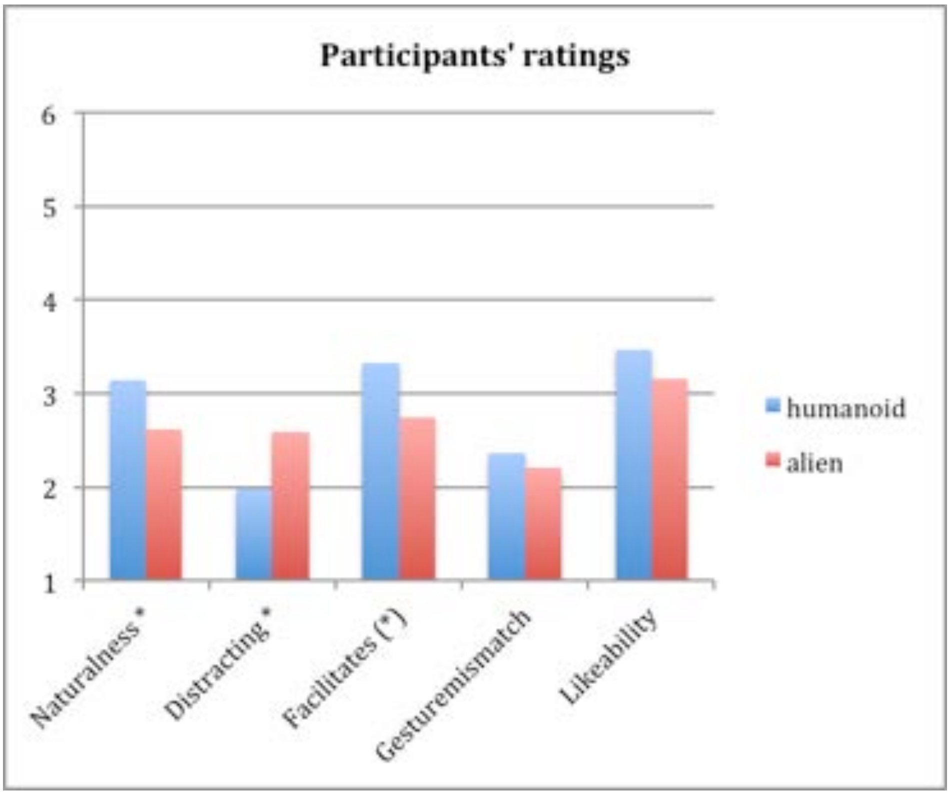


Fig. 2. Participants rated the humanoid significantly more natural and less distracting than the alien and tend to rate the gestures of the humanoid more facilitating on information uptake than those of the alien. indicates statistical significance. (*) indicates a strong trend.

be regarded as a pilot study, especially as we were constrained by time and material. However, we did find that size matters: the larger gestures, the better information uptake.

The statistical significance was marginal, though, and we suspect that this can be attributed to the 'normal gesture amplitude' having rather high baseline amplitude. The woman retelling the narratives was dressed in a black bodysuit equipped with reflectors all over the body and face and was instructed to talk and act in a natural, undirected way. But how natural is the situation in itself? It does not seem unlikely that, under these conditions, one becomes more conscious about one's gestures and tends to either restrict or exaggerate them. Actually, there were a few gestures that we could not amplify, as they, already in the 'normal' condition, reached the limits of the picture frame. Had time allowed, a 'minimized gesture amplitude' condition should have been added to control for this possible confound.

As with most untrodden grounds, our study raises more questions than answers. For instance, as time was scarce we had to use precast material that was based on animated comic strips for young *children* – on young *adults*! Would we have gotten the same result with adult narratives? And would a video of longer durance than 2 minutes and 20 seconds yield a more robust finding?

Also, for practical reasons the sample was constituted of predominantly young, male engineer students (which, of course, makes it harder to generalize the result to the population as a whole), who by definition may be more interested in the digital technique than in children's narratives and multiple-choice statements. Some participants actually expressed their professional views on the quality of the animations in the comment-box. Did their professional interest steal their attention away from the task?

Our second research question concerned a possible difference in information uptake in the humanoid and the alien conditions. We did not find that information uptake was affected by type of character. However, participants significantly rated the humanoid moving more naturally than the alien (interesting, considering that both characters were animated with the same mocap data!). One participant commented that it was difficult to create a mental representation of the narratives when the character (the alien) was far from realistic in appearance and did not have any facial expressions, even if the gestures seemed quite realistic. The humanoid was found significantly less distracting than the alien and we found a strong trend that participants rated the humanoid's gestures more facilitating on information recall than those of the alien.

There was no difference in rated likeability. Virtual characters or agents are being increasingly applied to digital educational tools and software developers might be tempted to focus on the likability or the humanness of the digital characters in an intuitive attempt to reinforce information uptake. However, our study shows that other parameters, like gesture amplitude, naturalness, and, degree of attention or distraction may influence information uptake to a higher degree than likeability.

Hopefully, future studies will contribute to the establishment of proper criterions and guidelines for developers to follow. There is a world of other parameters to be explored, especially regarding cultural stereotypes. For instance, will the same digital character have the same impact on information uptake in collectivistic societies as in the Western World? And does size matter everywhere in the world?

References

Ahn, S.J., Fox, J., Bailenson, J.N. (2012). Avatar. In W.S. Bainbridge (Ed.), *Leadership in Science and Technology: A Reference Handbook.* SAGE Publications.

Alibali, M.W. Hostetter, A.B. (2010). Mimicry and simulation in gesture comprehension. *Behavioral and Brain Sciences, 33,* 433-434.

Corballis, M.C. (2009). Language as gesture. *Human Movement Science, 28,* 556-565.

CVfabriken (n.d.). *Bemanningsföretag.* Retrieved 5 October, 2012, from CVfabriken: http://www.cvfabriken.se/tips/bemmaningsforetag.

Cutica, I. Bucciarelli, M. (2008). The deep versus the shallow: Effects of co-speech gestures in learning from discourse. *Cognitive Science, 32,* 921-935.

Draude, C. (2011). Intermediaries: reflections on virtual humans, gender, and the Uncanny Valley. *AI Soc, 26,* 319-327.

Gazzaniga, M.S., Ivry, R., B., Mangun, G.R. (2009). *Cognitive Neuroscience. The Biology of the Mind* (2nd rev. ed.). New York: W.W. Norton Company Inc.

Green, Rachel (n.d.). *Ten tips on what to do with your hands.* Retrieved 5 October, 2012, from http://www.rachelgreen.com/cgi-bin/a.pl?tips105.

Gullberg, M. Holmqvist, K. (2006). What speakers do and what addressees look at. Visual attention to gestures in human interaction live and on video. *Pragmatics Cognition, 14,* 53-82.

Gullberg, M. Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *J Nonverbal Behav, 33,* 251-277.

Habets, B., Kita, S., Shao, Z., Özyurek. A., Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension.

Hostetter, A.B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137,* 297-315.

Hubbard, A.L., Wilson, S.M., Callan, D.E., Dapretto, M. (2009). Giving speech a Hand: gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping 30,* 1028-1037.

Kelly Services AB. *Intervju Tips.* Retrieved 5 October, 2012, from Kelly Services AB: http://www.kellyglobal.net/eprise/main/dev/Backup/se/services/se/intervju_tips.

Kendon, A. (1980). Gesticulation and Speech: Two Aspects of the Process of Utterance. In M.R. Key (Ed.) *The Relationship of Verbal and Nonverbal communication.* The Hague, NL: Mouton Publishers.

Kendon, A. (1986). Some reasons for studying gesture. *Semiotica 62,* 3-28.

Kendon, A. (1994). Do gestures communicate?: A review. *Research on Language and Social Interaction, 27,* 175-200.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 3,* 350-371.

Mehrabian, A. Ferris, S.R. (1967). Inference of attitude from non-verbal communication in two channels. *The Journal of Councelling Psychology, 31,* 248-252.

Mori, M. (1970). Bukimi no tani. K. F. MacDorman T. Minato, (Trans.), The Uncanny Valley. *Energy, 7*, 33–35. Retrieved on 5 October, 2012 from Android Science: http://androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html.

Neff. M., Wang, Y., Abbot, R. Walker, M. (2010). Eveluating the effect of gesture and language on personality perception in conversational agents. In J. Allbeck et al. (Eds.): IVA 2010, pp. 222-235. Berlin and Heidelberg: Springer-Verlag.

Niedenthal, P.M., Mermillod, M., Maringer, M., Hess, U. (2010). The simulation of smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences, 33,* 417-480.

Rizolatti, G. (2007). *Mirrors on the mind.* New York: Oxford University Press.

Rizolatti, G. Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,* 169-192.

SKTF (n.d.). *Att framträda.* Retrieved 5 October, 2012, from SKTF:
http://app.sktf.se/utbildning/sktf/obj_temp/info.asp?object_id=1653modul_id=43comp_id=1bounce_object_id=1644bounce_modul_id=43.

Willems, R.M. Hagoort P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language, 101,* 278–289.

Appendix 1

Transcription of narratives

Words in brackets are cues to when gestures occur and to which type of gesture.

Narrative 1 (in Swedish)

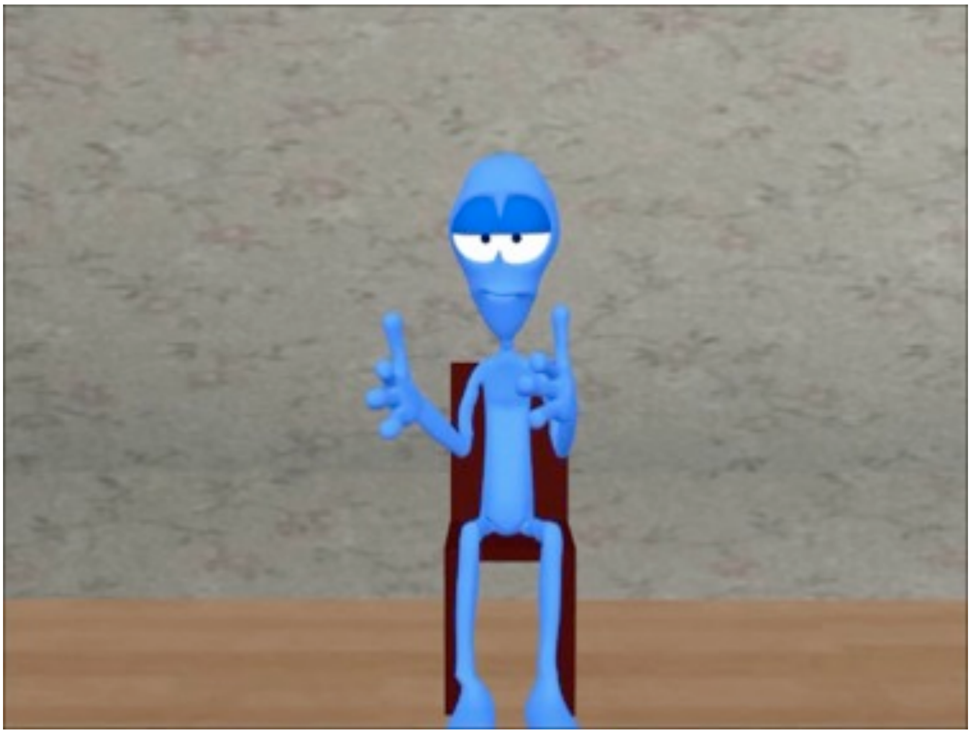
Det handlar om en mus. Musen är ganska stor[size]. Som står och sopar med sin lilla skyffel[shape] och sin lilla kvast[movement] på en ganska stor matta[size]. Så han har ganska mycket grus[size] på sin matta, så han måste stå där och sopa[movement] ganska mycket. Och så kommer hans kompis elefanten, som är ganska liten[size] och blå. Och så kommer[direction] han och suger[movement] på den här stora mattan[size]. Och den lilla musen blir jätteglad och uppmanar[movement] han att göra en skillnad, "komigen det är jättebra att du gör det". Så han fortsätter att dammsuga upp[movement] så att säga[iconic gesture for citations marks] grusen som är på mattan[pointing]. Men det blev nog lite för mycket[movement] för honom så han nyser ut allting[movement] igen så det blir ett stort grusmoln och den lilla musen[correcting herself] eller den stora musen[size] blir lite besviken[movement] och lite butter över att det gick så dåligt[movement].

Narrative 4 (in Swedish)

Dags för den lilla musen att motionera. Han e, har en ställning[painting in the air] med två pelare och en stång emellan[painting in the air] som han ska ta sig, så han ska träna[movement] lite på den. Så han börjar med att värma upp[movements] och så hoppar han upp[movement] och så får han riktig fart[movement] så snurrar jättemycket, flera varv tills han hoppar ner[movement] och ett väldigt stiligt hopp måste jag säga. Eh så kommer den lilla elefanten[direction] och tycker det var väldigt snyggt gjort[illustrating being impressed], han vill också testa på den här häftiga ställningen[painting in the air]. Men e, han är inte så stor[size], han får ta sats[movement] ganska länge och hoppa[movement] sig upp och så och det går inte så bra[movement] men till slut får han tag[movement] i stången med sin snabel. Men elefanten är ganska tung[direction], så han böjer[painting in the air] ner hela den här stången[shape], och musen[movement] blir väldigt arg[movement] det blir han och hotar[movement] den lilla elefanten att han ska slå honom, men, så han[movement] vad ska han göra nu då han måste ju rätta[painting in the air] till den här så den blir rak igen[painting in the air], så han ställer och försöker trycka upp[direction] den, men det går inte[movement], den är[painting in the air] alldeles[movement] för böjd. Så att de, vad ska han nu ta sig till[movement]. Och då helt plötsligt[movement] så kommer det en liten trollkarl[movement] inspringande och bara stöter[movement] till den här stången[painting in the air] med sin hatt[shape] och då rättar[painting in the air] den ut sig och blir alldeles rak. Och det tycker[movement] den lilla elefanten är ganska roligt och musen förstår ingenting[movement], det är jättekonstigt.

Appendix 2

Alien character



Humanoid character

Appendix 3

Distractor (meteorological description of the weather condition ‘hail’)

Hagelkorn är klumpar av is som kan variera i storlek från några millimeter till flera centimeter i diameter. Ibland definierar man dock iskorn mindre än 5 mm som annat än hagel. Hagel bildas i åskmoln, framförallt de som är höga, har mycket uppåtvindar och innehåller mycket vatten. Det ska även vara kallt – den bästa temperaturen för hagelbildning är -13 °C. Ett hagelkorn börjar som en vattendroppe lågt i molnet, men när den följer uppåtvindarna till högre och kallare höjder fryser den till is. Flera små hagelkorn kolliderar till större och på så sätt kan riktigt stora bildas. När hagelkornet blivit så stort att gravitationen vinner över uppåtvindarna faller den till marken. Tillväxtfasen brukar ta minst 30 min. Hagelkorn större än 2 cm kan orsaka väsentlig skada på exempelvis bilar och tak. Det största hagelkorn som hittats i USA var 20 cm i diameter och vägde 0,88 kg.

Appendix 4

Questionnaire in the digital survey

Vad tror du syftet med denna undersökning är?

Jag tycker att karaktärens rörelser gjorde att jag lättare kom ihåg berättelsen.

1	2	3	4	5	6
Stämmer inte alls					Stämmer precis

Jag tycker inte att karaktärens rörelser matchade berättelsen.

1	2	3	4	5	6
Stämmer inte alls					Stämmer precis

Jag tycker karaktären var distraherande.

1	2	3	4	5	6
Stämmer inte alls					Stämmer precis

Jag tycker om karaktären.

1	2	3	4	5	6
Stämmer inte alls					Stämmer precis

Jag tycker att karaktären rörde sig naturligt.

1	2	3	4	5	6
Stämmer inte alls					Stämmer precis

Har du några kommentarer?

Din ålder?

Du är? Man Kvinna

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

A Troublemaker as a Teachable Agent: How learning is affected by an agent with resistance

Christoffer Andersson, Simon Johansson, Anders Persson and Adam Wamai Egesa

In this paper a pilot study is presented about how different personalities in a Teachable Agent (TA) affects learning in newly developed educational game of history. A TA is a digital character that uses the paradigm of "learning-by-teaching", where the student has the teaching role, and an independent computer character the learning role. Previous studies have found positive effects of motivation, due to the socially interactive elements, using a TA compared to no-TA.

In this study we defined and explored differences between a (i) confirming and encouraging Normal-TA, and (ii) a critical and questioning "Troublemaker"-TA (TTA). Building on theory from Gärdenfors (2010), understanding is based on seeing patterns and casual relationships, and the TA/TTA was designed to exemplify positive/negative examples of casual relationships, respectively.

The study measured (a) Self-Efficacy of the students, a domain specific attitude towards own ability, (b) Likeability of the agent (LA) and the game (LG), and (c) a Memory test of how many facts they remembered a few days after. 38 subjects (aged 10-12 y.o.) played the game for 30 minutes. No significant difference could be found between TA and TTA. However, it is discussed, that the time to get social interaction effects might be much greater than the time the subjects had to play the game.

The study found a clear gender difference, females scored lower on the memory test, and is speculated to be due to adolescent males greater experience with similar computer games. The study also found significant positive correlations between SE and LA, but negative to LG. One interpretation of the regression analysis is that it is most important to like the teaching activity, rather than the whole game.

Introduction

How do we as individuals develop our knowledge more, by reinforcing our progress, or by getting our perception of the world questioned? Which way is better at achieving transfer from facts, to knowledge and understanding? With the help of a computer learning game with social interaction of digital characters, we constructed a pilot study examining what effect two alternative routes could produce.

One way to make this happen is to have an authoritative character in the game questioning what the subject is doing, like a teacher. However another role is to let the student be the teacher in the game. A Teachable Agent (TA) has been used in several experimental learning games the past decade, and is a digital character that the student is supposed to teach a new subject to, instead of learning it directly.

Previous research have shown pedagogical effects by using a TA in digital learning activities. Chase et al. (2009) concludes that learning with TA-games generates a motivating social component in the activity, which they mean produces two positive effects. The first is what they call an *Protégé effect*, that the student takes the role of a teacher while learning his computer character. Learning by Teaching is a known acknowledged pedagogical method to increase greater learning of a subject by providing a sense of responsibility and motivation. This is first and foremost shown by how students with TA:s spend more time on the strict "learning" parts of the games they play, rather then doing some other activity, like chatting och competing with others. The second effect they call an *ego-protective buffer*. The game scenario, the characters behaviour and interaction is constructed to give the impression of an independent, computer controlled character (agent), and it is the characters knowledge that is tested in the game. In other words, if it fails to perform it is likely to be interpreted as the *agent* failing to an higher extent than oneself. A incorporated TA have also shown to positively affect *low-performing* student more, and it is speculated to depend on the latter effect. Studies of Pareto et al. (2011) also showed a heightened gain in *self-efficacy*, a form of self-confidence in the specific domain.

Blair et al. (2006) discusses which aspects of *Learning By Teaching* (LBT) that are normatively desirable to include in a TA-game, namely; (i) an explicit, well-structured and shared visual representation of what is learned, (ii) independent action by the agent, (iii) the agents ability to create productive learning behaviour, and (iv) to put the agent in an environment supporting learning.

As the second aspects states, a necessary premises for the effects presented by Chase et al. (2009) is that the students accept the impression of the agent as *independent*. According to the authors in their example, this was confirmed by observing behaviour and chat messages as *social* interaction by the students with their TA.

The Troublemaker

In the TA from previously examined games and learning activities, the agent has always been fairly compliant. It has been able to ask questions to the student, as in "Betty's Brain" (Chase et al., 2009), but without much resistance the TA followed instructions and corrections from the student. It was however a more fundamental aspect of the original TA concept, as established by Brophy et al. (1999), to have greater and more distinct personality traits. Behaviour by the TA-module was there modeled to have a more distinct component of a personal *disposition*, leading to differentiated actions. For example, (i) confident or uncertain on it's own ability, demanding more or less repetition on an assignment, (ii) refusal in doing assignments or following instructions, or (iii) ask for explanations and clarifications (ibid, pp.7). This was supposed to, first and foremost, stimulate metacognition in the student's own attitudes towards learning. By observing the behaviour of the TA, and for example the lack of success of an agent refusing to do assignments, or being too confident in wanting to do tests before actual adequate learning, the student was supposed to be given the opportunity to reflect about it's own behaviour.

Another purpose of the authors (Brophy et al. 1999) was to try and remove itself from the traditional Platonic paradigm and view of learning; that "correct knowledge" would always produce "correct behaviour". Rather, the view they want to proclaim is that "correct knowledge" may lead to "incorrect behaviour", given "bad" dispositions.

According to A. Gulz and M. Haake (personal communication, Sept. 2012), that have developed and researched TA-learning, a not unusual expressed experience by students is that the agents seems *flat* or *impersonal*. They are not perceived as acting, functioning, and learning as students normally do. It was speculated that what was lacking, with reference to Brophy et al. (1999), was certain components of *resistance* in learning. Students, as well as adults, tend to rarely learn something new straight-forwardly, easy and comfortably, without encountering hinders while learning and understand new concepts. Rather, learning often seem to at times be a bit troubling.

Based on this we defined a concept of a (1) TA, a normal agent that is compliant and reaffirming, and a (2) Troublemaker-TA (TTA), with the disposition to give resistance, to question and be critical towards the information presented by the student.

The Lust of Understanding

A possible explanation and insight to why learning often seem to have resistance, and how some important element could perhaps be recreated, is found with Peter Gärdenfors "Lusten att förstå" (eng.: The Lust of Understanding) (2010). *Understanding* of something is, according to Gärdenfors (2010), about seeing *relationships* and *patterns,* of how things are causally related with each other. The deepest form of learning is not knowing or being able to repeat facts, but is the mentioned sense of *understanding*, what kind of "picture" these facts create together. Understanding is in this sense then rather about something abstract, the *underlying pattern,* than strictly about concrete facts.

This is in line with another thesis by Gärdenfors (2006), that one of the biggest drives in humans, and one thing that sets us apart from other animals, seems to be the drive towards relationships, purpose, connections and the *meaning* of things, and in the experience of life itself. A premise for this is speculated to be a (greater) human ability to plan for the future, which also creates a demand to understand the causal connections of how future events may come to turn out.

As Gärdenfors (2010) highlights, there is a sense of satisfaction in seeing relationships. The *Aha-epiphany* of things can be described as a dam breaking, a hinder passed, and gives us *pleasure*. The opposite, when the dam is holding, when you do *not* understand, is on the contrary highly frustrating. In other words, given a premise that someone want to feel pleasure, you also want to understand, and if you could tap into this process learning would be *motivation in itself*.

So how could this be incorporated in a TA/TTA. Gärdenfors (2010, ch.5) also highlights that *metaphors* and *analogies* is a very effective way of showing abstract relationships (in a concrete way). However, it can also be of help to highlight *bad examples* of metaphors, or highlight aspects that can *not* be transferred to the target object. In other words, bad examples may invoke reflection to what is a good or better understanding. This would then be a possible endeavour for a TTA, to question, to be critical of the students answers, to give resistance and give *negative relationships*, which could provoke reflection in the student and produce better understanding. And the normal TA could make statements of *positive relationships*, showing correct causal connections.

The History Agents

The TA-games developed previous to this study have generally been in the field of science or mathematics. How a TA can operate in social sciences has yet been let unexplored. It can also be speculated that a TTA could perform better in these areas, where for example critique, might play a more natural part of the subject.

In collaboration with another project group enrolled in the same course, the agents was incorporated in a game called “Historieagenterna” (eng: The History Agents), which targets historical inventions. The concept of the game was based on a time-traveling agency from the future, that had agents stuck in time and needed the help of the student to pinpoint where they were. With the help of finding facts from the specific time period, the agents would be able to calibrate their machine and go home. The students are asked to go into the virtual 3D-database of history knowledge, where you can interact with objects (inventions, famous paintings, etc.) and historical inventors (Johannes Gutenberg, Galileo Galilei, Leonardo Da Vinci, and Robert Hooke).

The teaching activity is where you interact with the agent (TA/TTA), as shown in figure 1.1, where the student is asked to fill in a matrix, where the inventors are the rows, and the columns are “Where”, “What”, “When”, “Effects”, “Today”, as representing facts about the inventor and the invention. The facts are found in the 3D-world/ database, and while it is filled in the Agent looks on and makes statements of two kinds. Either “horizontal relations”, summarizing facts for the same inventor, for example: “Yes, Gutenberg have meant a great deal for the media development.”. Or vertical relations, comparing facts between inventors/rows, for example: “Ah, so Gutenberg lived
before Da Vinci!”

The above are TA-statements, positive relationships (see previous section). A contrasting example for a questioning relationship by the TTA would be: “Are you sure Gutenberg lived before Da Vinci?”. The interaction was a combination of text and audio (swedish), and slightly different animations for the agent. In figure 1.1 the character is the troublemaking version of the agent, leaning forward on the table, where the TA-version was leaning back in a relaxed fashion.

For a technical report on the development of the agent types, see Appendix 1.

Figure 1.1. The History Agent that in the game needs help to calibrate his Time-machine.

Main focuses

The purpose of this pilot study is to give grounds to further investigations on how different personalities of a teachable agent affects students learning and the likeability of educational games. To do this we test any effects of: (i) a “nice”, easy to teach, and encouraging TA in contrast to a (ii) questioning, challenging, and critical TTA, to explore any differences or similarities between them.

The troublemaking disposition of teachable agents were discussed, by Brophy et al. (1999), in the early development of the TA framework, but seems since then to have been forgotten. Instead, research in the field of TA has been more interested in exploring effects of educational TA games in contrast to traditional learning, or educational games without TA:s. This study can not, therefore, rely directly on any earlier findings regarding the effects of different TA:s personalities, and the results of the study will hopefully shed light on a previously unexplored area of investigation. Due to the explorative nature of this study, no hypothesis will be given. However, the main focus of this study can be presented in the following two questions:

Q1: How does the TTA affect students learning compared to the regular TA?

Q2: Does students with lower self-efficacy (SE) in regards to history as a school subject respond different than students with higher SE to the TA- or TTA settings?

Method

The participants in this pilot study were two 5th grade school classes (age span 10-12 y.o.) in Lund, Sweden. Of the total 38 students, 21 were male and 17 female. The method for measuring the effects of the different kinds of TA:s can be divided into three stages. First the students self efficacy was measured, secondly the students played the game, and third and finally we measured the students information gain by giving them pen and paper memory tests of the historical information presented in the game.

The three stages of this pilot study will be examined in further detail below.

Self-efficacy-measurements

Self-efficacy (SE) regards a persons attitude towards one's own capabilities in a specific domain. In this study, since the game in which the teachable agents are incorporated is an educational history game, a questionnaire regarding the students attitude towards history as a school subject was created. Loosely based on an already validated self-efficacy questionnaire for the same age group regarding mathematics, the questionnaire for this study was developed to the the history domain by using good advice from Bandura (2006). The result (Appendix 2) being a seven question long form, which are answered using a 6-point Likert scale. By using this even numbered response scale, any biases towards middle options are constrained. The number of points of the scale was deemed a good mediation between nuances that can be obtained by the young participants introspective capabilities and the detail of the resulting data. I might be of importance to point out that the questionnaire was not validated.

The participants SE-data was used to divide them into two, in respect to the obtained measurements, equal groups, which would later test different versions of the TA. The division process was simply to give each participant either a TA- or a TTA code by first rank all SE scores and then give the first student on the list a TA code, the second a TTA code, and so on.

This division, however, does not include the gender variable, which will be discussed later.

Exposure of TA/TTA

The second stage of this study was letting the participants play the game. Because of the limited set of available computers, they were divided into four groups, letting them play the game one group at the time. The participants played the game in their own school environment, which made it difficult to control all interactions between the students while playing the game, even though there were few of these incidents. The game was also played with headphones, which disabled the participants to get audible input from other players game progress. The participants were let to play the game for approximately 30 minutes each.

After the game playing phase, the participants were told to stop and fill in a digital evaluation form (Appendix 3), making statements about how much they liked the different parts of the game. The response for the first five statements was given through a 7-point Likert scale, with two extra questions which participants could respond to trough free text.

Memory test

The third stage of the study was to test the participants acquired information from the game. This was done 2-3 days after the playing stage by simply letting them fill out a pen-and-paper memory test of the information presented in the game (Appendix 4). The test took approximately 5-10 minutes to do. The tests was then given total scores from a member within the project group. Every fact written in the test gave one point, except if the student remembered exact dates that where asked for, which instead gave 2 points. A single score point was instead given if the student was in a hundred year span of the year asked for.

Results

The 38 subjects were divided to TA or TTA-characters based on Self-efficacy (SE) score. With gender the following distribution was acquired: (i) 9 male with TA, (ii) 12 male with TTA, (iii) 10 female with TA, and (iv) 7 female with TTA. Since gender was accounted for in the assigning of TA/TTA, the distribution of SE was as seen in Figure 3.1. Since the distribution was unbalanced across gender, this variable was included with agent-type in the explorative, multiple-regression analysis that follows.
The distribution of the both gender's SE-estimations was also uneven as seen in Figure 3.2. Five subjects ranged from 25-28, where the main group ranged from 30-40, forming a more unimodal, normal distribution.

In the subsequent analysis the five subjects discussed were removed with the SE score (4 boys + 1 girl), unless stated explicitly. Across the analysis, stronger p-values were noted. A second motivation for this was post-study information of reading and writing disabilities in the class studied, which may follow by a low SE in school subjects.

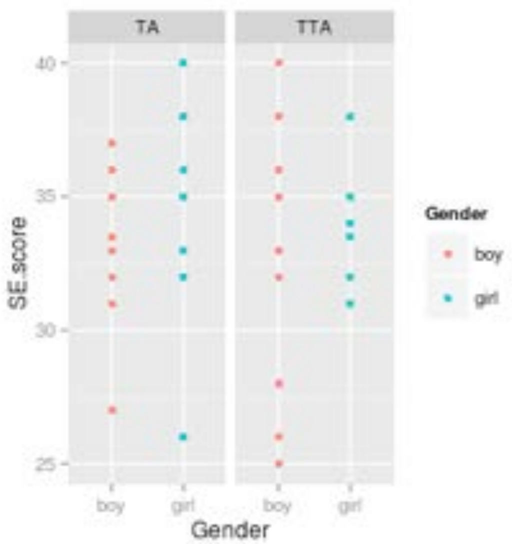


Figure 3.1. Distribution of TA/TTA and gender across Self-Efficacy estimations.

Figure 3.2. Distribution of Self-Efficacy scores.

After removal, an almost equal SE for male/female subjects was acquired, with a mean (M) of 35.03/34.74 and standard deviation (SD) 2.52/2.53, respectively.

Our main focus concerned differences or similarities in the memory test between agent settings. No significant difference could be found (p-value 0.96). The skewed gender distribution was accounted for, and there were however a significant difference in memory test result between genders (lm[Memorytest~Gender+Agent]), as shown in Figure 3.3. Males had M=5.61 (SD=2.52) and females M=2.96 (SD=2.17), with p-value 0.0032 (**) of the difference.

No significant correlation of memory test score and SE-score (p-value 0.44) was found, but accounting for SE made a stronger gender correlation (p-value 0.0012).
No significant difference was found in Likeability between either gender or agent. A p-value of 0.15 was noted for difference in “Likeability of the agent” (LA), where females had a M=5.44, and males M=4.61 (on an increasing scale of 1-7). Another p-value of 0.23 was noted for “Likeability of game” (LG) compared to TA/TTA, with a difference of 0.36 between the two.

There was however a significant correlation between SE-score and both Likeability indicators. As Figure 3.4 and 3.5 shows, higher SE was correlated to lower LG (4) , but higher LA (5), with p-values 0.057 (4) and 0.035 (5). Accounting for gender made the correlation stronger (p-value 0.048/0.29 respectively), but including agent-type made it go back to about the previous p-values.

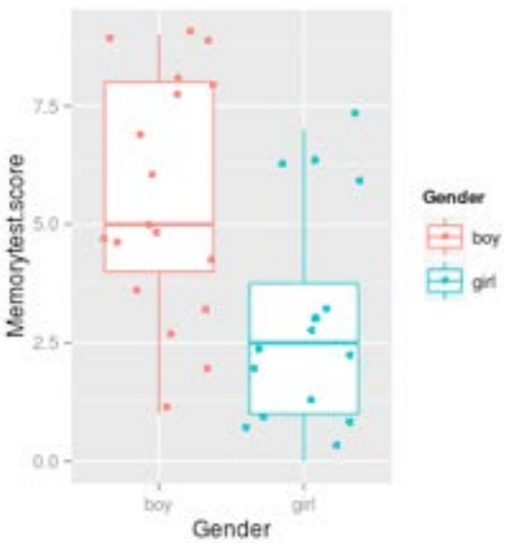


Figure 3.3. Gender difference of how much is remembered of the history facts in the game.

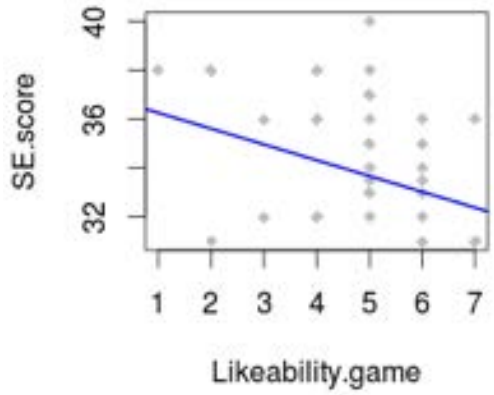


Figure 3.4. Self-efficacy correlated negatively to Likeability of the game.

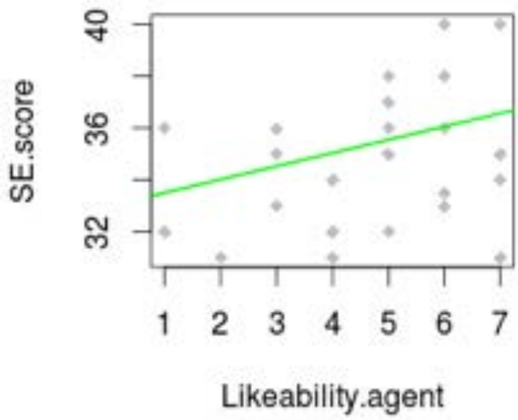


Figure 3.5. Self-efficacy correlated positively to Likeability of the agent.

The graphs seemed to indicate an interaction and analysis can be seen in Table 3.1 below. P-value of the interaction is 0.12 (n=33), but one possible interpretation otherwise is that LA, likeability of the agent, is the stronger variable of the two, where LG seems to follow on the first.

Table 3.1. Multiple-regression interaction analysis between

	Estimate	St.err	t-val	p-value
Intercept (SE)	29.868	3.7045	8.063	.86e-09 ***
LA	1.6541	0.7301	2.266	0.0314 *
LG	0.7268	0.8873	0.819	0.4196
LA:LG	-0.2658	0.1660	-1.601	0.1205

Likeability of the Game (LG) and the Agent (LA), correlated to Self-efficacy.

Further analysis showed a possible correlation between memory test score and LA, where subjects seemed to remember more depending on Likeability of agent (LA) (p-value 0.062, n=38). If including SE (n=33) and LG correlations were stronger (p=0.019), however including the gender variable made correlation weaker (0.12). The interpretation of this is further discussed in the next section.

Discussion

The Gender-distribution as shown in Figure 1 was a unbalanced, however including this variable in the multiple regression analysis that was executed, it should have been accounted for. However, since the study only included 38 participants, dividing them by 4 (on two variables) the resulting groups was quite small, so interpretations of significances throughout the report should perhaps have this fact kept in mind.

There was however a quite distinct difference in memory test scores between genders, which seems hard to account for, even with the lack of numbers. Males were much better at remembering facts, and received almost twice the test score. Why this is has perhaps two explaining variables, the limited time spent in the game, and a persistent better familiarity of 3D-games for boys. The game, more so when interacting with the historical characters, consisted of a quite complex maneuverability system to the 3D-game, as in a 1st-player-shooter, rotation with mouse, and multiple keyboard-keys for actions. The subjects only played the game for 30 minutes, of which in general 10 mins was in the Teaching-activity (with the TA). So perhaps girls had to use more cognitive resources on the spatial task rather than on remembering facts. With time, by playing the game more, it is hypothesized that females would catch up and familiarize themself with the maneuvering system.

Perhaps a bit counter-intuitive, opposite correlations were found for Likeability of the game and of the agent. There are perhaps two possible interpretations of this and the first is that what seemed to be the strongest variable of the two were the likeability of the agent (LA). LA in itself was supposed to capture how they deemed the different agents, but perhaps it is more likely that what they associated the agent with was the teaching activity in itself (as can be seen in figure 1.1). LA was also indicated to be positively correlated to how much they remembered. Perhaps a possible conclusion then, is that it is more important to have a stimulating teaching activity rather then the “fun” game overall, if student should remember more.

An alternative interpretation on the opposite correlations would perhaps be that it would make sense that to have stronger self-efficacy regarding history, will make for more and quicker stimulation by the teaching activity, and make the player subsequently less eager to play the game again, because they have done more of what the game was intended to do. The Likeability-question was intended to be a more precise question on how much they wanted to play the game again. In other words, they can be said to have already *“been there, done that”,* and to a higher extent wants to do something else next.

Future Troublemaking

The main focuses in this pilot study was to find differences between TA and TTA, but no significant numbers were found. One could see it as a result in itself, that confirming and questioning statements did not elicit much of a difference. However, this study is deemed a pilot and is centered on questions that are planned to be investigated further. One thing is certain, that a greater number of subjects is necessary for adequate data analyses. Out of the 38 subjects, 33 was left after they were removed by consideration to their self-efficacy scores, which seems too few to divide into further subgroups by different variables (e.g. gender and TA/TTA).

One important factor to keep in mind is the time. As already stated, the participants overall only spent 10 minutes with the Teaching Activity, where the agents were to be taught, and only 3 of 38 participants finished the game, as in finished the matrix themself and then taught/corrected the agents attempt to fill in the same matrix. Time-logs of the game also reveal that they pop in-and-out of the Teaching Activity, to the 3D-world getting the answers, then back in again and stayed anywhere from 10 seconds to 2 minutes. Which was intended, but it may give little time for social interaction with the agents specific personality and the agents overall may get an *insignificant* part of the game.

In previous TA-games, as for example “Betty’s Brain” which was used in studies presented by Chase et al. (2009), subjects played the game on and off for two months. Other studies also have a trend of having more sessions and longer time periods. Our goal was also to stimulate “understanding”, not to only be able to repeat facts, and for this it could also be argued that a longer period of time is needed.

Except changes in the study, there are also possibilities to change the game, and I will here discuss two. First, the social interaction part of the game could be made more important and play a larger role in the game. In our system one could ignore the statements of the agents altogether, as they were only that, statements. Another possibility is to force answers through multi-option questions, which was discussed during development, or if possible a chatbot to interact with and to explain questions that are asked. Also, the very limited amount of time subjects actually played the game may again account for the lack of difference. To draw any conclusions about if, or how, the different agent settings affect learning or likeability, subjects would probably have to play the game during a much longer time span.

Secondly, a different kind of teaching activity could produce clearer connections. The matrix is a simple way of showing lots of facts, but it is not necessarily intuitively clear how they are combined. The TA was supposed to be the mediator, connecting the dots of facts, but as already stated perhaps the part it played was too small. Other teaching activities were prototyped and started to be developed, with mind-map-, or timeline like designs, which would be able to make clearer "representations of connections", that would perhaps be easier to either be confirmed (TA) or questioned (TTA) by the agent.

Perhaps an overall problem when it comes to questioning-, as well as confirming relationships, is that nobody wants to hear something they already know or understand, or to have something they are certain is right to be (annoyingly) questioned just for the sake of it. Questioning and confirming relationships at random may have this effect and it is perhaps difficult to replace the function a teacher in real life can have. As Gärdenfors (2010) argues, one of the most important roles a teacher is to be an expert in its specific area, to be able to recognize where its students are in their own *understanding*, and to know what to trigger in them next. What new example of a positive relationship to present to them, or what negative relationship to present, question what they *believe they understand.*

References

Bandura, A. (2006). Guide for Constructing Self-Efficacy Scale. In: F. Pajares & T. Urdan (Eds.), *Self-Efficacy Beliefs of Adolescents* (pp. 307-337). Charlotte: Information Age Publishing, Inc.

Blair, K., Schwartz, D., Biswas, G., & Leelawong, K. (2006). Pedagogical agents for learning by teaching: Teachable agents. *Educational Technology & Society,* Special Issue on Pedagogical Agents.

Brophy, S., Biswas, G., Katzlberger, T., Bransford, J., & Schwartz, D. (1999). Teachable agents: Combining insights from learning theory and computer science. In S.P. Lajoie & M. Vivet (Eds.), *Artificial Intelligence in Education* (pp. 21-28). Amsterdam: IOS Press.

Chase, C., Chin, D. B., Oppezzo, M., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*.

Gärdenfors, P. (2006). *Den meningskapande människan.* Natur och kultur. Stockholm.

Gärdenfors, P. (2010). *Lusten att förstå.* Natur och kultur. Stockholm.

Pareto, L. , Arvemo, T. , Dahl, Y. , Haake, M. & Gulz, A. (2011). A Teachable-Agent Arithmetic Game's Effects on Mathematics Understanding, Attitude and Self-efficacy. *Lecture Notes in Computer Science*, Volume 6738/2011, pp. 247-255.

Seyama, J. & Nagayama, R. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence,* Vol. 16, No. 4 (pp. 337-351).

In Balkenius, C., Gulz, A., Haake, M. and Johansson, B. (Eds.), Intelligent, socially oriented technology: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 154, 2013.

Exploring the Use of a Teachable Agent in a Mathematical Computer Game for Preschoolers

Erik Anderberg & Anton Axelsson
Lund University Cognitive Science
Faculty of Humanities
Lund University

Sanne Bengtsson, Maja Håkansson & Lisa Lindberg
Department of Design Sciences
Faculty of Engineering
Lund University

Learn for oneself has long been the most prominent learning framework applied in educational institutions. Conversely, studies have shown that when teaching, we in fact excel at learning far more than when studying for ourselves (Bargh & Schul, 1980). This has recently been embraced within educational technologies, facilitated through Teachable Agents (TA). Studies carried out on primary school students using TA:s showed them to be beneficial for the students development (Biswas, Leelawong, Schwartz, Vye & The Teachable Agents Group at Vanderbilt, 2005; Blair, Schwartz, Biswas & Leelawong, 2006). With this pilot study, we wished to initiate research looking at whether TA:s can be used as far down as preschool age. A research tool in the form of a mathematical game was developed and tested in an experiment with 10 preschool children. The results showed that there were no apparent downsides in using TA:s with preschool children, and thus it is concluded that further research can and should be carried out.

Keywords: teachable agent, number sense, theory of mind, preschoolers

Introduction

Through a multitude of studies it has been established that humans treat computers and other media as social entities. This phenomenon has been dubbed the *Media Equation* (Reeves & Nass, 1996). For example, humans can be just as flattered by a computer's praise as when received from a fellow human (Johnson, Gardner & Wiles, 2004). Social integration is increasing within digital environments with the use of a wide variety of pedagogical agents (Haake, 2009). They are manifested, amongst other, as teachers, embodied chat bots, instructional tutors or digital representations of oneself (Avatar). Recently, pedagogical agents have also been incorporated in educational software for primary school students in order to facilitate *Learning by Teaching*. The purpose of this pilot study is to investigate whether the benefits of these pedagogical digital characters, called *Teachable Agents*, can be extended down to children of preschool age, more specifically, four-year-olds. This is an age where children are developing a *Theory of Mind* (Box 1). The question is whether a four-year-old can comprehend and make use of a Teachable Agent or if that would require higher social skills which are encompassed in a fully developed theory of mind.

Advantages with Educational Software

There are obvious benefits to digital teaching environments. First of all, teachers can more easily follow the progress of the students and carry out formative evaluations in order to assess what areas the student needs to improve upon. Educational software can be developed so that the competence level of the student can be adhered to by adjustments carried out by the teacher, or even automatically matched by the software itself. Another advantage is the possibility of carrying out virtual exploration that otherwise would be physically impossible or even lethal to perform, such as physics or chemistry experiments, but also alteration of characteristics in students that would not come about naturally in a classroom (c.f., *Proteus Effect*, Yee, Bailenson & Ducheneaut, 2009).

Teachable Agents and Learning by Teaching

A Teachable Agent (TA) can be described as an autonomous, digital student utilised in computer aided learning. The TA is incorporated in educational computer software, commonly mathematical games. The idea is that a student takes the role as a teacher in order to tutor the TA. This role switching encourages the student to take responsibility for someone else's learning (c.f., *Protégé Effect*, Chase, Chin, Oppezzo & Schwartz, 2009) within the Learning by Teaching framework (see e.g. Biswas et al., 2005; Brophy, Biswas, Katzlberger, Bransford & Schwartz, 1999). Thus, the student learns in order to teach. It has been shown that teaching others is a very effective way of learning for oneself (Bargh & Schul, 1980). This is because in order to teach someone else you will need to be well-read and prepared. Some great advantages of using a TA for this purpose is that no actual student suffers from a poor teacher, and also, this paradigm creates an ego-protective buffer for the real student because poor results can partly be attributed to the TA (Chase et al., 2009). The largest gains in using a TA have been observed in low achieving students (Chase et al., 2009).

General Purpose

With this study we aim to contribute to the development of educational games for preschool children, specifically games facilitating children in their acquirement of *Number Sense*. The purpose of our study is to examine whether preschoolers can make use of a TA, something that, to our knowledge, hitherto has not yet been explored. The studies carried out so far, on the subject of TA:s, have all had participants of primary school age or older. As discussed above, it has been established that TA:s are a useful means of assisting learning in digital educational environments. We therefore expect that the use of a TA in a mathematical game for preschoolers should yield positive effects on preschool children's educational development. The purpose of this precursory study was two-fold. Firstly, the aim was to develop a mathematical computer game targeted at four-year-old children to be used as a research tool. Secondly, the aim was also to enquire into the question whether children at the age of four can grasp the concept of a TA.

Preschoolers' Cognitive and Mathematical Development

The age of four is a very active period with regards to cognitive development in human beings. A prominent cognitive shift in children of this age, is their ability to symbolically represent concrete objects and actions through play (Piaget, 1952), for example pretending that a shoe box is a boat. In general, they are also beginning to develop problem solving skills by making use of physical features, and reasoning about cause-and-effect relationships. They also start developing an understanding for concepts, classes and hierarchical relationships; make distinctions between reality and dream or make-believe, as well as between introspective beliefs and public opinion (Seefeldt & Wasik, 2005).

Number Sense. Children of an early preschool age have a very vague understanding of numbers and their relations. When presented with two lines of sweets of equal amounts, it is not uncommon for children around the age of four to believe one of the lines to contain a larger amount of sweets if the sweets in that line are distributed further apart. For this child, larger (or longer) means more. The same applies for understanding of conservation of quantity where a child at this age believes a tall glass to contain a larger amount of juice than a wide one regardless of whether they in fact can hold the same quantity (Seefeldt & Wasik, 2005). Definitions of Number Sense are notoriously convoluted, and researchers have found it difficult to crystallise the term (Griffin, 2004a). There are also ambiguous views on how to interpret Number Sense (Howell & Kemp, 2009). From a cognitive perspective, Number Sense can be seen as an understanding of the meaning of numbers and the ability to make comparisons, as well as showing proof of fluency with numbers (Gersten & Chard, 1999) and an understanding that they relate to quantities (Griffin, 2004b).

Box 1. Glossary of terminology

Avatar A digital representation of a person. Most commonly a picture used on the internet to represent the person digitally, but can also be a character representing the player of a computer or online game such as World of Warcraft (Ahn, Fox & Bailenson, 2012).

Ego-Protective Buffer Learning by Teaching, in combination with a Teachable Agent, creates an Ego-Protective Buffer in the sense that the student does not have to blame him- or herself completely for failures. The blame can instead be shared with the TA (Chase et al., 2009).

Learning by Teaching A pedagogical framework where students take the role of a teacher in order to accelerate their own learning through the Protégé Effect (Bargh & Schul, 1980; Blair et al., 2006).

Media Equation The theory that humans attributes social features to computers is referred to as the Media Equation (Reeves & Nass, 1996).

Protégé Effect Students who are encouraged to teach others work harder to understand material because they feel a responsibility for their scholars (Chase et al., 2009). This phenomenon has been dubbed the Protégé Effect.

Proteus Effect The observed effect of game players taking on traits possessed by their avatars is known as the Proteus Effect (Yee et al., 2009).

Sally-Anne Test An experimenter uses two dolls — Sally and Anne —, two boxes and a marble. The marble is placed in one of the boxes. The experimenter hides Sally explaining to the participating child that Sally has left the room. Anne then picks up the marble from the first box and places it in the second box, whereafter Sally returns. The child is then asked in which box Sally will look for the marble. If the child says that Sally will look in the box the child knows the marble is in, it would indicate that the child does not have a fully developed Theory of Mind.

Theory of Mind (ToM) A meta-cognitive concept that other people have a mind, knowledge, and feelings of their own. This trait is gradually developed alongside language comprehension and is usually fully acquired around the age of four (Sodian & Kristen, 2010).

From an educational curriculum perspective, Number Sense can be divided into five categories: (1) counting — understanding one to one correspondence, order, and cardinality; (2) number knowledge — discriminating quantities and magnitude comparisons; (3) number transformation — addition and subtraction; (4) estimation; (5) number patterns — discerning numerical relationships (Jordan, Kaplan, Oláh & Locuniak, 2006). Most of these abilities usually emerges in children through normal interaction with parents and siblings. If they do not emerge, or if children do not develop them enough during their time at nursery school, difficulties in understanding more complex mathematics will most likely occur one the child starts primary school (for overviews on mathematical disability and its relation to Number Sense, see e.g., Berch, 2005; Gersten & Chard, 1999; Jordan et al., 2006). Number Sense can be taught (Griffin, 2004b) and for children who have not been exposed to numerical reasoning at home, formal training of Number Sense is essential (Bruer, 1997). Educational mathematical software should well fit the purpose for formal training of Number Sense at nursery school. The ability to customise such software is essential to meet the needs for individual tutoring due to preschooler's varying socio-economical backgrounds and upbringings. This in conjunction with the use of a TA could potentially be an enhancement in a child's acquisition of Number Sense. Partly due to benefits such as the ego-protective buffer, but also because the TA can provide consequence feedback, as well as mathematical reasoning geared towards the preschooler's level of Number Sense.

Theory of Mind. Children who, before the age of four, are told that there are equal amounts of sweets in two lines, or that two glasses can hold the same amount of juice, will still base any choice on what they consider the most salient feature, length and height. They will generally not follow the advice of others, because children of this age usually view the world from a highly subjective perspective (Piaget, 1969). The child eventually acquires a theory of mind (ToM) and shifts perspective to one including knowledge of others' feelings and desires. There is no fixed age at which this shift come about. Instead, it is dependent upon language development, as well as general social interaction and social guidance from people in the child's surrounding. However, generally all six-year-olds have a fully developed ToM and develops gradually from around the age of three. One possible caveat with introducing TA:s to preschool children is that those that do not have a developed ToM might not be able to reap the benefits from it. One important aspect that a student will need to understand is why a TA might reason incorrectly when solving a problem. This requires an understanding that others' knowledge and reasoning is separate from yours. Thus, preschoolers might not comprehend the intentions of the TA, and the child might instead be distracted by it. The first step was therefore to develop a research tool which could be used as a formal tool for supporting training of Number Sense, into which a TA could be integrated to carry out a pilot experiment investigating these concerns.

Tool Development

Overall Game Concept: The Lost Chicks

The game concept developed for this study revolves around a flock of chicks that are blown out of their nests and need help to get back up. The player helps the chicks return home via a lift, by pushing buttons on a lift panel. The chicks, one at a time, hold up a number of feathers representing the floor they live on, and the player matches this number with the correct button on the keypad of the lift panel. After a while, a TA is introduced and observes the player using the lift in order to learn how to operate it. Further on, the TA takes over and the player now corrects the TA any time the player believes the TA makes a mistake. The idea behind using a lift is that it represents a vertical number line. It gives a good representation of parts of the whole — branches as floors — and that higher numbers are further up. It is also important, in a child's numerical development, to use familiar concepts (Griffin, 2004b; Hannula, Mattinen & Lehtinen, 2005), and lifts are a common feature in our society with mathematical properties.

Pretest

A pretest was conducted in order to test the game concept on preschoolers using a prototype of the game. The pretest had three purposes: (1) to see whether the game concept engaged the children, (2) to find out if the children could grasp the idea of a TA, and (3) to test which number representation to use in the digital version of the game. Two experimenters carried out the pretest. An experiment leader interacted with the children and gave them instructions, whilst an experiment observer took notes.

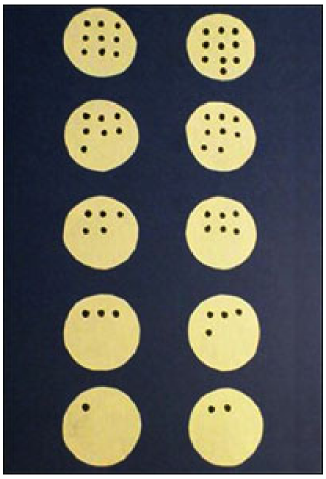
Method

Participants. Six children from the Rida Ranka nursery in Lund, Sweden, participated in the pretest. Four of these participants were 4-year-old girls, one was a 4-year-old boy, and one was a 3-year-old boy.

Material. A lo-fi paper prototype of the above described game was used in the pretest (Figure 1). The paper prototype consisted of a tree with ten numbered branches (Figure 1a), each holding a nest with a bird mother. There were also ten chicks, a lift (represented by a bucket held up by a rope), and three keypads with ten buttons representing a lift panel. Each keypad had separate number representations: (1) dots (Figure 1c), (2) hands holding up fingers (Figure 1d), and (3) Arabic numerals (Figure 1e). A soft toy — Stitch — was used as the Teachable Agent (Figure 1b). The observer kept a handwritten protocol.

(a) Tree with lift.

(b) The soft toy Stitch used as the Teachable Agent.



(c) Dot representation.

(d) Finger representation.

(e) Arabic numeral representation.

Figure 1. The lo-fi prototype game set-up used in the pretest. At the top: the paper tree, consisting of ten numbered branches holding a nest with a bird mother and the Teachable Agent. At the bottom: the three keypad lift panels.

Design. A between subject design was adopted. Two of the participants were tested with the Arabic numeral keypad, two were tested with the dots keypad, and two were tested with the fingers keypad.

Procedure. The participants were escorted by their nursery teacher to a secluded part of the nursery where the experiment took place. Before a participant entered the experiment room, one of the keypads were randomly selected. To relieve any scrutinising discomfort, the participants were told that the experiment observer was present in order to observe the experiment leader. The experiment leader then explained the game story and asked whether the participant could help the chicks home. The experiment was carried out in three modes with each participant: (1) self-playing, (2) TA-watching, and (3) TA-playing. Each mode started with the participant being asked to pick a chick. The experiment leader then held up a number of fingers representing the "floor" the chick would like to go to. In the self-playing mode the participant was asked to press the correct button on the lift panel keypad in order for the lift to carry the chick to the correct floor. After the participant had helped two chicks, the TA (Stitch) was introduced to the participant in the TA-watching mode. The experiment leader asked the participant if Stitch could watch whilst they were playing in order for Stitch to learn how to operate the lift and help the chicks. After another two rounds the experiment entered the TA-playing mode. The experiment leader explained that Stitch would try to use the lift, and if Stitch made a mistake for the participant to correct him. Stitch's movements and choices were controlled by the experiment leader. This mode was also played in two rounds, and in total each participant helped six chicks. At the end of the experiment, the participant was given stickers as an appreciation for taking part.

Results

Time-on-task for each of the modes for all participants is presented in Table 1. What became apparent was that the participants identified numbers from 1 to 3 with ease, without counting them, a principle known as *subitising* (Fischer, 1992). For larger numbers (> 6), however, two participants struggled and were somewhat reluctant to count dots and fingers on these versions of the keypads; expressing their annoyance by sighing or pretending to be thinking.

The participants found the layout of the lift panels confusing. The buttons were ordered from the bottom left to the top right (see e.g. Figure 1e). Problems occurred when participants wanted to correct a mistaken number. For example, if they wanted a button with a higher number they would choose the button directly above the one they had just pressed. However, higher numbered buttons were either situated directly to the right, or diagonally above the current button.

The participants tested with the Arabic numerals had no problem identifying the numbers, and the task was too easy as indicated by the measurements of time-on-task presented in Table 1, especially for the last participant.

No participant showed any sign of boredom, they all laughed or smiled at some point whilst playing the game. They all got excited when the chick was reunited with its mother. If they sent the chick to the wrong floor, they went silent, but they did not give up, instead they tried again without hesitation, until they got it right.

An observation made by both experimenters was that the two male participants were engaged and excited throughout the experiment. Conversely, the four female participants were more reserved initially, but grew excited the more Stitch was involved in the experiment, and they were as excited and engaged as the males in the last game mode when Stitch was playing by himself. Whether this difference is individual, mood or gender specific cannot be established by this experiment due to the small sample size.

All but one participant corrected Stitch when he was making mistakes.

Conclusions

In general, the children participating in the pretest seemed pleased with the overall game concept. They also had no problem interacting with the TA, and appeared willing to show, as well as help him. Some of the children's enthusiasm even grew when the TA was introduced.

From the pretest it was concluded that some adjustments to the digital version of the game had to be implemented.

- The use of Arabic numerals on the lift panel was too easy, whilst the use of dots was challenging for the children in this pretest. Therefore it was decided that finger representations was used. Although, the artwork of the fingers in the paper prototype was somewhat confusing and therefore had to be redesigned.
- The bi-columnal structure of the buttons with a vertical increment on the lift panels perplexed the children and had to be rearranged in a more conventional horizontal number line.
- The children struggled with numbers as large as nine and ten, therefore the number of branches was limited to eight.

Main Experiment

Method

Participants. Ten children, aged 4 to 5.5, from the Rida Ranka nursery in Lund, Sweden, took part in this experiment. Six of the participants were female and four were male.

Material. Utilising the results of the pretest, a digital version of the game was developed for the main experiment using HTML5 and JavaScript (Figure 2). A panda, named Panders, was incorporated in the game with the role of a TA. Usability tests of the digital version of the game, conducted on a convenience sample of three children aged 4 to 9, revealed that some children had problem counting numbers larger than 5. Therefore the application was developed so that the number of branches could be dynamically adjusted to suit individual participants mathematical abilities. Spoken instructions were recorded and incorporated into the game in order to make it as self-explanatory as possible. Sound effects were added to emphasise consequence feedback. For example, every time the lift passes a floor a pinging sound is played, and when the lift reaches the intended floor a different pinging sound is played to signal cardinality.

Design and Measurements. A between subjects design was adopted with TA as an independent variable with two levels, presence and absence, and False Belief with two levels, true or false, as well as Sex and Age. The dependent variables measured were Commitment and Concentration. Commitment can be described as how engaged or dedicated the participants are in playing the game, and would manifest itself through, for example, pointing or laughing. The opposite would be a participant who is perceived as bored. Concentration can be described as how focused the participant is on the task at hand, by showing signs of absorption in thought, such as starring or wide open mouth. The opposite would be a participant who is perceived as engaging in activities irrelevant to the task. During the experiment, the participants were filmed with an unobtrusive web camera situated above the experiment laptop screen. All mouse events were logged. In order to ensure homogeneity between the two levels of

Table 1
Time-on-task for each participant in the pretest measured in minutes and seconds.

	Dots		Fingers		Numbers	
Participant	1	2	3	4	5	6
Self-playing	05:54	04:30	04:18	06:03	03:35	01:26
TA-watching	03:03	02:30	02:30	02:30	01:40	01:29
TA-playing	01:19	01:25	01:13	01:30	02:03	01:26
Total	10:16	08:25	08:02	10:03	07:18	04:23

Figure 2. The final digital version of the game showing the tree with nests, the lift going up, and the TA watching.

the independent variable, a priority algorithm was developed. This algorithm had to be developed because the experiment was to be carried out in an average day of the preschoolers when they were all engaged in daily activities. Therefore the participants could not be balanced in advance. Our biggest concern was that children lacking in ToM might not grasp the concept of a TA. Therefore, it was of utmost importance that non of the two groups had a predominance of children with underdeveloped ToM. It was concluded that a quick and dirty way of balancing the groups between children with and without ToM was to divide the children equally over age (Priority 1). Due to the indications in the pretest that the female participants were more affected by the introduction of the TA, we next balanced the participants by sex (Priority 2). Lastly, we balanced children over a simple false belief task (FBT) as second indicator of ToM (Priority 3). If all above priorities were balanced we simple let the participant use the TA version of the game (Priority 4). A protocol was designed to facilitate the balancing during the experiment (Appendix A).

Procedure. The participants were escorted by their nursery teacher to a secluded part of the nursery where the experiment took place. Before playing the game, the participants were tested in a FBT. For this purpose, an adoption of the Sally-Anne test was used (Box 1). A giraffe and a fox were used as substitutes for Sally and Anne. In order to pass the FBT, the participants would first have to point at the correct box, and also give a coherent account for their choice. A small assessment on the participants counting skills were then carried out. The experiment leader held up eight fingers and asked the children to tell her how many fingers she held up. Those who struggled with counting passed five were assigned to play the game with a maximum of six branches, or floors. Those who were able to provide an answer with more ease were assigned to play the game with a maximum of eight branches. The rationale behind this was for the children not to feel discouraged by the games level difficulty. Four children ended up playing with six branches, and six children played with eight branches. After these pre-experiment tests, the participant was delegated to play the game either with the TA or without. Those participants who played with the TA first helped three chicks to their nests. Then the TA entered and watched the participant help three more chicks. Lastly, the TA tried to help three chicks by suggesting which button should be pressed. The participants then had to press a green button with a tick if the TA was correct, or a red button with a cross if the TA was incorrect. Whenever participants pressed the red cross, they were asked by the TA to show him which button he should have chosen. These three different game modes are depicted in Figure 3. Participants who played without the TA simply helped nine chicks in total, by themselves. When a participant finished the game, the preschool teacher was called back into the room, and the child was asked to explain to her what the game was about. Those who played with the TA were also asked to explain its role in the game.

Coding

Overall Commitment and Overall Concentration was subjectively rated by the experiment leader and observer (henceforth experimenters) during the experiment, and later, recordings of the participants were rated by three control observers (henceforth controllers). Overall Commitment and Overall Concentration were both rated on an ordinal scale from 1 to 7 where 1 represents fully unfocused/uncommitted respectively, and 7 represents fully focused/committed respectively. Two controllers rated three videos and one controller rated

(a) Self-playing mode. (b) TA-watching mode. (c) TA-playing mode.

Figure 3. The digital version's three different game modes.

four videos. The recordings were mute so that the observers could not tell whether the participant was playing with a TA or not. One of the controllers then split each video into three clips. The three clips for participants playing without the TA consisted of them helping three chicks by themselves. Regarding participants playing with the TA, the clips matched the three game modes (Figure 3). Each clip was given a filename of a random letter so that participant order and clip order could not be recognised. The resulting 30 clips were then randomly divided between the experimenters, whilst all of the clips were given to the two remaining controllers. The rationale behind this was that no rater should be able to tell how far a participant had progressed in the game, and also to make it more difficult for the raters to recognise whether a participant was playing with a TA or not. The experimenters and controllers then rated participant Commitment and Concentration for their allocated clips on the same 7-point ordinal scale as used with Overall Commitment and overall Concentration.

Data Screening

When analysing inter-rater reliability for the 30 clips, Spearman's rho revealed that there were no consensus among the raters of what to consider Concentration. This variable was therefore excluded from any further analysis. In contrast, agreement as how to rate Commitment was sufficiently correlated. Subsequently, a mean of Commitment was calculated for each clip from all raters and utilised in the succeeding analyses. The results of the inter-rater reliability analysis is presented in Table 2. From the three Commitment ratings, resulting from the clips of each participant, a mean was calculated and used in conjunction with the original Overall Commitment ratings in order to analyse overall inter-rater reliability for commitment. Spearman's rho revealed that there was a significant correlation between these measures (Table 3). It was thus concluded that the mean of Commitment for the three clips of each participant was a sufficiently reliable measure of Overall Commitment.

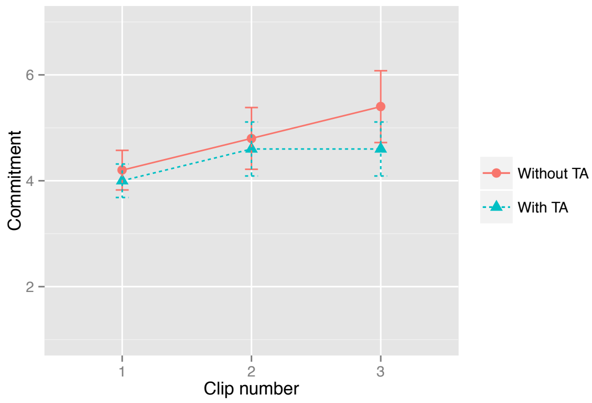


Figure 4. Mean Commitment of the ten participants for each of the three clips (± 1 SE).

Results

Quantitative Analysis. An analysis of covariance revealed that there were no significant differences of Commitment in participants playing with or without a TA (Figure 4). Commitment generally increased throughout the game. There was, however, a slight tendency that the older participants were less committed to the game than their younger peers as indicated by the plot in Figure 5. There was a marginally higher Commitment in the group which played without the TA when compared to the group which played with the TA. It should be noted that the group which played with the TA also was slightly older (M_{age} = 4.72) than the group which played without (M_{age} = 4.44). This could account for the difference. The indications of the pretest that female participants would be more committed was not corroborated in this experiment.

Qualitative Analysis. The participants were in general pleased with playing the game and seemed very much to enjoy it. There was a lot of laughter and surprised faces during game play. Though, one of the oldest participants was quite bored of the game, and was not shy to make that clear when

Table 2
Spearman's rank correlation coefficient for ratings of the experimenters and two controllers for the dependent variables: Commitment and Concentration.

		Experimenters	Controller 1
Commitment	Controller 1	.585*	
	Controller 2	.676*	.640*
Concentration	Controller 1	.434	
	Controller 2	.330	.399

*$p < 0.01$ (two-tailed).

Table 3
Spearman's rank correlation coefficient for Overall Commitment mean of ratings over the 30 clips, and ratings of the two experimenters and three controllers.

	Experimenters	Means
Means	.681*	
Controllers	.721*	.759*

*$p < 0.05$ (two-tailed).

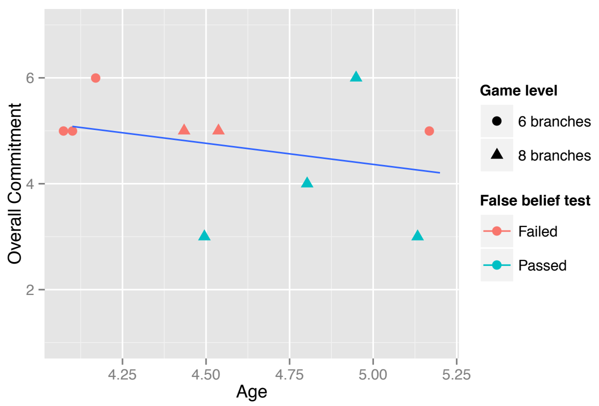


Figure 5. Overall commitment of the ten participants distributed by age. The blue line represents the fitted linear model.

asked. The participants gave very vague answers to the role of the TA, and the answers were very similar to what they had heard the TA explain. This is, on the other hand, expected of children of this age. A few answers were quite satisfactory, such as "I was supposed to help him", "The panda was thinking right or wrong" and "Panders was supposed to help out". One participant claimed that she did not remember what the TA was there for. This participant seemed afraid that she would give an incorrect answer. The youngest participant who played with the TA, seemed very reluctant to correct the TA. He continuously pressed the green tick button in the TA mode, even when the TA had guessed incorrectly. This participant struggled with counting the experiment leaders finger above the number five, and did not pass the FBT. To the post-test question of the role of the TA, the participant answered that "he watches". From observing the children during the experiment, there was a notion that at least three of the participants where more motivated to play once the TA was introduced. However, this did not surface in the quantitative analysis of Commitment.

Conclusions

It can be concluded that the game itself holds good quality and it targets its audience very well. The intended target group is four-year-olds, and we saw indications that it can be too boring for children who are older, or that have more developed math skills. The use of a TA did not impede on commitment to the game and was not obtrusive for the players. In general the children seemed at ease in interacting with the TA. Some developments of the game are needed for it to be used in further research. For example, the game should probably be built to work on tablet rather than a computer because the children had some problem using the mouse to navigate in the game, or at least a touch pad should be used which seems to be more intuitive for children. It might also be beneficial to integrate overt reasoning in the TA when it is making choices in order to elevate the preschoolers' mathematical and logical reasoning.

Discussion and Overall Conclusions

As expected, it was generally the older participants who passed the FBT, although one five-year-old did not (Figure 5). Thus, the rationale of using a balancing protocol was warranted, and two fairly homogenous groups were achieved. Although, as previously acknowledged, the TA group did have a slightly higher age mean. A within subject design could possibly have been used with balancing of the order that the TA was introduced. In that way all participants could have played with the TA, and also this would enable you to measure differences within the group of playing with

Table 4
Summary of participant statistics.

	Participant											
	1	2	3	4	5	6	7	8	9	10	Mean	SD
Age	4.4	5.2	4.1	4.2	5.1	4.9	4.8	4.5	4.1	4.5	4.58	0.39
Sex	F	F	M	F	M	F	F	M	M	F	-	-
Passed FBT	No	No	No	No	Yes	Yes	Yes	No	No	Yes	-	-
Game	TA	TA	Self	Self	TA	Self	TA	Self	TA	Self	-	-
Time-on-task	09:29	10:46	08:28	08:30	05:38	08:27	05:28	05:41	10:35	6:35	07:58	01:56

and without a TA. This may have yielded stronger results. A within-subject analysis could not have been carried out with the present data, because it is not balanced and there is therefore no way of controlling for carry-over effects. Carry-over effects seems to be present when looking at the plot in Figure 4, which shows that the mean Commitment increases during play. This could possible be explained by the difficulties children had in operating the mouse, which led the experiment leader to explain to the participants how to use it. It can of course also be a factor of the participants becoming more familiar with the experimenters and the game. As previously mentioned, one child expressed that he found the game boring, another child seemed quite bored whilst playing. This could possibly be rooted in the fact that children today are used to computer games from home and expect faster interaction and more stimuli. Therefore, if the game is to be used with older children, and perhaps also with children who are more mathematically advanced, further assessments must be made, and the game must be developed to be dynamically adjusted to the age and mathematical skills of the player. The TA developed for this particular study is very limited in any artificial intelligence. Albeit it serves the purpose for a research tool, for a more formalised educational purpose it will have to be improved. It should, however, be noted that a properly developed TA for preschoolers will have to look and behave somewhat differently from your average TA used for primary school children. This in order to make sure it levels with the preschoolers cognitive development. Regarding the participant that kept pressing the green tick, it could be argued that it was because he had trouble counting. However, he pressed it even when presented with a chick wanting to go to the third level which should not be too difficult through mere subitisation, and he made no errors prior to the TA mode even when presented with the numbers 5 or 6. There can be several explanations as to why the participant had problems with the interaction of the TA. One explanation could be that the interaction design was flawed, in the middle of the participant counting the chicks feathers, the TA started talking, asking if he had chosen the correct button. This could have distracted the participant. Another explanation might be that the participant could not inhibit the urge to press the green button, which can be a sign that executive functions are not yet fully developed. A third explanation could be that he does not want to offend the TA, or the experimenters, by correcting the TA which can be a sign that ToM is not yet fully developed, which, for this participant, seems to be the case judging from the failure on the FBT. Naturally we cannot draw any conclusions from the present study as to what, if any, level of development of ToM a child needs in order to fully benefit from a TA. Though, just by appealing to our innate ability to prosocial behaviour, a child might benefit regardless of whether the child understands the intentions of the TA, or whether the child can reason about the TA:s knowledge. No participant in our study was reluctant to help or interact with the TA, or with the chicks, and all did their best to complete the mathematical tasks. The media equation phenomenon most likely have a part in this.

Limitations

One obvious limitation of this study is the very low sample sizes of both the pretest and the main experiment. Another is the use of preschool children in Lund, which is an academic city with a high socio-economical status. The preschool children of this city is known to be more than a year ahead in cognitive development to their peers in the rest of Sweden. These are convenience factors and this study should only be viewed as a pilot study aimed at developing a research tool for the intended target group. This research tool will be used in future studies facilitating educational technology research in the use of TA:s for preschoolers.

Implications and Future Research

This study has indicated the possibility to continue researching potential benefits of using TA:s with preschoolers. The research tool developed for this project can now be modified to be used in a wide range of studies. One apparent study that should be conducted is a more longitudinal study with a summative evaluation of preschoolers mathematical development where one group continuously plays with a TA and one without. The current tool must then be further developed to house a number of levels that the preschoolers are able to progress. Another interesting study could be to evaluate focus of attention in preschoolers. During the age around four, executive functions are developed. Thus, the tool can be adjusted to facilitate a study to look at how well preschoolers can inhibit distractions in order to keep focused on the TA. The tool could possibly also be developed in such a direction that it can facilitate the development of executive functions and ToM.

Acknowledgements

Special thanks are directed towards Layla Husain who has been serving quality control to our study through valuable comments and proof reading of preliminary reports, and whose valuable advice on interaction design for children we could not have been without.

The authors wish to thank the Rida Ranka nursery school and Annika Janz for helping in the work of this study. Special thanks go to the six children who participated in the pretest, and to the ten children who participated in the main experiment.

Special thanks also go to our supervisors, Professor Agneta Gulz and Dr. Magnus Haake, for making this study possible, and for their valuable insights.

References

Ahn, S. J., Fox, J. & Bailenson, J. N. (2012). Avatar. In W. S. Bainbridge (Ed.), *Leadership in Science and Technology: A Reference Handbook*. SAGE Publications.

Bargh, J. A. & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology*, *72(6)*, 583–604.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, *38(4)*, 333–339.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N. & The Teachable Agents Group at Vanderbilt (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, *19(3-4)*, 363–392.

Blair, K., Schwartz, D., Biswas, G. & Leelawong, K. (2006). Pedagogical agents for learning by teaching: Teachable agents. *Educational Technology & Society*. Special Issue on Pedagogical Agents.

Brophy, S., Biswas, G., Katzlberger, T., Bransford, J. & Schwartz, D. (1999). Teachable agents: Combining insights from learning theory and computer science. In Lajoie, S. P. & Vivet, M. (Eds.), *Artificial Intelligence in Education* (pp. 21–28). IOS Press.

Bruer, J. T. (1997). Education and the brain: A bridge too far. *Educational Researcher*, *26(8)*, 4–16.

Chase, C. C., Chin, D. B., Oppezzo, M. A. & Schwartz, D. L. (2009). Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning. *Journal of Science Education and Technology*, *18(4)*, 334–352.

Fischer, J. P. (1992). Subitizing: The discontinuity after three. In J. Bideaud, C. Meljac & J. P. Fisher (Eds.), *Pathways to number: Children's developing numerical abilities* (pp. 191–208). Hillsdale, NJ: Lawrence Erlbaum.

Gersten, R. & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education*, *33(1)*, 18–28.

Griffin, S. (2004a). Building number sense with Number Worlds: a mathematics program for young children. *Early Childhood Research Quarterly*, *19(1)*, 173–180.

Griffin, S. (2004b). Teaching Number Sense. *Educational Leadership*, *65(5)*, 39–42.

Haake, M. (2009). *Embodied Pedagogical Agents*. PhD thesis, Lund University.

Hannula, M. M., Mattinen, A. & Lehtinen, E. (2005). Does social interaction influence 3-year-old children's tendency to focus on numerosity? A quasi-experimental study in day care. In E. De Corte, G. Kanselaar & M. Valcke (Eds.), *Studia Paedagogica, 41* (pp. 63–80). Leuven University Press.

Howell, S. & Kemp, C. (2009). A participatory approach to the identification of measures of number sense in children prior to school entry. *International Journal of Early Years Education*, *17(1)*, 47–65.

Johnson, D., Gardner, J. & Wiles, J. (2004). Experience as a moderator of the media equation: the impact of flattery and praise. *International Journal of Human-Computer Studies*, *61(3)*, 237–258.

Jordan, N. C., Kaplan, D., Oláh, L. N. & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, *77(1)*, 153–175.

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: International Universities Press.

Piaget, J. (1969). *The psychology of the child*. New York, NY: Basic Books.

Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. Cambridge: Cambridge University Press.

Seefeldt, C. & Wasik, B. A. (2005). *Early Education: Three, Four, and Five Year Olds Go to School* (2 Ed.). Pearson.

Sodian, B. & Kristen, S. (2010). Theory of mind. In B. Glatzeder, V. Goel & A. v. Müller (Eds.), *Towards a Theory of Thinking* (pp. 189–201). Berlin: Springer-Verlag.

Yee, N., Bailenson, J. N. & Ducheneaut, N. (2009). The Proteus Effect: Implications of Transformed Digital Self-Representation on Online and Offline Behavior. *Communication Research*, *36(2)*, 285–312.

Appendix A: Balancing Protocol

A = Med agent
S = Själv

Deltagare	Ålder (3-5)	Kön (P/F)	False belief (J/N)	Spel (A/S)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				

Prio 1

Ålder	Agent	Själv
3		
4		
5		

Prio 2

Kön	Agent	Själv
Pojke		
Flicka		

Prio 3

False belief	Agent	Själv
Ja		
Nej		

Prio 4 = Placera i Agentgruppen