Intelligent, socially oriented technology V:

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

Editors

Birger Johansson

Agneta Gulz

Magnus Haake

Mattias Wallergård

Günter Alce

Diederick Niehorster

Jens Nirme

Eva-Maria Ternblad

Betty Tärning

Lund University Cognitive Studies, 180

Johansson, B., Gulz, A., Haake, M., Wallergård, M., Alce, G., Niehorster, D., Nirme, J., Ternblad, E. M., & Tärning, B. (Eds.) (2021). Intelligent, socially oriented technology V: Projects by teams of master level students in cognitive science and engineering. *Lund University Cognitive Studies, 180.*

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

Table of Contents

Epi - Give a robot a voice

Edvin Boyner
ed6808bo-s@student.lu.se,

Simon Grendeus
simon.grendeus.223@student.lu.se

Pierre Klintefors
bev15pkl@student.lu.se

Alexander Pettersson
har14ape@student.lu.se

This project aimed to give the humanoid robot Epi the ability to communicate emotion through non-verbal vocalizations. These are things like screams and sighs, and "mhm", "huh?", "oh", etc. The vocalizations were created by collecting non-verbal vocalizations from humans, validating them, and synthesizing them using the Soundgen package for R. The vocalizations were compared to typical robot sounds, referred to as non-linguistic utterances. The initial plan was to compare these using an experiment where participants would interact with Epi in person. However, COVID-19 made this impossible. Instead, the comparison was done through an internet survey. The results showed that the affective content of the non-linguistic utterances was slightly easier to interpret, while the vocalizations were perceived as less creepy. However, the good performance of the non-linguistic utterances might be due to a practice effect, and the performance of the vocalizations was brought down by a few particular calls. If these were replaced, the performance of the vocalizations would likely be better.

1 Introduction

Vocal communication has been essential for human evolution. Humans can communicate with each other vocally in many ways, for example without utilizing any language. With that in mind, there are certain vocal sounds that we can understand that are not actual words. The project presented in this report will look further into how to capture and utilize these non-verbal sounds.

Further, to be able to improve the audio feedback of a humanoid robot without increasing the risk of perceiving it as disturbing, there is an opportunity to utilize these non-verbal sounds. Adding these non-verbal sounds to the feedback of the robot can be seen as a way of making the audio feedback rewarding without resulting in an unsettling resemblance to a human.

Additionally, by utilizing these non-verbal sounds it is possible to analyze if they are comprehensible on their own. The project will set out to investigate whether the sounds can be used to improve the feedback of the humanoid robot or not. Further, the purpose of the project is to attain non-verbal communication robot communication.

Background

You might have heard the adage that verbal communication amounts to only 7 percent of our total communicative behavior while 93 percent is non-verbal. Non-verbal communication can take the form of body language, gestures, and facial expressions; prosodic features like tone, pitch, rhythm, and stress; and lastly, through vocalizations like laughs and screams. The adage was originally created from the results of Albert Mehrabian's book Silent Messages (1971) in which he discusses the scope of non-verbal communication.

While the adage might be an exaggeration the exact number is irrelevant. The essential part is knowing that most of our communication is non-verbal. These cues can be facial expressions, posture, body movements of all sorts, and auditory non-verbal cues. Non-verbal cues or utterances include vocal sounds as well as the behavior of the voice. These different behaviors include pitch variation, volume, pauses, hesitancy, speed variation, tone qualities, and responses confirming understanding (e.g. "uh-uhm", "huh", "aaah") (Anikin, 2020) .

It has been proposed that humans will apply a certain humanistic approach to a sociable robot and will, in many cases, initialize the interaction with a preconceived expectation based on their empirical interactions with humans (Duffy, 2003). Non-verbal communication has been shown to facilitate this tendency (Bates, 1994). Experiments have also shown that emotion and expressive behavior helps when trying to establish a smooth flow in the interaction (Breazeal, 2003), and many robots are designed to make expressions with their faces (Bates, 1994; Breazeal, Takanishi, & Kobayashi, 2008; R. Brooks, Breazeal, Marjanovic, Scassellati, & Williamson, 2002; Saunderson & Nejat, 2019). On the other hand, according to A. G. Brooks and Arkin (2006) it has not yet been proven that humans respond in the same way, with regards to body language, when talking to a humanoid robot as when talking to another human being. Brooks and Arkin continue to address the futuristic opportunities in how humans and robots can successfully communicate with each other if the non-verbal mechanisms can be fully interpreted as humanistic behavior.

When humans interact with other humans all these discrete non-verbal mechanisms are used to share information more efficiently with each other. We constantly get reminded of how frustrating it can be when we remove one or more of these non-verbal mechanisms in our day to day life. Technology has made it possible for us to communicate with each other using less and less of these non-verbal tools. Conversations over the phone, chat, and video all strip away parts of these tools and can make it more difficult to understand the actual intent behind the communication. What happens when we want to communicate with a robot using only a limited option of our communication tools, is it possible to successfully share information between a robot and a human using only non-verbal utterances?

Today we interact more and more with robots and autonomous machines of some kind. Technologies that are using different voice assistants are being used on a regular basis making the segregation gap between humans and robots smaller. When the demand for an improved human-to-robot-interaction is growing the implementation of non-verbal vocal sounds in robots becomes more relevant.

Within Social Robotics, non-verbal vocal communication is sometimes called *semantic-free utterances* (SFUs) (Yilmazyildiz, Read, Belpeame, & Verhelst, 2016). Semantics, as a branch of linguistics, is the study of meaning outside of its context. An example would be dictionary definitions. This is often contrasted to pragmatics, which is a branch of linguistics dealing with meaning in context. Thus, if one considers the utterance "Is the Pope still Catholic?" a semantic analysis would claim that the utterer is genuinely curious if the leader of the Catholic Church has had a change of faith. A pragmatic analysis, on the other hand, would claim that the utterer answers "yes, obviously". SFUs are, thus, utterances that lack semantic information and only communicate pragmatic information. For example, a particular scream does not have an entry in a dictionary: the context determines the meaning. Semantic information is more detailed than pragmatic information. However, pragmatic information is still informative. Pragmatic information also stands for the majority of human communication and all non-human animal communication.

There has been an increasing interest in SFUs in recent years. While natural language is one of the main forms of communication in human-robot interaction (HRI), this mode of communication has flaws. The ability for programs to understand natural language, to manage dialogue, and to spontaneously generate speech, is still limited and challenging to implement (Yilmazyildiz et al., 2016). Errors are commonplace and when they occur they often propagate and can lead to a breakdown of the entire interaction (Yilmazyildiz et al., 2016). Further, if the robot communicates successfully, this might create the illusion that the robot is at a similar cognitive level as another human. However, given the present state of social robotics, no robot can reach that bar. Thus, this only sets the interaction up for disappointment and failure further down the line. Communication using natural language also depends on language comprehension. If the listener comes from another country, they might have a poor understanding of what the robot is communicating. Some listeners, like infants or people with serious mental disabilities, might not understand any language. Thus, if this is the only venue used by the robot to communicate, these people are cut-off from communicating with the robot. SFUs have the benefit of being a far simpler form of communication. Thus, it is less error-prone and easier to implement. Also, since the robot is using a simpler form of communication, this lowers the expectations of the robot. SFUs are also language-independent. This does not mean that SFUs should replace natural language in HRI. For example, when it comes to communicating detailed information, this is easier to do with natural language than with SFUs. However, in communication where this level of detail is not needed, an increasing number of people are looking at SFUs as a solution.

SFUs are sometimes split into four categories: gibberish speech, musical utterances, non-linguistic utterances, and paralinguistic utterances (Yilmazyildiz et al., 2016). *Gibberish speech* (GS) refers to speech-like utterances that are meaningless or unintelligible. The idea is that in a speech signal, the semantic meaning is not the only thing being communicated. For example, if the speech signal has high amplitude, the words are uttered quickly, there is a lot of stochastic noise coming from the voice, etc., this might indicate that the speaker is angry. GS, thus, aims to remove the semantic parts of the speech signal while keeping the rest (for a review of implementation in HRI, see (Yilmazyildiz et al., 2016)). *Musical Utterances* (MUs) are short pieces of music used for communication. Music is known to communicate emotion. For example, major being associated with happiness, and minor being associated with sadness. Thus, the idea is that one can take what is already known about emotion in music and apply it to HRI to communication (again, see (Yilmazyildiz et al., 2016) for a review).

More relevant for this paper are the categories: non-linguistic utterances and paralinguistic utterances. *Non-Linguistic Utterances* (NLUs) are basically any sound not produced by a vocal tract used for communication in HRI. However, they generally consist of chirps, beeps, squeaks, etc., and are associated with prototypical robot sounds: for example, the sound made by R2D2 from Star Wars (Read & Belpaeme, 2016). When used in interaction design as a whole, they have gone by a bunch of other names: for example, auditory icons and earcons. MUs can be seen as a subset of NLUs. However, the design of NLUs do not generally take as much inspiration from music theory. They have been shown to be very effective at communicating affect in HRI. For example, Read and Belpaeme (Read & Belpaeme, 2016) investigated whether humans discriminated between NLUs and whether they categorized them by affective properties. Their NLUs consisted of sine waves that would modulate in frequency. They used five categories: flat, rising, falling, rising-falling, and falling-rising. These utterances were further embodied in a Nao humanoid robot. The utterances were tested in a standard categorical perception experiment. To assign an affective category, the participants used a facial gesture tool, the AffectButton, where the user can move their mouse to dynamically, in real-time, change the gestures of a face on the screen. Read and Belpaeme found that the participants did interpret the utterances as affective and that they categorized the sounds by affective properties (Read & Belpaeme, 2016). Work of this kind makes NLUs seem promising for future HRI research and implementation. However, NLUs have also been shown to be difficult to interpret for listeners and they seem to only pay attention to a small set of cues: for example, modulation of pitch (Khota, Kimura, & Cooper, 2019). Further, some research indicates that NLUs are even ignored when the listener is also interpreting the gestures of the robot (Frid, Bresin, & Alexanderson, 2018). It has been shown that the contrast between the physical appearance of a robot and the NLUs needs to be in balance for the NLU to have an effect (Read, 2014). The same work shows that when presenting NLUs to different people, the perception of affection is the same, however, the meaning of the utterance is different from people to people. Continuing, these results show that the situational context is having a big impact on how utterances are perceived and interpreted in various situations.

There are two definitions of *Paralinguistic Utterances* (PUs), one narrow and one broad (Yilmazyildiz et al., 2016). The narrow definition includes all non-verbal vocalizations, while the broad definition also includes non-vocal signals like gestures and postures. In this paper, the narrow definition is used. PUs have been the object of a decent amount of recent research: for example, cataloging the call repertoire of humans (Anikin, Bååth, & Persson, 2018) and the development of a synthesizer for PUs (Anikin, 2019). However, in the domain of HRI, the amount of work on PUs is very limited (Yilmazyildiz et al., 2016). There is a growing body of work on PUs that takes inspiration from non-human animal vocalizations. For example, the Haptic Creature (Yohanan & MacLean, 2011) uses purring to communicate positive emotion, and MiRo (Moore & Mitchinson, 2017) vocalizes in real-time using a mammalian vocal synthesizer. Other studies using dog-like vocalizations showed that participants could perceive effect in these as well (Gácsi et al., 2016; Korcsok et al., 2020). Moving on to human vocalizations, people seem to experience robot-produced laughter as natural if the context is appropriate (Becker-Asano, Kanda, Ishi, & Ishiguro, 2009, 2011) and may find the situation more amusing if the robot can respond in this way (Niewiadomski et al., 2013). However, given the sparse research, it is hard to say whether PUs may prove effective for HRI. Be that as it may, PUs should, in theory, possess many benefits over NLUs for HRI. For example, in areas outside of vocalizations, robotic gestures like changing eye color have proven to be less effective for communication than more human gestures (Rosenthal-von der Pütten, Krämer, & Herrmann, 2018). Other research showed that PUs were interpreted as more natural for a humanoid robot than mechanical and animal sounds (Read & Belpaeme, 2010). Further, biologically inspired sounds like PUs should be more ecologically valid since humans evolved while interpreting affect from these types of sound, not while interpreting robotic sounds.

2 Method and Implementation

In the present project, we want to contribute to research on PUs and compare their effectiveness in HRI to NLUs. The humaniod platform Epi (Johansson, Tjøstheim, & Balkenius, 2020) was used as the robot, from whom the auditory communication was emanated. We have recorded human non-verbal vocalizations to serve as a basis for synthesized vocalizations that were implemented in a humanoid. These fall under PUs but will be referred to as vocalizations. In order to investigate the communicative value of these non-verbal vocalizations, an experiment has been designed where the human-based non-verbal vocalization will be compared to NLUs.

Vocalizations

In order to implement non-verbal vocalizations in the humanoid, we first needed to collect data from human participants that would serve as the base for the synthesized vocals that were implemented in the humanoid. We collected data from 12 participants who were instructed to make non-verbal representations of words presented in front of them. The words were: affirm, decline, pleased, displeased, bored, satisfied, confused, disgusted. Together with these words, they were also instructed to represent emotional transitions in a non-verbal manner. The emotional transitions were the following: Satisfied→Attracted, Satisfied→Disgusted and Surprised→Attracted.

The recordings from the participants were later validated through a survey on prolific where new participants categorized the recorded sounds according to their perceived suitability. This validation process was done by Andrey Anikin, a researcher at Lund University with expertise in non-verbal vocalizations and a central contributor for this project. By doing this, we controlled that the recorded sounds actually were being interpreted as intended. The recordings that received the most accurate categorization became the prototypes to base synthesized vocals on that would be implemented in the humanoid. At this stage, it was noted that the emotional transitions were too arbitrary to be categorized and were thereby dropped from the implementation process. The words satisfied and displeased was also dropped due to their poor result in the categorization validation. This left us with 6 different categorical responses to be synthesized (affirm, decline, pleased, bored, confused, disgusted).

The synthetization was done by Andrey Anikin with the R package SoundGen (Anikin, 2019). The reason for using synthesized vocals for the humanoid was to minimize confounding factors for human vocalization. Where people might have a preference for a male or female voice and or other vocal attributes. Synthesized vocals also bring the advantage of easily manipulating vocal characteristics such as intonation, amplitude, and rhythm. This creates an opportunity to create variations of each sound, to avoid a static presentation of the exact same sound. Every synthesized vocal response got 5 different prototypes to originate from where three intensity levels were created for every prototype with 10 slightly different variations for each level. This resulted in 150 vocal variations per response category, in terms of the 6 words presented earlier.

The experimental design

By the time of execution of the initial experiment, the COVID-19 pandemic was so severe that the public restrictions made it extremely difficult to fulfill the requirements on the experiment set by the group. During this time, it was highly suggested to maintain a social distance from each other and to not meet new people. The latter led to a change in the structure of the experimental design which had to be made to be able to execute the experiment. The result of this is two versions of experimental designs, one initial intended and one restructured due to the limitations of COVID-19.

Experimental design 1.0

Initially, the group got a document with a preliminary experiment design from Andrey Anikin. It was set up in the following manner: A participant (P) is expected to show objects to Epi and based on Epi's reactions, in the form of non-verbal vocalizations, P is instructed to perform certain actions. Table 1 shows an adaptation of the table found in that document.

Context 1 is not complicated. If nothing elicits a change of state in Epi, Epi will eventually react with vocalizations representing 'bored'. P is then expected to do anything that would elicit a change of state: show Epi an object if no object is being shown or remove an object if an object is already being shown. Sometimes Epi will not properly see an object because the object might be too far away, or upside down, or in some other state which makes it hard to see. In this case, Epi will try to guide P with 'pleased' or 'displeased' sounds until Epi can properly see the object. This is what is happening in contexts 2 and 3. Once the object can be seen, Epi will react with a 'satisfied' sound. After this, depending on whether Epi likes or dislikes the object, Epi will react with either an 'attracted' or 'disgusted' sound. P will then either keep showing or remove the object depending on Epi's reaction. This is what is happening in contexts 4 and 5. In the last context, 6, Epi does not recognize the object, in which case Epi reacts with a 'surprised' vocalization. This is followed by an 'attracted' sound, perhaps because Epi is curious about the novel object, and P is expected to do the same action as in context 4.

This was the preliminary design and, as such, it had a few issues. Firstly, it became apparent that contexts 2 and 3 would prove difficult to implement. Epi would need to track the BCH-code as it moved through a 3D environment. This requires algorithms that predict the trajectory of the code. It would be an enormous task to implement, and an unnecessary one given the fact that these aspects of the interaction were not directly tested. Thus, this part of the experiment was scrapped. Another problem was in regards to the 'satisfied' and 'surprised' sounds. In the original document, these had no associated expectations on the actions of P. The design was also such that once these sounds were made, they would quickly be followed by either 'attracted' or 'disgusted'. 'Satisfied' and 'surprised' were only appendages for these later sounds. In other words, they were not being investigated in the experiment. Lastly, the document did not specify what parts of the experiment would serve as the data, and there was no hypothesis. For a while, the idea of time as a parameter was entertained due to how P was expected to wait after hearing the sound for 'attracted'. However, it was unclear how this would be applied to the rest of the sounds. The document also only detailed a single condition. Thus, once the data had been collected there would be nothing to compare it with. The above issues prompted a revision of the experiment. The hypothesis is that human-based non-verbal vocalizations have a positive effect on the communication between humans and robots. In other words, the type of sound coming from the robot is the variable. Further, the experiment has two conditions: In the first condition, Epi uses human-based non-verbal vocalizations. In the second condition, Epi uses earcons from Windows operating system. What follows is an outline of the procedure, the procedure is the same for both conditions:

1. P shows Epi a BCH-code representing the question "Are you ready?" and Epi can:
   (a) Affirm, in which case P moves to 2.
   (b) Decline, in which case P waits for a bit then repeats 1.
2. P shows Epi a BCH-code representing an object and Epi can respond:
   (a) Pleased, in which case P waits until Epi makes the sound for Bored then repeats 2.
   (b) Disgusted, in which case P removes the code and repeats 2.
   (c) Confused, in which case P goes back to 1.

Where previously Epi had to guide P towards the correct actions, this is now replaced by a simple yes-no question. This is to still test the 'pleased' and 'displeased' sounds (renamed after the validation to 'affirm' and 'decline'). The 'satisfied' sounds did not make the cut. Since it might get annoying to keep asking whether Epi is ready, the bulk of the experiment will be 2a and 2b. However, this will occasionally be interrupted with a 'confused' sound, where P is instructed to go back to 1. This will also ensure that the 'affirm' and 'decline' sounds will occasionally be tested, and it also serves as a unique action associated with the 'confused' sounds. During the second condition, the paralinguistic vocalizations are replaced with earcons with similar communicative functions. These were chosen based on criteria found in Brewster et al. (1999), particularly the criterion that earcons with similar or related meanings should have similar sounds.

Nr.	Contexts	Epi reacts with:	P is expected to:
1.	Nothing is happening	Bored	Do anything.
2.	P is doing the right thing	Pleased	Continue current action.
3.	P is doing the wrong thing	Displeased	Stop current action.
4.	Object is recognized and perceived as good	Satisfied → Attracted	Show object until Epi gets Bored.
5.	Object is recognized and perceived as bad	Satisfied → Disgusted	Remove Object.
6.	Object is not recognized	Surprised → Attracted	Show object until Epi gets Bored.

Table 1. A table of the contexts in the preliminary design, together with Epi's reactions and P's expected actions.

Experimental design 1.1

The concept of the experimental design at this stage in the process was to implement as much as possible from the Experimental design version 1.0 and convert it into a digital experiment. The design was made so that the participant could communicate with EPI from a distance using the video communication software Zoom from their computer. To keep the principle of the interaction between the participant and EPI the same setup was used as in version 1.0 in which EPI was placed at the opposite side of the participant, only the participant was replaced with a computer running Zoom. EPI could then scan the BCH-codes that were showing on the computer selected by the participant from Google slides. Meanwhile, the test participant could see EPI from the computer's camera. Using this setup, we maintained the initial idea of letting the participant control the interaction using BCH-codes. The digital layout was structured using Google Slides and Zoom. Before the experiment could start both the participant, the examinator and the computer that EPI viewed were to be logged in on Zoom. The computer in front of EPI was then connected to Google Slides and remotely controlled by the participant. Different BCH-codes were displayed, and EPI could read and respond to them as shown in figure 1. The experiment was able to be conducted using this setup thus converting the experimental design of 1.0 to a digital version.

Pilot testing

Two pilot tests were conducted to ensure a good experience for the participants and to give a chance of fair results. Both participants were given the same information before and during the test. The test participants were asked to use the 'Think aloud' method to give more data for the group. During the experiment, the participants' thoughts, actions, and results were noted. Additional information was given to the group collected from a minor interview at the end of each experiment. Each participant answered a couple of questions about how they perceived the experiment. After the two tests were conducted it was shown that a few adjustments to the experiment had to be made. The instructions given before the test were simply not enough for the participant to perform the experiment in the intended way. Both participants confessed that they did not remember all the responses in which EPI could respond after having completed the experiment, this was also shown in the results. Another important discovery made was the fact that having EPI respond with a 'bored' sound after a fixed amount of time created confusion in the participant. This sound was interpreted as an additional response to the object displayed rather than its intent; to display boredom from EPI. To improve the experiment, using the feedback from the pilot test, the amount of information given to the participant before the test was adjusted. The same information was also given during the test to remind the participant about the different responses available. A new more pedagogical visual layout was used in which the participant could, in a more efficient manner, navigate through the menu. Another adjustment that was made was that all the NLUs were changed, from what we in the project group perceived, to more appropriate sounds. We expect with this change that the boredom sound would give more clarity to the situation. Two more tests were carried out with new test participants. The experiment ran more consistently with the new changes, however, confusion regarding EPI showing boredom was still present. The test participants still thought the boredom sound was connected to the object. The problem seemed to be that having two responses linked to the same object was simply too confusing.

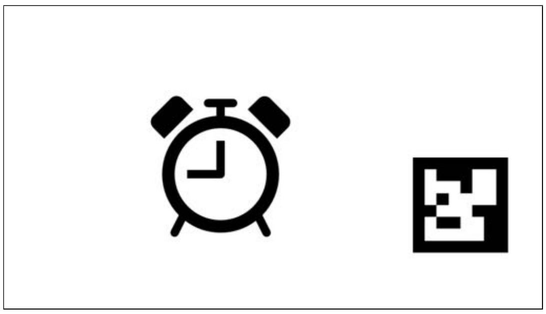


Figure 1. Object with BCH-code

At this point, we were struggling with perfecting the digital experiment and a meeting with the projects' supervisor was made. At this meeting, one last test was conducted and feedback was given. It was decided that the digital experiment did not reflect the initial experiment well enough and had to be scrapped. Instead of trying to imitate the original design of the experiment, the group changed the direction of the experimental design more suited for a digital version.

Implementation

This section describes the work of implementation of necessary code for the initial experiment. This was not used in the digital experiment's but the sections are part of the report to be of use for potential future work. The detailed code and used audio files are available in this Github repository: https://github.com/pierreklintefors/GiveARobotAVoice.

In order to conduct the initial experiment, the humanoid Epi (Johansson et al., 2020) was needed to be equipped with necessary programs to be able to express context-specific sounds, targeting to guide the interactor's behavior. For the experimental design in this project, Epi needed to be able to elicit utterances that would be triggered by some external cue. The design of Epi of that time was limited in its sampling of the external environment, it could only be done by visual input. It was not equipped with any audio recording device or tactile system so the utterances needed to be triggered by visual cues, in terms of BCH-codes that were mentioned in the section about the experimental design.

Technical implementation

The cognitive modeling infrastructure IKAROS (Balkenius, Johansson, & Tjøstheim, 2020) was used as the control architecture for Epi as well as the platform for interpreting the input stream, controlling the signal chain, and generating the sound output. This was done by integrating a network of task-specific modules. The chain was constituted by the following modules: EpiTorsoMinimal, MarkerTracker, VocalReaction, SoundOutput, the presented order of the modules corresponds to the signal chain, from input to output.

The module EpiTorsoMinimal was used to capture the visual input by grabbing the video stream from one of the two cameras, mounted in the head as the humanoid's eyes. This was done via the module Epivision which was subordinated to EpiTorsoMinimal. The video stream was serving as the input of the next module in the chain, named MarkerTracker.

MarkerTracker used the software library ARToolKitPlus (Rojtberg & Walton, 2014) in order to scan and identify the BCH-codes. The object identification, in terms of the identified BCH-code, was sent to a module called VocalReaction. VocalReaction was a new module that did not exist in the Ikaros repository prior to this project. This module aimed to interpret the BCH-id and send the appropriate input data to the final module, SoundOutput, which was able to play pre-recorded audio files (i.e. .mp3 .wav) which can be listed in one of its parameters.

The SoundOutput module took a binary vector as input. The vector needed to be of equal size as the number of audio files listed in the SoundOutput parameter. If an element of the vector was set to 1, SoundOutput would play the audio file that which position in the list corresponded with the position of the element. The VocalReaction's task was thereby to generate the appropriate input vector to SoundOutput in order to play the appropriate audio based on the scanned BCH-code. Due to the variational intonation and intensity aspect of Epi's vocal responses, different audio files with these variations were created for each response. When a BCH-code, associated with a specific response, was scanned, the playback of one of all the audio files associated with that BCH-code needed to be randomized. This randomization process took place in VocalReaction where one element within an appropriate span of elements in the outgoing input vector for SoundOutput could be set to 1. The lower and upper boundaries of that span were decided by the scanned BCH-code.

The choice of using BCH-codes as visual input rather than other visual cues as facial recognition or body gestures was motivated by convenience. The MarkerTracker module was already implemented in Ikaros which meant that BCH-codes could be used to code for different "objects" that would elicit different vocal responses. This was suitable for the aim of this project which was to evaluate how different utterances can guide humans in interaction with a humanoid. Facial recognition and other visual cues can be implemented in a later stage in order to equip Epi for a more external valid setting to act in.

Experimental design 2.0

A new approach was implemented which would be the final design. In this version, the design was made more comparable to a survey than to a real-life interaction. A short video of EPI was used in which the robot mimicked a sound using small head movements and displayed a light from the mouth for a brief moment of time. The NLU and vocalization were layered in the video in synchronization with the light from the mouth which would reflect the different responses in which EPI could give (the audio files can be downloaded from: `https://github.com/pierreklintefors/GiveARobotAVoice/tree/main/Sounds`). One sound from every vocalization prototype was used for every response category to create a more human-like image of EPI. A total of 25 videos with vocalizations (5 categories * 5 vocalizations) and 5 videos with NLU were made giving a total of 30 different videos. Multiple-choice questions were given after each video about what the test participants thought that EPI was expressing. The survey was split into two parts: one survey had only videos showing EPI reacting with vocalizations and the other with only NLUs. To have the same condition the survey with NLUs had to be expanded to show 25 videos instead of only the 5. This was done by simply adding each NLU video five times. Both survey parts had their videos in a randomized order and each test participant had a 50/50 chance of doing either the part with vocalizations or the part with NLU. This was simply implemented using a question at the beginning of the experiment whose answer forwarded the test participant to either the survey with vocalizations or NLU with a 50/50 chance of either. Using this type of experiment meant less interaction with EPI but presented the opportunity to have more participants conduct the experiment since there was no need to plan, book, or monitor each experiment.

Since the experimental design 2.0 used multiple-choice questions to collect the data this would also yield different results from the former experiments. Parameters like correct classification, perceived level of intelligence, and perceived level of creepiness in EPI were compared.

3 Result

A total number of 43 participants completed the survey and was used for the analysis. These 43 participants were randomly assigned to the two conditions, whereof 20 were assigned to the NLU-condition and 23 to the Vocalization-condition. The parameters of interest (i.e. proportion of correct answers, the robot's perceived level of intelligence, and the robot's perceived level of creepiness), was tested with two-sample Wilcoxon tests because normality was not obtained in the distributions.

The means, medians and the standard deviations of the parameters are presented for the conditions in Table 2. Participants in the NLU-condition had a higher proportion correct answers (Mdn = 96%) than the participants in the vocalization-condition (Mdn = 84%), this difference was significant, $U = 144$, $p = .034$. The participants in the NLU-condition also rated the robot as more creepy on the 5 point likert-scale ($Mdn = 3$) than the participants in the vocalization condition($Mdn = 1$), , $U = 140.5$, $p = .022$. There was no significant difference in the ratings of perceived intelligence, $U = 253.5$, $p = .556$.

The vocalization received significant lower correct classifications and by comparing the confusing matrices in Figure 2 respectively Figure 3 a more detailed picture of the misclassification emerges. A darker square indicates a higher frequency of that classification. The dark squares in diagonal, in both Figure 2 and Figure 3 shows that the majority of classifications were correct, as shown by the descriptive metrics of Table 2. By looking at Figure 2 it seems like the most common misclassification in the vocalization condition was to mistake an 'Agree' vocalization for a 'Disagree' vocalization, happening 22% of the 'Agree' classifications. Confused was also wrongly mistaken for being a 'Pleased' vocalization 17% of the time. The classification rate of these two most commonly misclassified prototypes is presented in Table 3.

Parameter	Mean	Median	SD
Correct answers			
Vocal	84.87%	84%	10.92%
NLU	90.4%	96%	13.32%
Intelligence			
Vocal	2.91	3	.9
NLU	2.75	2.5	.97
Creepiness			
Vocal	1.65	1	.83
NLU	2.4	3	1.1

Table 2. Metrics of the parameters of interest

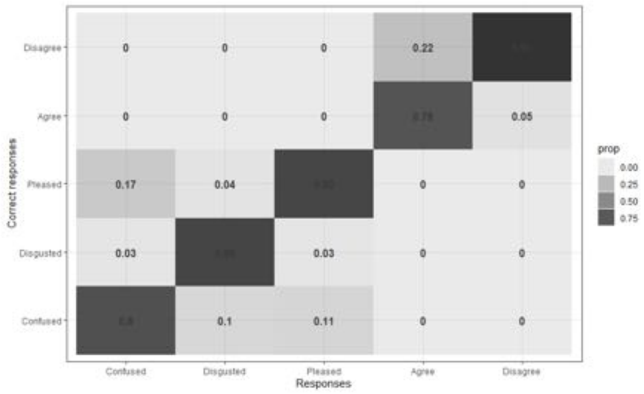


Figure 2. Confusion matrix of responses in the vocalization-condition

The prototypes 'Confused 3' and 'Agree 4' have especially low correct classification rates, 47.83% respectively 56.52%.

In contrast, shown in Figure 3, there was no ambiguity of the NLUs' representing 'Agree, 'Disagree' and 'Pleased', being correctly classified 100%. The most ambiguous NLU sounds were 'Disgusted' and 'Confusion'. 'Disgusted' being classified as 'Confused' 27% of the time and 'Confused' being classified as 'Disgusted' at a rate of 17%.

4 Discussion

Results

The NLUs were more correctly classified to the targeted semantic emotion/response than the vocalizations. This is interesting because the vocalization sounds had been validated but NLU sounds had not but still were more successful in mediating semantic meaning for the chosen response categories. However, only 5 sounds were used for the NLUs, 1 per response category, while the biologically inspired sounds had 5 different vocalizations per category adding up to 25 sounds in total. This makes the NLU-condition more prone to a

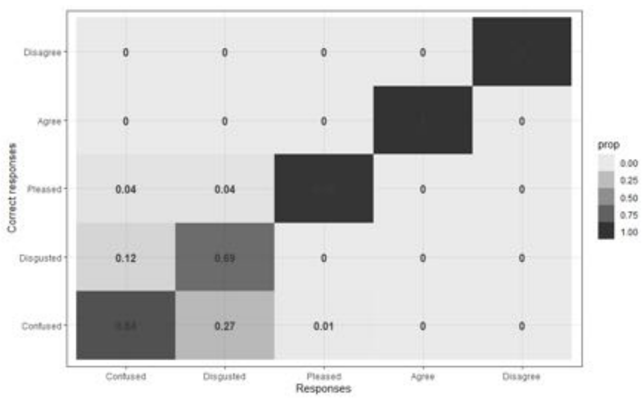


Figure 3. Confusion matrix of responses in the NLU-condition

Prototype	Correct classification
Agree 1	82.61%
Agree 2	91.30%
Agree 3	69.57%
Agree 4	56.52%
Agree 5	91.30%
Confused 1	100%
Confused 2	95.65%
Confused 3	47.83%
Confused 4	73.91%
Confused 5	82.61%

Table 3. Proportion correct classifications of confusion and agree prototypes

practice effect, where the participants were exposed to all 5 sounds quickly. This could have contributed to the high performance in the NLU-condition. Uncertainty and ambiguity of a sound could be resolved by hearing the other sounds and thereby not repeat misclassification in the subsequent trials that was potentially made in previous trials.

Even though the vocalizations were more misclassified than the NLUs, Mdn=84% versus Mdn=96%, it is not a poor performance considering that the chance level would be 20%. Furthermore, it was mainly a few prototypes that contributed to the poorer performance of vocalizations sounds compared to the NLUs. Considering that the other prototypes have relatively high performance, it might be appropriate to remove these poor performing prototypes that causing the misclassifications for further studies.

Difficulties during the process and failed approaches

Looking back at the project process there were some complications and pitfalls. The first category of difficulties was related to communication. The main issue was that the project group was initially a bit unsure about what the goal and project scope actually consisted of. This situation created a condition where some initial work had to be reworked, as it was not relevant to the actual goal of the project. With the outcome in mind, it would most likely have been beneficial if the project group prepared a set of questions in the early stages of the project to get a better understanding of the project.

Furthermore, there were also some technical difficulties during the project process. A relatively big part of the technical problems was related to the setup with different operating systems and configuration management. In retrospect, it would presumably be beneficial if the division of work was performed in an earlier stage of the process.

Moreover, there were initially some minor issues with finding a robust design for the experiment that included all the generated sounds. The reason why this approach was not optimal could be that the content of the sound library was defined before the actual experiment. This resulted in a situation where the experiment had to be adapted to the given sounds. Looking back, it would possibly have been more suitable to define the design of the experiment and then starting collecting sounds that were appropriate for that design. In the real context project, one could argue that the steps of the project were not in a logical order.

Sources of error

Further, with the design of the project defined and sounds library available there are some aspects that are important to keep in mind for the preparation of further development. For instance, the sounds that were generated are not perfectly adapted to the design of the experiment. This situation can potentially create a gap between the sound generated by the robot and the state that is supposed to be communicated. To be able to handle this difficulty it would probably be beneficial to test the experiment design a few times before the actual testing.

Secondly, the generated sounds are based on a relatively small set of sounds. This can result in the sounds lacking in strength since a larger set of sounds would possibly capture diversity better. In addition, the number of people validating the sounds was also limited to a relatively small group of individuals. Increasing the number of people in the validation process would most likely ensure the level of quality to a greater degree.

Looking further into the NLU sounds, there was no variation of the sounds. As mentioned earlier in this report, the vocalizations all had a set of different variations for each sound. In the case of NLU sounds the lack of variation could have favored the practice effect, meaning that the test person most likely started to associate the sounds with the answers during the test. Analyzing the structure of the test, it would probably be fairer if the NLU sounds also had variations.

Further, the NLU sounds were never officially validated. Unlike the NLU sounds, every vocalization sound was validated. In the case of the NLU sounds the project group was responsible for validating the sounds. Comparing this process to the vocalization validation process one could argue that it could be seen as a source of error. The results would probably have been more equitable if the NLU sounds were validated under the same conditions.

Reflection on the work process

Additionally, looking at the overall work process there are a few key takeaways. Firstly, the order project phases can affect the outcome. It is presumably beneficial to begin the process with comprehensive planning. This will favor the work process, by minimizing redundant and unnecessary work.

Further, working with unfamiliar technology can result in difficulties. With that in mind, getting comfortable with the technical domain is very important for the project process, especially if there is a given for delivery.

As previously mentioned, the order of the tasks included in the process was not optimal. In the case of implementing similar projects in the future, it would probably be wise to define the design of the experiment in the early stages of the project process.

Suggestions for future work

For future work, there are a few suggestions that should be highlighted. Firstly, if to use the same prototype sounds that were used in this case, there are some sound variations that should be avoided. The prototypes with the highest misclassification ('Confused 1', 'Confused 4', and 'Agree 1') should be excluded.

Furthermore, it could be interesting to look further into comparison with an actual language. In this case, the comparison with a language was never included, as the assumption was that it would probably be very easy to understand. Apart from that, there are some social aspects of this comparison that could be interesting to look further into, for example, the level of social discomfort.

References

Anikin, A. (2019). Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavior research methods*, *51*(2), 778–792.

Anikin, A. (2020). *Human nonverbal vocalizations* (Doctoral dissertation). Lund University.

Anikin, A., Bååth, R., & Persson, T. (2018). Human non-linguistic vocal repertoire: Call types and their meaning. *Journal of nonverbal behavior*, *42*(1), 53–80.

Balkenius, C., Johansson, B., & Tjøstheim, T. A. (2020). Ikaros: A framework for controlling robots with system-level brain models. *International Journal of Advanced Robotic Systems*, *17*(3), 1729881420925002. doi: 10.1177/1729881420925002

Bates, J. (1994). The role of emotion in believable agents. communications of the ACM. , 122-125.

Becker-Asano, C., Kanda, T., Ishi, C., & Ishiguro, H. (2009). How about laughter? perceived naturalness of two laughing humanoid robots. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 1–6.

Becker-Asano, C., Kanda, T., Ishi, C., & Ishiguro, H. (2011). Studying laughter in combination with two humanoid robots. *AI & society*, *26*(3), 291–300.

Breazeal, C. (2003, 03). Toward social robots. *Robotics and Autonomous Systems*, *42*, 167-175. doi: 10.1016/S0921-8890(02)00373-1

Breazeal, C., Takanishi, A., & Kobayashi, T. (2008). Social robots that interact with people. *Springer Handbook of Robotics*, 1349–1369.

Brooks, A. G., & Arkin, R. C. (2006). Behavioral overlays for non-verbal communication expression on a humanoid robot.

Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B., & Williamson, M. (2002, 03). The cog project: Building a humanoid robot. *Lecture Notes in Artificial Intelligence*, *1562*. doi: 10.1007/3-540-48834-0_5

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, *42*(3-4), 177–190.

Frid, E., Bresin, R., & Alexanderson, S. (2018). Perception of mechanical sounds inherent to expressive gestures of a NAO robot-implications for movement sonification of humanoids. *Sound and Music Computing*.

Gácsi, M., Kis, A., Faragó, T., Janiak, M., Muszyński, R., & Miklósi, Á. (2016). Humans attribute emotions to a robot that shows simple behavioural patterns borrowed from dog behaviour. *Computers in Human Behavior*, *59*, 411–419.

Johansson, B., Tjøstheim, T., & Balkenius, C. (2020). Epi: An open humanoid platform for developmental robotics. *International Journal of Advanced Robotic Systems*, *17*(2). doi: 10.1177/1729881420911498

Khota, A., Kimura, A., & Cooper, E. (2019). Modelling synthetic non-linguistic utterances for communication in dialogue. *International Journal of Affective Engineering*, IJAE–D.

Korcsok, B., Faragó, T., Ferdinandy, B., Miklósi, Á., Korondi, P., & Gácsi, M. (2020). Artificial sounds following biological rules: A novel approach for non-verbal communication in hri. *Scientific reports*, *10*(1), 1–13.

Mehrabian, A. (1971). *Silent messages*. Wadsworth Publishing Company.

Moore, R. K., & Mitchinson, B. (2017). A biomimetic vocalisation system for MiRo. *Conference on Biomimetic and Biohybrid Systems*, 363–374.

Niewiadomski, R., Hofmann, J., Urbain, J., Platt, T., Wagner, J., Bilal, P., ... Shehory, O. (2013). Laugh-aware virtual agent and its impact on user amusement. *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, 219–226.

Read, R. (2014). *A study of non-linguistic utterances for social human-robot* (Doctoral dissertation). Plymouth University.

Read, R., & Belpaeme, T. (2010). Interpreting non-linguistic utterances by robots: studying the influence of physical appearance. *Proceedings of the 3rd international workshop on Affective interaction in natural environments*, 65–70.

Read, R., & Belpaeme, T. (2016). People interpret robotic non-linguistic utterances categorically. *International Journal of Social Robotics*, *8*(1), 31–50.

Rojtberg, P., & Walton, D. (2014). Artoolkitplus. *Github repository https://github.com/paroj/artoolkitplus*.

Rosenthal-von der Pütten, A. M., Krämer, N. C., & Herrmann, J. (2018). The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *International Journal of Social Robotics*, *10*(5), 569–582.

Saunderson, S., & Nejat, G. (2019). How robots influce humans: A survey of nonverbal communication in social human-robot interaction. *International Journal of*

Social Robotics, *11*, 575—608.

Yilmazyildiz, S., Read, R., Belpeame, T., & Verhelst, W. (2016). Review of semantic-free utterances in social human–robot interaction. *International Journal of Human-Computer Interaction*, *32*(1), 63–85.

Yohanan, S., & MacLean, K. E. (2011). Design and assessment of the haptic creature's affect display. *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 473–480.

A system for Bidirectional Touch-Emotion Interaction for a Humanoid Robot

Axel Holmqvist
tfr14ah2@student.lu.se

Martin Karlsson
martingkarlsson@gmail.com

Mariam Mirström
maryammirzaei22@gmail.com

A project has been carried out with the purpose of constructing a somatosensory system for a humanoid robot, including the identification of different types of touch and a related response. The detection of touch on an Epi humanoid robot head is made from conductive paint on the inside on the head shell, the electrical signal produced is processed into a digital representation of touch, a classification of touch is made through the use of machine learning, and a representation of related emotions is presented as a response. A review of literature has been performed, from which a theoretical model of 11 different touch types was created and applied in the identification of touch of the somatosensory system. A study was conducted to investigate the relation between types of touch on an Epi head and their expected emotional response, where the results from this were applied in producing the representation of an emotional response from touch. Enabling the system to include multiple types of touch, the resolution of touch representation was modified through the use of a continuous signal of high frequency, and Support-Vector Machine machine learning was applied to provide a sufficient certainty of prediction and learning with a validation accuracy of 89%. The system could thereby detect and identify 11 different types of touch and produce a related response, enabling the bi-directional touch-emotion interaction with a humanoid robot.

1 Introduction

Constructing a humanoid robot, the different modalities used in interaction with humans are essential. Historically, a sense of touch has often been discarded, therefore the development of a somatosensory system could provide a relevant contribution to humanoid and social robotics.

This project aims to investigate the possibilities of constructing an easily applicable somatosensory system for a humanoid robot. First of all, we would like to evaluate whether the detection of touch on the surface of a robot head can be created and if the signal produced can be processed and analyzed to obtain information about the type of touch that has been applied. We will also investigate a possible response to touch, in the form of triggered emotions related to specific types of touch. We have therefore reviewed related literature and performed a study on how touch is connected to different emotional expressions. All together, the goal is to construct a prototype of a somatosensory system that fulfills the requirements stated byvan Erp and Toet (2015) when it comes to creating a meaningful human-computer interaction and closing the loop of bidirectional interaction:

> "A robot that has the ability to 'feel,' 'understand,' and 'respond' to touch in a human-like way will be capable of more intuitive and meaningful interaction with humans."

In order to create and design a humanoid system corresponding to such requirements, enabling human-robot interaction (HRI), the system should *1)* determine where the touch occurs, *2)*, distinguish what type of tactile stimulation took place and *3)* and evaluate the affective quality of the touch (van Erp & Toet, 2015). However, the system does not necessarily have to describe a one-to-one mapping with the human somatosensory system but produces a convincing likeness to human behavior and social interaction.

The project also aims to contribute to one of the research topics suggested by van Erp and Toet (2015) in their compilation of social touch within HRI. The authors review recent findings and conclude that a distinct somatotopic mapping between tactile sensations and emotional feelings could exist. It is of interest to determine the basic definitions of such a map as a map of responsiveness to interpersonal touch. Thus, a significant part of the project will focus on the mapping of touch-to-emotion.

The project's scope includes affective touch within social touch interaction, where referent social conventions are taken into consideration. Our study on touch-emotion considers possible relative differences between cultures and genders regarding the perception of and expectation regarding social touch, as the incorporation of related factors is suggested by van Erp and Toet (2015). Other social touch fields are not considered, and factors such as context, previous experiences, or personality traits are not part of the study.

Finally, the purpose of the project is to investigate different alternatives and suggest ways of implementing a the system for Epi (Johansson et al., 2020), an open humanoid platform for developmental robotics, and the Ikaros (Balkenius et al., 2010) platform, a cognitive modeling infrastructure with robot functionalities.

Background

Touch can be deemed necessary for interaction with embodied social agents, as discussed in (van Erp & Toet, 2015), where the review of current research shows that the lack of tactile interaction results is generally connected with a negative appraisal of the robot. Touch matters, without doubt. It is central to human experience, culture, and communication. Touch is the first sense through which humans apprehend their environment, and it is central to our development (Field, 2001). The different aspects of touch may not yet be sufficiently studied, but it is clear that touch provides essential information about the world, as it is of importance for tool use (Fulkerson, 2014) and communication: "Just as we 'do things with words' so, too, we act through touches" (Finnegan, 2014). Indeed, knowing how to infer meaning from touch is considered a basis of social being (Dunbar & MacDonald, 1998). Today touch is at the center of Human-Computer Interaction (HCI), and computer science's imagining of digital sensory communication makes the question of 'digital touch', touch that is digitally mediated, significant for communication. Touch screens are arguably 'transforming our embodied experience of sociality and material culture' in various contexts, including in the home (Richardson & Hjorth, 2017). The emerging arena of what might be classified as digital touch also extends to other forms and sites of touch-based interfaces and haptic technologies, both in face-to-face and remote interaction. Given that touch is so fundamental and intimately tied to the human experience and communication, it can be argued that a social science perspective could also be applied on such touch in order to gain an understanding of the societal impact of emerging touch technologies and that it 'should not be left to technicians/scientists alone' (Wilson et al., 2014).

The Human Somatosensory System

In order to understand touch, an intuitive way to begin is to start with the somatosensory system, which provides a person with information about the environment and acts as input in carrying out motor action. Indeed, it helps a person to distinguish what external events are taking place from the actions that she executes. In humans, the sense of touch is processed in the primary somatosensory cortex, in the parietal lobe postcentral gyrus, which is somatotopically ordered, where neighbouring bodily areas are neighbouring brain regions (Banich & Compton, 2011). The human body contains multiple different somatosensory receptors that can be categorized into groups for contributing to the detection of pressure, temperature, and pain. They may be adaptive to stimuli, and habitation can cause them to become less sensitive or become sensitive to slighter touch. Different parts of the body have different receptor density, purveying different sensitivities to touch. The somatosensory system is involved in creating the imaging of embodiment and provides information about body position. The orbitofrontal cortex is connected to certain parts of the somatosensory cortex and is functionally linked to social functions, affect and reward. Therefore, touch may be considered a vital part of social interaction. Together with bodily action and awareness of the body, the somatosensory system may be involved in social cues and social coordination (Kolb et al., 2016).

Affective Touch

Touch comes in several forms, and sometimes needs to be narrowed down in order to be manageable in a project likte this. Thus, the scope of this project has a main focus on one; affective touch. Touch applied in social interaction and emotions is called affective touch and can be contrasted with discriminative touch, mainly connected to information transmission and processing in communication. Humans have many non-verbal ways of communicating, usch as facial expressions, prosody, gesture, and touch. In contrast, according to van Erp and Toet (2015), affective touch is the primary way of conveying emotions and emotional states. Research carried out by Olausson et al. (2002) suggests that humans have a specialized neurophysiological system for affective touch alone, separate from the properties of discriminative touch, where the optimal type of stimulation is a gentle stroke and an optimal speed for touch is approximated to 1-10 cm/s (Löken et al., 2009). van Erp and Toet (2015) further states that as (affective) touch implies direct physical interaction, it naturally mediates and evokes emotions and feelings of social presence and Suvilehto et al. (2015) seems to be in agreement and claims that affective touch "[...] provides a unique contribution to the formation and maintenance of social bonding.".

However, the topography of interpersonal social touch plays a huge role in interpretation since it's meaning differs depending on the location of where it is executed. Furthermore, whether an area of the body is deemed available or allowed to touch or not differs greatly in dependence on what emotional bond exists between the person administering and the person receiving touch (Suvilehto et al., 2015). Affective and social touch is hence highly dependent on the context in which it occurs.

Social touch in humans and HCI

When it comes to this project the goal was to study and implement affective touch as an interaction, and should therefore be seen as a form of social touch. van Erp and Toet (2015) have compiled research about humans and social touch within communication, and states that it is used for *a)* well-being, *b)* to mediate emotion, *c)* to trigger emotion, and *d)* as a behaviour modulator. In other words, it is strongly connected to our emotional state, either for mediating or triggering, and seems to be related to a human experience of well-being. It is also a behavior modulator, meaning that it can provide effective means of influencing people's attitudes toward either other persons, places, or services. However, this project only aims to investigate triggered emotion(s).

When it comes to human-computer interaction (HCI), and social touch, the authors claim that it *a)* reduces the machine likeness, *b)* increases the communication of emotions, *c)* helps in creating a relationship to social agents, *d)* mediates a higher degree of friendship, *e)* mediates a higher degree of social presence. In summary, it can be concluded that social touch will make the agent/machine more human-like and help in creating a suitable relationship – which are both two aspects that we hope this project will contribute with to some extent.

Cultural and Sexual Differences in Social Touch

When studying social touch, and especially when including participants in surveys and studies, we have to be aware of the cultural and sexual differences that exist, and have therefore tried to take these aspects into consideration throughout this project. The degree to which social touch is differentiated between cultures and sexes is not clear from the literature, but there is support for such effects and it is therefore of importance to be aware of that in carrying out a related study (Hertenstein et al., 2009; Field, 2001). Differences exist, but the size and scale of these differences are not yet succinctly quantified. Distinctions between what is universal and what is cultural or gender-specific can, according to Remland and Jones (1988), be described in terms of 'contact cultures', where touch through interaction is relatively frequent, and 'non-contact cultures', where it is less so. The authors review prior research indicating that e.g. Americans tend to have less interaction through touch than Mediterranean people do and that American men are typically less likely to touch each other relative to Mediterranean men. Their own study shows that men are generally more avoidant of touch in same-sex interactions than women tend to be. Furthermore, a difference between American and Far eastern cultures is noticeable, with touch being clearly less frequent in Far eastern culture, a difference that is pronounced between the sexes. However, Hertenstein et al. (2009) mention that there are no existing studies on the cultural and sexual differences that systematically include the communication of emotion, a field that thereby requires the attention of researchers.

Human Emotions

Since a big part of this project includes triggered emotion(s), when it comes to the social aspect of touch, and since emotional expressions play such an important role in human communication McNeil (2005), we wanted take a closer look at basic human emotions. Research shows that the are multiple emotional expressions, but fewer basic emotions that occur regardless of factors such as cultural background and education. The basic emotions are classified according to a set of criteria, and different psychological theories can apply different such criteria. An agreement however exists between theories, in that a primary emotion is affiliated with a particular universal nonverbal expression and in which distinct components are involved on a neural and physiological basis. The basic emotions are essential guidelines in the design of emotional expressions. Six emotions, also known as "The Big Six," are classified based on ubiquitous recognition of emotions from facial expressions. These emotions are, according to Ekman and Friesen (1969), 'happiness' (sometimes 'enjoyment'), 'sadness', 'fear', 'surprise', 'anger', and 'disgust' – which gives us a fundamental foundation to work around throughout this project. Other emotions may in turn be considered amalgams of these basic emotions, and are thereby be referred to as complex emotions.

Cognitive and Social Robotics

In order to design and implement a system for touch, that can be connected to other senses, we have to be aware of how a cognitive model of a social robot works. First of all, perception in robots is a cognitive problem relating to how the environment is processed by robot systems (Balkenius et al., 2008). In mapping the environment, the perceptive input provides the basis for the robot's orientation and behavior and this may therefore be largely considered a perceptual problem. A robots detection of the environment could involve infrared sensors and cameras, providing the basis for such robot behaviors as motor control and locomotion, which in turn make planning possible. In this way, hierarchically ordered behavioral layers produce the robot's actions, in dependence on which sensory stimuli are present. In robot learning, robots can discriminate different objects in the environment and learn through imitation or demonstration, including learning from interaction with humans. Robots can be designed as *humanoid*, with features and behaviors approaching those of humans. This may also be considered an increase in social interaction competence (Balkenius et al., 2008). Robots can replicate human movement and expression in a behavior called mirroring. In gaze following, eye movements in the robot are used to track the human use of eyes in directing attention. Speech synthesis and speech processing enables the robot to participate in communication through speech, where the basic structure for human verbal communication and dialog, such as the shifting engagement of the partners, is implemented. The sensory input also enables a spatial orientation and the detection of events occurring in the environment as well as providing a foundation for adapting to social interaction. Social cues can, in regulation, be used for producing social coordination in a robot and enable it to adapt to a social system. Intention movements means that robotic behavior includes expressions corresponding to an awareness of the presence of other social agents, and the robot's actions can in this way be accompanied by signals alerting humans about their purpose. The design of facial and bodily features makes the integration of facial expressions, eye movements, gestures, and pointing in communication, into behavior, possible. A humanoid robot can thereby convey emotion and emotional expressions can strengthen bonds with humans. Overall, embodied action can provide important contributions to social coordination.

Robotics and Somatosensation

Furthermore, somatosensation within robotics was studied in order to gain inspiration that was useful for a project regarding this. To begin with, haptic interfaces for computers have been the object of much recent development while the somatosensory aspects of social interaction is often neglected in creating humanoid and social robots. Somatosensorial information is used as a defining sensory input specifically for robots that have socially assistive, companion, and learning behaviors for children and the elderly (Leite et al., 2013; Limbu et al., 2013; Feil-Seifer and Mataric, 2005). We have found a limited number of studies involving the general investigation of touch in robot interaction such as Kerpa et al. (2003), Cooney et al. (2012b), and describing specific materials for the detection of touch (Zhang et al., 2017), but the literature on robotics is however overall directed towards other ends. Somatosensation in robotics could provide the robot abilities to detect the environment and provide the basis for actions in motion and its spatial orientation. In social robotics, somatosensation could be of interest in robots bodily relating to humans, in direct interaction and communication involving touch. The increased embodiment made possible by a somatosensory system could further enhance abilities for regulating actions, the mirroring of humans, and intention movement signaling, in social interaction. This functionality could be continued to social coordination and for more specific purposes in assistance, companion, and learning behaviors where a robust affective component is part of the interaction. Here a somatosensory input could be of fundamental importance in humanoid and social robots.

Artificial Neural Networks and Support-Vector Machines

Enabling the creation and implementation of the processing and classifying affective social touch for a humanoid robot, the inclusion of a cognitive model for analysis of touch is required. This could be an artificial neural network or a pure classifier in the form of a support-vector machine, which both use a technique that is often referred to as machine learning.

Artificial neural networks (ANN) is a machine learning system that is inspired by the biological brain, with inputs and outputs and where a distributed parallel processing updates weights related to nodes in network connections, corresponding to synaptic connections in the brain (Zha, 2003). The ANN can be used for the classification of data through the detection of characteristics and the consequent generation of pattern recognition, making predictive processing possible. In supervised learning a mapping of input and output is used, and in unsupervised learning no such mapping is applied (Wong & Hsu, 2006).

The ANN may in some respects be limited, and another method of machine learning is provided by Support-vector machines (SVM), which could be more appropriate for identification regarding movement direction (Lau et al., 2008), including in relation to touch patterns. SVM machine learning has been introduced as an alternative to ANN computation as a technique for solving problems regarding learning, classification and prediction. The SVM system originates from a implementation of the Structural risk minimization principle by Vapnik (1995). The fundamental principle of an SVM is to first transform the data into a higher dimensional space R^{N}, where N is the number of features, and then find hyperplanes that can distinctly classify the data points (Gandhi, 2018). When the number of input parameters is N=2, the hyperplane is just a line and for N=3, the hyperplane is a 2D plane, as illustrated in Fig. 1, where a 2D hyperplane separates three parameterizations of data into two classes. The further cases, where N>3 are difficult to illustrate visually, but work according to the same principle. The SVM derives its name from the support vectors describing the data points that are close to the hyperplanes, and therefore have the most influence when optimizing the positioning and orientation of the hyperplane. It is possible to use different hyperplanes, where the system is optimized when the margin between the data points of different classes is maximized.

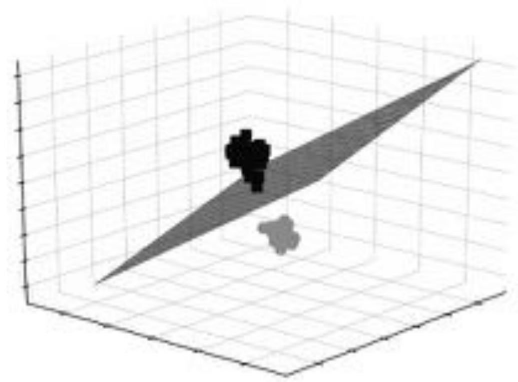


Figure 1: Example of a 2D hyperplane classification.

Empirical testing and studies, as well as the study carried out by Wong and Hsu (2006), has shown that SVM outperforms ANN in both classification and regression and this regardless of what coding scheme is used. In Wong and Hsu (2006) it is also stated that an SVM is easier to optimize due to the few (2) parameters required when creating the classifier, in contrast to an ANN, for which a specific classification can be produced by an indefinite number of connection weight combinations. Finally, it requires less computation power (Gandhi, 2018).

2 Method and Implementation

The design and implementation of a somatosensory system for a humanoid robot was divided into two parts, with different focus.

Part 1: Somatosensory/Touch theory, Touch identification and classification, Prototyping We first worked with with examining and evaluating prior research regarding the somatosensory system and social touch in general. This laid the foundation for the selecting a method and constructing a setup for classifying touches from available hardware components and using machine learning. From there, the production of a prototype for the detection of touch on an Epi head was carried out and connected to an interface for touch and analyzed through machine learning. In our project, we carried out this part first, as it enabled us to know what was possible with regard to the technical side of creating a somatosensory system.

Part 2: Touch definition and categorization, Touch-Emotion Mapping, Optimization of touch identification and classification, Connection to a Cognitive Modeling Infrastructure The second part of the project meant taking part of the literature available and developing a theoretical model of touch and emotions. An in-depth literature study of definitions and categorizations of different types of touch was followed by a mapping of touch in relation to possible emotional expressions. A study was carried out as to provide new information on expected emotional reactions in humanoid robotics. The cognitive model created for processing and classifying touch was also restructured, redeveloped and optimized as to provide a corresponding identification. Finally, the results were implemented in the formerly created prototype. We also worked as to provide a possible implementation of this system into the elaborate cognitive modeling infrastructure of Ikaros, which would allow for real-time interaction and an extended integration with a robotic response function.

The Humanoid Robot

The goal of our project was to create a somatosensory system for robot interaction. This is carried out as an investigation of touch in humanoid and social robotics. We have worked with an Epi robot, which is an open humanoid platform for developmental robotics created by Johansson et al., 2020. The Epi design is suitable for our purposes and working with Epi is made possible by the Lund University Cognitive Robotics group. The work with constructing a somatosensory system was done on a prototype of the Epi head shell, on which development and testing could be carried out. This provided us an opportunity to to optimize the functionality for our purposes. A cognitive model for Epi can be constructed in Ikaros (Balkenius et al., 2010), which is compatible and often used with Epi.

Touch Identification

A previous project had created a functioning prototype for touch detection from the Epi head, using conductive paint on the inside of the Epi head shell, connected via cords to a board with capacitive electrodes. We continued with a similar construction and developed new versions of the Epi head to optimize the detection as the first part of our project.

We further took interest in what the detection of touch confers to interaction and therefore reviewed related literature as to evaluate different techniques for this. Due to the limitations in time and budget of the project, as well as in technical knowledge within the group, we agreed that the usage of an Arduino board, with associated sensors, was within the scope of the project. Based on this, the *Components-Sensors* part of the Arduino Store (Arduino, n.d.) was thoroughly browsed and the available sensors evaluated. As discussed above, the somatosensory system and the method for recognizing touch does not necessarily have to describe a one-to-one mapping with the human somatosensory system, wherefore different types of available sensors could be investigated. The ones identified as compatible with our system and providing a detection making possible an identification of touch were:

- Temperature and Humidity Sensor
- Button
- Temperature Sensor
- Ultrasonic Sensor
- Digital Light Sensor
- Infrared Proximity Sensor
- Touch Sensor
- Light Sensor

Temperature and humidity would be interesting to work on since interpersonal and social, affective touch tends to involve touch through human skin, where changes in humidity and temperature could be tracked. We however disregarded this approach due to temporal limitations as humidity and temperate are not sufficiently measured within the short time intervals related to the touch interaction considered in our project. Light and sonic sensors have a processing and response time that could be used for these purposes. The limitations of light sensors is in their dependence on available light and in that different light settings would have an effect on how patterns of touch are detected. While not working in the dark, a light setting would require a specialized calibration and different light settings may require new calibrations. The ultrasonic and infrared sensors, were concluded to not be useable for the distances involved the detection of touch and in differentiating between different touches, where a resolution in the order of millimeters can be considered necessary for a reasonable measurement. Using a button as detector would be a mechanical solution involving a pressing that would have to be directed in a limited area and grosser in its application, touches that are executed with lower pressure would not trigger it, making it directly unsuitable for affective touch. The button would further have to be placed on the outside of the robot head, and this would be detrimental to the humanoid appearance of the Epi robot.

The touch sensor would also make possible a placement on the inside of the robot *skin* and works through the measurement of capacitance, the charge of an area relative to electrical potential, which could be used to detect the patterns of touch of interest, including gentler touch. The measurement of capacitance is not directly related to the pressure applied, but depends on the surface of contact, which is limited to the sensor size, of 20x20 millimeters thus requiring multiple sensors to detect touch over the area of the robot head inside surface. Each sensor would also require a separate connection to the Arduino touch board used for processing. We therefore considered it outside the scope of our project to create an implementation of touch detection using such touch sensors including a high economic cost, perhaps in the end any way covering an insufficient number of points to provide a detection of necessary quality.

We therefore decided to employ a detection of touch from electrical charge through the use of conductive paint, where we could decide the size of the detector areas used, as well as the shape of different areas. Conductive paint can specifically be used as a potentiometer sensor that provides a measurement of capacitance, meaning the charge of an area relative to its electrical potential. These areas can be connected to electrodes for the tranmission of an electrical signal to be processed. We settled on the use of Bare Conductive's Electric Paint (Bare Conductive, n.d.) together with Bare Conductive's Touch Board (Bare Conductive, n.d.) (see Fig. 2).

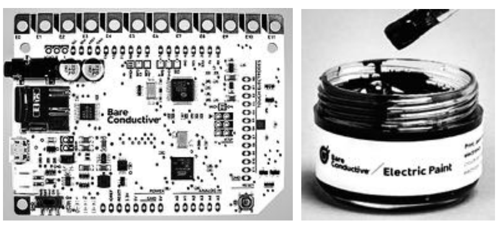


Figure 2: Chosen hardware

The Bare Conductive touchboard consists of 12 electrodes (E0-E11), where each electrode connects to an MPR121 capacative touch sensor, thus providing touch sensing for further processing. Furthermore, the touch board has a serial output for both receiving and transmitting data, with a micro USB connection. The board is based on an Arduino Leonardo board and therefore uses the ATmega32U4 microcontroller (Bare Conductive, 2020). This meant that the board could be reprogrammed for our purposes, using the Arduino IDE, connecting through a micro USB cable to a computer.

At Part 1 one of the project, the board was programmed to enter a triggered state when the capacitance increases above a certain threshold, in at least one electrode. At the triggered state, the values of E0-E11 are sampled at t=0 and do so every 0.5 seconds over a time window of 3 seconds. These values are converted and samples as binaries, meaning that if an electrode is touched, it is represented as a 1, and if an electrode is not touched, it is represented as a 0. If none of the electrodes are high on a sample, the triggered state will end before the time window has finished setting the remaining sample to 0. Accordingly, a touch in our system was represented as a vector of 84 values (7 samples times 12 electrodes) in the first prototype in Part 1 of the project.

The simplified picture of the setup can be seen in Figure 3. The electric paint is painted on the surface inside of the robot head (i.e., the *skin* of the humanoid) and connected via cords to one of the board's electrodes. Once a conductive material, such as the skin of a human hand, touches the surface, there is a change in capacitance in the electrode on the board, thus causing a triggering that can provide the grounds for identification. Triggering can also occur when conductive material is in the proximity of the paint, and an electric field can be detected from a distance.

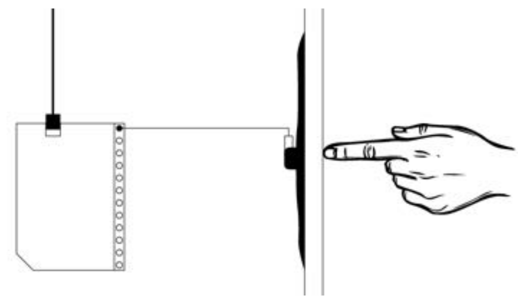


Figure 3: Simplified setup of identifying touch.

Each of the electrodes are connected to an area of conductive paint on the inside of the robots head. The number of areas where set to twelve due to there being 12 electrodes on the board. The topography of the different areas are shown in Fig. 4 and Fig. 5. Note that the model represents the view from the inside of the head, and therefore has to be seen as a mirrored image of the outside. The technical setup of a prototype will be further shown below.

With different areas on the head that potentially triggering touch, we can detect a pattern of touch within a time window, and different types of touches can thereby be represented. An example of how touch

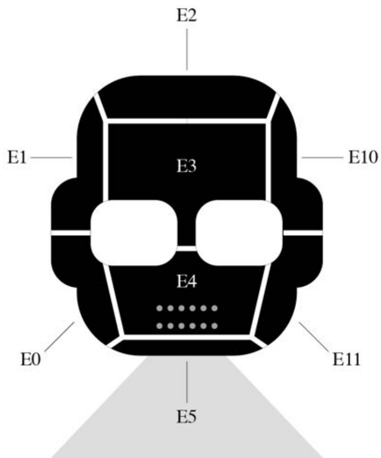


Figure 4: Topography of areas and connected electrodes, front.

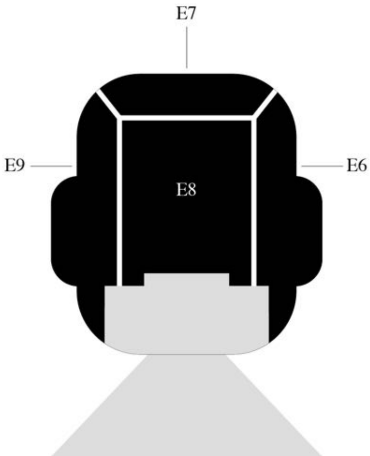


Figure 5: Topography of areas and connected electrodes, back.

is represented in this way can be seen in Table 1, which we treat as a vector with 84 places, where the exemplified touch is a measurement of a 'stroke', over multiple areas.

t\E	0	1	2	3	4	5	6	7	8	9	10	11
0	0	0	0	1	0	0	0	0	0	0	0	0
0.5	0	0	1	0	0	0	0	0	0	0	0	0
1.0	0	1	0	0	0	0	1	1	0	0	0	0
1.5	0	0	0	0	0	0	1	0	0	0	0	0
2.0	0	0	0	0	0	0	0	0	0	0	0	0
2.5	0	0	0	0	0	0	0	0	0	0	0	0
3.0	0	0	0	0	0	0	0	0	0	0	0	0

t rows represent time stamps (s),
E columns represents electrodes

Table 1: Binary representation of a touch (84 data points).

In this first proof of concept implementation, in Part 1 of the project, 4 types of touch were initially applied, which we refer to as 'short poke', 'long poke', 'stroke' and 'hold'.

The technical implementation was upgraded in Part 2 of the project. With an increased number of types of touches, the former representation of touch had to be expanded and optimized further. This meant changing from binary to continuous values in the processing of the signal, in order to enable an increased differentiation of touches. This could also potentially enable a measurement corresponding to the pressure of the touch applied, where pressure as force per unit area may be related to an increase in the total charge of an area and where changes in capacitance would therefore increase with the area of touch on the surface. We investigated different methods for doing this and settled on a technique based on three aspects: *1)* since the capacitance can differ based upon were the unit is placed (in connection to other conductive units or material) a baseline is first defined for every electrode, *2)* this baseline is subtracted from the measured value in order to get a delta-difference in capacitance, and *3)* setting very low deltas to zero in order to discard noise and disturbance. After applying such a scheme we can now differentiate values in "strength of touch" in a range of 0-250 for every sensor.

Furthermore, we decided to increase the sample rate to ∼10 Hz from (∼2 Hz) in order to get a more detailed measurement of the length and timing of a touch. In order to optimize our classification model and get a higher accuracy in the prediction of touches this was eventually increased further, to ∼28 Hz, (see *Touch processing and classification*). The maximum duration of the time window was kept at 3.5 seconds.

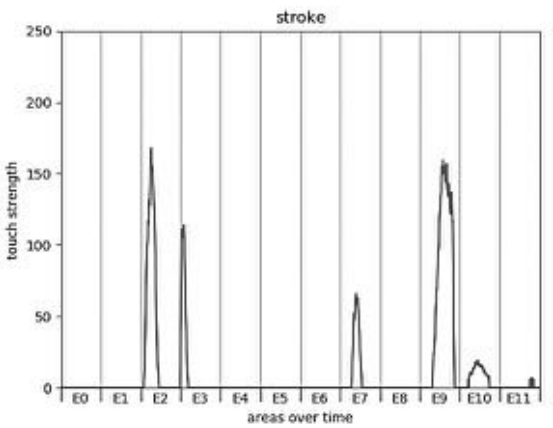


Figure 6: Continuous representation of a touch (1200 data points).

A final representation of a touch based on continuous values at a high sample rate was thereby developed. An example of the representation of 'stroke' using the new approach is illustrated in Figure 6. Examples of the different types of touch included in the study can be found in Appendix A.

Touch Processing and Classification

In order to analyze and classify the different types of touch, we considered multiple alternatives and concluded that the implementation of an artificial neural network (ANN) could be used. A goal of the project was to use and integrate the ANN as new functionality in Ikaros, as cognitive modeling infrastructure Balkenius et al. (2010) and which is compatible with Epi. In Part 1 of the project, it proved time-consuming to carry out this implementation as it required extensive programming in Ikaros, which uses C++ as its programming language. We came to consider this a bottleneck to the progress of the project and worked on a standalone system for analysis and classification system in Python instead, where the machine learning platform of Tensorflow (Martın Abadi et al., 2015) was employed, for testing the conceptual design. Tensorflow is described as "an interface for expressing machine learning algorithms and an implementation for executing such algorithms" (Martın Abadi et al., 2015).

The interface, developed in the form of Python scripts, came to consist of three main parts: *a)* a data generator, *b)* a trainer and *c)* a predictor, which were all used in the process of creating, handling and distributing data to the ANN. The interface received data that was streamed from the touchboard via a USB connector to a computer, on which the interface was run. Tensorflow was used to create an ANN model, which was then run through this setup.

To train the ANN, the robot head was touched according to approximate patterns for different types of touch, with the corresponding triggers being streamed through the serial interface, and handled digitally within our systems. The different representations of touch were formatted and outputted to a CSV file suitable for storing data. Training was a time-consuming process, involving the repetition of different types of touch on the robot head, where at least 50·4 examples of touch had to be recorded to create a proper classification model for the first prototype in Part 1 of the project.

The design of the trainer allowed us to feed the network with the data set generated. The ANN automatically generated and mapped out characteristics for every touch. The trainer's output was thereby a model of a network that had attained the knowledge and ability to classify an alien touch on the robot head. The trained model produced could now be applied to the system and the predictor. The process of learning in ANN can in certain ways be considered analogous to teaching a child about the world in order for it to later use that knowledge on its own.

The predictor design allowed us to use the trained model on data that was directly streamed from the board, following triggering from its head, to predict and classify the type of touch. The identification of touch based on this setup proved successful, as the somatosensory system prototype could process and classify touch, discriminating between the different types of touches of 'short poke,' 'long poke,' 'stroke' and 'hold'. We enabled the predictions to be printed in simple graphics on screen, directly following the application of touch and prediction. Fig. 7 shows the output thereby produced, where the vector presented represents a predicted value of a specific touch. The indexing applied is here 0 for 'short poke', 1 for 'long poke', 2 for 'stroke' and 3 for 'hold'. In this way, and by applying these touches, we could produce a proof of concept for the somtatosensory system. The second value in the vector (index 1) shows a prediction value of 0.99851, meaning that the model predicts it to be a 'long poke' with a certainty of 99.9%.

The network architecture of the ANN was created as a Keras sequential model from the Tensorflow library. This architecture could be regarded as basic, but may be considered resourceful in applying dense, fully connected, layers and allowing for complex computation in determining the classification output from touch input. Two hidden layers, between output and input, with 100 neurons each, were applied with *relu*, rectified linear unit activation, in our model. The training phase used a standard Adam optimizer, a sparse categorical cross entropy loss function, and ten epochs, number of runs in training. A classification accuracy of 98.03% was achieved on the training data model, meaning that a model was made that could include and classify 98.03% of the input values, corresponding to touch patterns, in the training data set. The actual classification accuracy of an unknown data set was yet to be determined since we did not have a data set for testing at this stage in our project. However, when manually introducing new values, by touching and checking which prediction was made, with approximately 9 out of 10 predictions being correctly classified, we regarded this a simple estimation of accuracy.

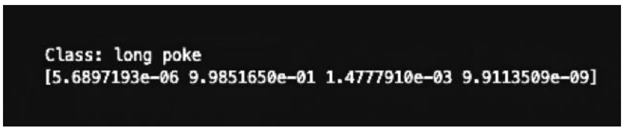


Figure 7: Touch prediction in the early stages of the project.

In Part 2 of the project, we proceeded with further optimizing the representation of touch and the ANN. However, when applying the ANN model for classification of the final representation of touches, using 12·35 (later increased to 12·100) data points of continuous values and 11 different types of touch in classification, it was concluded that the model could no longer be considered optimal. It now had an accuracy of approximately 27%, i.e. better than chance, but not usefully applicable for our classification. As described earlier, the ANN structure can contain an indefinite amount of combinations and may in this respect be considered difficult to configure for optimization. Based on that the SVM can provide a better performance in classification, and being simple to tune, we continued by studying this new technique and were able to apply and later test it. Scikit-learn (Pedregosa et al., 2011), a Python library for machine learning, was used to create the classifier of the SVM, using an 'rbf' kernel, and other settings of an optimized C, gamma values and balanced weights. We now also use 85% of the data set for training of the model, and 15% for the validation and testing, as preferred in machine learning. The average accuracy from 10 epochs of training was now approximately 43%, a relevant increase compared to the former ANN, but not yet at a satisfactory level for our purposes. In applying a normalization to our data set, by subtracting the mean and division with the standard deviation, the average accuracy from 10 epochs increased to approximately 81%, which confers a reasonable classification. By evaluating the confusion matrix, of how touches tend to be misclassified, we could discern a major problem in how the classifications of 'poke' and 'tap' tended to overlap. The reason for this could be the small difference in time scales between these touches, and we therefore proceeded by increasing the sample rate used from 10 Hz to 28 Hz. The touch data now comprised 1200 data points from 12 electrodes and 100 samples compared to the previous 420 (12·35) data points, providing a better possibility of distinguishing touches 'poke' and 'tap'. We can assume that this is since the analysis provided by the SVM was given a higher resolution and an increased space in defining a separating hyperplane. Continuing with the project and given the new prerequisites, a new training data set had to be produced. An extensive session for generating data for training was therefore carried out (Fig. 8), applying and recording 100 entries of the different touch types, producing in total 1100 touch patterns. The new data set was saved, structured and later used for training a new classifier based on the SVM. This finally resulted in an average accuracy of 89%, after validation testing on 15% of the data set, 10 times (i.e. 165·10 touches).

Figure 8: Creation of data set for training the cognitive model.

We further developed the setup so that we could get a screening of the prediction probabilities, enabling also the possibility of processing a combination of touches. The output from applying a touch on the surface of the humanoid head is viewed in Fig. 9. The different types of touch applied are shown in Fig. 9. The selection and definitions of touch are based on our studies of related research, carried out in the purpose of expanding and increasing the types of touch known to Epi. We consider our investigation of touches to provide grounds for the use of the different types in the somatosensory system, which we will expand on below.

Figure 9: The prediction of touch after elaborate optimization of the setup.

Moreover, time and effort was spent on applying the concept of this setup and the classification of touch developed, into the cognitive modeling infrastructure of Ikaros, as we will be elaborate further below.

Prototypes

Prototyping and producing a functional somatosensory system were the primary focus of Part 1 of our project. We wanted to investigate what was technically possible, given the material available and test which conceptual approaches were feasible, before advancing the objectives of the project.

We started from a standalone plastic head, a prototype that was created in a former project. The Epi head shell was 3D-printed in a thin plastic material, with conductive paint on areas on the inside of the head, with cords attached to the different areas and glued, soldiered and taped to them (see Fig. 10). It was unfortunately in pretty bad shape and in many respects fragile. We therefore applied a basic undercarriage out of cardboard to its design to make it more stable, filled in the areas of electric paint further and could together with the gluing and taping of the nestled cords, use it as our first prototype, create our first data set and do prediction testing.

There were issues regarding the robustness of the paint in this shell, and an additional reproduction of the head, with a new thicker shell, was also made available to us, where graphite paint was applied, but it had problems in conduction. With the printing of a third shell, in an effort to refine the design, we applied the conductive paint and attached the cords so that the problems found in previous constructions would not occur, where we took note to separate areas of conductive paint properly to avoid interference due to fragments of paint between areas, and fastened the cords using screws and screw threads submerged into the shell. The setup of the third version of the Epi prototype is shown in in Fig. 11 and Fig. 12.

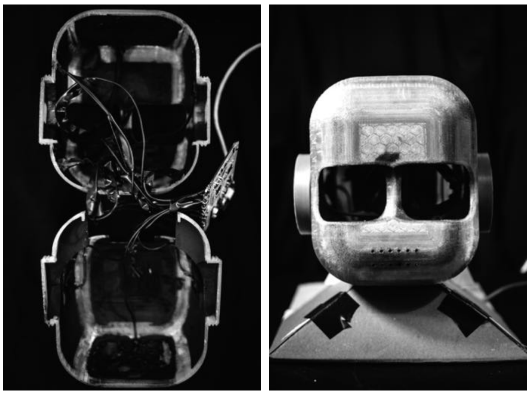


Figure 10: First prototype.

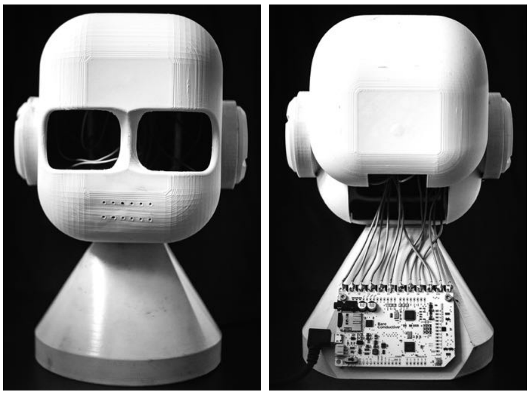


Figure 11: Upgraded prototype, outside.

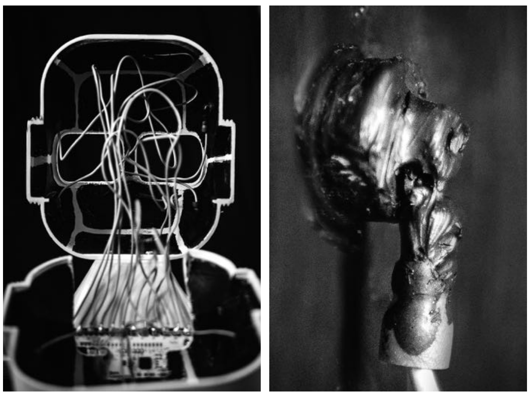


Figure 12: Upgraded prototype, inside.

Touch Definition and Categorization

In Part 1 of this project, the types of touch were confined to 'short poke', 'long poke', 'stroke' and 'hold' and the choice of these types were based on an assumption that they could be considered simple and basic types of touch, that could be used to implement and test the technical aspects of the project. In Part 2 of the project, a primary focus was on basing the design of the somatosensory system on prior research and studies. Therefore, an extensive review of the literature available on work related to the categorization and classification of types of touches, was carried out. Roughly 50 articles, and a few books, were scanned by the group members before concluding that a standard in the classification of touch had not been established and that we should create a classification based on a combination of the studied material, but that additional input was also needed.

We came to use 5 of the studied articles that contained relevant breakdowns of affective and social touch of different types and collected the different types of words used in classifying touches, from van Wingerden et al. (2014), Alonso-Martín et al. (2017), Sun et al. (2017), Huisman (2017) and Cooney et al. (2012a), and compiled these in Table 2. Names of touches specifically involving a location were taken by removing the location word, so that e.g. 'stroke cheek' became 'stroke', and 'rub back' became 'rub'. From here, only the words for touches that were applicable on a humanoid robot head with non elastic skin and movement constraints were extracted. Words that had the same meaning were merged.

As the touches described in the literature sometimes provided a useful definition for our purposes, but in other cases were not clearly defined, we used a combination of the definitions, together with their definition in Merriam-Webster dictionary (Merriam-Webster, n.d.), together with our own opinions and interpretation, in order to define every type of touch as clearly as possible. The location of the applied touch were chosen and defined so that the exclusion of this factor in the study should not prove detrimental to our study. These definitions are presented in Table 3, together with the related results.

Touch-Emotion Mapping

As we had an interest in possible reactions to touch, and could in our review of literature find a lack of related research (van Erp & Toet, 2015), we therefore gave this a significant focus in Part 2 of the project. To map specific types of touch to specific emotions, a thorough study was carried out regarding people's expectations of what emotion(s) a specific touch triggers on a humanoid robot. The study contains 41 participants with a wide range of backgrounds. The participants were almost half males, half females, with ages varying from 18 to 58 years old, and they originate from 12 different countries (including three continents). However, a clear majority of participants originate from Sweden, with ages between 21-30 years old. See Appendix B for more information about the participants. There has to be a clear transparency with the origin of the participants of the study since it has

Touch	1	2	3	4	5
Abstract				■	
Contact				■	
Cover					■
Grab	■		■		
Handshake				■	
Holding hands				■	
Hit	■				
Hug				■	■
Kiss				■	■
Massage	■				
Minimized touch					■
Move					■
Pat	■				■
Pinch	■				
Poke	■		■	■	
Press	■		■	■	
Pull					■
Push			■		■
Rub	■				■
Scratch	■		■		
Shake					■
Slap	■	■			■
Stroke	■	■	■	■	■
Squeeze	■		■	■	
Tap	■	■			
Tickle	■	■		■	
Touch					■

1: van Wingerden et al. (2014), 2: Alonso-Martín et al. (2017), 3: Sun et al. (2017), 4: Huisman (2017), 5: Cooney et al. (2012a).

Table 2: Found type of touches

been shown in previous research that cultural belonging has a significant impact on how we interpret social touch (Hertenstein et al., 2009; Field, 2001). Therefore, the results of this study should be considered an average with regard to culture, provided the majority of the participant come from, or belong to, a Swedish culture. Results based solidly on the Swedish group are attached in Appendix C.

In the study participants were shown a recorded video of a touch being applied on a humanoid robot's head. All of the defined touches in Table 3 were included. After each video the participants were asked:

> "What emotion(s) do you **expect** Epi to express **after** the applied touch?"

The focus was on the participants' expectations after the touch had been applied, thus the bolded parts.

As an answer to the question above, the participants should rank each of the six basic emotions (sadness, anger, enjoyment, disgust, surprise, and fear) according to three options; 'none' (0.0), 'weak' (0.5) and 'strong' (1.0) where 'none' was the default. If unsure about a specific touch, the participant could answer 'I do not know' and skip it. After completion, the participants had the possibility of commenting on the study and what they thought about their participation in it, as the last part of the survey.

The data from the study was compiled and displayed based on the mean value of a specific emotion, for a specific touch. The compiled data from the study can be seen in Table 11 under *Results*, where expected emotions, and their strength, have been mapped to each touch studied in this project. The strongest emotion(s) of each touch, with a mean greater than 0.40, were compiled in a more compact format and are presented in Table 5. The amount of inconclusive answers were also collected and compiled, as presented in Table 6, followed by the participants comments.

Connection to a Cognitive Modeling Infrastructure

An initial goal for our project was that the data produced from touch should be used as input for the Ikaros program, to be run on a standard computer, to further analyse the signal. Ikaros is a system level cognitive modelling program that was originated by the Lund University Cognitive Science robotics group (Balkenius et al., 2010). The focus of the program is on functionalities for simulating brain computation and neurological functions, for the investigation and study of the brain and how its functions process input provided the program, in the purpose furthering the understanding of neurological processes. Cognitive modelling here means to create a simulation which functions as the corresponding cognitive system does with regard to how information is processed. The Ikaros program further includes the possibility of controlling robots and prescribing robot actions based on the processing taking place within the system and is well integrated with the Epi humanoid robot (Johansson et al., 2020).

The infrastructure of Ikaros is based on inputs and outputs, as formalized in XML protocols, including specifications of their types. The touchboard output could be accessed as the Arduino card is attached to a USB port, which could be made an input to Ikaros. Ikaros processing is comprised of modules for specific functionalities, that are written in C++ code. A graphic interface is also made available to act as display for the data streams produced, which allows a monitoring of the signal within Ikaros.

In Part 1 of the project, a simple rule-based classification of touch, from a previous project, was already implemented and we investigated the possibility of using additional functionality for machine learning in Ikaros instead of such a rule-based approach. This however proved difficult as the stand-alone machine learning modules in Ikaros were examples of a perceptron and of back-propagation, which would be too simple networks for us to apply to the analysis of touch in the somatosensory system. We here considered the possibility of constructing a new module in Ikaros, and through programming create a more extensive ANN. We therefore investigated the literature on possible ANN:s, but instead came to continue using the Tensorflow library in a Python setup.

In Part 2 of our project, we came to further consider the implementation of the somatosensory system and machine learning in Ikaros, where this could be achieved through the use of external libraries. Specifically, we investigated the possibility of implementing the Tensorflow model applied in our Python setup, either through the adaption of code, the inclusion of a related library in our Ikaros setup, or through the inclusion of the model as a standalone object. In this investigation we came to install the Tensorflow API and a Tensorflow code library for use in an implementation in Ikaros, in a common environment. This was a difficult setup, which proved time-consuming to make functional, as Tensorflow does not provide support for C++. Further issues regarding setup included its implementation on a Linux system, using a computer with a limited processor, and incompatibilities with certain related program packages, as well as the performance of the computer used in providing a functional compilation in a reasonable time-span. We were however eventually able to use the Tensorflow functionality for saving and exporting a trained model and including it as an object in an Ikaros module, accessing the Tensorflow setup. A module in Ikaros was developed as to allow an input corresponding to that of the pretrained model, and in which the model object from Ikaros could be included. Running on this module, with a corresponding input, we were able to produce the related prediction, in Ikaros. As the Python setup was extended to use an SVM, however, and with features of extended numbers of touches, the relation of touch to emotion and an increased resolution of signal already implemented in the existing Python setup, the implementation of these functions into Ikaros proved beyond the scope of our project.

The implementation of an external library in Ikaros could however be further developed as to provide a consistent framework for cognitive modelling in connection to the somatosensory system. This would allow for integration with other modules in Ikaros, including the various aspects of cognitive modelling, and make available interaction in real-time. It would also simplify learning so that it could take place in real-time, be directly updated and processed in the machine learning, which is not possible in our current Python setup. This would make the process of producing discrimination of touch a continuous part of the interaction process with the robot. With an Ikaros implementation, also, the responses of the system could be channeled to suitable actions of as performed by an Epi robot. The robot's response could be related to touch, with Ikaros allowing the connection of output from one modules functionality to the input of another module, so that a touch module could be linked to a response module. With an Ikaros integration we could hence have the robot provide reactions to being touched, such as in motor action through e.g. a tilt of the head, a change of the colour of the eyes, or through using different auditory expressions produced from sound synthesis. This could further lay a foundation for performing a study in which learning is carried out by research subjects applying touch to the somatosensory system, gaining a more varied data set. This could include supervised and in an advanced stage also unsupervised learning. The interaction with the robot would then also involve direct responses from the robot. The responses of the robot could in turn be evaluated by the research subject and related to the type of touch applied, allowing for a more elaborate bidirectional interaction and a more advanced evaluation of the somatosensory system.

3 Results

Types of Touch

From examining related research, as described earlier, possible different touch types were compiled and defined. These types were merged and narrowed down to fit our purposes and to be applicable to touches on a humanoid robot head like that of Epi. We thereby defined 11 different types of touch; 'hold', 'rub', 'pat', 'pick', 'poke', 'press', 'scratch', 'slap', 'stroke', 'tap', and 'tickle'. These touches are compiled in Table 3.

Touch	Definition	Area
Hold	Grasp with two full hands, no movement.	7+9+11, 8+10+12
Rub	Touch with flat hand, move repeatedly back and forth.	5
Pat	Touch repeatedly with the flat of the hand, fingers closed.	1
Pick	Touch repeatedly with one finger.	5
Poke	Quickly jab or prod with one finger.	6
Press	Jab or prod with multiple fingers.	6
Scratch	Touch with fingernails, move repeatedly back and forth.	1
Slap	Quick hit with flat hand.	10
Stroke	Touch with several fingers or flat hand, move slowly over an area.	5→10
Tap	Quick and very gentle touch with one or several fingers.	10
Tickle	Touch gently and repeatedly with several fingers after each other.	12

Area simply states the chosen area used in this project and study.

Table 3: Definition of touches

Furthermore, the area on which we applied a specific type of touch in our study are listed under *Area* in the same Table 3, where Fig. 16 shows the numbers, from 1 to 12, assigned to different areas on the Epi robot head design. These areas describe the location and movement of touch and are not identical to the conductive paint areas on the inside of the head, used for touch detection (See Fig. 4 and Fig. 5).

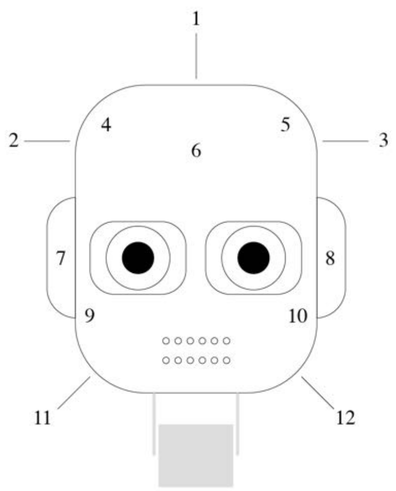


Figure 13: Areas of touch.

Touch-Emotion Map

In Table 11, the results from the study on mapping of touch-emotion, are presented. For every touch, the expected triggered emotion and the mean strength of assigned to it, are shown. The scale for strength ranges from 0.0, for the choice of 'none' to 1.0, for the choice of 'strong', where 0.5 could be considered the value for 'weak'. Emotions are in this table color-coded for strength, with a darker shade meaning stronger and a lighter shade meaning weaker, as to provide an overview of this variable. We see from this table that 'enjoyment' has a relatively high representation among the emotions selected, and that 'anger' and 'surprise' are also often considered expected emotions to be triggered from touch. 'Sadness', 'disgust', and 'fear' are less related to touches in our study, except for special cases, such as a 'slap' (that triggers 'sadness') and a tap (that triggers 'fear').

Certain touches were not clearly connected to an expected emotion. As an overview of this, the amount of *I don't know* flags on a specific touch are listed in Table 6, where we can observe that 'press', 'tap' and 'tickle' are the three types of touch that are unclear with regard to reaction, among the different types.

Final thoughts were given from the participants after completing the study, which are summarized below:

- Context is very important when it comes to what emotion a type of touch triggers. This made it difficult for many participants to select an expected emotion from touch.
- It was hard for participants to **not** make a statement about *when*, and in *what* emotional state, the specific type of touch is used, rather than focus on what emotion it *triggers*.
- Some participants took the questions as meaning how they themselves would react to a specific type of touch, when mapping emotions, rather than what they would expect from a humanoid robot. This could be considered a lack of expectation on particular reactions from a robot.
- The participants had some difficulties with mapping types of touch they associate with emotions and their related meanings such as 'being loved' or 'being cared for', with one of the 6 basic emotions used in the survey, and could issue comments like "Enjoyment does not fully cover this.".
- Surprise was considered both positive and negative by the participants, and a wish for distinguishing between being positive and negative surprised was expressed in the comments. This may also correspond to a selection of either *Surprise+Enjoyment* or *Surprise+Anger*.
- Participants commented that *Surprise* is expected when a touch is uncommon, which many of the touch types were.

Cultural and Sexual Differences

When investigating cultural differences, the Swedish group alone (N=19) 14 male and 5 female, born *and* living in Sweden) was compared to the complete universal group (N=41, Swedish group included), due to the lack of a large enough homogeneous group to compare with. The results are presented in Table 7, where differences with values greater than 0.1 are shown. There are no entries for 'enjoyment' in the table, indicating small differences in the evaluation of this emotion. Swedes overall connect the different types of touch less to 'anger', 'disgust' and most clearly to 'fear'. The only relative increase in emotion for Swedes was for 'pick', which was to a larger extent related to 'surprise'. Notably, and not shown in the table, Swedes expect the exact same grade of 'anger', with a value of 0.89, from a 'slap' as the universal group. Conclusions should however not be considered to be succinct as half of the universal group is constituted by the Swedish group. The split out of the Swedish group can, on the other hand, be useful data if a homogeneous group is needed, since previous studies tend to show that there are cultural differences in touch. The complete results from the Swedish group can be found in Appendix C.

When separating the data between males (N=19) and females (N=22), i.e. two almost equally sized groups, we could observe clear differences in expected emotions from touch. Differences in emotional strength greater than 0.10 are presented in Table 8. Overall, it can be stated that females expect stronger emotions when it comes to all types of touch, which is also true for all emotions. In the 40 presented differences exceeding ±0.10, females expect stronger emotions than males in 35 cases, while males only expect a

	Sadness	Anger	Enjoyment	Disgust	Surprise	Fear
Hold	0.16	0.16	**0.61**	0.14	0.58	0.34
Rub	0.21	0.05	**0.63**	0.06	0.29	0.10
Pat	0.21	0.29	**0.40**	0.16	0.38	0.18
Pick	0.19	0.46	0.03	0.31	**0.65**	0.31
Poke	0.19	**0.59**	0.01	0.29	0.45	0.28
Press	0.22	**0.56**	0.00	0.36	0.43	0.47
Scratch	0.04	0.10	**0.67**	0.13	0.46	0.15
Slap	0.45	**0.89**	0.00	0.41	0.63	0.60
Stroke	0.28	0.01	**0.73**	0.08	0.38	0.05
Tap	0.08	0.31	0.00	0.15	**0.61**	0.37
Tickle	0.03	0.21	**0.54**	0.24	0.38	0.13

Scale ranged from *None* (0.0) to *Strong* (1.0), with an option of *Weak* (0.5) in between.
Shading of emotion strength: [0.0-0.2], [0.2-0.4], [0.4-0.6], [0.6-0.8], [0.8-1.0]
The strongest emotion on every touch is **bolded**.

Table 4: Complete map of mean strength of expected emotions triggered by touches. (Universal group)

Touch	Emotion(s)
Hold	**Enjoyment**, Surprise
Rub	**Enjoyment**
Pat	**Enjoyment**
Pick	Anger, **Surprise**
Poke	**Anger**, Surprise
Press	**Anger**, Surprise, Fear
Scratch	**Enjoyment**, Surprise
Slap	Sadness, **Anger**, Disgust, Surprise, Fear
Stroke	**Enjoyment**
Tap	**Surprise**
Tickle	**Enjoyment**

Emotions that have a mean > 0.40.
The strongest emotion from each touch is **bolded**.

Table 5: Most expected emotions from touches

Touch	I don't know
Hold	7 %
Rub	2 %
Pat	2 %
Pick	5 %
Poke	2 %
Press	12 %
Scratch	0 %
Slap	0 %
Stroke	2 %
Tap	24 %
Tickle	17 %

Table 6: Touches that were hard to interpret.

stronger emotion than females in 5 cases. The largest differences seem to be when a 'press' and 'stroke' is applied, where the female expected value is 0.38 higher in 'fear' and 0.48 higher in 'enjoyment'. However, some examples where males actually expected a stronger emotional reaction than females are +0.18 'enjoyment', for a 'pat', +0.24 for 'enjoyment' for a 'tickle', and +0.10 'disgust', following a 'stroke'. The complete results from the male and female group can be found in Appendix C.

A System for Bidirectional Social Touch-Emotion Interaction

In this project we have produced a functional prototype of a somatosensory system, as illustrated in Fig. 14. Results from literature and the study on touch and emotions carried out have been applied and implemented to the system. The somatosensory system comprises the detection of touch on a robot head, the processing of a continuous signal into a digital repre-

Touch	Differences
Hold	-0.14 Surprise, -0.17 Fear
Rub	-0.16 Surprise, -0.10 Fear
Pat	-0.10 Fear
Pick	-0.12 Disgust, +0.13 Surprise, -0.12 Fear
Poke	-0.11 Disgust, -0.10 Fear
Press	-0.13 Disgust, -0.17 Fear
Scratch	-0.13 Fear
Slap	-0.10 Fear
Stroke	-0.17 Sadness
Tap	-0.17 Anger, -0.15 Fear
Tickle	-0.10 Anger

Differences $\geq$ 0.10 are included.

Table 7: Swedish group compared to the universal group.

Touch	Differences
Hold	+0.19 Enjoyment, +0.14 Surprise, +0.14 Fear
Rub	+0.10 Sadness, +0.10 Anger, **+0.30 Surprise**, +0.19 Fear
Pat	+0.19 Sadness, +0.17 Anger, -0.18 Enjoyment, +0.12 Surprise
Pick	+0.10 Disgust, +0.15 Surprise, +0.14 Fear
Poke	+0.17 Anger, **+0.20 Disgust**, **+0.22 Fear**
Press	+0.12 Sadness, +0.10 Anger, +0.11 Disgust, +0.18 Surprise, **+0.38 Fear**
Scratch	+0.19 Anger, -0.13 Enjoyment, +0.11 Disgust, +0.18 Surprise, +0.18 Fear
Slap	+0.16 Disgust
Stroke	+0.14 Sadness, **+0.48 Enjoyment**, -0.10 Disgust, **+0.22 Surprise**
Tap	+0.12 Sadness, **+0.25 Anger**, **+0.21 Disgust**, -0.10 Surprise, **+0.24 Fear**
Tickle	**+0.21 Anger**, **-0.24 Enjoyment**, +0.19 Disgust

Differences $\geq$ 0.10 are included.
Differences $\geq$ 0.20 are **bolded**.

Table 8: Female group compared to the male group.

sentation of a touch and the classification of the type of touch, where a response is given as a representation of emotion, closing the loop of bidirectional interaction.

Summarizing, the somatosensory system is applicable to a humanoid that has an input of at least 100 samples from 12 different areas over 3.5 seconds where the signal created is interpreted as touch and related emotion. The cognitive model has learned the patterns and characteristics of a 'hold', 'rub', 'pat', 'pick', 'poke', 'press', 'scratch', 'slap', 'stroke', 'tap' and 'tickle'. In the machine learning applied, validation provides an 89% accuracy in classification of touch among the 11 types of touch.

The system also makes possible the combination of touch and emotion. The outputs from the identification of touch and related emotion are the probabilities for different touch and the strengths of the related emotions. A touch can be interpreted as for example a 'hold' with fraction of 'rub', and the triggered emotion could be 'enjoyment' with a fraction of 'surprise'. This example is presented in Appendix C, where the humanoid's head was touched with a rubbing hold and the output was 'hold'=53.23%, 'rub'=22.65%. This may be due to an interpretation of a 'hold' from the two hands covering the robot head cheeks for an extended time, but with the characteristics of a 'rub' as the hands are moved back and forth during the holding of the cheeks. If a combinatory scheme was constructed, this could be classified as a 'rubbing hold'.

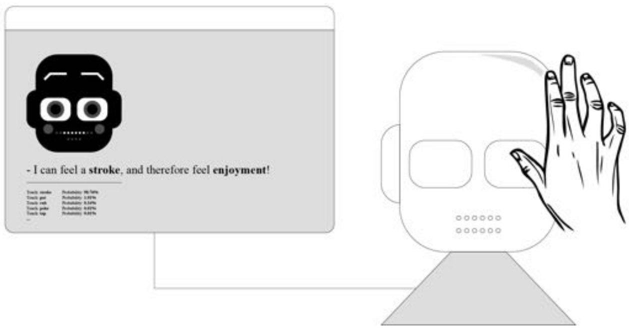


Figure 14: Final system prototype.

4 Conclusions

After having investigated alternative ways of developing a somatosensory system for a humanoid robot, we have produced a substantiated and functional way of *a)* detecting touch, *b)* classifying different types of touch and *c)* relating touch to an emotion as response. Our system could be improved in many ways, but can nevertheless be considered as comprising touch interaction in humanoid robotics.

Using electric paint, a touch board and machine learning (e.g. SVM) allows for a cheap and relatively simple somatosensory system. The electric paint is however fragile and tends to start flaking with time and risks the construction short circuiting when the paint becomes dysfunctional or overlaps between areas. This was helped by applying lacquer and one should be very meticulous and careful when applying the paint not to cause short circuit. In this setup the 12 electrodes available restricts the topography of touch on the robot head, and makes difficult the detection of the location of touch. Location is of particular importance in affective touch. An alternative setup or technique has to be applied to, for example, distinguish between affective touch on the lips and the chin of the humanoid.

As a summary of our findings from studying the mapping of touch and emotion we can say that:

- When considering affective touch as a whole, it is mainly correlated with enjoyment. There is an expectation that touch will produce pleasant emotions.
- Long lasting types of affective touch seem to be more strongly associated with enjoyment.
- Shortly lasting types of affective touch seem to be more strongly associated with either anger or surprise (or both).
- A tap or tickle type of touch seems to be hard too associate with an emotion.
- It is difficult to propose which emotions are expected from touch, when little is known of context, history or personality of the social agent. Affective touch is related to social bonding and this is a parameter that could be added to the ones, described by van Erp and Toet (2015) in the Introduction, of what is needed in order to have a natural interaction in HRI.

From the experience we have gained from working with the classification of different types of touch, we can conclude that an improved foundation of definitions is sorely needed. These definitions would benefit from including at least the following factors: *1)* part(s) of hand in contact, *2)* movement pattern, *3)* duration in time, and, *4)* pressure or force.

5 Reflections

Sources of errors

In this study, there are several sources of error. One of the major sources of uncertainty was the trivial possibility of random answers from the participants in the study carried out, which would have a big impact on our system as a whole since it is based upon the conclusions from the study.

Another questionable aspect is the data set from which the cognitive model that was trained, since this was a data set that contained 1100 man-made examples of touch, for this particular purpose. This data may be skewed due to performing the different types of touch 1100 times can have meant changing the application of touch, for example by speeding it up, and touch thus becoming less ordinary or non-representative. There is therefore a risk that the cognitive model is biased towards stressed types of touch.

When carrying out the study, the different types of touch were only applied to a specific area of the humanoid's head. However, the related emotions, derived from our study, were applied to that specific touch as a whole, without regard to the area of the head. Thus, it has to be taken into consideration that the location of touch used in the study could have biased the results. The triggered emotions used in the system should therefore be seen as approximate, and only completely accurate when applied to the specific areas described in this report.

Regarding the results and conclusion from the comparison between males and females and their expectations of triggered emotion(s) from touch, we need to consider that that cultural differences could have contributed to this, as most of the participants from the male group belonged to the homogeneous Swedish group. In contrast, a large part of the female group was non-Swedish and did not belong to a homogeneous group.

Difficulties during the process

Unfortunately, this project was carried out during the epidemic of Covid-19. Thus, the issue of *a)* not being able to fully meet in person, *b)* not being able to access the LUCS Robotics Lab, and *c)* not being able to carry out physical studies, have to be addressed.

Reflection on the Work Process

The first part of the project began with a primary focus on technical issues and the production of an optimal prototype. This since we were eager to see results, rather than first gaining a solid theoretical foundation to base our decisions on. It was not until the second part of the project that we thoroughly studied the available research, and thereby could construct a system with a firmer foundation in theory.

An alternative approach would have been to start with an extensive study of previous research in order to ensure that our project was based on the available knowledge within the area, before constructing a prototype. In hindsight, it could have been more productive to lay a theoretical foundation and proceed from there in constructing our setup. If there is something that we could have done differently in our work process, it is this; to spend more of our time in the first part of the project on the review of literature, compiling previous research and carrying out studies, and more time in the second part of the project on technical implementation and prototype building.

Suggested Future Work and Research

Researching this area and constructing a system for touch has enabled us to understand what possible next steps in this process could be and what further development could be carried out. Areas that would be of great interest to study further are:

- A way of recognizing the location of a touch without the limitation of a using multiple, yet finite, number of areas.

- Definitions and standardization of a set of different types of touch that could be used for application in a system similar to this. Characteristics included in such definitions could be *1)* part(s) of hand in contact, *2)* movement pattern, *3)* time duration, *4)* pressure or force.

- The relation between vision and touch and its effect on touch classification, both for humans and robotics. Studies can e.g. *a) for human subjects* involve carrying out experiments with subtraction and addition of vision when trying to interpret touch, and *b) for humanoid subjects* involve creating a module with both visual and somatosensory input to classify and predict types of touch from combinatory patterns.

- The possibilities of integrating context and social bonds in a humanoid cognitive model, in order to apply them as factors in the classification of touch.

- An investigation of how culture and sex should be a part of a humanoids interpretation of a touch. *Idea:* Maybe humanoid robots should develop their own culture and own way of interpretation; "The humanoid way"?

- Further optimization regarding identification and differentiation of touch using capacitive sensors. E.g. how does sampling rate, capacative delta scale, number of electrodes and data set format affect the accuracy of the classification model (in an ANN or SVM) when it comes to the prediction of a type of touch?

- "When is a touch over?" E.g. is a 'pat' just an incremental part of a 'stroke'?

- Is age a parameter that should, or should not, be included when evaluating touch? In what way does age have an impact?

- Experiments on cultural differences, when it comes to affective touch, are needed. Our study was, unfortunately, not able to find two homogeneous groups big enough to draw succinct conclusions regarding this.

- Investigate the potential for identifying combinations of touch in the constructed system. How can the system use the probability information (after prediction) for every type of touch in order to create new combined touches? For example, how does a 53% 'hold' plus a 23% 'rub' equal a 'rubbing hold'? What/How many types of touch are needed in order to enable the creation of combinations that can describe human touch interaction thoroughly?

- An implementation of the somatosensory system into the Ikaros framework could allow for a consistent system in relation to other cognitive modelling functionality. This would widen the possibility of touch-based interaction with a humanoid robot in real-time, including different modalities of response. The inclusion of robotic responses in the system would also make possible a future study investigating research subjects in interaction with such a humanoid robot.

Acknowledgments

We would like to acknowledge contributions and attribute certain participation to this project, when it comes to people beside the authors of the report:

Contributions: A special thanks to our former project member Anne Line Gjersem for your work on Part 1 of this project, our supervisor Birger Johansson for giving us input on our work and for printing 3D models, and lastly to Filip Cederquist (master thesis student in applied machine learning at LTH) for guidance on classifying the data.

Attributions: The hand gesture in Figure 3, 14 were supplied by the user 'macrovector' on Freepik.com (macrovector, 2018) and later modified. All is under Freepik License, which allows commercial use (in both digital and printed media, with modifications) under attribution. Thank you for the illustration, 'macrovector'.

References

Alonso-Martín, F., Gamboa-Montero, J., Castillo, J., Castro-González, Á., & Salichs, M. (2017). Detecting and classifying human touches in a social robot through acoustic sensing and machine learning. *Sensors*, *17*(5), 1138. https://doi.org/10.3390/s17051138

Arduino. (n.d.). *Arduino Store sensors*. Retrieved January 4, 2021, from https://store.arduino.cc/new-home/components-sensors

Balkenius, C., Johnson, M., & Johansson, B. (2008). Understanding cognition by building robots. In P. Gärdenfors & A. Wallin (Eds.), *A smorgasbord of cognitive science* (pp. 99–122). Bokförlaget Nya Doxa.

Balkenius, C., Morén, J., Johansson, B., & Johnsson, M. (2010). Ikaros: Building cognitive models for robots [Informatics for cognitive robots]. *Advanced Engineering Informatics*, *24*(1), 40–48. https://doi.org/https://doi.org/10.1016/j.aei.2009.08.003

Banich, M., & Compton, R. (2011). *Cognitive neuroscience*. Wadsworth cengage Learning.

Bare Conductive. (n.d.). *Bare Conductive shop*. Retrieved January 4, 2021, from https://www.bareconductive.com/shop/

Bare Conductive. (2020). *Touch Board technical data sheet*. Retrieved January 4, 2021, from https://www.bareconductive.com/shop/touch-board/

Cooney, M. D., Nishio, S., & Ishiguro, H. (2012a). Recognizing affection for a touch-based interaction with a humanoid robot. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1420–1427. https://doi.org/10.1109/IROS.2012.6385956

Cooney, M. D., Nishio, S., & Ishiguro, H. (2012b). Recognizing affection for a touch-based interaction with a humanoid robot. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1420–1427.

Dunbar, R., & MacDonald, I. (1998). *Grooming, gossip, and the evolution of language*. Harvard University Press.

Ekman, P., & Friesen, V., W. (1969). The repertoire or nonverbal behavoir:categories,origins,usage, and coding. *Semiotica*, *1*(1), 49–98.

Feil-Seifer, D., & Mataric, M. J. (2005). Defining socially assistive robotics. *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.*, 465–468.

Field, T. (2001). *Touch*. Cambridge, MA: MIT Press.

Finnegan, R. (2014). *Communicating: The multiple modes of human communication*. Routledge.

Fulkerson, M. (2014). *The first sense: A philosophical study of human touch*. MIT Press.

Gandhi, R. (2018, June 7). *Support vector machine — introduction to machine learning algorithms*. Retrieved January 13, 2021, from https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

Hertenstein, M. J., Holmes, R., McCullough, M., & Keltner, D. (2009). The communication of emotion via touch. *Emotion*, *9*(4), 566–573. https://doi.org/10.1037/a0016108

Huisman, G. (2017). Social touch technology: A survey of haptic technology for social touch. *IEEE Transactions on Haptics*, *10*(3), 391–408. https://doi.org/10.1109/TOH.2017.2650221

Johansson, B., Tjøstheim, T. A., & Balkenius. (2020). Epi: An open humanoid platform for developmental robotics. *International Journal of Advanced Robotic Systems*, *17*(2), 1729881420911498. https://doi.org/10.1177/1729881420911498

Kerpa, O., K, W., & H, W. (2003). Development of a flexible tactile sensor system for a humanoid robot. *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, *1*, 1–6.

Kolb, B., Whishaw, I., & Teskey Campbell, G. (2016). *An introduction to brain and behavior*. Worth Publishers.

Lau, H.-Y., Tong, K.-Y., & Zhu, H. (2008). Support vector machine for classification of walking conditions using miniature kinematic sensors. *Medical & biological engineering & computing*, *46*(6), 563–573.

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, *5*(2), 291–308.

Limbu, D. K., Anthony, W. C. Y., Adrian, T. H. J., Dung, T. A., Kee, T. Y., Dat, T. H., Alvin, W. H. Y., Terence, N. W. Z., Ridong, J., & Jun, L. (2013). Affective social interaction with cuddler robot. *2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, 179–184.

Löken, L., Wessberg, J., Morrison, I., & et al. (2009). Coding of pleasant touch by unmyelinated afferents in humans. *Nat Neurosci*, *12*, 547–548. https://doi.org/https://doi.org/10.1038/nn.2312

macrovector, F. (2018). Touch interface gestures icons free vector. https://www.freepik.com/vectors/design

Martın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, ... Xiaoqiang Zheng. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems [Software available from tensorflow.org]. https://www.tensorflow.org/

McNeil, D. (2005). *Gesture and thoughts*. The University of Chicago Press.

Merriam-Webster. (n.d.). Dictionary. *Merriam-webster.com dictionary*. Retrieved November 21, 2020, from https://www.merriam-webster.com/dictionary/

Olausson, H., Lamarre, Y., Backlund, H., & et al. (2002). Unmyelinated tactile afferents signal touch and project to insular cortex. *Nat Neurosci*, *5*, 900–904. https://doi.org/https://doi.org/10.1038/nn896

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Remland, M. S., & Jones, T. S. (1988). Cultural and sex differences in touch avoidance [PMID: 3217204]. *Perceptual and Motor Skills*, *67*(2), 544–546. https://doi.org/10.2466/pms.1988.67.2.544

Richardson, I., & Hjorth, L. (2017). Mobile media, domestic play and haptic ethnography. *New Media & Society*, *19*(10), 1653–1667.

Sun, J., Billing, E., Seoane, F., Zhou, B., Högberg, D., & Hemeren, P. (2017). Categories of touch: Classifying human touch using a soft tactile sensor. http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-13838

Suvilehto, J. T., Glerean, E., Dunbar, R. I. M., Hari, R., & Nummenmaa, L. (2015). Topography of social touching depends on emotional bonds between humans. *Proceedings of the National Academy of Sciences*, *112*(45), 13811–13816. https://doi.org/10.1073/pnas.1519231112

van Wingerden, S., Uebbing, T., Jung, M., & Poel, M. (2014). A neural network based approach to social touch classification. *Proceedings of the 2nd international workshop on emotion representations and modelling in human-computer interaction systems, erm4hci 2014* (pp. 7–12). Association for Computing Machinery (ACM).

van Erp, J. B. F., & Toet, A. (2015). Social touch in human–computer interaction. *Frontiers in Digital Humanities*, *2*, 2. https://doi.org/10.3389/fdigh.2015.00002

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer-Verlag, New York.

Wilson, T. D., Reinhard, D. A., Westgate, E. C., Gilbert, D. T., Ellerbeck, N., Hahn, C., Brown, C. L., & Shaked, A. (2014). Just think: The challenges of the disengaged mind. *Science*, *345*(6192), 75–77.

Wong, W.-T., & Hsu, S.-H. (2006). Application of svm and ann for image retrieval. *European Journal of Operational Research*, *173*(3), 938–950. https://doi.org/https://doi.org/10.1016/j.ejor.2005.08.002

Zha, X. (2003). Soft computing framework for intelligent human-machine system design, simulation and optimization. *Semiotica*, *7*(3), 96–183.

Zhang, Y., Laput, G., & Harrison, C. (2017). Electrick: Low-cost touch sensing using electric field tomography. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1–14.

Appendices

A Continuous representations of types of touch

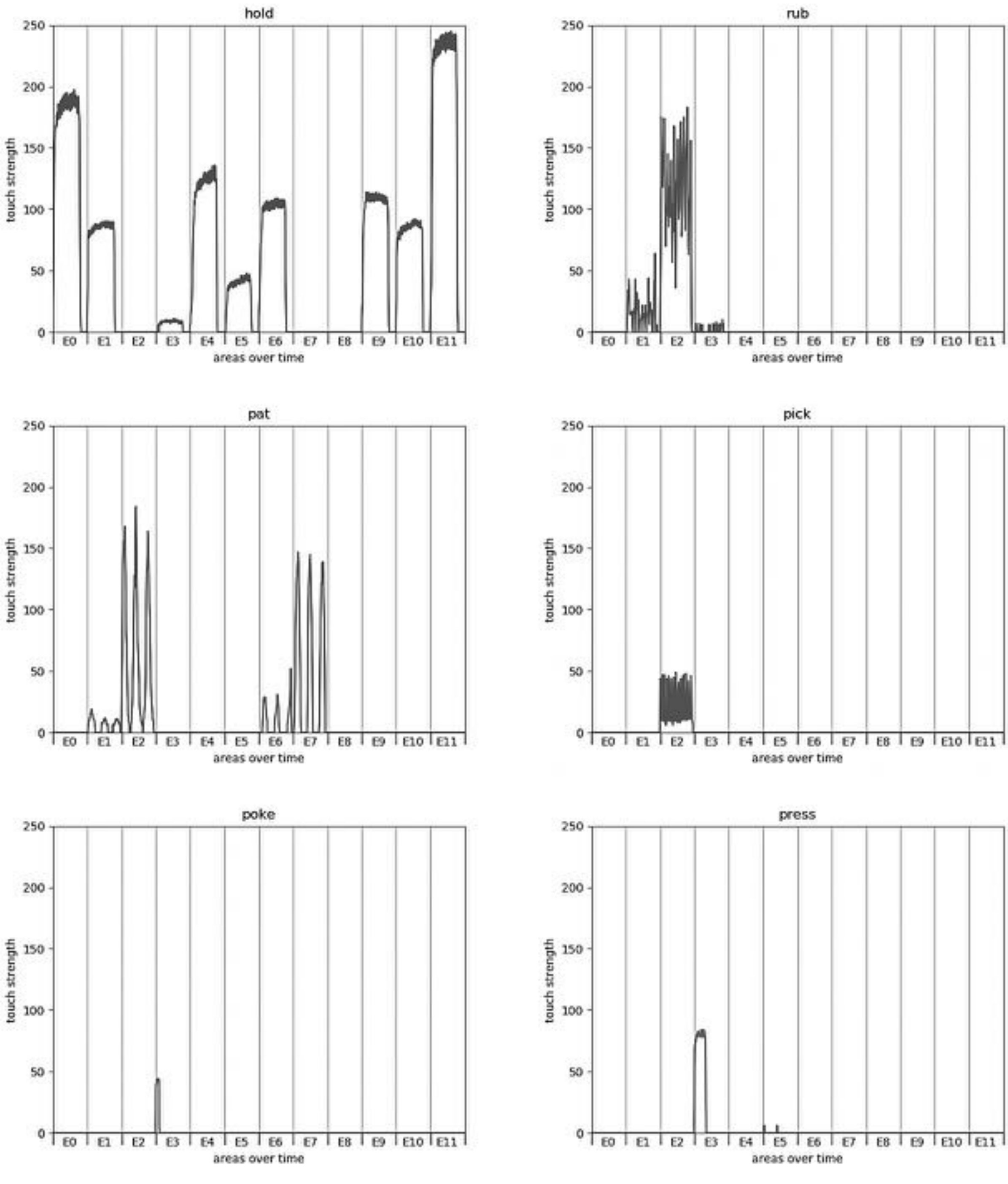


Figure 15: Examples of analog representation of types of touch (Pt. 1)

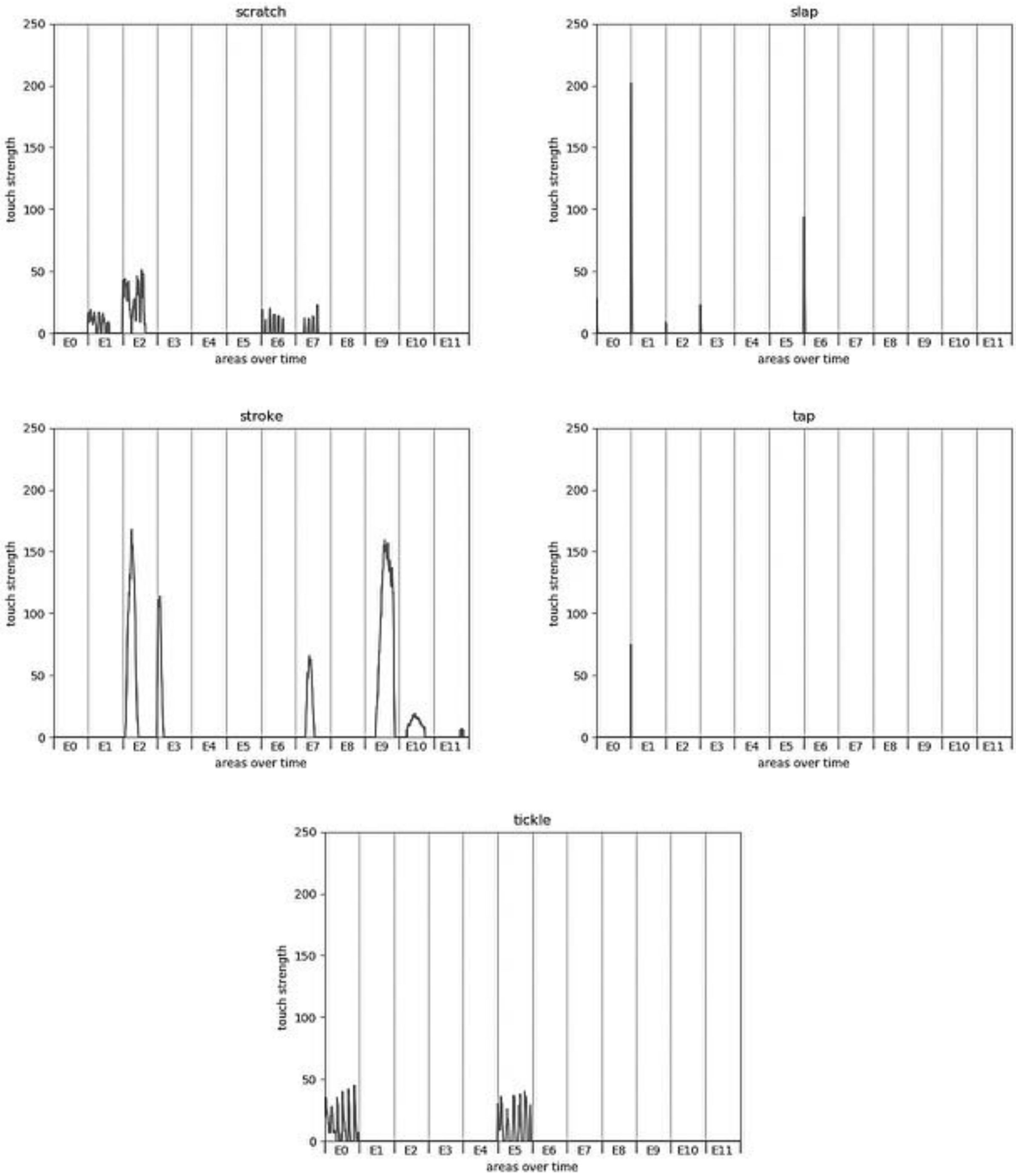
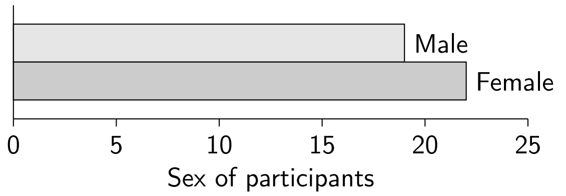
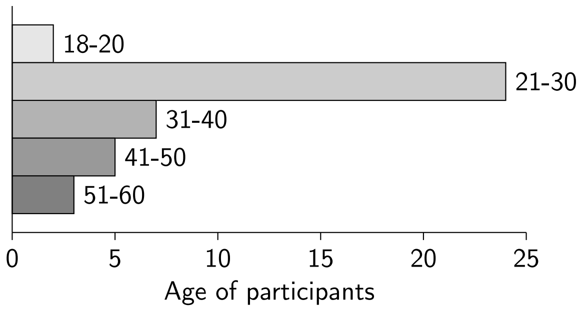
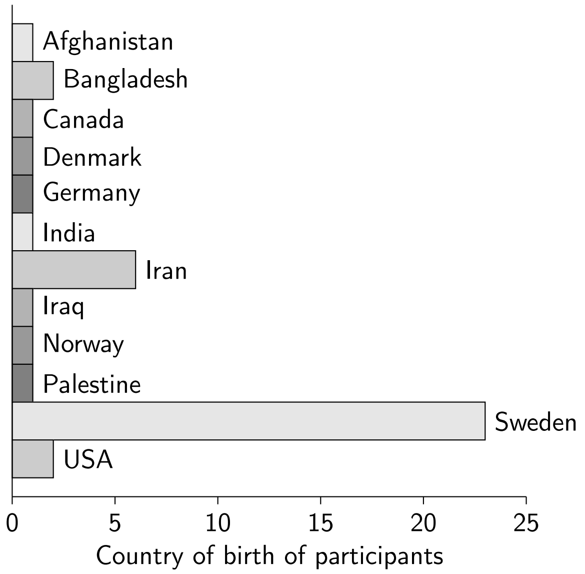


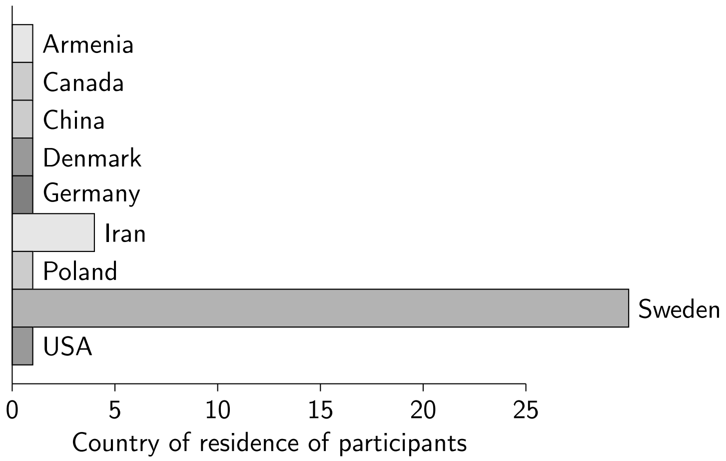
Figure 16: Examples of analog representation of types of touch (Pt. 2)

B Detailed information about participants in study









C Complete maps of mean strength of expected emotions triggered by touches for different groups

	Sadness	Anger	Enjoyment	Disgust	Surprise	Fear
Hold	0.18	0.08	**0.65**	0.15	0.44	0.17
Rub	0.13	0.00	**0.73**	0.05	0.13	0.00
Pat	0.08	0.26	**0.45**	0.18	0.32	0.08
Pick	0.09	0.46	0.03	0.19	**0.78**	0.19
Poke	0.16	**0.53**	0.02	0.18	0.45	0.18
Press	0.20	**0.57**	0.00	0.23	0.37	0.30
Scratch	0.03	0.05	**0.71**	0.07	0.34	0.02
Slap	0.39	**0.89**	0.00	0.29	0.66	0.50
Stroke	0.11	0.03	**0.55**	0.11	0.32	0.05
Tap	0.03	0.13	0.00	0.09	**0.69**	0.22
Tickle	0.00	0.11	**0.46**	0.32	0.43	0.14

Scale ranged from *None* (0.0) to *Strong* (1.0), with an option of *Weak* (0.5) in between.
Shading of emotion strength: [0.0-0.2], [0.2-0.4], [0.4-0.6], [0.6-0.8], [0.8-1.0]
The strongest emotion on every touch is **bolded**
Group: Born and living in Sweden, N = 19 (14 males, 5 females).

Table 9: Complete map of mean strength of expected emotions triggered by touches for the **Swedish group**.

	Sadness	Anger	Enjoyment	Disgust	Surprise	Fear
Hold	0.21	0.18	**0.50**	0.18	**0.50**	0.26
Rub	0.16	0.00	**0.66**	0.03	0.13	0.00
Pat	0.11	0.19	**0.50**	0.11	0.31	0.28
Pick	0.18	0.50	0.03	0.26	**0.59**	0.24
Poke	0.16	**0.50**	0.03	0.18	0.45	0.16
Press	0.16	**0.50**	0.00	0.31	0.34	0.28
Scratch	0.03	0.00	**0.74**	0.08	0.37	0.05
Slap	0.50	**0.89**	0.00	0.34	0.66	0.61
Stroke	0.21	0.03	**0.47**	0.13	0.26	0.03
Tap	0.03	0.21	0.00	0.06	**0.68**	0.26
Tickle	0.00	0.08	**0.69**	0.12	0.42	0.08

Scale ranged from *None* (0.0) to *Strong* (1.0), with an option of *Weak* (0.5) in between.
Shading of emotion strength: [0.0-0.2], [0.2-0.4], [0.4-0.6], [0.6-0.8], [0.8-1.0]
The strongest emotion on every touch is **bolded**
Group: Males, N = 19.

Table 10: Complete map of mean strength of expected emotions triggered by touches for the **male group**.

	Sadness	Anger	Enjoyment	Disgust	Surprise	Fear
Hold	0.12	0.14	**0.69**	0.12	0.64	0.40
Rub	0.26	0.10	**0.60**	0.10	0.43	0.19
Pat	0.30	0.36	0.32	0.20	**0.43**	0.30
Pick	0.21	0.45	0.02	0.36	**0.74**	0.38
Poke	0.21	**0.67**	0.00	0.38	0.45	0.38
Press	0.28	0.60	0.00	0.42	0.52	**0.66**
Scratch	0.05	0.19	**0.61**	0.19	0.55	0.23
Slap	0.41	**0.89**	0.00	0.50	0.64	0.59
Stroke	0.35	0.00	**0.95**	0.03	0.48	0.08
Tap	0.15	0.46	0.00	0.27	**0.58**	0.50
Tickle	0.05	0.29	**0.45**	0.31	0.36	0.17

Scale ranged from *None* (0.0) to *Strong* (1.0), with an option of *Weak* (0.5) in between.
Shading of emotion strength: [0.0-0.2], [0.2-0.4], [0.4-0.6], [0.6-0.8], [0.8-1.0]
The strongest emotion on every touch is **bolded**
Group: Females, N = 22.

Table 11: Complete map of mean strength of expected emotions triggered by touches for the **female group**.

D Example of an output from a combined touch

Figure 17: Combined 'hold' and 'rub'.

Illustrating emotion AI through simulation and interaction with an affective robot

Erik Haglund, Jonas Tvedt

Erik91_Haglund@Hotmail.com
jsb.tvedt@gmail.com

Here we present the development of two prototype applications that allow the user to interact with a robot that we have equipped with emotion recognition technology and simulated emotive expressions. One of the applications is a guessing game where the user and the robot take turns guessing each other's emotional expressions and the other is a simulation of supervised machine learning where the user helps the robot to learn to express emotions. One of these applications, and of course the robot, will be used in an AI science exhibition. Our goal is to demonstrate aspects of emotion AI in affective robotics through a communicative interaction where visitors get to experience and evaluate artificial affective expressions and as well have their own facial expressions analyzed by emotion recognition technology. In addition to presenting the applications, we here report results from user tests of the artificial emotive expressions, the emotion recognition technology, and indicative results from a pilot user test of the prototype applications. The results from testing the artificial expressions suggest that the only expressions (from our selection of 6 basic emotions) recognizable to test persons above chance level is Anger and Sadness. The results from tests of the emotion recognition technology suggest that it is reliable to predict any of the 6 basic emotions in varying conditions. The pilot test of the applications suggests that the guessing game is the preferred interaction, due to it being more "fun", possibly because it is simpler and more dynamic. However, further tests have to be conducted.

1 Introduction

Vattenhallen Science Center is a science and technology museum with a focus on interaction and experimentation. It is run by LTH and the faculty of science at Lund University, and its ambition is to be a venue for communicating science to the public.

Since the fall of 2019, Vattenhallen has been planning an exhibition about Artificial intelligence and machine learning which will be based on actual research and technological development from various fields and institutions, including Lund University. The goal of the exhibition is to address the following questions: What is AI? How does it work? What can it be used for? How does research relate to ethical questions concerning AI? Another goal is to illuminate the width of these questions and show the utility of the fast-growing area that AI and machine learning is. This paper introduces a specific contribution to the exhibition, in development by a group of students in technology and cognitive science.

Our purpose and goals

There is a rapidly growing approach within AI that has the potential to define our interaction with machines. Computational systems and intelligent agents are increasingly given the capacity to alter their behavior according to our emotional expressions and to use affective behavior to influence our affective and mental states. The alterations may entail under-the-hood adjustments to the parameters that govern the interaction, like lowering the level of difficulty in a game or customizing the information available to the user, or they may be in the form of perceivable affective expressions made by the machine itself, like a robot smiling back at you or a chatbot changing its tone and words. Whether explicit or implicit, emotion AI is increasingly becoming a core part of human-machine interaction. Moreover, emotion AI is transcending from a sub-branch to a central theme of general AI (Yonck, 2020).

Given this relevance, we aim to introduce central aspects of emotion AI to the visitors of Vattenhallen Science Center within the constraints of *affective robotics*; a field in which emotion AI has gained solid scientific grounds. Affective robotics is a developing field and a scientific field of study that explores how features of affective computation and behavior, like emotion recognition and expression, can be used to influence the interaction between human and robot (Tao and Tan, 2005; Kirby, Forlizzi, and Simmons, 2010). Our goal is to demonstrate a few key features of affective robotics through simple interaction with a social robot.

It is important to note, as a disclaimer, that the robot used in the interaction will not be equipped with any AI other than simple emotion recognition through face tracking. The robot will not be able to perform any autonomous evaluations or generate any autonomous responses, thus, the interaction will be more or less scripted. However, through framing and progressive responses, we might be able to create the illusion of a somewhat dynamic AI. Nevertheless, the point of the interaction is not to impress with the technological features of AI, but rather to illustrate and stimulate thoughts about limitations and constraints as well as the possibilities of emotion AI in affective robotics.

Engaging the visitors in this sort of interaction entails certain challenges. Especially when considering that the main target group of the exhibition are teenagers at the age of fourteen. Hence, in addition to the informative goals, we have engagement goals aimed at making the content and dynamics of the interaction entertaining and fun without appearing childish or condescending.

Through the interaction, the user is invited to discover the meaning and significance of a robot's emotive expressions that differ in appearance from human expressions but that still have the potential to communicate at the affective level with humans. Furthermore, the visitor is introduced to the idea that human-robot interaction can be mediated by affective features; by emotional expression made both by the visitor and by the robot.

The robot

The robot, Epi, comes in two versions, one including a torso and arms and one which is only a head on a neck (Johansson, Tjøstheim and Balkenius 2020). The full-size Epi has a greater variety in physical and motoric attributes, but due to certain practical disadvantages and the full-size Epi being needed at the robotics lab, we only have access to the Epi-head. We have therefore modeled emotional expressions using only a few of Epi´s physical features, namely alteration of eye color, pupil dilation, eye movements (vertical), head movements (vertical and horizontal rotations), and mouth display. In addition, we make use of the cameras in Epi's eyes to track and recognize visitors' emotional expressions. Using these features have designed affective expressions, displayed by Epi, which can be activated by the visitor through a screen-based user interface, and by emotion recognition data.

An interactive station

When the visitors arrive at the station they will come face-to-face with the Epi head, which sits in a glass cubicle. They will also find the screen below the cubicle through which they can interact with Epi. The interaction starts when the visitors have read the instructions displayed on the screen and pressed "start". After this follows a set of steps that involves back-and-forth interaction with Epi, that is, the visitor and Epi respond to each other's actions in a step-by-step manner until a result is obtained.

The interface and program used to mediate the interaction between the visitors and Epi runs as an application where button presses dictate the flow of the interaction.

Emotion recognition through face tracking

To achieve our interaction goals we have included a face tracker API, which runs on the Tensorflow machine learning platform, in the application. The face tracker utilizes the cameras in Epi's eyes to allow the visitor to take photos which can then be used by our application, via the face tracker, to predict what emotions the visitor tries to convey. This opens up the possibility for the visitor to have fun and make exaggerated faces if they want to. Ther face tracker allows us to utilize as much of the build-in functionality in Epi as possible to better showcase its potential and better meet the expectations of the visitors. Allowing Epi to see, and at some level understand the visitors also adds to the interactivity of the station, hopefully engaging the visitors by pulling them into a more dynamic interaction.

An important point to clarify is that the system will not save any photos taken of the visitors, something that we intend to inform the visitors about before taking the picture.

Adding emotion recognition does not end with implementation, it's an important aspect that the data provided by the face tracker are reasonably precise, and because of this, reliability tests have been performed. This was in part to ensure that even if the visitor disagrees with the AI, the results still seem plausible and should therefore not diminish the experience.

In terms of usability and inclusion, we have added an alternative to face tracking in our application (for the visitors that may be uncomfortable with taking their picture or having trouble with expressing faces) that still allows for a fully interactive experience. The gallery consists of images of faces expressing different emotions which can be analyzed with the same method of emotion recognition.

Two alternative designs

To increase the chances of meeting the goals of the interactive station, we have developed two alternative prototype applications that differ in their interactive form and narrative. One is a simple guessing game where Epi and the visitor take turns at guessing each other's emotive expressions, and the other will put the visitor in the center of a simulated training program where the robot learns to improve its responses.

The reason we have developed two different applications is due to uncertainty about which form of interaction will better meet our goals and Vattenhallen's wishes, in terms of engaging the visitors and being fun and informative.

Only one application will be used in the exhibition, thus we have developed a procedure (appendix 4) for testing which one will better meet our demands. We have piloted the test on both applications with adult users, and hence obtained indications for weaknesses and strengths in both applications, but further user testing with representatives from the actual target group is needed before we can exclude one or the other, which of course is problematic due to the current constraints imposed by the pandemic.

Limitations

Most boundaries set by the team can be attributed to limited time and resources. The constraint of time has a significant impact on the project due to the reduced workforce. Furthermore, the project relies on third-party software, delayed communication with stakeholders, and lack of knowledge of technical conditions. Due to the lack of knowledge of the technical conditions, a platform-independent approach has necessarily been developed, which will be discussed later on. Additionally, the current state of the world, the pandemic, have caused delay and time-consuming efforts relative to an ideal situation.

2 Theoretical and scientific relevance

Emotion AI

Emotion AI is a relatively young subset of AI, introduced in the mid 90's by RW Picard (1995) in the developing area of *affective computing*, which concerns the study and implementation of affect in computer systems and intelligent agents. Emotion AI constitutes the capacity of artificial intelligent systems to recognize, measure, interpret, simulate and respond to human emotions (Tao and Tan, 2005).

The field of AI is beginning to catch up with the fact that human cognition is fundamentally intertwined with affect, i.e., emotions, moods and feelings, and therefore, emotion AI is increasingly becoming a core part of general AI, particularly in the branches of AI that attempt to recreate human intelligence. In the tech industry, emotion AI has many purposes that increasingly are becoming part of our lives. For example, it is being developed and implemented to: monitor and manipulates attention in educational software (McStay, 2020); adjust difficulty level in games and education technology (Harley, Lajoie, Frasson, Hall, 2017); enhance engagement and entertainment when interacting with intelligent agents (Brooks, Gray, and Hoffman, 2004); enhance interaction and immersion (presence) in virtual reality (Riva et.al., 2007); and generally to effectivise human-machine interaction and collaboration across situations (Tao and Tan, 2005).

A weakness with emotion AI is that many of these systems are based on Ekman's (1996) model of universal basic emotions, which constitutes joy, sadness, anger, disgust, fear, and surprise. It is easy to imagine the appeal of this idea in the AI industries - that some emotional expressions do not vary across cultures - because it seems to (falsely) promise generalizable and reliable predictions of emotions on large-scale populations. However, subsequent studies have shown that there is much more variability in how people express emotions than Ekman's research accounts for; it varies across contexts and situations (Barrett, Adolphs, Marsella, Martinez, and Pollak, 2019). Moreover, there is not much solid evidence for the consistency and reliability of the connection between facial expressions and affective states. Especially when it comes to the connection between facial expression and the *cognitive* implications of affective states, which, in extension, is often what emotion AI is used to make predictions about (Fernández-Dols, Sanchez, Carrera, and Ruiz-Belda, 1997; Barrett, et.al., 2019). In other words: emotions may entail certain mental states, but inferring mental states from emotional expressions alone is an unreliable practice, as there often is no more than weak (if any) congruence, and thus decisions (e.g., made by AI) that might affect the lives of people (e.g., the tech user) should not be based on such inferences (Barrett, et.al., 2019; Crawford, et.al., 2019). However, whether emotion AI can be used to accurately decode our inner states or not does not seem to stop it from becoming a regular part of our everyday interaction with computers and intelligent agents.

Our project is not about showing how accurate or reliable emotion AI can make predictions, but rather to let the visitor get familiarised with this kind of technology on a fundamental level, with both its potential and its shortcomings. This entails introducing technology that may read our facial expressions and attempts to classify our emotions to make deliberate alterations of their behavior.

Social and affective robotics

While emotion AI has its clear limitations in systems whose appropriate function is dependent on accurate emotion recognition, it has a more promising venue in fields that attempt to make computers into better collaborators and companions. This idea is of the essence in social human-robot interaction. Social robots are not designed to make life-affecting decisions based on interpretations of facial features, instead, emotion AI is used for facilitating natural, fun, effective, and intuitive interaction (Tao and Tan, 2005). For example, in its simplest form, emotion AI may be used by the robot to identify a smile, whereupon the robot might respond by mirroring the expression (simulating natural human behavior) or illicit another context-appropriate response.

Bartneck and Forlizzi define a social robot as: "*...an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact.*". Since social and behavioral norms frequently are communicated through non-verbal affective cues (Hareli, Moran-Amir, David, and Hess, 2013), one can argue that some form of "emotional intelligence" that can pick up on and act on affective signals is essential for a social robot adhering to this definition. This assumption is at the heart of the field of *affective robotics*.

Affective robotics studies how human-robot interaction can be made more natural, effective, and engaging by employing models of human affect and emotion AI, which entails recognition and simulation of human affect (Yonck, 2020). Research suggests that such affective behavior is beneficial even in general robotic assistance when the primary goal of the robot is to help humans solve problems, i.e., not only to be good social companions. For example, robot assistants that make use of affective behavior have a higher success rate in guiding human interlocutors in gameplay (Leite, Pereira, Martinho, and Paiva, 2008).

On the *affect simulation* account, studies have found that robotic emotive expressions can be used to facilitate natural and intuitive communication (Breazeal, 2001) and that humans (and the interaction) are generally sensitive to changes in a robot's emotive expression. For example, slight changes of expressions have been found to influence how people perceive and interact with robots; positive expressions increase the likelihood of common ground between robot and human and negative expressions decrease willingness to engage (Kirby, Forlizzi, and Simmons, 2010).

On the *affect recognition account*, studies suggest that a robot's ability to communicate recognition and understanding of the human interlocutor's affective state is more important for the interaction than its ability to simulate affective reactions or display emotive expressions (Leite, et.al., 2013). There are potentially many psychological explanations for this, perhaps because humans like to feel understood, or it could be because achieving authenticity of affect simulation with robotics is challenging and the result may even be rather unsettling (e.g., see Sophia the humanoid robot at hansonrobotics.com/sophia/, which looks more like an animate corpse than an emotional being).

Affect communication with appearance constrained robots

With the rising importance of affect in human-robot interactions, the challenge of how robots can communicate emotive expression effectively becomes more apparent. One approach is to construct robots with anthropomorphic human-like appearance, that can simulate emotive expressions with artificial facial contractions, however, there are good reasons why this might not be the best approach. A human-like appearance can lead human interlocutors to have unreasonably high expectations of the robot's cognitive abilities; when their expectations are not met, their engagement in the interaction deteriorates (Leite, et.al., 2013), and common ground is harder to establish (Kiesler, 2005) (in this regard we may be lucky to have only the Epi head). Reasonably, it would make sense to design the robot's emotive appearance in a way that reflects its capabilities and purpose.

Epi could be considered a somewhat appearance constrained robot in comparison to other humanoid robots (e.g., Sophia, at hansonrobotics.com/sophia/). For example, Epi has no eyebrows and has no moving facial features other than eye movements and pupillary dilation. Epi do, however, have color displays, which have been found to be an effective tool in communicating affect with robots (Song and Yamada, 2017; Löffler, Schmidt and Tscharn, 2018; Tärning, Tjøstheim, and Johansson, 2019). Research suggests that some colors are more strongly associated with specific emotions than others. For example, one study confirmed some conventional associations between colors and emotions illustrated in Plutchick's wheel of emotion: anger-red, anticipation-amber, joy-yellow, trust-green, fear-violet, surprise-red, sadness-blue, and disgust-purple (Terada, Yamauchi, and Ito, 2012). Another study found similar strong results for anger-red, sadness-blue and joy-yellow, but found that surprise was even more commonly associated with yellow than joy. They also found that dark green was associated with disgust and fear with black (Johnson, Cuijpers, and van der Pol, 2013). In accordance, a study at Lund University using Epi has found that red-colored eyes are strongly associated with anger and blue eyes with sadness. The same study found that turquoise green was moderately associated with joy (Tärning, Tjøstheim, and Johansson, 2019). Taken together, the results of these studies indicated that some colors (red and blue particularly) are more reliably associated with specific emotions than others. However, in the study with Epi, the associations could be modulated by color intensity, so that lower intensity red was associated with disgust, and higher intensity turquoise with surprise (Tärning, Tjøstheim, and Johansson, 2019).

Certain head movements have also been linked to specific affective states. For example, one study linked fast circular movements to joy; slow turning away movements with sadness; shaking movements with anger; and jumpy avoidant movements with fear (Löffler, Schmidt, and Tscharn, 2018). Another study with a robot lacking facial expressions found that both children and adults associated an upwards oriented head with positive emotions like pride, joy, and excitement; a downwards oriented head with negative emotions like anger and sadness; and a straight oriented head with mixed emotions like anger-excitement, and in some cases pride (Beck, Cañamero, Hiolle, Damiano, Cosi, Tesser, and Sommavilla, 2013).

Common for all the cited studies above is the notion that colors and movements are more or less prone to subjective interpretations and not distinctly associated with one emotion, particularly when they occur in isolation. Thus, combinations of different modalities may be more reliable when communicating affect. Accordingly, another study using Epi found that sadness could easily and distinctively be identified when blue eyes were displayed in combination with eyes pointing inwards (towards Epis nose, if she had one) and head moving downwards to the side, and that anger was easily and distinctly identified when eye color was red, the head made a shaking movement, and pupils were contracted. Happy was easily identified (not so distinct, however) by green eyes in combination with the head moving upwards and pupils dilated (Lindberg, Sandberg, Liljenberg, Eriksson, Johansson, and Balkenius, 2017). In accordance with studies using Epi, and other studies on robotic affective displays, the safest bet when designing emotions in appearance constrained robots is to use specific combinations of different modalities, like color and movement.

3 Conceptual design: Two prototypes

Due to the constraints imposed on our group, it was deemed necessary to begin implementing our concept parallel to designing it. We anticipated that in order to succeed with such an approach, we had to quickly come up with a simple *fundamental concept* that could be turned into a functioning product straight off, yet also have the potential to support various extensions and different narratives (different framings and storylines dictating the interaction) so as to be promoted into more advanced concepts that better meet our goals. When implemented, the product would also have to be dynamic enough to be adjusted and expanded in accordance with the evolving concept (within certain boundaries of course).

The fundamental concept was described in detail in the previous report, but since it then was under development, a few changes have been made to the implemented version. An important difference is that the fundamental concept now incorporates an emotion AI that detects emotional expressions from facial features. The face tracker has the capacity to detect joy, sadness, anger, fear, surprise, disgust and neutral. The face tracker was only a property of what we in our previous report called the *advanced concept,* but due to success in implementing it is now a fundamental feature in our application corresponding to our *new* fundamental concept. We will briefly give a functional description of what we may call *the fundamental application* in simplified steps:

1. The visitor shows an emotional facial expression to Epi, either by demonstration, i.e., by taking a picture through the cameras in Epis eyes or by choosing a photo of a from a gallery.
2. The expression is analyzed through an emotion recognition face tracker and results are given to the visitor framed as Epi's evaluation. The results of the tracker may be given like this: joy 60%, surprise 30%, neutral 10%.
3. The visitor is offered a choice to start over with a new emotional expression or to proceed.
4. Epi responds to the visitor's expression by demonstrating its own emotional expression.
5. The visitor evaluates Epis emotional expression by categorization.
6. A comparison between the visitors and Epis expression and evaluation is given as a result.

This fundamental application serves as a basis that can easily be converted to support more advanced interactions, simply by copying or removing code (with some additional new code of course) and adding or altering the content of the app. For example, steps (or states) may be repeated or rearranged and Epis responses may be altered in various ways simply by requesting an alternative response from a library of pre-recorded movements (which will be extended to include voice recordings). Instructions may be altered and added to the various states to frame the interaction as wished.

From this fundamental application, we have developed two apps that differ in their form and complexity of interaction (see flow charts in appendix 1). We call the simple one *the guessing game (see appendix 6)* and the more intricate one *the learning game (see appendix 2).*

4 Conceptual design: The guessing game

The interaction starts with the visitor being prompted to demonstrate an emotional expression so that Epi can try to guess it. They are also informed that Epi is going to show an emotive expression for the visitor to guess and that they will take turns in guessing each other's emotional expressions. The visitor is told that the point of the guessing game is for Epi to improve her emotional intelligence by playing.

Demonstrating emotive expressions

The visitor may show any expression to Epi, but the tracker will only evaluate it in terms of the basic six emotions and neutral, thus Epi will only give one of these as a guess, or combinations like 50% joy, 20% surprise, and 30% neutral.

Epi may demonstrate any emotive expression we wish to model, but for now, we have only modeled the 6 basic emotions, plus neutral, that correspond to the capacity of the face tracker.

Guessing

Epi gives her guess as a percentile based on the data from the face tracker. The visitor guesses by choosing from an array of words that describes an emotion (e.g., joy, nervous, grumpy). The visitor confirms or disconfirms Epi's guess by pressing a button on the screen (a more interactive alternative to be explored is if the visitor can give a sign like a thumbs up or a thumbs down that Epi can be made to read through the cameras). Epi confirms or disconfirms the visitors guess by a vocal reply, an appropriate movement, and a sign on the screen. Points are collected for every correct guess.

Rounds and results

When both the visitor and Epi have guessed once, the visitor is prompted to demonstrate a new emotion, and another round starts (see appendix 1 and 6). The rounds go on for a predefined number of times and ends with a scoreboard where points for correct guesses is displayed. In addition to the scoreboard, the different expressions and guesses throughout the game, from both the visitor and Epi, are displayed for comparison.

Scoring system

Because the goal of an affective social robot should be to clearly communicate specific emotions and to be understood, it does not make sense to give the visitor points when she successfully guesses Epi's emotional expression and give Epi points for successfully guessing the visitors expression and have them compete for the best guess like in a traditional guessing game. Because a game like this would not give any incentives for the robot (nor the visitor) to try to be understood. Then it would make more sense for the robot (with AI) to try to obscure its expression to make the visitor fail in their guess.

Instead, a correct guess should be attributed to the communicator's effort in expressing an emotion clearly, and not to the guesser's ability to guess. However, if the scoring system is as such, and the visitor and Epi are in competition, it would obviously make more sense for both parties to intentionally guess incorrectly to keep their competitor's score down. The most constructive way out of this clammer is to eliminate the competitive factor. After all, good communication is rewarded because of collaboration, not competition.

Thus, in the guessing game, Epi and the visitor have a joint goal to score as high as possible by adding their points together; their collective effort in communicating clear emotional expressions and their collective success in guessing each other's expressions is what determines the score. If Epi is able to guess the visitor's expression, the visitor gets points that contribute to the collective score. Thus, the visitor has an incentive to put on their best face, so to speak. Likewise, if the visitor is able to guess Epi's expression, Epi is given points that contribute to the collective score. Thus Epi has an incentive to "learn" to communicate emotions as clearly as possible to make it easy for the visitor to understand them. This is how the effort on Epis behalf will be framed in the text content of the game.

A part of the challenge in this interaction is to figure out the scoring system as they go along. That is, realizing that they do not compete against each other, and the collaboration and clear communication is the key to success. Of course, hints will be given, but a strength with "the guessing game" compared to the "learning procedure" is that it is not so sensitive to whether the visitors understand the point of the interaction at outset. It is also more robust against visitors who try to game the system.

5 Conceptual design: The learning game

The visitor is informed that Epi is an affective robot in training, and needs supervised learning in real interaction to become better at communicating with emotions. The visitor will get to take part in this training.

The interaction starts with the visitor showing Epi an emotional expression that Epi will try to recognize, respond to and learn to simulate with the help of feedback, or evaluations, from the visitor

Epis learning progress

In the learning game, Epi has a goal: to learn to simulate the expression that the user has demonstrated. Epi starts off towards that goal by generating an emotive expression limited to one module, for example, pupil size or head orientation. The limited expression is the first part of a predefined emotive expression (e.g., joy) split into several accumulating parts that mounts to a complete emotive expression over the course of the interaction (see appendix 1 and 2 for illustration). Let's say that the visitor have demonstrated "sadness", then Epis first expression may be dilated pupils + eyes turning; the second expression may be dilated pupils + eyes turning inwards + head orienting downwards; the third (and final) expression may be dilated pupils + eyes turning inwards + head orienting downwards + eye color changing to blue.

The visitors role

Every time Epi has demonstrated an extension of the expression, the visitor gets to respond (give feedback) by evaluating the expression in terms of its affective properties (see appendix 1 and 2). The evaluation will be done on the screen by selecting one or more buttons displayed in an array. The buttons (framed as nodes in the narrative) are marked with affect-laden words that the visitor may associate with the emotional expression (e.g., grumpy or playful). Each word belongs to a superordinate category of 6 basic emotions (e.g., grumpy and annoyed both belong to anger): joy, sadness, anger, fear, surprise, and disgust. *We have chosen these emotions despite the flaws they represent in emotion AI because our emotion recognition software is tuned to these, for better or worse.*

The point of this step-by-step accumulation and evaluation of the emotive expression in terms of nuances (subcategories) is to invite the visitor to think about how these different expressive factors contribute to the communication of emotions with an affective robot.

Results of the learning session

After the simulation is completed, a summary of the evaluation from all steps are displayed so that the visitor can see how it changed as Epi's emotive expression changed. The results are given as percentiles (see appendix 2) of the 6 basic emotions, corresponding to how many subcategories (buttons) the visitor has selected from each category. For example, in the first evaluation (step one) the visitor may have been selecting 4 buttons in the category joy, 2 in the category of sadness, and 2 in surprise. Thus the results show that the visitor evaluated joy at 50%, sadness at 25%, and surprise at 25%. In the next evaluation (step 2) of Epi's has altered response, the visitor may have changed the selection of buttons (i.e., the evaluation) to 4 in the sadness category, 1 in anger, and 2 in surprise, so the results show that sadness is evaluated to be at 57%, anger at 14% and surprise at 29 %.

In addition, the results of Epis' evaluation (data from the face tracker) will be given as a comparison and illustration of how Epi has categorized the visitors expression. This way the visitor will be able to see how their categorization of Epi's expressions (judging how far Epi is from the goal) differs from Epi's evaluation (judging what the goal is by categorizing the visitors expression). Epi's goal is to simulate the expression that the visitor has demonstrated to her, so if the visitors evaluations are close to Epi's evaluations, that indicates success. Furthermore, the results illustrate how AI makes decisions; by calculating the likelihood of all the alternatives and acting on the most weighted alternative.

After the results have been given, the visitor is invited to try again with a new expression (i.e., take a new picture). Now that Epi has "learned" to simulate an emotive expression, this can be stored as the default response to whatever emotional expression was the goal (the visitors demonstration), so that if the visitor makes the same expression again as in a previously completed trial, Epi responds directly with the complete emotive expression; if the expression is novel, a new trial starts.

The visitors (actual) role in Epis learning progress

The evaluation (selecting buttons) is framed (in the narrative) as feedback-input into Epi's learning mechanism, so when the evaluation is completed, Epi in turn responds with a new, altered, expression (demonstrates the newly learned skill) which is an extension of the previous emotive expression (e.g., adding a new part to the accumulation of sadness). The user is made to believe that her evaluation causes Epi to "modify" her expression (remember that every modification is actually scripted). For example, if the visitor demonstrates the expression of joy to Epi, and the first response Epi gives is dilated pupils and a raised head, the visitor may, in her evaluation, think that Epi's expression looks both a bit joyful but a bit more surprised. Epi responds by adding another element to the expression, perhaps green or yellow eye color, and the visitor is made to believe that her feedback made Epi adjust her expression.

We anticipate that the visitor will believe that her evaluation will have a stronger causal relation to Epi's learned responses than it actually has. We believe that with proper framing, we can take advantage of people's tendencies to believe that there are causal relationships between their actions and timely events, as long as these events are perceived as contingent on their actions and plausible consequences of their actions (Jenkins and Ward, 1965; Ono, 1987; Rudski; 2000. Hence, we hope to induce an enhanced (false) sense of agency perceived in oneself and the belief that there is some actual machine learning going on in Epi.

5 Development and design technicalities

Thinking through prototyping

Prototyping is known to be an essential medium for communicating and developing ideas in interactive collaborative teams, moreover, it facilitates concretization of ideas and leads to unexpected realizations (Schrage, 1996). Through this reflective practice, we have tested our ideas, shaped the conceptual design, identified problems and solutions, and generated new ideas.

Low-fidelity wireframes

We started the design process by creating low-fidelity wireframes. The purpose of these wireframes was to clarify the basic structure of the conceptual model as well as the user interface and to identify basic elements and design principles that were necessary for the user to be able to navigate through and use the product. As is usually the case with lo-fi wireframes, all design elements that could potentially be changed or removed were consciously left out of the design. The lo-fi design was then used more as a tool to explore different scenarios, for brainstorming purposes, and to better get an understanding of the user flow and interaction with the system.

Mid-fidelity wireframes

Whenever our concept got an update, we prototyped it using Adobe XD, which is a design tool that enabled us to quickly create realistic and interactive prototypes. Hence, we could move through the steps of the interaction to get a feel for how our solutions would hold up in the real interaction. This helped us identify aspects of the interaction that made little sense or were illogical and to make corrections in line with conventional design principles.

In the XD prototypes, we modeled the interaction with both the screen-based interface and aspects of Epis behavior. However, note that the information, symbols, and labels seen on the elements in the mid-fi prototype in appendix 2 are just examples (i.e., “lorem ipsum” stuff), as is the amount of buttons and elements (e.g,. at the evaluation/feedback screen) and their layout. The purpose of these prototypes was to test the flow and functions of the interaction and to identify what elements to include and how they should relate to one another. Thus, layout, shapes, sizes, labels, etc. are different in the hi-fi prototype applications. Similarly, the purpose of modeling Epi’s responses in these prototypes were mainly to get a feel of how they would function in relation to actions made on the screen-based interface and not as templates for designing the real responses. Thus, the responses seen in appendix 2 are merely hastily made examples, more for the purpose of testing how a response fits in at a certain point of the interaction, and less about testing exactly how that response should unfold (for example if pupils should dilate or contract in the sadness simulation, or whether the movement should precede eye color or visa versa). Nevertheless, note that the prototypes *have* given us important insight and ideas about layout, labels, graphics, Epi’s responses, and more, even though this was not the main purpose or focus.

Web-UI

Due to limitations with the existing web-UI implemented in the platform for controlling Epi, IKAROS, we decided to create our own web application using the javascript framework React, with the goal of having it communicate with IKAROS. This choice was made primarily to meet the demands for interactivity. Creating a highly interactive interface in compliance with our engagement goals is simply not possible in IKAROS alone.

In essence, what we need is to establish a connection between our application and IKAROS, both to activate Epi’s pre-recorded responses and to use the cameras in Epi’s eyes. IKAROS will also be used to record the expressions which Epi will display.

Because we decided early on to develop our own application, we did not have to plan our work based on the constraints of an already implemented system, and therefore we could focus more on the possibilities of javascript as a tool for making the WebUI simple and interactive.

The layout, feedback models, and navigation of the Web-UI have been influenced by the seven fundamental principles of design described by Norman (2013). The mapping is based on western conventions, for instance, forward navigation through the application is made by selecting buttons whose placement is consistently in the bottom right corner; backwards navigation is consistently in the bottom left corner; and cancel is in the top right corner. Hence the web-UI developed to ensure a clear conceptual model of usage that is familiar and easily learned. Furthermore, feedback of actions follows the immediacy and visibility principles and the functions and flow of the web-UI is constrained so that no action can lead to the user getting lost or setting the system in a complex state. There are minimal options available to the user that could interfere with the flow in a way that might cause unintended interruption.

Code

The code is based on object-oriented programming to a certain limit which JavaScript and the React framework allows. An overview of the technical implementation and how it works is broadly considered to be relevant to the report (see appendix 3), further details are not considered essential. React is a state-based framework, where a parent component controls child components. Methods do exist in several child components but the main methods updating the general state are implemented in the parent component. Methods and parameters are passed from the parent to the child. The child component can use the methods and parameters received from the parent, hence reduce redundancy and simplify the understanding of the code. Furthermore, the children components can have their own methods used only within the class, thus isolating the responsibility of every child component.

Designing emotive expressions

As mentioned before the features which are available for modeling emotive expressions with Epi are eye color, color intensity, pupil dilation, eye movement, head movement, mouth color, and sound. Since the station will be located in a rather noisy environment, the expressions are not dependent on sound (nor is the interaction), but sound will be used to meet our engagement and entertainment goals.

Based on the theoretical overview introduced earlier in this paper, specifications have been set for the design of certain emotive expressions, while further investigation (user studies and further literature reviews) has been conducted for defining other expressions. The expressions that has been modeled this far (presenting here only the full expressions, i.e., excluding the incomplete expressions from this list):

Anger - High intensity red eyes, small pupils, facing visitors, shaking head tilted slightly down (anger is metaphorically conceptualised as “heat”).
Sadness - Blue eyes, large pupils, eyes turned inwards, head tilted downwards and facing slightly away (sadness may be metaphorically conceptualized as “down” and “turning inwards”).
Joy - Turquoise eyes, large pupils, facing visitors, head turned upwards, circular head movements, mouth blinking slightly (joy may be metaphorically conceptualized as “up” and “active”).
Surprise - Low intensity red eyes, dilating pupils, head moving up and back.
Disgust - Low intensity green eyes, contracting pupils, head facing away while eyes look straight.
Fear - high intensity white eyes, rapidly dilating pupils, head moving rapidly up, back and facing away, with jumpy avoidant movements, and eyes facing straight.

An additional 12 reduced, or incomplete, expressions have been designed using only a few and a minimum of Epi’s features.

Previous work on visual displays using Epi and the Ikaros system includes Sandberg et. al., 2017, who designed the following mental states in their experiment: thinking, angry, confused, happy and sad. Sandberg et. al., shows that the mental states angry and sad were by far the easiest to detect.

Designing the evaluation items

The items (affective adjectives) the visitors will be able to choose from when evaluating the emotive expressions will be encompassing many words for emotions and feelings more or less related to a superordinate category of a basic emotion (e.g.anger is the category, but the evolution may include sub-items like grumpy, annoyed, furious, etc.). This is partly because of the subjectivity and variability of interpretations, and as Sundberg et. al., suggests, to not have strictly predefined items since that could create a bias. Furthermore only being able to choose from a few items could both make the interaction less fun and make the user aware of the options beforehand. One way to solve this problem would be to implement an input function, where the user could write the emotion she thinks Epi displays, however, we will not do this for the sake of the robustness of the interaction and problems it would entail for the categorization of words in the application; misspellings and unexpected word choices would be difficult to account for.

6 Testing Epi's emotive expressions

Procedure and materials

To test whether the modeled expressions are perceived as intended/hypothesized, we construed an online survey where 11 participants of mixed ages and gender (no personal information was obtained) were asked to categorize Epi's various expressions.

The survey consisted of 21 video recordings of Epi's expressions that each could be categorized by selecting one or more descriptive affective words from a list of 20 items. All the recordings can be viewed here: https://www.youtube.com/playlist?list=PLbh8VoxmYtTyYwOI9ngPBDNhb1lAR6TQ7. The clips are coded with numbers but a code sheet with names of the emotions corresponding to the numbers can be found in appendix 5.

Of the 21 video clips, 7 recordings depicted complete expressions, that is, fully composed of all modularities (head movements, eye color, eye movement, pupil dilation). Of these was one recording every complete basic emotion and two of joy. The reason there were two versions of joy is that this expression was judged beforehand to be particularly ambiguous, thus we wanted to see if one of the alternatives were stronger than the other. The remaining 14 recordings were incomplete expressions (for the learning procedure app), varying in their degree of completeness. 7 of these recordings (two of joy) depicted expressions consisting of only two of Epi's communicative features, for example only eye color and head movement. The remaining 7 recordings (two of joy) depicted expressions consisting of only one feature, e.g., only eye movements.

The list of items to be selected from when categorizing the expressions consisted of the following words:

Busig	Arg	Ängslig
Hoppfull	Förvirrad	Äcklad
Glad	Rädd	Överraskad
Butter	Nyfiken	Skyldig
Skamsen	Irriterad	Ledsen
Chockerad	Upprörd	Uppspelt
Stolt	Neutral	Annat

The participants were told that they could select more than one item if they had difficulties deciding between the items. They could also add their own words if they didn't find a descriptive word in the list. The words added were:

Empatisk	Avvisande	Oangagerad
Tänkande	Ångest	Skeptisk
Frågande	Generad	Fokuserad
Osäker	Uttråkad	Trött
Motvillig	Besviken	

Results

Anger:
The complete expression of anger (red eyes, small pupils, shaking head) was associated with "Arg" 82% of the time, with "Upprörd" 36% of the time and 18% of the time with "Irriterad".

Anger part 2 (only shaking head and contracted pupils; without red-eye color) was associated with "Ledsen" 46% of the time, with "Chockerad" 27% of the time, with "Skamsen" 27 % of the time, and with "arg" only 10 % of the time.

Anger part 1 (only shaking head movements) was also strongest associated with "Ledsen" 36% of the time and with "Arg" 18% of the time.

Sadness:
The complete expression of sadness (blue eyes, large pupils, head moving down, eyes moving inward) was associated with " Ledsen" 55% of the time, with "Skamsen" 46% of the time and with "Skyldig" 36% of the time.

Sadness part 2 (eye color, eye movements and pupil dilation only; without head movements) was strongest associated with "Ängslig" 36% of the time and with "Ledsen" 27% of the time. The expression was not associated strongly with any other words (below 18%) and the ratings were spread out.

Sadness part 2 (only eye color and pupil dilation) was strongest associated with "Ängslig" at 27% and only 9% with "Ledsen". The rest of the ratings were distributed among the items and none above 18%.

Disgusted:
The complete expression of disgust (green eyes, head-turning away, pupils contracting, eyes facing straight) was strongest associated with "Irriterad" at 27%. It was only associated with "Ecklad" 18% of the time, and scored equal to "Redd" and "Upprörd".

Disgusted part 2 (head movements, eye movements, and pupils only; no eye color) was strongest associated with "Rädd" at 27 % and not at all associated with disgusted (0%). The rest of the ratings were distributed and not associated with any item above 10%

Disgusted part 1 (head and eye movements only) was strongest associated with "Skamsen" and "Skyldig" at an equal 45% on both items. It was associated with "Ledsen" at 27%. It was associated with "Ecklad" at only 9%.

Surprise:
The complete expression of surprise (head moving upwards, large pupils, low intensity red eyes) was strongest associated with "Chockerad" at 27%, and not associated at all (0%) with

"Överraskad". The rest of the ratings were distributed and below 18% on any item.

Surprised part 2 (head movement and pupil dilation only; no eye color) were associated with "Överraskad" at 27%, but equally associated with "Chockerad" and "Förvirrad" (also 27%). The rest of the ratings were distributed below 18%.

Surprise part 1 (head moving upwards only) was strongest associated with "Förvirrad" at 27%, and with "Överraskad" at 18%, equal to happy also at 18%. The rest of the ratings were distributed at below 10%.

Fear:
The complete expression of fear (rapid avoidant head movement, dilating pupils, eye movement, white eye color) was most strongly associated with "Chockerad" with 36%, and secondly with "Redd" at 27 %. The rest of the ratings were distributed and below 18% on any item.

Fear part 2 (dilating pupils and head moving back) was equally associated with "Chockerad" and "Överaskad" at 27%, and only 10% associated with "Redd". No other item was above 10%.

Fear part 1 (dilating pupils and white eye color) was strongest associated with "Hoppfull" at 36%, and 27% associated with both "Nyfiken" and "Neutral", and 18% with "glad". The rest of the items were under 10%.

Joy:
(The expressions of joy were modeled in two versions. Here we only report the versions/recordings of joy strongest associated with "Glad")
The complete expression of joy (turquoise eye color, upwards rotating head movement, dilated pupils) was strongest associated with "Upspelt" at 46%, and secondary 27% associated with "Glad" and equally 27% associated with "Upprörd".

Joy part 2 (yellow eye color and dilated pupils only) was 36% associated with "Glad" and equally 36% associated with "Upspelt". It was 27% associated with "Hoppfull", and the rest of the ratings were under 10% on any item.

Joy part 1 (yellow eye color only) was strongest associated with "Nyfiken" at 27%. It was 18% associated with "Glad" but equally 18% associated with "Förvirrad" and "hoppfull". The rest of the ratings were distributed and under 10% for any item.

Discussion

Many new words (items) were added to describe the expressions but none were rated higher than 10%, which we have set as the threshold for not using these items in the app as evaluation words (button labels). 18% was set as a threshold to filter out the weakest associations.

As predicted by the literature, Anger (complete) was the strongest expression in terms of recognisability, followed by Sadness (complete). These two were the only expressions recognized above chance level. Red eye color seems to be the most significant factor for recognizing the expression as anger, and shaking head movements was a stronger factor than contracting pupilles. For sadness, downward head movements seem to be a strong factor.

The weakest expression was disgust, which was only recognized as such by two persons in its full version, and by one person in its least complete version. None of the two more complete disgust recordings were strongly recognized as any particular emotion, but the least complete one was recognized as guilt and shame. All associations were at least in the negative dimension.

The expression of surprise (all) was mostly recognized as shocked and confused, but these still may be found on a spectrum of surprise. The low intensity red eye color (as modeled after suggestive findings from Tärning et.al, 2019) seemed not to add anything to the recognisability of this expression as surprise.

Fear (complete and second part) was also most strongly recognized as being shocked (and confused), and only weakly as being scared, but chock is at least as related to fear as a surprise. The head movements seem to be important for fear modeling because the recording without these was associated with hopefulness.

The modeled expression of joy seemed to be mostly recognized as excitement, but also equally but moderately associated with joy when head movements were removed. All recordings were recognized in the positive dimension, and co-occurrence of joy, hopefulness, and curiosity with excitement (as the strongest association) suggests that the perceived excitement was positive rather than negative.

There is no clear tendency for recognisability of a given emotion (except for sadness) to linearly grow stronger the more of Epi's features are added to the expressions. This may be a problem for the "learning procedure" app because every time a feature is added Epi should be perceived by the visitors as becoming better at simulating the emotion.

7 Testing the face tracker

Reliability tests of the face tracker were conducted in part by analysing random stock photos found online of various people demonstrating predefined emotional expressions, and in part by analysing photos of ourselves, attempting to show specific emotions, taken through the cameras in Epi's eyes. All 6 basic emotional expressions plus neutral were represented at least 5 times each in both the photo test (with random people) and the camera test (with our own faces), with various and wide variations in facial expression, angle, size, luminance, clarity, color, and quality. The faces in the random photos found online also varied in skin tone (from white to dark), age (from ca 5 to 60+), ethnicity and gender.

The face tracker predicted the expressions of the random photos with an estimated 70-80% accuracy on average of adult faces, and with 90% accuracy on the majority of adults, despite variabilities. The face tracker also predicted the expressions made by ourselves through Epi's camera with an estimated 90% accuracy on average and 100% accuracy for the majority of the expressions. For young children (ca 5-12), the tracker is much less reliable with an estimated accuracy of 50% on average, which is a common problem with emotion AI (McStay, 2020).

8 Piloting the applications

The general aims of the user tests are:

1. To test whether one or the other of our applications better meet our goals in terms of being fun, interesting and informative.
2. To investigate if the "learning procedure" app is too intricate and if the "guessing game" is too simple.

3. To assess to what degree test persons think they are interacting with AI.
4. To identify any usability issues.

Materials and procedure

The test consists of two schemas for guided observations. Thus two observers focus on different things while the test subjects go through both applications (half of the test persons start with the “guessing game” and the other half with the “learning procedure”): one of the observers focus on crossing off predefined questions while the other observer freely records the test person's actions and thoughts by taking notes. The intention behind this strategy is to maximize the information we can get from the test persons in the least possible amount of time. See appendix 4 for further details. The documents marked V1 (appendix 4.2) and V2 (appendix 4.3) contain the predefined questions that we want to keep track of during the observation. The info document is common to both. The document called "After the test" (appendix 4.4) is a post-test that contains questions that we want answers to after completing the interaction, for example "how fun did you think the interaction was on a scale from 1 - 10?". Furthermore, The test is designed so that it places more demands on the observers than on the test person.

Pilot results

Due to the constraints imposed by the pandemic, we have so far only had the possibility to pilot the test with a few adult users. Thus we have only obtained mere indications about strengths and weaknesses of our apps, as well as indications on how well our test questions and procedure is suited for obtaining the measures/observations needed to identify said strengths and weaknesses.

The pilot indicates that the test procedure works as intended, and we did not discover any problems that would give reason to modify the test.

Furthermore, the pilot indicates that the “learning procedure” is so intricate that it becomes confusing, as all test persons understood all steps of the interaction as well as the Epi’s behavior. However, the test persons responses and open comments indicate that the “learning procedure” might be too complicated for its own good in terms of being fun and engaging. The pilot test persons rated the “learning procedure” lower than the “guessing games” on being fun and engaging, and free observations of the users' reactions, as well as the user's explicitly stated preference for the “guessing game”, supports this. Conversely, the test persons rated the “learning procedure” as more interesting than the “guessing game”. The level of complexity of the learning procedure seems to outweigh how fun one (adults at least) can possibly have with such a simple robot.

Regardless, these observations are mere indications and further testing with the actual target group (in collaboration with Vattenhallen) will be conducted before we decide which of the applications will be presented at the exhibition.

9 Further development

Improving the emotion recognition

The visitor may show any expression to Epi, but the tracker will only evaluate it in terms of the basic six emotions plus neutral, thus Epi will only give one of these as a guess, or combinations like 50% joy, 20% surprise, and 30% neutral. As an alternative to overcome this limitation, we are considering the possibility of combining data from the face tracker to be able to identify more nuanced emotions, for example could 50% sad and 50% scared perhaps be combined to indicate anxiousness? However, this demands much testing and is beyond our current scope.

Scoring system in the guessing game

For added gamification, the one (the visitor or Epi) who has the highest contribution in the collective score, i.e., who has made the most clear emotional expression, could be given extra credit in the end when results are displayed.

Illustrating “learning” in the learning procedure

When the selection of buttons, i.e., evaluation feedback, has been made, the screen may display the buttons as nodes in an animated neural network to illustrate that a learning process involving the selected words is taking place (or some better medium of illustration yet to be thought of). The result of the evaluation and subsequent “learning” is that Epi alters her next response accordingly, that is, Epi progresses towards the goal (i.e., simulating the visitors demonstration).

Illusions of agency in the learning procedure

What might strengthen visitors' illusions of agency (and belief that there is actual machine learning) could be to have two different outcomes of the button presses *depending on patterns of buttons* that are pressed. Either Epi proceeds with a response that is an extension of the previous expression (shows improvement towards the goal of simulating the emotion defined by the face tracker, e.g., joy), or Epi responds with an alternative expression that is *not* a part of the emotion that is being simulated (obstructing the goal). This alternative expression can have different forms, but the point being that it should demonstrate that Epi has not learned or is unwilling to demonstrate the learned skills. In the case of the former, a robot that shows that it understands but chooses to act defiantly can be a means of promoting entertainment in the interaction and thereby engagement (Brooks et.al, 2004). Exactly what combination of buttons will trigger an alternative response is yet to be defined, but the idea is to identify patterns of responses (specific combinations of buttons to specific emotive expression) that unreasonably would lead to Epi learning to improve the expression.

In the case of visitors deliberately pressing whichever buttons (trolling), an alternative response could also be what is needed to keep these visitors from seeing through the script, i.e., discovering that Epi does not not actually learn anything through the evaluations. However, a consequence of certain button combinations leading to interruptions in the script is that some visitors may actually believe that the combination is a reasonable evaluation (Our user test indicates that there are wide individual differences in how people experience Epi’s expressions). This poses a challenge that will be explored in further user tests with the target group, where the aim is to be able to anticipate such evaluations. However, we also believe in the possibility to frame the interaction so that evaluations that unexpectedly lead to an alternative expression (throws the interaction of the script) will illustrate to the visitor that the results of training

AI may not always be as expected, and then encourage the visitor to try again. Unexpected results are, after all, often the reality when dealing with AI.

10 References

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, *20*(1), 1-68.

Bartneck, C., & Forlizzi, J. (2004, September). A design-centred framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759)* (pp. 591-594). IEEE.

Beck, A., Cañamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., & Sommavilla, G. (2013). Interpretation of emotional body language displayed by a humanoid robot: A case study with children. *International Journal of Social Robotics*, *5*(3), 325-334.

Breazeal, C. (2001). Affective interaction between humans and robots. In *European Conference on Artificial Life* (pp. 582-591). Springer, Berlin, Heidelberg.

Brooks, A. G., Gray, J., & Hoffman, G. (2004, September). Robot's play: interactive games with sociable machines. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology* (pp. 74-83).

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., ... & Raji, D. (2019). AI now 2019 report. *New York, NY: AI Now Institute*.

Fernández-Dols, J. M., Sanchez, F., Carrera, P., & Ruiz-Belda, M. A. (1997). Are spontaneous expressions and emotions linked? An experimental test of coherence. *Journal of Nonverbal Behavior*, *21*(3), 163-177.

Hareli, S., Moran-Amir, O., David, S., & Hess, U. (2013). Emotions as signals of normative conduct. *Cognition & emotion*, *27*(8), 1395-1404.

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education*, *27*(2), 268-297.

Norman, D. (2013). *The Design of Everyday Things revised & expanded edition*. New York: Basic Books

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological monographs: General and applied*, *79*(1), 1.

Johansson, B., Tjøstheim, T. A., & Balkenius, C. (2020). Epi: An open humanoid platform for developmental robotics. *International Journal of Advanced Robotic Systems*, *17*(2), 1729881420911498.

Johnson, D. O., Cuijpers, R. H., & van der Pol, D. (2013). Imitating human emotions with artificial facial expressions. *International Journal of Social Robotics*, *5*(4), 503-513.

Kiesler, S. (2005). Fostering common ground in human-robot interaction. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* (pp. 729-734). IEEE.

Kirby, R., Forlizzi, J., & Simmons, R. (2010). Affective social robots. *Robotics and Autonomous Systems*, *58*(3), 322-332.

Leite, I., Pereira, A., Martinho, C., & Paiva, A. (2008). Are emotional robots more fun to play with?. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 77-82). IEEE.

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, *5*(2), 291-308.

Lindberg, M., Sandberg, H., Liljenberg, M., Eriksson, M., Johansson, B., & Balkenius, C. (2017). The expression of mental states in a humanoid robot. In *International Conference on Intelligent Virtual Agents* (pp. 247-250). Springer, Cham.

McStay, A. (2020). Emotional AI and EdTech: serving the public good?. *Learning, Media and Technology*, *45*(3), 270-283.

Löffler, D., Schmidt, N., & Tscharn, R. (2018). Multimodal expression of artificial emotion in social robots using color, motion and sound. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (pp. 334-343).

Ono, K. (1987). Superstitious behavior in humans. *Journal of the experimental analysis of behavior*, *47*(3), 261-271.

Picard, R. W.: 1995, Affective Computing. MIT Technical Report #321.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., ... & Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, *10*(1), 45-56.

Rudski, J. M. (2000). Effect of delay of reinforcement on superstitious inferences. *Perceptual and Motor Skills*, *90*(3), 1047-1058.

Sandberg, H. Liljenberg, M, Eriksson, M. Lindberg, M. (2017). *Designing mental states in a humanoid robot. Intelligent, socially oriented technology III*: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Science.

Song, S., & Yamada, S. (2017). Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI* (pp. 2-11). IEEE.

Tao, J., & Tan, T. (2005). Affective computing: A review. In International Conference on Affective computing and intelligent interaction (pp. 981-995). Springer, Berlin, Heidelberg.

Terada, K., Yamauchi, A., & Ito, A. (2012). Artificial emotion expression for a robot by dynamic color change. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication* (pp. 314-321). IEEE.

Tärning, B., Tjøstheim, T. A., & Johansson, B. (2019). Communicating Emotional State and Personality with Eye-color and Light Intensity. In *Proceedings of the 7th International Conference on Human-Agent Interaction* (pp. 214-216).

Schrage, M. (1996). Cultures of prototyping. *Bringing design to software*, *4*(1), 1-11.

https://www.hansonrobotics.com/sophia/ Retrieved 2020-10-30

Green Island:
An Asymmetric Collaborative Learning Game

Astrid Jansson, Jacob Kristensson, Olof Larsson,
Jens Jakob Tonn-Petersen & Victoria Wandsleb

The report describes our research and conclusion in the process of designing a concept of a digital learning game about the environment using a user-centered design approach. A user study is performed along with a literature study of different learning approaches. From this, the Green Island game concept is formalized. Our prototype representing these results conceptualize this as an asymmetrical collaborative learning game framework. The framework structure with minigames and modules allows for further creation of games with additional content and topics other than the environment and the process used in this study can be reused for this purpose. The framework further allows for additional exploration and testing of the practical approach of the asymmetrical collaborative game mechanics.

1 Introduction

Designing a learning game is a challenge as there is a need to find a compromise between the interests of the students and the actual learning goals. Learning about the environment and environmental issues is something that is a part of the Swedish curriculum but the subject might appear boring or forgettable if not presented in a way that caters to the students' interests and reality.

Presenting this subject through a game might increase the engagement of the students. However, the games that do exist in the area aren't very thorough and in-depth. With how the world looks today we feel it's more important than ever to delve deeper into this subject.

With this in mind, we aim to design a digital learning game that can engage a target group of seventh to ninth-grade students in the Swedish school system. Here, we want them to get a deeper understanding of the complexity of sustainability-related issues, while still maintaining a solution-oriented attitude. Therefore, we approach sustainability through a positive narrative and teach the students insights about how things are interconnected. This is done by designing a game setting that reflects societal processes that can provide a transformation towards a realistic understanding of sustainability on a global scale as well as in everyday life. To ensure that the knowledge gained from the game will be memorable we also want students to engage in active reflections when playing the game. With the vision of an interactive game in mind, we conduct a user study as part of our human-centred design process as well as carry out literature research in relation to digital learning through games.

1.1 Human-Centered Design

About our design principles. When we refer to design we use the definition of design as a process where an artefact's properties are determined and relate to the needs of the end-product users. A human-centred design process is used in the development of this game. Using a human-centred approach is beneficial for the user by making the product easier to use, improving user experience, and in making it more accessible for the user. In using this approach one reduces the risk that the product is rejected by its user.

The general principles of human-centred design include having a design based understanding of users and their needs, making sure that users are involved in the process, that the process is iterative and is driven by user-centred evaluation, etc. (ISO, 2010). There are certain steps in the user-centred design that are followed in this project. The target group is already defined, therefore we move straight into the empirical research on our target group. As the requirements and the context of the game are not set for this project we have a need for investigating the context and pedagogical methods before continuing to follow proper design steps in creating user scenarios, user cases, and finally prototyping.

1.2 Structure of the paper

The structure of the paper follows the development process. Part 2 of the paper presents empirical research which includes how we design our study and results with an analysis. Part 3 of the paper lays out the theoretical research and background. Part 4 presents the theoretical game concept of "The Green Island". In part 5 we describe how we formalised this concept into a software requirements specification. Part 6 describes the prototype design process and part 7 describes the prototype itself. In part 8 we outline how the prototype testing could be designed and performed, both relating to the usability and relating to the learning theories. In Part 9 we present a summary and discuss process, results, asymmetric collaborative learning and future work. Finally, in part 10 we present our conclusion.

2 Empirical Research: User Study

As a human-centred design approach was used in the study, there was a need to understand the user and the user's need in the context of creating an environment learning tool. The target group was already known therefore we moved into the design step of understanding the user as well as clarifying the context of the game. As this was known from the start, the context and the content needed to be explored to provide the setting before creating a conceptual design. In understanding the context and the user, a basis for game design decisions is created. The planning of the user study was prepared in *2.1 User Group And Context workshop.* The motivation for why a context study wasn't performed can be found there as well. From the results of the workshop, a form was created and tested in *2.2 Form creation*. The results of the form are presented in *2.3 Data Presentation and Result Analysis* and allow us to get acquainted with the user.

2.1 User Group and Context Workshop

The human-centred design process should revolve around the user needs and be driven by user-centred evaluation (ISO, 2010). In order to design from a user perspective, there was an initial idea of contacting students in the target group grade seven to nine. Furthermore, the idea of contacting teachers to provide context and setting of the learning was considered. Before starting the empirical research we wanted to make sure whether or not the answers we sought could be provided from the resources considered: students and teachers. Therefore we had to answer two initial questions:

1. What do we want to accomplish by contacting teachers and what do we want to know about the learning context?
2. What do we want to accomplish by contacting young students and what do we want to know about the users?

The initial questions were answered using individual brainstorming, presentation, and group discussion in a facilitated and workshop focused meeting.

The result from brainstorming initial question 1 can be summarized in the following areas:

- There might be limitations in the classroom (time, equipment, etc.)
- We need to find learning goals for the students
- We need to know the students' current knowledge base
- We need to have an understanding of students' individual needs

The result from brainstorming initial question 2 can be summarized in the following areas:

- We design for a user, specifically a student. This is reason enough to get to know the target group
- Target groups' interest in games, game type, game stories
- What are the students' preferred gaming platform
- Target groups' views and preferences on learning and lectures

Using the described brainstorming and discussion method on both these questions, we came to different conclusions which will be presented below.

The reason for contacting teachers revolved around curriculum, classroom setting, and equipment, student learning behaviours in a classroom, and their knowledge base. Limitations in the classroom and the knowledge base of the student are not reasonable to consider and analyze due to great differences between classrooms and schools. We chose not to research this ourselves due to the limited time-frame. We assumed that most schools have access to computers or computer rooms and that most students have smartphones. Figuring out the specific medium will be left to the early stages of creating the conceptual design. Furthermore, instead of making our own study of learning behaviour and pedagogical approaches, we will do a literature study on the topic related. The information about student goals can be found in the curriculum. Finally, we found there to be no reason to contact teachers directly, for the information we strive to use for the context and setting of the game. We can find the necessary information from related research and the Swedish school curriculum.

In brainstorming and discussing question 2 we found that the areas revolve around the user interest, what platform they prefer, what makes them enjoy learning, and that the reason for getting data from users is important for the human-centred design approach we are using. From this, we considered whether collecting focused data through interviews would be preferable to reaching as many students as possible. We found it preferable to have both quantitative and qualitative data to back up our user study. Using a digital form provided us with the means to use open questions, rating scale, and multiple-choice questions. The questions of the form are designed to reflect the result of the initial question 2 regarding what we want to know from the user (i.e the students).

Decisions made regarding the brainstorms:

- No collection of data from teachers. We find what we need from the curriculum and literature.
- Final decision to make a form and collect quantitative and qualitative data from the target group.
- Discussion points that reflect what we want to know about the target group will lay the basis for creating form questions.

2.2 Form Creation

In order to create a form for the students, we performed a workshop on the topic. The aim of this workshop was to create as many questions as possible that reflect the result of the discussion points in 2.2.1 and to get varied questions that will produce the information that we need from the target group to start a user-centred design process. As the game is centred around the environmental concepts we also added this as a possible area to include questions.

From the individual brainstorming, we created three areas of questions and topics: gaming, environment, and learning. As a form should not be of a too-large size, we needed to limit the number of questions that we added to the form. The topic around student preference on learning is something that will be difficult to use and as research has been done on the area a decision was made on finding related sources from the literature search. Therefore the learning topic with its questions was eliminated and decided not to be used in this user study. The environmental questions are focused and kept for the next step in the form of preparation. The gaming questions are many, some are similar and therefore grouped into topics. By voting the team chose important topics to focus on that reflect the goal of the workshop. Every attendee gave points to important issues. The topics in gaming without votes are eliminated.

The topics were combined and the questions were translated for the benefit of our young Swedish target group. The aim of the language was to reflect the target group level and the form was created and pilot tested on a small number of testers. The aim was for the feedback to relate to content, level, and language appropriate for the target group. Pilot test feedback was applied accordingly before the actual user study.

See Appendix A for the complete set of questions and the layout of the form.

2.2.1 Method of Workshops

The workshops were a central part of the progression of the decisions made throughout the empirical research. The method of individual brainstorming was used. Many methods of generating ideas belong beneath the heading of brainstorming where two major rules are followed. These being the generation of numerous ideas and being creative without regard for constraint and not removing ideas during the brainstorming (Norman, 2013). The brainstorming results were presented to the group and discussed after the presentation. In the discussion, we used Norman's third rule of questioning everything (Norman, 2013) to find out the intent of all the brainstorming content created. At the end of discussions, the aim was to align the result with our user study goals and from there the results were sorted and refined.

2.3 Data Presentation and Result Analysis

In this section, we highlight findings from analysing the data collected with the form. The data is important for creating an understanding of the potential users of our digital learning game, namely the students. This section will present some of the interesting findings from the form by firstly presenting general information about the respondents and the questions, secondly present game-related data, and thirdly present environmental-related data.

2.3.1 Participants

The respondents were students in grade 7 and 9 in the Swedish compulsory school. Most of these students were based in Malmö (58%) and Hjärup (31%), all of them based in Scania County (see Figure 1).

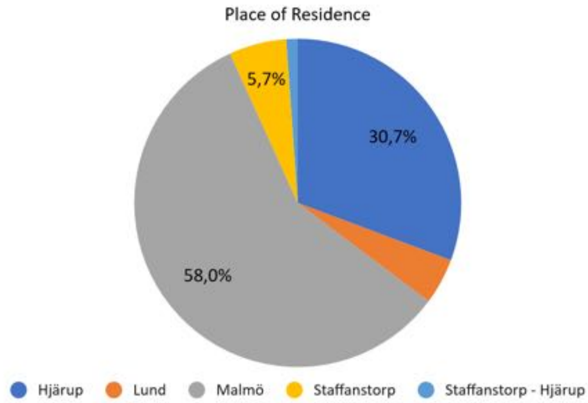


Figure 1: Pie chart overviewing the respondents' place of residence.

In total, there were 88 respondents (N=88) of which 66 are attending year 9 and 22 attending year 7 (see Figure 2).

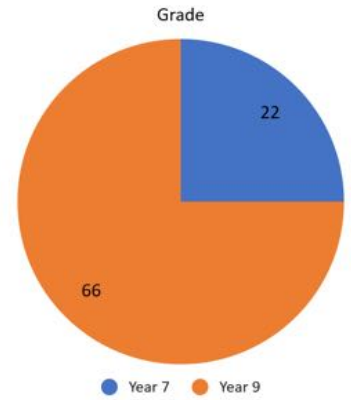


Figure 2: Pie chart overviewing the respondents' grade .

2.3.2 Apparatus and Procedure

The questions were distributed to and answered by the students using the online tool Google forms (google.com/forms/about/). The distribution of the Google form-link was handled by the students' teachers which we had come in contact with through our contact at Malmö Museum. The form is structured in a way that produces both qualitative and quantitative data. There are two main categories of questions, the first relating to gaming and the second relating to the students' environmental concerns.

The collected data was exported from Google form as a comma-separated value file (.csv) and this data was cleaned-up and analysed using R (RStudio) and Microsoft Excel. Some of the questions in which answers were not suited to be put into categorical variables were instead summarised in text.

2.4 Results of User Study

A number of points of interest are identified when analysing the data, a sample of these that are relevant for our project will be presented in this section. If we start out by looking at the game-related questions we can find a number of central points.

2.4.1 Collaboration or Competition in the Game

The respondents were asked if they preferred games that were focused on collaboration or competition. The answers, as seen in figure 3, show that 49 informants preferred a collaborative game style and 39 preferred competition. That means that a little more than half of the students (55,6%) preferred a collaborative game. If we divide the data up by grade (which presumably roughly also corresponds to age) we get the following results. Among the sub-group (n=22) which attended grade 7 the data shows that 14 prefered a collaborative game style and 6 preferred a competitive game style. This comes out to a percentage of 63.64% in favour of a collaborative game style. Among the presumably older students in grade 9 (n=66) the division was more even with a slight favorability towards collaborative game style. The numbers here were 35 in favour of a collaborative game style and 31 in favour of a competitive game style resulting in a 53.03% in favour of a collaborative game style. As a larger number of the respondents were in the grade 9 sub-group the overall favourability of collaboration is skewed towards the results shown by this sub-group.

One lesson that we learned is that it is not always clear-cut what a collaborative game is and what a cooperative game is. One popular game among a number of students is CS:GO (Counter-Strike: Global Offensive). In this game, a multiplayer first-person shooter, you are working in a team against another team (your team's objective can for example be to defuse a bomb). The question now is if CS:GO is a collaborative or a competitive game?

We would probably say that the game has both components. By comparing the answers given here to the answers given to the question regarding what the students' favourite game is we can perhaps see that the students might have a different view of what counts as a collaborative game. Games often have a mix of the two elements. What we can take with us from the question is a notion, if not to any high degree of certainty, of how the distribution looks concerning the preference of collaborative and competitive game style.

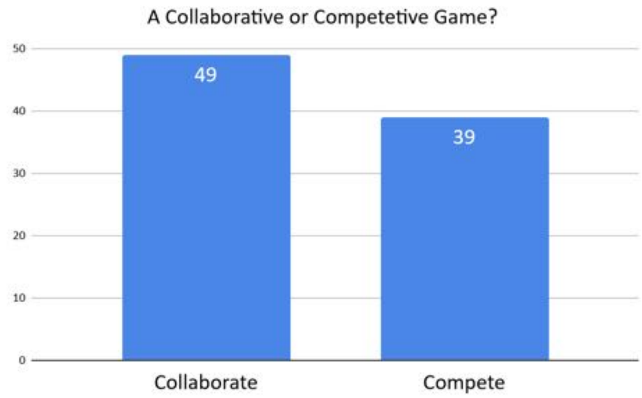


Figure 3: Bar chart giving an overview of the respondent's preference for either a collaborative or competitive games

2.4.2 Preferred Game Theme

The informants are asked what kind of theme they prefer in their computer/video games. They are given both a set of alternatives to choose from (realism, fantasy, Sci-fi) as well as an alternative to writing their own answers. They can also choose more than one alternative, for example, fantasy and realism (an answer which might be understood as corresponding to a thematic setting where fantasy **and** realism are combined such as the Harry Potter universe or as an answer that the student prefers a game which is **either** a fantasy **or** a realistic game **but not** a sci-fi game).

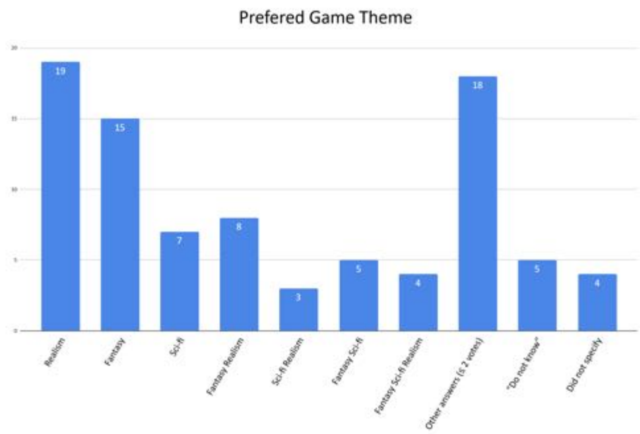


Figure 4: Bar chart giving an overview of the respondents preferred game theme

The collected data suggest that realism and fantasy are overwhelmingly the two thematic categories that most of the students prefer. 19 of the informants chose just the alternative "Realism" and 15 choose just the alternative "Fantasy". These two themes also occur in a number of other combinations as seen in figure 4. The combination of fantasy and realism is the third most popular answer given (n=8). Together these three bars represent 42 of 88 respondents' answers amounting to almost half of all answers. Regarding the two single largest thematic categories, realism and fantasy, the data divided by grade were as follows: Of the grade 9 students, 15.15% (n=10) checked the fantasy option and 22.73% (n=5) of the grade 7 students. The realism category was checked by 19.7% (n=13) of the Grade 9 students and by 22.73% (n=6) of the grade 7 students. As the division shows the differences between the subgroups divided by grades are not very large, at most differing with 7.61 percent points in the fantasy category.

A question arises whether the students' conceptions of words such as ‘realism’ and ‘fantasy’ in this context is aligned to how the words were meant to be perceived by us. One might argue that ‘realism’ and ‘fantasy’, in the context of alternatives alongside the alternative ‘Sci-fi’, is providing the respondents with the needed information to align their conceptions with how the questions were designed. But we also know that people are not always totally focused and giving a survey the totality of their attention and that they might answer questions based on preconceived notions of what the included words mean. Perhaps it would have been a good idea to anchor these abstract terms in concrete prototypical examples of the different instantiations of the themes such as J.R.R. Tolkiens ‘Lord Of The Rings’ as an example of the fantasy theme.

2.4.3 Preferred Game Style

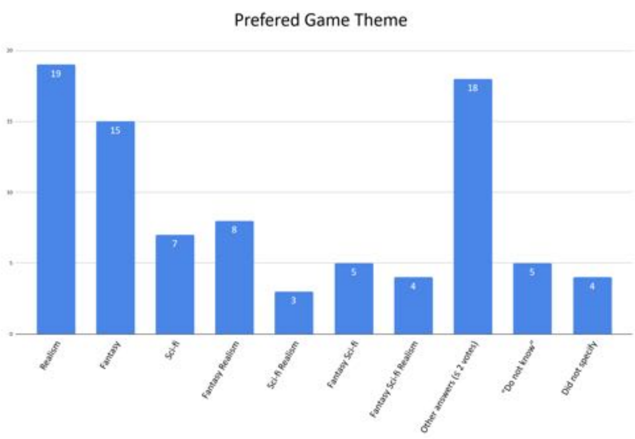


Figure 5: Bar chart giving an overview of the respondents preferred game style

Moving from the preferred overarching theme of a game the students are asked what kind of game style they prefer. Just as in the question about the theme the students are presented with a number of prewritten categories (action, strategy, Role-playing games, and puzzle games) which can be chosen individually or combined with other alternatives, there is also an alternative to write another answer. The alternative “action”, as can be seen in figure 5, is the most popular answer (n=24) and the puzzle genre is the least popular (n=1). “Strategy” comes in second place with 11 answers on it as a stand-alone answer and 18 times in combination with another answer (such as “strategy action” or “strategy puzzle action”).

2.4.4 Learning Through Games

To gauge the students' views on digital learning through a game they were asked the question “what do you think about learning things through playing games? Why?”. They gave us a number of interesting answers.

One student noted that it's good to have a game in which you learn things because the learning can be used as an argument to get to have more “game time”. Another student noted that he loves to play games, mostly action, and shooters, but puzzle games and adventure games are also “high on the list” - games where “you get to use your brain a little”. The concept of cooperation was also touched upon by yet another student who mentioned, “[i]n a game of cooperation where one has to think a lot, the brain develops” - the student writes.

Many students took this time to tell us about specific things they learn when they play games. Because many games are in English you learn a language, one student said. Another student echoed this idea and wrote that you “often learn English”, a third one said that “it is fun to learn English via a game”, yet another student noted that when one is playing with English speaking persons and talks to them one learns English. One of the students wrote “English and strategy. Both are fun”. The point of learning strategy was pointed out by a couple of other students, one who wrote that when one is learning one can find new strategies - probably referring to strategies in the gaming scenario. Some other concrete things the students said they are learning through games are “stress-management “, about “traffic signs”, about “money”, about “building things”, about “the wild west”.

Some students were concerned about the idea of learning through games, one explained that he/she feels a bit sceptical, but with the addendum that “perhaps one could learn a little, but not too much”.

Regarding a classical educational setting *versus* a digital learning game, we got a number of opinions and views from the students. Contra a classical educational scenario, one student said that learning through a game is a preferable method. This is because “one often learns more when one thinks something is fun”. Another student said that learning during a game is more fun than to “sit and listen” and that variety makes it more fun. This was also expressed in comments like “it is much more fun [to learn through a game] than to learn in, for example, school. One can learn a lot and it is fun at the same time as it is good”. A student wrote, on the same theme, that it’s good with learning games because then there are different environments where one can learn new things, “to be able to study in new and entertaining ways [which] also is more effective”. One student seemed to be more sceptical about mixing fun and learning, he wrote that learning isn't the important thing during a game and to learn things “is something you do in school. I think that one should have fun during a game". Here seems to be the implicit idea, that school and learning can't also be fun. This was also suggested by a couple of other students, of which one said that learning in games is "boring, games should be fun and competitive, not a lecture". Finally, two quite long and elaborate answers were given by two of the students that provide food for thought. The first student concluded that:

"If you are talking about, for example, learning a school subject through [a game], I think it is neither fun nor boring. It's just something one takes in and don't think much about, which I think can be good, but if it's loaded with [learning] it can be a little hard and repetitive"

The second student wrote:

"I think that it is good because it is taught in an effective, smart and fun way. If one can balance the entertainment of a game and be able to bake facts about, for example, a specific subject into a mission, then the probability that will remember the fact be higher because one is associating the fact with the fun in the mission."

How well do these results align with previous research into children's opinions about learning games? In a study with 74 children in primary school, researchers Iten and Petko found that the children had a "generally positive attitude towards games for learning" (Iten & Petko, 2016, p.156). This general positive feeling seems to align quite well with our students' self-reports. The two researchers also found that the anticipated usefulness and simplicity were more relevant to the children than the "fun" factor of learning with the game (Iten & Petko, 2016, p.160). As our results here are qualitative rather than quantitative we can't directly compare the data. But many of our participants confirmed in their answers that "fun" was an important factor. A few however also spoke to the need that learning games run the risks of being too "hard and repetitive" which seems to be more in line with the usefulness and simplicity aspects.

All of these answers are interesting and give us a certain level of insights into how the students view learning while gaming and related claims about what kind of things the students think they learn while gaming. It seems reasonable that for a digital learning game to be able to reach and engage the students it has to, to a reasonable degree, cater to their interests and perceived needs of the students. To have this data means that we do not need to second guess what the students think about learning and gaming.

However, it should also be noted that self-reporting of data of this kind can be unreliable. Self-assessment through introspection, in certain situations, can have its limits and as early as 1977 Nisbett and Wilson claimed that evidence showed that there "may be little or no direct introspective access to higher-order cognitive processes" (Nisbett & Wilson, 1977, p. 231). Even if this claim by Nisbett and Wilson might be too strong, there is, for example, newer data that indicates that there might be a mismatch between experiment participants' introspective reports and their actual choices made under certain conditions (Johansson et al., 2006).

One thing that we could have included in our study was a question relating to if the students had any experience form using an existing learning game, and if they had, what that game was and if they liked it. This kind of data could help to anchor the students' self-reports in more concrete facts and in that way provide some grounding for their introspective reports. For example, as discussed by Iten and Petko, how should one interpret the word "fun" in this context (Iten & Petko, 2016, p.153).

2.4.5 Environmental Views

Regarding the data, relating to the students' views on the environment and how we as humans are affecting the environment, we also want to gauge the students' views. Two of the questions can be highlighted here. One more descriptive and individual in which the students on a five-level Likert scale were asked to estimate how often they thought about their own environmental footprint. Another one more normative and collective question we asked was if the student has a positive or negative view of how humanity is working with environmental issues.

Regarding the first of these questions, we can see, in figure 6, that the answers nicely fell into a normal distribution with close to half of the respondents (n=39) in the middle of the five-level Likert item. Is this result surprising? Perhaps a little, if one is to believe the news that tells us about how the climate movement is prevalent among young people today (the so-called Greta Thunberg-effect). This "surprise" however does not take into account that the distribution might have been skewed to the left before Greta Thunberg so the result of the Greta Thunberg-effect now is a normal distribution.

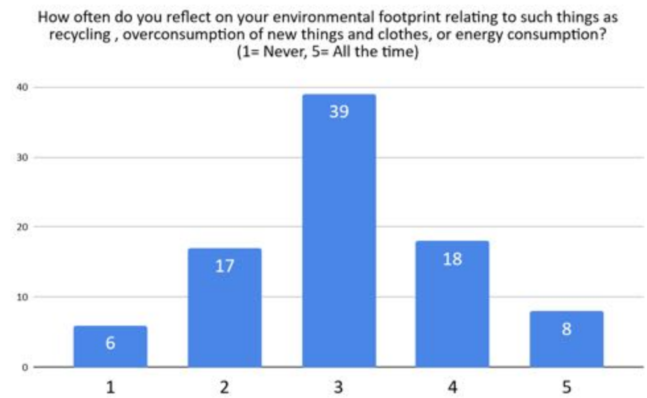


Figure 6: Bar chart giving an overview of the respondents self-evaluation of their own environmental footprint

This might be interesting to speculate on but for our present aim, it's more important to note this distribution shows that the students think of themselves as quite aware of the environmental footprint they have in regards to recycling, overconsumption of for example clothes and relating to energy consumption. Only a few students (n=6) report that they never think about their own environmental footprint. If this overall quite high awareness also translates into an interest in one individual environmental footprint this might be a good point to incorporate in the content of our digital learning game.

The second question of the environmental part aims to gauge how the students look upon the environmental future and humanity’s role in this development.

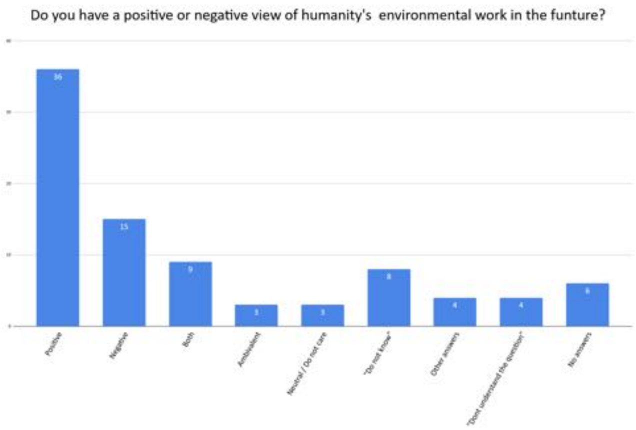


Figure 7: Bar chart giving an overview of respondents' views on humanity's future environmental work.

The answers, as presented in figure 7, show that the amount that has a positive view (n=36) are more than twice as high as the ones that have a negative view (n=15). A lot of the students (n=37) however are not giving a binary answer for one side of the positive/negative dichotomy for different reasons.

Some interesting points can also be reached by modelling some of these answers on answers given by the students in other questions. One example is by cross-referencing the data of students who have a positive view of humanity's environmental work in the future with the answers the same students are giving on the gaming question relating to if they had a preference for a collaborative or combative game. This shows that out of the students that prefer a collaborative game style (n=49) 25 has a positive view of the future as opposed to 7 of them who have a negative, the others (n=17) have not given a clear-cut answer on the positive/negative question. The same effect is however not as strongly present when looking at the students that are preferring competition in their games - here the distribution is more even.

3 Theoretical Background

The theoretical background provides the pedagogical requirements of the game that will be reflected in the next step of the user-centred design process. The pedagogical paradigms we are taking into account are “social constructivism” and “situated learning”. They will be further elaborated in the chapter. We continue with a paragraph about environmental education and educational gaming theories. After that, the state of the art is presented along with how the Swedish curriculum is used.

3.1 Pedagogical Paradigms

The first pedagogical paradigm we are following is one of social constructivism. Social constructivism in the context of learning is the idea that we learn in social contexts and the students have a mutual communication with the teacher and with each other. This ties into Schwartz’s (1999) idea of collaborative learning. It stresses that the involved people are rather contributing to knowledge than just borrowing it. Collaborative learning enables people to build a shared understanding step by step. In order to collaborate, individuals must enter a relationship. Here they have to produce ideas, decide whether to communicate and to compromise their goals (Schwartz, 1999). The social constructivist way of teaching is based on Vygotskij’s zone of proximal development. The idea is that the student’s learning process can be mapped onto three different zones: a “current understanding zone”, followed by a “zone of proximal development”, and an “out of reach zone” (see Fig. 8). In the “current understanding zone”, the student has no problems mastering the content individually. The “zone of proximal development” is characterized by the students learning and progressing with material that is on the next step of what they already know. This is a zone that requires interaction with a more competent other in order to achieve assimilation. The last zone is the “out of reach zone”, often causing anxiousness or a lack of motivation in the student because of too difficult content.

In order to target the zone of proximal development, students need to engage in the content they can only solve with the help of more competent others like fellow students, teachers, or pedagogical agents, etc. The students will then experience that they are later able to solve these assignments on their own and hence every zone expands (Løw & Skibsted, 2012).

Figure 8: The zone of Proximal Development - Vygotskij, Lev.

The second paradigm is one of situated learning. It means that learning is not something localized within the inner consciousness but arising in relations between people, in cultural contexts and practices. These complex relations will change together with the learning processes and continually become new tools for further learning. With our relations being present throughout our whole life, learning will follow and keep changing. Learning is not something we notice before we can do more than we could before (Løw & Skibsted, 2012).

In practice, this requires one looking at the relations between a person (student/teacher), activity, and situation as embedded in social practice to understand the conditions of learning and its inequalities. Learning is depending on all these relations and if the students have not gotten anything out of the teaching then it is necessary to analyze and improve the relations (Løw & Skibsted, 2012).

Furthermore, situated learning focuses on studying people's participation in and across social practices. According to Lave and Wenger (1991), learning is a movement from legitimate peripheral to full participation in different activities and different social practice communities. The legitimate peripheral participation includes someone who is legitimate because one is acknowledged and seen but peripheral because one is not yet participating in the practice communities' most important and challenging assignments. In the state of full participation, one has an understanding of the assignments and different ways of approaching them and can also teach others (Løw & Skibsted, 2012). By teaching others, one reflects on what has to be done and gets aware of the necessary processes to solve a specific assignment. The switch from being a learner towards becoming a teacher can cause a feeling of empowerment and engagement (Schwartz, 1999). This can cause students to find the assignments more meaningful and take responsibility when teaching their peers (Bargh & Schul, 1980). The incentive to explain things in a new way can give new perspectives (Bargh & Schul, 1980). This can cause recursive feedback where students have to reflect on how the teaching was received by the other students (Bargh & Schul, 1980). Therefore, this concept of learning by teaching should be taken into account when creating a proper learning environment.

3.2 Environmental Education

Environmental Education covers a specific facet of education which we want to take into account for our work as well. "As environmental problems are becoming some of the big challenges to democracy and humanity it is of the greatest importance critically to reflect upon their pedagogical implications." (Jensen & Schnack, 1997). This quote wraps up quite well why it is important to have a pedagogically valuable approach to sustainability. An approach we got inspired by is the action competence approach of Jensen and Schnack (1997). They are stressing the necessity of finding an alternative to the tradition of a science-oriented approach to environmental education, since "environmental problems are not only a matter of quantitative changes (less consumption of resources, less transport by car, less electricity consumption, etc.) but also (and maybe more so) of qualitative changes" (ibid., p. 164). These behavioural changes must be tackled at two levels: the societal and the individual.

3.3 Gaming and Learning

When wondering about how to approach specific learning goals, gaming is an increasingly popular option but has also met some resistance in later years (Linderoth, 2012). In current research of gaming and learning, there is a general consensus that learning through the medium of games offers some stimulating possibilities. According to Gee (2009), games are good problem-solving spaces that use learning and give a pathway to mastery and entertainment (Gee, 2009). One of his main points concerns the motivational aspect of learning. He says, that according to the work of Demasio, (1994, 1999, 2003) people learn more deeply when there is an emotional attachment to their learning and problem-solving (Gee, 2009). Games can in this case be a medium for providing the learner with a framework that can elicit these emotions. Also, Lee and Hammer (2011) point out that motivation is a driving factor in games and learning. Their idea is to use the motivational power of games and apply them to real-world problems of e.g. schools. From an emotional point of view, the games have the ability to expose learning through trial and error. This promotes experimentation and the positive relationship with failure is maintained through rapid feedback cycles and minimum penalties for mistakes (Lee & Hammer, 2011). In du Boulay's article on Evaluation Methods of Adaptive Systems (2020), he claims that the satisfaction of the learner has been degraded as a trustworthy indicator for learning. This is problematic, as it has an influence on the development of their metacognitive and meta affective insight and their long-lasting motivation.

Another possibility games offer is immersion. According to Lee and Hammer (2011), games allow players to try out new identities and roles, where they have to make in-game decisions. These roles can also be less fictional, and allow players to explore new sides of themselves. Gee (2009) refers to this as micro-control. He claims that it gives rise to embodied intimacy or a reach of power and vision. This could be the ability to control characters and other elements in the video game. This kind of micro-control has, according to studies, an interesting effect on humans. When we melt into the characters, our cognition becomes embodied. Research shows that humans learn best when we think and problem-solve through an experience we are having as embodied beings in the world (Gee, 2009). Gaming also has the ability to put the players into worlds, where the experiences can recruit learning. These experiences could in some ways represent the practice communities that Lave and

Wenger (1991) describe as being a vital part of learning. According to Gee, new studies claim that people learn through experiences that are stored in memory and are later used to run simulations in their minds to prepare for action and problem-solving situations. They then perform hypotheses based on these simulations and about how to proceed in new situations based on past experience (Gee, 2009).

One general issue with educational games is the fine balance between entertainment and learning. According to du Boulay (2020), there is a potential tension between capturing interest and inciting engagement and creating learning. This is because engagement and interest are necessary but not required conditions for learning. The reflective part of learning can sometimes be neglected if engagement and fun don't leave room for it. This is often seen in educational games, where the main reason for their use is their ability to engage (du Boulay, 2020).

As mentioned earlier, in order to move from the "current understanding zone" to the "zone of proximal development" the student must receive feedback from a more competent other in order to be guided in the right direction (Løw & Skibsted, 2012). Feedback signals the learner whether their answer is correct or not. Furthermore, the student can be presented with information that can help them assimilate or accommodate their knowledge. The students can also patch knowledge gaps and correct misconceptions by processing feedback (De Bra, P. et al 2008).

This promotes another problem when creating an educational system (ES). The system and its feedback have to be tailored to the students learning style. These are all factors that have been observed by De Bra et al. (2008) to have a significant influence on the students' performance in tests.

3.4 Summary

The results of our literature have been synthesized into these points of suggestion for the context in the designing of the game:

- Collaborative learning
- Learning by teaching
- Action competence approach
- Maintaining the gamer's motivation and immersion, while keeping the balance between entertainment and reflective learning
- Feedback

3.5 State of the Art

In order to design something that is new within the field of learning games, we performed a state of the art. Our main approach to the collection of literature is through the assistance of our supervisors, keywords on Google and Google Scholar, and previous knowledge of learning theory.

3.6. Design Theory – User Interface Heuristics

For the user interface (UI) design of The Green Island learning game, we are applying the UI heuristics defined by Jakob Nielsen and Rolf Molich. These heuristics for UI design, which are quite general, will be used as guidelines in our UI design. They range from reducing the cognitive load to promoting system transparency all to give a better user experience while using the system.

As our working memory and short-term memory is limited - something that famously was pointed out by Miller in his article 'The magical number seven, plus or minus two' (Miller, 1956) - it is important to minimize the memory load that is placed on the user. Nielsen and Molich suggest that good design should include both simplifying complicated instructions and presenting dialogue in a way so the user doesn't have to remember all the information presented to them (Molich & Nielsen, 1990, p. 339). Nielsen points out that recognition in the form of recognizable elements, for example, is one thing that reduces the need for recall which lessen the cognitive effort that is needed from the user (Nielsen, 2020). These points are of extra importance for us in our project when designing a learning game where we wish that the cognitive effort of the user should be directed towards learning the actual content and subject rather than navigating and understanding the system that aims to facilitate rather than obstruct the learning process.

This last point about cognitive and memory load ties in with another of Molich and Nielsen's design heuristics that states that the design should try to mirror familiar language and concept of the user rather than "in system" vocabulary. The system should also give feedback to the user so the user is kept informed about what is happening with the system as changes are made. The general principle of transparency is also captured in the heuristics of providing the user with possible "emergency exits" or undo buttons wherever the user finds themself clicking into the wrong menu, window or making a mistake (Molich & Nielsen, 1990, p. 339). This helps the user to avoid ending up in dead ends which, among other things, produce a frustrating and perhaps even broken user experience.

Along the lines of transparency the system should also, according to Molich and Nielsen, provide the user with good error messages when needed (and another perhaps quite obvious related design principle is to try and keep errors out of the system to begin with they note) (Molich & Nielsen, 1990, p.339).

These heuristic principles seem to go well with the views of Norman when he is writing about reflective cognition and external aids, he writes that the "external representations have to be tuned to the task at hand if they are to be maximally supportive of cognition" (Norman, 1993, p.25).

4 Conceptualising the Learning Game

The purpose of this chapter is to give an overview of how we applied the insights of our research onto an actual game concept. The concept finally combines the three areas that are important when designing the learning game: pedagogics, game design and sustainability. Firstly, the learning goals from the Swedish curriculum are presented. After that, we describe the setting of the game broadly. The chapter continues by explaining both the frame we set for the game and the content we want to fill this frame with.

4.1 Curriculum

While we want to meet the interests of the target group, we also want them to learn, so we searched for relevant learning goals in the Curriculum for grade seven to nine. We found learning goals that fitted our ideas in several subjects: civics, biology, home and consumer studies and geography. The learning goals are the following:

“How food and other goods are produced and transported, and how they impact the environment and health.” (Home and consumer studies)

“Where different goods and services are produced and consumed, and also how goods are transported.” (Geography)

“Opportunities for individuals and groups to affect decisions and development of society, and also how decisions can be affected within the framework of the democratic process.” (Civics)

“Impact of people on nature, locally and globally.

Opportunities for consumers and citizens of the society to contribute to sustainable development.” and “[...] Relationships between populations and resources available in ecosystems.” (Biology)

(Skolverket, 2018)

4.2 Green Island

The aim of the game is to teach students about environmental issues while creating awareness about the relation between the environment and human behaviour as well as human action.

The game will be set on an island. Since there is only a limited amount of natural resources, the variables having an impact on this small ecosystem should become prominent. The way of living mirrors the western society which the target group (i.e. the students) are familiar with. This decision of a setting that mirrors the students real world is a consequence of the students preferences for a realistic theme in the game as shown by our user study. The island theme also leaves open space for incorporation of fantasy which was another highly ranked theme by the students.

The game is designed for two players working in a team. The task for them will be to create a more sustainable society by solving different problems. The players have the option to act as different societal entities, for example, the government, business people or a private person. Thereby they should get an insight into different decision processes on various societal levels and learn how those levels both interact with each other and can have an impact on a transformation towards a sustainable society.

4.2.1 Research and Game Concept Brought Together

The main pedagogical approach is one of situated learning. Focusing on the unit “person-activity-situation” the student should get new insights by getting engaged both in a new surrounding (the game setting) and in relation to the other student they are interacting with. The method of learning by teaching should ensure that the students are reflecting on the input they get.

The game is set up with cooperative asymmetrical mechanics. Two players are building a team and work together to solve the problems within the game. Each player gets exclusive access to certain information. They need to mutually exchange the information they get to be able to solve the challenges of the game. This asymmetric relationship produces a need for collaboration, communication and mutual teaching between the two students. This connects to the collaborative learning and learning by teaching approach described in part 3.1 about the pedagogical paradigms. For example, Schwartz (1999) postulated in his article, “people appropriate knowledge when they are given opportunities to produce knowledge” (p. 4). This also aligns well with a majority of the students' interests as they preferred a collaborative game style as described in the results of the user study presented in part 2.3.2.

In order to ensure direct feedback and an immediate change of the setting, the game will be implemented as a digital learning game. A digital surrounding offers fluent options of visualising progress within the story of the game, which should motivate the students to continue playing. As mentioned previously, Gee (2009) claims that micro-control increases the learning experience. Therefore the students can act out decisions as a person from a certain societal level.

Moreover, the game should contain aspects of Environmental Education (EE). "In traditional EE, the emphasis is often put on knowledge, attitude, and behaviour change at the individual level that fails to consider actions taken in the social and political contexts" (Chen & Liu, 2020, p. 2). Therefore we want to follow the action competence approach for EE by Jensen and Schnack (1997). This “is a breakthrough in differentiating action from behaviour and stressing critical thinking for transformative learning” (Chen & Liu, 2020, p. 2). The aim with the action competence approach is to provide students with authentic settings where they get “opportunities of planning and taking action (Chen & Liu, 2020, p. 2). This should empower their social skills of problem solving in order to prepare them for being active citizens later on. “Studies indicate that action-oriented and transformative pedagogy cultivate students to be active participants, […] and construct their visions for finding strategies toward the problems” (Chen & Liu, 2020, p.1). While Jensen and Schnack see the teacher as the facilitator of the students’ learning process, we want the game to be the student’s tool for becoming facilitators for each other. Our learning goals are orientated towards the Swedish Curriculum.

4.2.1.1 The learning situations as a socio-technical system

To get a full overview of the learning situation that is created around our design we can utilise a model of distributed cognition to describe the different aspects of this socio-technical system. Distributed cognition, as championed by Edwin Hutchins, is the theory that the outcome of a cognitive process (or perhaps task) can not be solely determined by looking at the single individual’s internal system. Rather we need to look to the individual in connection with its surrounding social and technological environment as a distributed cognitive system. As Hutchins puts it, we should take the “socio-technical system rather than an individual mind as its primary unit of analysis.” (Hutchins, 1995, p.265).

This necessity to see the total system is made extra clear by the situation that we have set up in our learning game. By dividing up the information on two different screens for each student, neither of them has by themselves the needed information to form the appropriate representations to solve the learning module at hand. The socio-technical system can here, in its first state, be described as dysfunctional as the asymmetrical distribution of information in the system hinders the system from reaching its end state (goal), namely to solve the learning module in the game.

So, to give an overview of the learning situation we can identify a number of internal and physical representations that propagate, transform and interact with each other in the system. Following the flow of information, we can first identify that we have two computer screens which the students are watching. On these screens, we have physical representations in the form of words, numbers, pictures and so on, these are accessed by, and propagated to, the students by ocular inspection. The students then produce internal mental representations of the information. The students can now, in their mind, perform transformations on these complex internalised representations, trying out different combinations of facts and try to draw conclusions from these. These internal representations can also be propagated by verbal communication from one student to the other student, and vice versa. The verbal communication, presumably most often in the form of words, is another example of physical representations in the distributed system. By relaying information to each other the two students can expand their own internal representation of the active scenario and problem with new information the other student provides. By doing this the two students' respective models become more aligned.

The asymmetry of the distributed system is reduced the more aligned the two respective mental representations of the two students are. Thus, in a perfect scenario, after discussing and “teaching” each other, the two students would end up with somewhat isomorphic mental representations and, in consequence, the needed combined information to solve the module.

One thing to note is that the set-up of the learning situation is such that all of the above-mentioned elements and the interactions between them have to be in place for the problem solving socio-technical system to work. If we, for example, remove one of the students, or one of the computers, or the possibility for communication etc. the whole system breaks down. It is thus an optimised system given the constraints we have imposed upon the learning situation through our game design. Optimised is here simply to be understood as there not existing any redundant components in the system, so perhaps ‘minimal’ might be a more fitting word than ‘optimal’.

This is of course a formalised and idealised model of what is going on in the learning situation. The real world is messier, for example, both of the students come into the situation with both explicit and tacit knowledge (representations) about a diversity of things that will come in handy in solving the learning module. The students are not copies of each other and will apply unique skills when solving the problem at hand. This above description does not take into account the details of feedback the computer system gives, and that the students might give each other when the learning module is completed (or not completed satisfactorily).

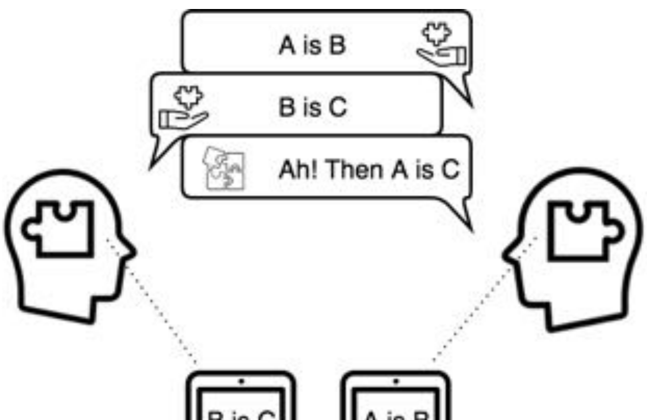


Figure 9: The created learning situation as a socio-technical system

To make an analogy we can view the problem-solving process as a puzzle-solving exercise. We have two non-overlapping sets of puzzle pieces that are respectively represented in the minds of the two students. By the transfer and propagation of representations through verbal communication the students hand each other previously unavailable pieces of the puzzle which are needed to complete the set of puzzle pieces the students started out with. Through discussion, teaching and thinking the students can now solve the puzzle.

4.2.2 Thematic Concept and Environmental Content

To make the challenges of the game close to reality the game will reflect the environmental and political challenges that our world is facing. These challenges will be simplified but the relations between them will be reflected in the game. The goal of the player is to make choices to improve the environmental conditions on the island.

The idea is that players can choose between two different narratives. Following the governmental narrative, players are making decisions and solving issues based on the whole governmental ecosystems. They should learn about the challenges of sustainability in the context of the base functionalities of the world and the tools that exist today to improve environmental efforts on a global scale. In the citizen narrative, a player follows and solves environmental issues in everyday life while learning about the impact everyday choices have on the environment.

4.2.2.1 The Governmental Perspective

The goal of the player will be to achieve sustainability through top-down processes within society, e.g. the governing of laws. This is done by protecting nature, the wild and the oceans while still working for the economy to be stable. This constructive take seems to be in line with the larger part of the students which had a positive view of humanity's future environmental work (as presented under the heading *2.3.2.5 Environmental Views*).

A stable island economy leads to more resources for the player to use in sustainability efforts on the island. Furthermore, a growing population will affect the environment of the island and the player will need to increase opportunities for the population as research shows that a population eventually stops growing and even decreases as the living standard of the population increases. Resources are the game's representation of money and these resources are gained by the player when decisions are good for the economy. Decisions regarding environmental efforts will initially be costly in the game. As the game progresses the companies will move towards sustainable economic growth and resources can be earned on these as well. The game will in this way reflect the challenges that the world today stands before, where many environmental efforts are costly and therefore many choose less sustainable options. However, as the world progresses and becomes more aware of sustainable solutions the easier and cheaper it becomes to run a sustainable green island.

4.2.2.2 The Citizens' Perspective

The goal of the player will be to achieve sustainability through bottom-up processes within society. The citizens' perspective will focus both on decision making in everyday life (e.g. consumption) and on opportunities to come into action and have an impact on political decisions (e.g. demonstrations).

5 Requirements

Requirements are necessary from both a design and a software product point of view. The purpose of creating requirements is to clarify the criteria necessary for building the product (in our case the game). This benefits developers, designers and stakeholders as it should create consensus on what is to be developed. The requirements were compiled in several workshops and assembled as Appendix B. The appendix contains an introduction, general game description, and system features containing requirements on game content, game structure and scoring.

5.1 Preparation

Two requirement workshops were performed to ensure that the result of the research done this far in the paper is central in the creation of the game prototype, the result of these are presented in appendix B.

The first workshop was conducted in the form of a literature study where requirements theory was researched individually and then presented in the group. The focus of the workshop was to clarify possible structures and uses of requirements. The second workshop was combining the learnings from the first requirement workshop and the research performed in chapter 2-4 in this paper. The game was described and results were rewritten in the form of requirements, see chapter 1, 2 and 3 in Appendix B. The result was grouped into game content requirements, game strategy content requirements and user requirements.

5.2 Game structure

The game structure workshop was performed to ensure that the game design will reflect the requirements that are the result of the work in chapter 2, 3 and 4 in this report.

We chose to structure the game into modules to allow for many of the ideas and theories to be included in the game more comprehensible. A module can be seen as a thematic collection of environmental topics. By using modules, different topics are easily structured and the design of scenarios can make use of feedback mechanisms and collaborative learning theories.

Modules can be split into multiple submodules. As presented in figure 10 a submodule could be split into phases. Problems are formulated and through the presented information players discuss and then make a decision of some sort, for example in answering a quiz. Based on their answer, players receive feedback.

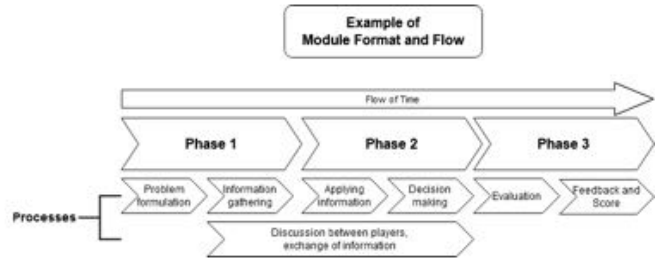


Figure 10: Example of module format

Scoring system was further discussed along with the importance of feedback: visual, quantitative, and qualitative. The results of the game structure workshop were written down as requirements in chapter 4 in Appendix B.

5.3 Requirements iteration

In iterating requirements the game design was started and iterated before the actual creation of prototypes, saving much time as requirements are iterated instead of prototypes. Changing requirements is less time consuming compared to changing prototypes. When the prototyping starts the game structure is already iterated and thought through.

6 Prototype Design Process

The purpose of this chapter is to present the design process that concretises the conceptual design and how the game will implement a collaborative asymmetrical learning theory. We present the proof of concept, how topics were chosen to be implemented in the prototype and the importance of the learning goals in the creation of the prototype.

6.1 Proof of concept

To initiate the design and start the creative process a proof of concept was introduced building on the ideas around the game structure that had been discussed during the requirements iterations. The purpose of the proof of concept is to concretise the ideas into something more palpable where the structure can be iterated, tested and improved.

Two versions of the main island overview were created. One was in a more cartoonish style and the other more realistic, see figure 11.

Figure 11: Two versions of the Green Island design

We decided to go with the realistic theme, allowing for the themes in the game consisting of pictures rather than drawn cartoon pictures taking more effort to create compared to using photos as a base.

The idea of the game is for the player to choose a topic and from there open a module where the asymmetrical collaborative learning theory will be implemented in terms of different textual and pictorial information being presented to the players. Figure 12 presents a rough example of how this could look like.

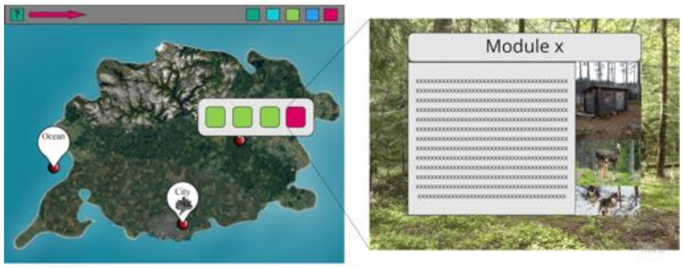


Figure 12: Opening a module - here, they are presented with information in the form of text or pictures

As can be seen in figure 12 we further expect an information bar in the top to present the player with scores, progress bar and help button. At the end of each module, a player is presented with a test based on the learning content in the module as presented in figure 13. Furthermore, whenever a module is completed the player is presented with feedback on progress and scoring.

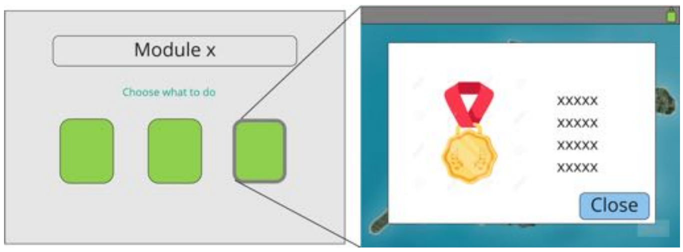


Figure 13: Module end-screen

6.2 Choosing environmental topic for prototype

As we are creating an educational game with a focus on the environment there are many topics in the area that can be included. We aim for this to be reflected in the prototype along with choosing a topic to demonstrate how the learning theories can be implemented and the game structured.

The topic brainstorming produced about 90 subtopics that were grouped into main topics, see appendix C for full list:

- Waste
- City
- Agriculture
- Ocean
- Energy
- Mining
- Forest

These main topics will be visible from the island main page in the game, and by clicking on them related subtopics will be presented.

In order to elaborate on the broadly discussed areas, we brainstormed specific criteria that should be covered by the topic we want to exemplarily prototype:

- The topic should be broad, covering as big of a span as possible of the game.
- The topic should enable the player to approach the problem from different angles.
- The topic should be able to be broken down to submodules allowing for a lot of prototyping
- The topic should be relatable for the students while being a good representation for the game concept

Each member of the project group chose and presented three of the 90 subtopics reflecting the four criteria the most. After group discussions, voting was performed using the criteria. As a result, the environmental topic chosen was waste, specifically electronic waste. The topic is furthermore suitable as it is anchored in the everyday life of our target group – digital natives. When we were comparing the topics, we wanted to include the broad learning goals of the school curriculum. The learning goal from the subject *Technology* aligns with our choice of topic: “Transforming raw materials into finished products, and managing waste in some industrial processes, such as the manufacture of paper and food stuffs” (Skolverket, 2018).

6.3 Learning Goals

The prototype will contain 4 modules where each one of these needs learning goals before they can be implemented. We first present learning goals for the topic selected in the previous section and then the learning goals for the introduction module. Finally, we reflect on the importance of the learning goals for creating the prototype and further effort in developing the game.

6.3.1 General waste and electronic waste

On the basis of the broad learning goal from the subject *Technology* and content on the topic of electronic waste we formulated concrete learning goals that are used in the prototype design:

- The student should know why electronic waste should be recycled.
- The student should know what kind of products are considered electronic waste.

Even as electronic waste is considered a sub-topic there is still much knowledge that could be considered as important. Learning goals above are to be considered as part of the introduction to electronic waste. The player can then enter more detailed modules that aim to produce a more detailed understanding of the topic. Warranty (for electronic products) is prototyped as one of these more detailed knowledge topics with the learning goal:

- The student should know how laws on warranties of electronic products work in theory and in practice.

6.3.2 Introduction module learning goals

The introduction module contains two submodules. The welcome module introduces the rules of the game and how the modules are set up. The learning goal for the welcome module is formulated as:

- The student should be familiar with the rules of the game and the way it is structured.

The learning goal of introducing the general environmental topic is formulated as:

- The student should be able to give a broad definition of the term "sustainability" and "environmental thinking".
- The student should be able to name what is needed to have solutions taking effect in improving sustainability on a global scale.

6.3.3 Learning goal conclusions

In the creation of the full game, there is a need for evaluation of what learning goals are important for each topic. The method used above identifies content and knowledge that is considered important in order to cover the broader learning goals. Learning goals for environmental topics need to be weighted from the perspective of society. Over time some of the more detailed topics will be outdated if the content depends on facts or changing research results. However, this is hard to get around as specific content is needed to use the asymmetric collaborative theories as presented in the next chapter.

Our recommendation to content designers (implementing the learning theories using facts and environmental knowledge) is that learning goals and the knowledge content always are determined before modules are developed.

7 The Prototype

On the basis of our general game concept, we designed a testable prototype. The aim of the prototype is to provide an example of one branch the players could follow from the very start of the game until a specific submodule and its corresponding mini-games. Creating the prototype allows testing the usability of the design along with game structure. Furthermore, testing the game application of the asymmetric collaborative learning theory is needed to verify whether the different concepts work together. In this chapter, we present an overview of the prototype in appendix D, how modules are structured and finally give examples of how the content of a module is presented through mini-games and a test.

We recommend a walkthrough of the slides in appendix D while reading this chapter.

7.1 Prototype Overview

Appendix D contains the prototype. There are 8 documents with slides, that contain 4 storylines (D1, D2, D3, D4) where each storyline have two slides, one for each player (player A and player B) as some slides are different for each player. The documents are named appendix D1_A, D1_B, D2_A, etc. An overview of how the prototype should be played can be seen in Figure 14. In the top is the main-menu where arrows going down indicate that the player has clicked to enter a module that is presented vertically. As can be seen, 4 modules have been created.

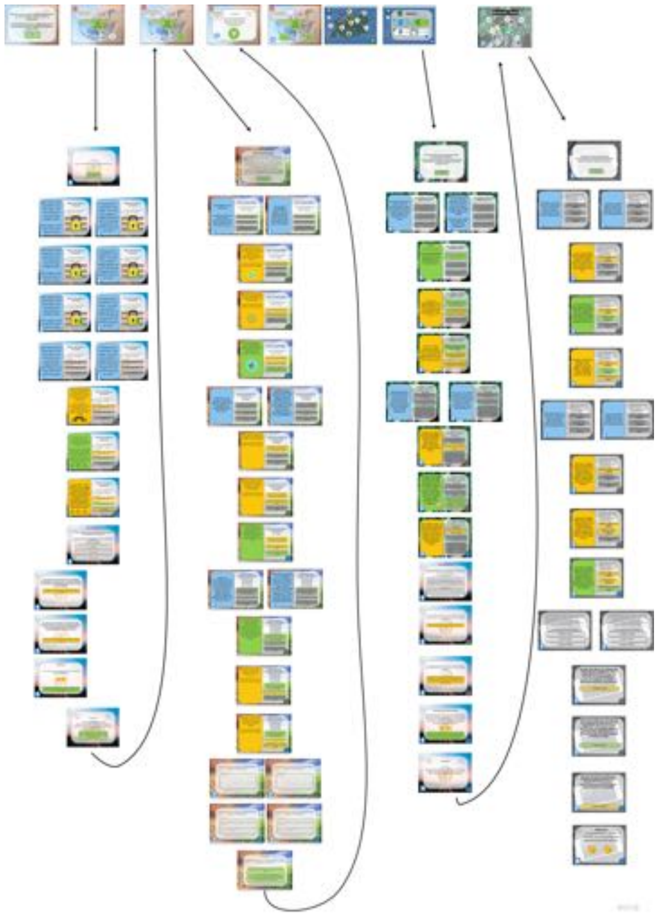


Figure 14: Prototype flow, see full detailed figure in appendix D0

A player needs to know how the game works and be supported initially. Therefore, we have added an introduction tutorial that embodies the same structure as the game modules. The game is initiated through this intro.

The game introduction is found in Appendix D1 and D2. The introduction contains two parts, one module explaining rules around how the modules should be played and one module introducing the environmental topic on a very general level.

After the introduction module is finished the player enters the island. Once the player has access to the island, modules are available for unlocking, see figure 15.

Figure 15: Island after intro and waste area

Progress within the topics is shown by colour indication and an increasingly filled circle in the right corner of each topic reflecting the current progress. Storyline D3 contains the general electronic waste module. When a player has finished the general electronic waste module, detail modules are unlocked entering the electronic waste area in the game as seen in figure 16. Storyline D4 contains the module on how warranty on electronic products affects the consumers.

Figure 16: Electronic waste module

When an entire area is completed the progress indications turn into a green medal along with the dot on the island map going from red to green. Example of the finished waste module is presented in figure 17. On the island main menu, there are always two buttons in the top left corner, one book icon to track finished and unfinished modules and a help icon.

Figure 17: Completed area

7.2 Module Overview

Asymmetric collaborative learning is implemented through the minigames and test of each module. The base module structure that has been used in D3 and D4 presents knowledge through 2 minigames and then has a final module test that is scored. Essentially they are building blocks that we have chosen to use. In the further development of the game, it is possible to add or remove mini-games and even change the final test structure. In D1 we had one mini-game and one final test. In D2 we decided to add an extra mini-game. Here, the student should write a short presentation of what they just learned from the content, instead of doing the final test.

Knowledge is, in the modules, presented as mini-game quizzes where the outcome of the player is not scored. The learning goals mentioned in 6.3 and the environmental content provides the foundation for the mini-quizzes to be created.

The purpose of this is that the two players discuss the differing information they are given and answer a multiple-choice question. They get the same question but will have to answer individually on their respective screens. They need to combine the information they have been given to answer correctly. Behind each possible answer, we present feedback that provides more information guiding the player. Discussion can further be made around why certain issues are less right than others.

7.3 Mini-game and test example

As an example, we use the intro module in Appendix D1 to present how the quiz works. The learning goal for the introduction is that the "Student (player) should be familiar with the rules of the game and the way it is structured".
As can be seen in figure 18 the players are presented with different information on how the game works.

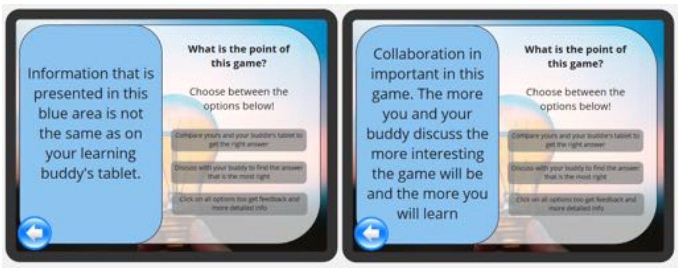


Figure 18: Asymmetrical information

In the quiz, more clearly viewable in figure 19, the two faulty options provide more information on how the game should be played. In this example, the feedback adds information on how the players could sit to not be able to see each other's tablets.

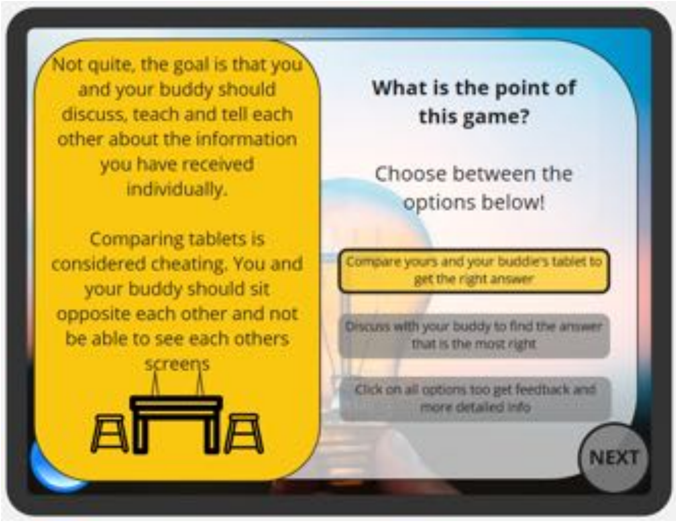


Figure 19: Feedback

After the mini-quizzes, the players enter the test based on what they have just learned, see figure 20. The test is scored and the players answer individually. If they answer correctly they receive a good score and if both answer the same then a bonus.

Figure 20: Scoring test

8 Prototype testing

Rather than developing a prototype that tests out every aspect of our conceptual design, time constraints and pandemic considerations, made us focus the attention of the prototype towards testing and evaluating the asymmetrical game mechanics. Future research and testing could complement this by thorough prototype testing and incorporating the details of the learning theory developed in this paper. Some considerations regarding prototype testing will be represented in this section.

8.1 Purpose

There are several techniques for building in usability. We have already found our focus group, sent out user study, built an idea based upon loads of research and now created a prototype. The next step in the process is to test the prototype in order to make sure it is intuitive and that it meets the criteria of a human-centred design. Rubin and Chisnell (2008) emphasize iterative testing on a product intended to expose possible usability deficiencies and gradually improve the prototype over time. This kind of testing would have to be done in order to go from a prototype to a first release of the product.

8.2 Our prototype

There are several things that could be focused during testing of our prototype, the more important one would be the use of asymmetrical gameplay and collaborative learning. Several design choices could also be looked into during the testing, things such as colour coding, button placements, icons and font size.

8.3 Testing the prototype

Testing of our prototype has been prepared in the form of PowerPoint slides where several screens can be swapped between to simulate a working prototype. Testing should include two people, preferably in the age interval of our intended users. During testing a test moderator along with someone taking notes and perhaps someone filming the test would be preferable to gain as much information as possible, these are roles suggested by Rubin and Chisnell (2008). Testing results vary between different test participants which is why a bigger sample size of participants might give a better understanding of what should be changed.

However, when evaluating the prototype test it is important to be aware of the complex situation that is created. Thus, there are limitations to how generally representative an evaluation is. The socio-technical system includes two tablets and two students, each interacting with the game in front of them and also interacting with one another. Every single student interacting with the game contains – roughly said – the game as a controlled variable and the student as an uncontrolled variable. Since the students should also interact with each other, the situation gets very dynamic, because two uncontrolled variables are now interacting as well. As Schwartz (1999) mentions in his paper about collaboration, small group behaviours are challenging to analyse: “Unlike sociological data, small group behaviours are not smoothed by averages over the thousands of people involved in a social movement. And unlike individual psychological analyses, the multiple interacting agents of a group cannot be turned into the simple equation of a single individual interacting with a single stimulus.” (Schwartz, 1999, p. 5) That behaviour might emerge dialectically and might not be causally traced back, must be taken into account in that case.

8.4 Testing the learning theory of the prototype

In this section, we will focus on how the applied learning theories of the prototype could be tested. In our theory chapter, we distilled four different ‘points’ that we wanted to implement in our design. These were collaborative learning, learning by teaching, motivation and immersion, and a balance between entertainment and reflective learning, and feedback.

The testing should include subjects at the age of our intended users. The theoretical points could be tested in different ways. The motivation and immersion could be tested through post interviews where the subjects give their opinions on their experience of these parameters during the testing.

The collaborative learning and learning by teaching should be qualitatively evaluated through observations by an observer. Here parameters like; what is the level of interaction, discussion, and negotiation during the game, should be noted.

When testing formal and informal feedback the overarching difficulty of controlling the experiment arises. The approach could be a quantitative method where thirty subjects were presented with content without formal feedback. Thereafter, they would be presented with similar content but including formal feedback. In the end, statistical descriptions of the data could be made, and a regression analysis could be performed in order to see if there is any correlation between our formal feedback or without. It is important, however, to remember that a low R^2 would be accepted due to the nature of social science. Testing the informal feedback is even more difficult. A way of doing it could be conducting qualitative observations of the experimental setting.

In order to test the actual content, we need to have an awareness of the learners’ baseline knowledge of the content. This can be measured through an interview or a test. After the game has been completed, we compare their new answers to their old ones and assess if there has been any assimilation of knowledge.

8.5 Prototype testing conclusions

All these test-principles should be used to get as good of information as possible which is why we have decided to not perform this test due to time constraints and Covid-19. Instead, we leave this task to a possible successor of our work.

9 Discussion

There are a lot of points and threads to follow up on relating to what has been presented in this paper. These include both concrete results and suggestions about future work as well as more theoretical discussion points in relation to cognition and digital learning. Some of the points that stood out to us, and which notability we wish to relay and discuss, are in short presented below. These points are also tied in with suggestions so they can be explored in future work.

9.1 Summary

In this paper, we have given a description of the process of designing a learning game that had two distinct purposes: (i) to better the understanding about environmental questions in school children attending grade 7-9 in the Swedish school system and (ii) to design a learning game in such a way that this environmental knowledge can be learned in a way that produces 'deeper learning' about the subject. To reach this level of description we have gone through a number of steps. Here is a summary of the main points.

This process was based on a human-centred design perspective where the need of the user was taken into account. This resulted in us performing a user study with students in grade 7-9 in the Swedish school system. The goal of the user study was to get the opinions of the students on topics relating to gaming and to environmental questions. This with the aim to learn more about the actual user of the learning game.

In parallel with collecting the data from the user study, we looked into the theory regarding learning in general and regarding digital learning games specifically.

The combinations of these two sets of data lead us to the concept of 'The Green Island'. The concept of the green island is that of a learning situation where two players, using different sets of information, have to collaborate to achieve the goals within the game.

The result of this was a theory-driven paper along with a prototype (see appendix D) which, together, aim to illustrate this novel attempt to design a good learning situation which can accommodate for deeper learning in regards to environmental questions.

As different factors, such as time and the ongoing situation regarding the Corona pandemic, hindered specific aims in the testing and evaluation process of the prototype we decided to instead document as much as possible in this paper and prepare it for future work and research.

We believe that the cooperative asymmetric game mechanic described in this paper gives rise to an interesting learning situation and learning process which could be explored further. Some of these points of interest are explored in the section below.

9.2 Process reflections

The process that we have used to come to the conclusions and the results that are presented in following sections can be seen in figure 21. Much of the work has been done through workshops, see workshop and process sheet in appendix E.

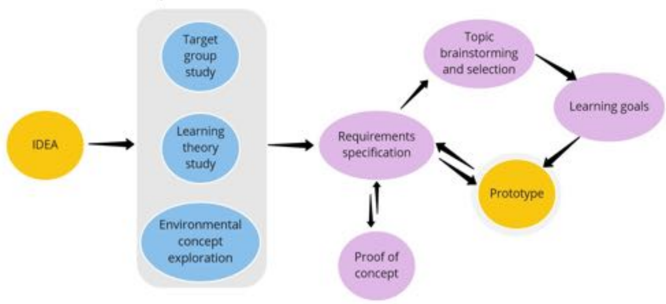


Figure 21: Process summary

From the initial idea of creating environmental learning, we did general research on the target group, learning theories, and exploring the environmental concept that we envisioned for the game. The result of these was synthesized into a requirements specification and the practical design process initiated after. By doing these preparations we aligned the purpose and goal of the project through researching the areas relevant for the vision of the initial idea. From the initial vision and the requirements specification, we continued using this value central process management and prototyped a proof of concept, chose the topic for the prototype, created learning goals, and designed the prototype. Throughout this work, the requirement specification was iterated and updated to further align with the concept of this specific game.

Finally what we find is that what has been created here is not only an environmental game. Instead, we find that the prototype presented in this project is a representation of a framework for asymmetric collaborative learning. Within this framework, using a formalized structure of the prototype any content or difficulty level can be presented, not only an environmental one.

9.3 Results reflections

Essentially the general results collected in this paper and the appendices can be formulated as follows:

- Requirement specifications for The Green Island and recommendations for how to develop further
- Modulated framework for an asymmetric collaborative learning game
- Environmental game with much content that can be extended further
- Testable prototype illustrating the green island

Concrete results are presented through the creation of the prototype in regards to the requirements documents. Below we compare the prototype and its features with the set of requirements listed in the Software Requirements Specification document (see appendix B for the full list of requirements). As these requirements are a formalized embodiment of both the empirical findings from the user study (section 2 of this paper) and of the theoretical background relating to the learning theories (section 3 of this paper) they can work as a check-list of how, and to what degree, the empirical and theoretical results were implemented in the prototype.

Some of the more notable or broader points of comparison will be mentioned here.

Regarding the Initial requirements (section 3 in appendix B) the following observations can be made. In the sub-section ‘content design’ (section 3.1 in appendix B) its states as requirements, that the learning game should be played by two players and be based on learning by teaching and collaborative learning and that the in-game decision processes teach the user about the impact of action towards a sustainable environment. All of these points have been accommodated for in the prototype in the sense that it is divided up into two “screens” for each player and the information is presented in such a way that neither player has access to all information. But it is worth mentioning that to actually see how well the learning by teaching and the collaborative dimension of the game works would require observation in an experimental setting.

Section 3.2 in the requirements document states, among other things, that the game should be simplified but based on reality and that in-game decision processes reflect reality at various societal levels of western society. By choosing the in-game scenario of warranty and electronic waste for the prototype these requirements can be said to have been accommodated for.

Regarding the requirements for the visual part of the game (section 3.3 in appendix B), we can note that the first requirement, that the game theme should have a realistic theme, is met in the sense that we opted for photographs and a composite of satellite images for constructing the graphics rather than going the route of hand-drawn graphics. The second requirement relating to the visuals, that the game should give the player visual feedback to choices being made is not accommodated for in the prototype in any deeper sense.

Moving on to an evaluation of how well the more specific requirements pertaining to the game’s structure were implemented in the prototype the following observations can be made. The structure requirements (section 4.1 in appendix B) dictates, inter alia, that the workings of the game should be clearly presented to the player at the start of the game and that the main game screen consists of a representation of the island. Both of these requirements were fully implemented in the prototype, the first by presenting the player with a tutorial of the game at the start, the second by having a rendering of the Green Island at the main screen workings as a hub for choosing the type of content to engage with (e.g. societal, forest or ocean-related tasks). Requirements only partially implemented includes granting the player with an extra reward at the end of modules will be given for successful cooperation, that the island’s progression should be clearly presented to the user as they progress, and that “Tips” should be presented to the user should be relevant to how the game should be played. The first two points were only partially realised in the prototype as we did not implement a scoring system for generating another level of feedback to the players. The “tips” requirement is partially met by the inclusion of the tutorial module, but a help system that can provide “tips” for the player at any stage in the progression of the game was not implemented. As scoring was not implemented the specific requirements relating to it was not fulfilled.

Judging the technical requirements (section 4.2 in appendix B) produces some difficulty as the game was not implemented as an executable program but as a sequential set of pictorial representations of each game screen simulating the flow and progression of the game. But the dimensions and aspect ratio of the images representing each game screen were chosen to reflect and comply with the dimensions and aspect ratio of the most common tablets currently on the market and used in the Swedish educational system.

9.4 Reflections on game mechanics: asymmetric collaborative learning

The aim of the project is to implement collaborative learning through asymmetric game mechanics. By splitting up the presented information on two different screens we want to engage the players in communicative exchange and discussion. It should not be possible to solve the quizzes in a first try as a single player. However, the current set up of the quizzes is “Player 1 gets information A, Player 2 gets information B, the correct answer contains A + B”. If player 1 tried to solve the quiz on their own, they could also scan the answers, looking for the option where information A is found with some little extra info and infer that this is probably the right answer containing the extra info that is found on the other screen. In order to really make the players collaborate, exchange information and avoid recognition, one could possibly also work with rephrasing the given information.

Another issue that we face in our educational system is that the students act as peers, student-teacher, and teacher-student. This occurs while they both remain, students, all the time. As creators of the ES, we can control the formal feedback but not the informal feedback between the students. The formal feedback being the feedback they gain from the game and the informal one being the discussion between students. The informal feedback can’t be controlled entirely because we want the students to discuss without much guidance.

9.5 Future work

In the theory chapter, we discussed what learning was in general and what positive effects and implications games can bring to the field of learning. But how does our vision of a collaborative environmental learning game correlate with these theories? One of the main issues is ensuring that the students remain within their zone of proximal development. This requires previous knowledge and data concerning the individual students' academic level to ensure that the content in the next module is adjusted to fit the students' individual needs.

One way to achieve this in future work is through a program architecture that is able to collect the output of the individual student, assess it, and provide new altered input from a database of the content of different difficulties. This iterative process is then continued throughout the modules. A possible inspiration for future work can be found in the work of Grawemeyer et al. (2017) "Affective learning: improving engagement and enhancing learning with affect-aware feedback". In this article, an architecture for an adaptive system is presented. This system could in future work be included in the creation of an adaptive system for the Green Island design. Of course, to implement this is an ambitious project which would take a lot of development time and resources.

Furthermore, future work could test whether our theoretically based approach for creating an environmental game suitable for deep learning is actually achieved. When creating an experimental setting where two players can interact with the game and each other, one should try to get them set up as close as possible to the actual learning environment. Under section 4.2.1 we have given a description of the learning-situation as a socio-technical system. A *post-factum* analysis of this situation based on the theory of Hutchin's socio-technical system could be enlightening. Did the description in section 4.2.1 give an account of the full learning situation? Did any unforeseen feedback processes emerge between the two participants, within the created distributed system? Utilising this model one could set up hypotheses about interesting patterns of information flow that might emerge in the learning situation.

Moreover, to analyse the learning situation based on a zoomed out aggregated socio-technological perspective can have its advantages. With this shifted focus the occurrence of "gaming the system" in the learning situation can instead be described as the result of the non-optimised design of the socio-technical model's architecture. If the system design allows for gaming the system as a way of reaching the system's end-state goal this may, realistically, very well be methods that are utilised if they lessen the work-load of the socio-technological system. The blame, in this case, is not on the students, but on bad or non-optimised design choices made by the developers and/or flaws in the settings surrounding the learning situation.

Furthermore, the module structure of the game leaves much room for filling it with actual content. The prototype we worked on should exemplarily show how a submodule could look like. Since there are six modules besides the waste module and each can include several submodules, the current structure contains an immense potential that can be built up in the future. For the sustainability-related content to be created the project would need to include someone with expertise in the subject matter, i.e. someone with broad and deep knowledge about environmental issues and how human behaviour affects and interacts with nature.

10 Conclusion

Since our planet and consequently humankind is facing climate change and future generations are those that will be affected, we wish to give a more solution-oriented awareness of the environment in the early stages of school learning. The goal of the project is to design a digital learning game in sustainability using a human-centred design process.

The results of this report include a game design that combines research from multiple fields, with a focus on learning theories and design and are the following:

- A user study on the target group for the game has been performed and provided knowledge on target group preference on games and the environment.
- The literature study on learning theories resulted in the idea to implement asymmetric collaborative learning theory into the game.
- The learning game conceptualization resulted in the general concept and the game idea and purpose is combined from the empirical and literature research.
- The requirements creation combines the previous results into a specification that allows creation and effective iteration of the game design as well as a concretization of the game structure.
- The prototype design included a proof of concept, choice of topic for prototype development, along with learning goals that will be used to create knowledge content in the game.
- The prototype is thereafter created using the previous results as well as iterated requirements and design decisions. The game prototype contains 4 modules presenting a walkthrough from the introduction of the game and the general environmental knowledge into the more detailed.

The game structure with mini-games and modules allow for testing other pedagogical approaches as well along with exchanging the content to include any kind of knowledge. The project was initiated as an environmental game creation using modern learning theories that now can be extended as a full framework that can include any learning theory or any learning topic. Our prototype is created using environment and sustainability as a topic with cooperative asymmetric learning as the pedagogical approach. In further development of this game, we recommend extending the details of the requirements, always creating learning goals before the development of modules as well as iterating ideas as soon as possible saving time in the overall design process.

References

Chen, S. Y., & Liu, S. Y. (2020). Developing Students' Action Competence for a Sustainable Future: A Review of Educational Research. *Sustainability*, *12* (4), 1374.

du Boulay, B. (2020). Jim Greer's and Mary Mark's reviews of evaluation methods for adaptive systems: A brief comment about new goals. *International Journal of Artificial Intelligence in Education*, 1-14.

De Bra, P., Vasilyeva, E., Pechenizkiy, M., (2008). Adaptation of Elaborated Feedback in e-learning

Ergonomics of human-system interaction: part 210: human-centered design for interactive systems (2010). International Organization for Standardization (ISO), Geneve, Switzerland. Retrieved from: https://www.iso.org/standard/52075.html

Gee, J. P. (2009). Deep learning properties of good digital games: how far can they go?. In *Serious games* (pp. 89-104). Routledge.

Grawemeyer, B. et al (2017). Affective learning: improving engagement and enhancing learning with affect-aware feedback.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive science*, *19*(3), 265-288.

Iten, N., & Petko, D. (2016). Learning with serious games: Is fun playing the game a predictor of learning success?. *British Journal of Educational Technology*, *47*(1), 151-163.

Jensen, B. B., & Schnack, K. (1997). The action competence approach in environmental education. *Environmental education research*, *3* (2), 163-178.

Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and cognition*, *15*(4), 673-692.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lee, J. J. & Hammer, J. (2011). Gamification in Education: What, How, Why Bother? Academic Exchange Quarterly, 15(2).

Linderoth, J. (2012). Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming & Virtual Worlds*, *4*(1), 45-62.

Løw, O. & Skibsted, E. (2012). Psykologi for lærerstuderende og lærere 1. Udgave, 3. Oplag. Akademisk forlag, København

Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review. 63 (2): 81–97.

Molich, R., and Nielsen, J. (1990). Improving a human-computer dialogue, Communications of the ACM 33, 3 (March), 338-348.

Nielsen J. (2020, November). 10 Usability Heuristics for User Interface Design. Nielsen Norman Group. Retrieved from: *https://www.nngroup.com/articles/ten-usability-heuristics/*

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological review*, *84*(3), 231.

Norman, D. (1993). Things that make us smart: Defending human attributes in the age of the machine. Perseus Books.

Norman, D. A. (2013). The psychology of everyday things. Basic books.

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing how to plan, design, and conduct effective tests*. Indianapolis, IN: Wiley Pub.

Schwartz, D. L. (1999). The productive agency that drives collaborative learning. In P. Dillenbourg (Ed.), Collaborative learning: Cognitive and computational approaches (pp. 197-218). NY: Elsevier Science.

Skolverket (2018). Curriculum for the compulsory school, preschool class and school-age educare. Retrieved from: https://www.skolverket.se/publikationsserier/styrdokument/2018/curriculum-for-the-compulsory-school-preschool-class-and-school-age-educare-revised-2018

Data Literacy - The Rescue of Dataville

Principles of an educational game for teaching data literacy to grades four through six.

Mårten Bengtsson, Andreas Berntzon, Malin Cronquist,
Joel Jonsson, Michael Rihovsky

The abundance of information that each person meets on an everyday basis is growing at a fast pace. This development has led OECD, in their 2030 Learning Compass, to include data literacy, the capability to read, explore, analyze, and argue with data, as a core competency across the curriculum. Our report describes the development of and theoretical framework for an educational game aimed at providing students aged 10 through 12 with the cognitive skills fundamental to data literacy. With the Swedish National Curriculum as a starting point, using theories in experience design, educational technology, and cognitive science, we created an educational game revolving around several mini-games, set in the fictional world of Dataville.

1 Introduction

More and more data is created every day and presented to us in various forms and media. It is an inundation that does not appear to be going away anytime soon. The general public is exposed to staggering amounts of numbers and graphical representations on a daily basis (Ridsdale et al., 2015). In a study from 2016 (Börner et al., 2016), in which young and adult museum visitors were tested for their ability to read and understand visual data, e.g., charts, graphs, maps, and network layouts, it was found that said ability was poor across the board. This was despite most of the visitors seeing themselves as having an interest in the areas of math, natural science, and social science. All of the participants also reported encountering similar data visualizations regularly in their life, whether at work, in books, on the internet, in the news, or in school. This also demonstrates the rise in the number of sources that produce data (OECD, 2019). It has also been found that students generally do not pay attention to graphs without being prompted to do so (Shreiner, 2019). Additionally, a study found that people often disregarded graphs when there was corresponding text (Chevalier et al., 2018). When they did look at it, they mostly appreciated its aesthetic qualities whilst overlooking its meaningful content.

Digitization has further led to children being exposed to this abundance of information which necessitates the development of capacities to navigate the flow of information early in life. These developments are elaborated on in the OECD Learning Compass for 2030 where it is established that the skills necessary for data literacy, defined as the abilities to read, work with, analyse, and argue with data as well as to derive meaningful information from data, will be essential for all people in the current decade (OECD, 2019). To be data literate, according to the OECD report, further includes the comprehension of "what data mean, including how to read charts appropriately, draw correct conclusions from data, and recognise when data are being used in misleading or inappropriate ways" (OECD, 2019) (Carlson et al., 2011).

This paper proposes an educational computer game, based on the Swedish curriculum in mathematics, social science, and the natural sciences, to attempt to provide students in grades four through six with the cognitive skills fundamental to data literacy. The game aims to be both engaging and educationally relevant for the said age group. In the following pages, we describe our exploration of the concept of data literacy, how we developed our working definition, the theoretical background, and our brainstorming, storyboarding, and iterative design processes.

2 Approaching Data Literacy

Literacy and numeracy are defined by the Oxford Dictionary of English as "the ability to read and write" and "the ability to understand and work with numbers," respectively (Dictionary and Street, 2019). But as Boy, et al. (Boy et al., 2014), write, this definition is no longer "tied to its textual dimension." The meaning of having numerical and literary abilities is instead now related to a broader concept of information processing. It means being able to apply these abilities in order to orient oneself in a digitized world.

This is expressed in extended definitions of literacy and numeracy such as digital literacy, computer literacy, information literacy, and media literacy (Boy et al., 2014). In general, these skills can be summarized as the abilities to use computers, media, and other knowledge resources effectively, being able to analyze, evaluate, and put to use the information one comes across.

Though there are many and wide definitions of the concept of data literacy, the basis of an understanding of the larger theoretical definitions begins with statistical literacy (Gould, 2017). As such, we find a logical place to start with young children is a grounding in statistical literacy and proportional reasoning.

Educational Technology researchers Joan Garfield and Dani Ben-Zvi propose that *statistical literacy* is "a key ability expected of citizens in information-laden societies, and [that it] is often touted as an expected outcome of schooling and as a necessary component of adults' numeracy and literacy" (Garfield and Ben-Zvi, 2007). Statistical literacy involves understanding and using the basic language and tools of statistics, i.e., "knowing what basic statistical terms mean, understanding the use of simple statistical symbols, and recognizing and being able to interpret different representations of data" (Garfield and Ben-Zvi, 2007).

Garfield and Ben-Zvi further state that statistical literacy should be kept apart from related terms (Garfield and Ben-Zvi, 2007). For example, it is defined separately from statistical reasoning, or reasoning with statistical ideas, interpreting statistical results, and explaining statistical processes. It is also separate from statistical thinking, the advanced statistical reasoning utilized by researchers within the field of statistics, e.g., evaluative methodology (Garfield and Ben-Zvi, 2007).

When exploring the concept of data literacy, two adjacent concepts become relevant: mathematical literacy and visual literacy.

Mathematical literacy is framed by the Program for International Student Assessment (PISA) as "learners [being able to] identify and understand the role of math in today's world" (for International Student Assessment and service), 2006). Developing mathematical literacy has also been framed as a way to combat emotional issues, such as math anxiety (Niemi et al., 2018). *Visual literacy* is defined as "the ability to use well-established data visualizations, e.g., line graphs, to handle information in an effective, efficient, and confident manner" (Boy et al., 2014).

It becomes interesting to relate visual literacy to data literacy due to the prevalence of what has been termed an *information-telling perspective* in graphical design in which it is believed that "any technically correct graph is adequate for conveying relevant quantitative information" (Shah et al., 1999) (Bereiter et al., 1982). Shah, Mayer, and Hagerty (Shah et al., 1999) found, on the contrary, that when individuals studied a graph used in a social science textbook, different visual presentations of the same quantitative information affected the meaning extracted by the reader of the graph. In their study, they also detailed two main cognitive modes of interpreting a graph:

- A relatively simple pattern perception and association process in which viewers can associate graphic patterns to quantitative referents.
- A more complex and error-prone inferential process in which viewers must mentally transform data.

They further found that graph comprehension could be improved by "minimizing the inferential processes and maximizing the pattern association processes required to interpret relevant information" (Shah et al., 1999). They also found further evidence for visual chunking, e.g., having relevant quantitative information close together as in a bar graph, being conducive to understanding trends in data when reading a graph.

2.1 Approaches to teaching Data Literacy

Using interviews with statistics professors, Garfield (Garfield, 1995) summarized several guidelines and objectives to be followed when aiming to teach data literacy. These refer to all the knowledge the professors wanted to convey to their students aside from the ability to solve mathematical problems. Even though the focus lies mainly on university students, the principles apply to the research goals of the game development at hand. Some of the guidelines and objectives are as follows:

- Fundamental statistical knowledge
- Statistical knowledge is for everyone and anyone.
- Learning statistics means being able to communicate and reason with it.
- A statistical problem can be solved in different ways.
- This means that people may come to different conclusions based on the same data depending on assumptions as well as methodology.

Garfield (Garfield, 1995) is critical of the way that students in general learn to solve only certain types of problems, or use only certain conceptions of problems, without learning the underlying principles or being able to connect them to the guidelines and objectives mentioned above. She (Garfield, 1995) brings up *statistical fallacies* as a concept important to the data literacy skills of students, and mentions how the use of statistical fallacies seems to vary with context; the shifting of context changes the ability to reason with statistics. This furthers the argument for the need to provide students with a general sense of statistical reasoning, as well as to connect the context-specific use of data, as in the summary of surveys in popular media, to a broader context.

As a possible way to provide this knowledge, Garfield (Garfield, 1995) (Garfield and Ben-Zvi,

2007) cites activity-based courses and group exercises. These, she holds, seem to be useful when learning statistics, especially with regard to statistical fallacies, as they provide students the opportunity to apply different formulations and conceptions of a problem in order to either reach the right conclusion, or to broaden their understanding of the problem.

D'Ignazio is a teacher, software developer, and researcher in data literacy who developed five areas of focus for teaching data literacy to children with little technological or mathematical background or practical use for data literacy(D'Ignazio, 2017). One is to make data messy, which forces students to sort the data themselves as well as to think about problems with the categorization of raw data. A related focus is the writing and discussion of data biographies, i.e., the contextualization of data as to how, why, and by whom it was collected. This is often missing when data is presented in the news and social media or in Google searches. As D'Ignazio writes, this situation poses a challenge for new learners. This is because new learners tend to see information organized systematically in a spreadsheet as "true" and complete, especially if the data is not noticeably missing values (D'Ignazio, 2017).

This is mirrored in other approaches for teaching data literacy. For example, in Carlson and Johnston's book, *Data Information Literacy: Librarians, Data and the Education of a New Generations of Researchers* (Carlson and Johnston, 2015), they summarize 12 of what they call *data information literacy competencies* that mirror the data life cycle, some of which are: discovery and acquisition of data, data visualization and representation, data management and organization, data quality and documentation, metadata and representation, and cultures of practice.

3 Learning

Digital technologies are used more and more in- and outside of the classroom. Students need to be taught foundational data literacy skills as a consequence, and a game is one potential way of doing that. To this end, the following are taken into account: the classroom environment, evidence-based teaching practices, and the Swedish curriculum for grades four through six in mathematics, social science, and the natural sciences as they relate to data literacy.

3.1 Classroom Environment

Digital devices are becoming increasingly incorporated into the classroom. At the time of writing, people are sitting at home with their devices due to the worldwide spread of the coronavirus. The trend is clear that more and more education takes place with the use of digital media. In the classroom, computers and tablets are used both for the purpose of instruction and recreation.

One of the authors of this report worked as a teacher in California, USA, with students with special needs ranging in age from five to 22 years in various classroom and community settings. With every age group, digital technologies were used for both instructional and recreational purposes. Often, computers were used as supplements, primarily for math instruction, and games were played that had some instructional basis, though preferably somewhat veiled to maintain student engagement. The curriculum was regularly enhanced with digital platforms, e.g., used after traditional instruction to reinforce prior learning.

From personal experience, said author found there to be an overstimulating presentation of information surrounding students. Posters taped and stapled to the walls, though with the best of intentions on the parts of the teachers, often reached a point of oversaturation that led to the posters appearing almost as wallpaper, and giving as much information as such. Students were seen to be forced to adapt, blocking out all the potentially helpful diagrams and charts to simply be able to focus on any one thing at all. This blocking out of diagrammatical stimuli can lead young students to ignore graphical representations in textbooks as well. However, when students work with their own context-specific graphical representations, their awareness of them can be increased (Luís, 2018). An aspect of the aim of the game proposed here is to increase student awareness of graphical representations. If they don't pay attention to them, they can't gather information from, nor critically appraise, them.

3.2 Evidence-Based Practices

There are a number of evidence-based practices that have become go-to's for many educators. Cross-curricular teaching emphasizes the importance of not isolating subjects from one other. The more subjects melt into each other, the more students are exposed to topics in various formats and environments, which can lead to deeper and increased learning. This kind of more integrated approach strives to break free from traditional compartmentalization, and is therefore theoretically more in line with how children actually learn (Kerry, 2015).

Cross-curricular teaching is similar in theory to the use of multiple modalities in instruction. Multiple modalities employ the use of diverse presentations and experiences when presenting information to students. Students learn in different ways and can benefit from being exposed to the same content in a variety of ways (Puccini et al., 2013). For example, students with an abundance of energy may benefit from jumping up and down during a counting lesson. Another example would be doing a science lesson on plants in the outdoors, allowing students to physically touch the plants they are learning about.

Another way to increase students' learning is to boost their motivation. This can be done in many ways, including using both student choice and interest (Patall, 2013). When students have a choice in the selection of learning materials and methods, they can be more engaged and take a larger stock in their own learning. This can be as simple as giving students the ability to choose which activity to do first, or just the choice of what they will do for their break if and when their task is completed. Incorporating students' interests into instruction can increase their motivation and engagement. An example could be using a topic of interest such as superheroes or cartoons for counting. However, while relatively simple to incorporate into special education settings where student-to-teacher ratios are small, the application of such strategies is difficult in larger general education settings.

Though it would conceivably be ideal to have students actively engaged and motivated in the content being taught, the reality in the classroom is often something different. If increasing the inner drive and motivation to learn is neither realistic nor desirable, then one can at least ask for student engagement; that students are taking part in learning activities, paying attention, though not necessarily enthusiastic and impelled to do so.

Realistically, the proposed game seeks to engage students. Interest in the storyline and the three minigames will hopefully lead to more engagement and more learning.

Many young students enjoy using screens (Papastergiou, 2009). Instructional games with an aspect of age-appropriate interest, e.g., sports, RPGs, and arcade styles, are inherently more interesting and engaging for many students than are pencil and paper tasks. Due to this, there is often a slight blurring of the lines for students about what constitutes a break and what constitutes instruction. It can be easier to engage students with the material in this way, and thus more time can be spent in the interaction.

There is evidence for the importance of breaks for student learning (Jensen, 2000). Many benefit from physical activity throughout the day as a means of expelling excess energy accumulated from sitting still at a desk. In special education settings, breaks can also take the form of screen time. This does not have to be framed negatively. If breaking up instructional time using screens leads to increased student engagement and learning before and after the break, the function of the break is still achieved.

Another way for students to learn, according to Gee, is to explain and interpret experiences together with others (Gee, 2009). He believes it is important to use previous experience in new, though similar, situations. This can be implemented in games, for example, by asking the same question but changing the criteria it depends on.

3.3 Swedish Curriculum

The Swedish national curriculum (201, 2019) for grades four through six for general education students, that are relevant to data literacy, are described in the following sections. They were pulled from the areas of mathematics, social science, and the natural sciences. The language used often sounds much more complex than the reality of the classroom, and whether these expectations are realistic for children of this age is subject to debate. That being said, they nonetheless form the basis for the educational content of the game.

Mathematics

In mathematics, students are expected to learn numbers in both fractional and decimal forms, as well as how to apply them in everyday situations. Regarding probability, they are expected to learn chance and risk assessment based on observations and simulations as well as statistical material taken from everyday situations. Additionally, they are expected to learn the probability of the outcomes of different randomized trials (201, 2019). Simple examples of everyday chance are the likelihood of a four when rolling a die, or heads when flipping a coin.

Expectations for graphical learning include using diagrams and charts to describe the results of surveys and tests, both with and without the aid of digital tools. Students are also expected to learn to interpret the data presented in charts and diagrams. Furthermore, they are expected to be able to come up with strategies for mathematical problem solving in everyday situations, as well as to formulate mathematical problems based on said situations. Finally, they should understand and be able to use the statistical concepts of mean, median, and mode (201, 2019).

Social Science

In the curriculum for social science, general education students in grades four through six are expected to have a general understanding of the dispersion of information, advertising, and opinion formation in different forms of media. They are expected to be able to separate the message from the sender and their intent in all forms of media. They are to consume media with a source critical point of view. They are also expected to know how to responsibly handle media based on legal, ethical, and social principles, as well as to understand the impact that the growth of digitalization has, and will have, on the individual and society, e.g., the increased ability for communication (201, 2019).

Natural sciences

As regards the natural sciences, students are expected to be able to understand simple systematic research, i.e., the process of planning, executing, implementing, and evaluating. They are to be able to represent simple research with charts, graphs, images, and simple written reports, with and without digital tools. They are to cultivate the ability to interpret and examine, or review, information with regard to chemistry, biology, and physics, e.g., how information is expressed in newspapers and digital media (201, 2019). For example, students are to search for scientific information in different sources and provide reasoning for the source's usefulness.

3.4 Cross-disciplinary approach

The aspects of data literacy education focused on in the game are mirrored in the national curriculum in at least three disciplines: the natural sciences, social science, and mathematics, elaborated on in section 3.3.

Our focus on the secondary effects of statistical literacy, more specifically on communication and argumentation, is merited by its proven effectiveness in prior research.

In the *Thinking with data* approach (TWD), researchers "take seriously the fundamental requirement that data literacy bridges disciplinary domains" (Vahey et al., 2012), and in so doing has cross-disciplinary applications in math, social science, the natural sciences, and English. Vahey et al.(Vahey et al., 2012), the originators of the TWD approach, also found some core abilities that were teachable and applicable across the aforementioned fields, namely: "formulating and answering data-based questions; using appropriate data, tools and representations; and developing and evaluating data-based inferences and explanations" (Vahey et al., 2012).

Vahey et al. also interviewed seventh grade teachers in the above-mentioned fields, who found it possible to integrate real-world complex data into their lectures. The real-world examples that are used in these different classes all provide different examples and contexts that develop general abilities connected to data literacy. The kind of understanding of data that comes about through the TWD approach, it is argued (Vahey et al., 2012), also gives the student a more global understanding of data literacy. As such, data literacy is not only needed to understand material across disciplines, but the learning process can also have synergistic effects that enhance data literacy skills overall.

Vahey et al. also point out that seventh grade teachers found it useful and applicable to use real-world complex data in their teaching (Vahey et al., 2012). This brings the focus to an interesting place, as using authentic problems to formulate mathematical problems seems to be conducive to the grasping of solutions. This is something that has been studied before, e.g., within the *Realistic Mathematics Education* approach (RME) (Gravemeijer and Doorman, 1999). In the RME approach, teachers use the actual, outside of school, experiences and contexts of students' lives.

3.5 Proportionality

The RME approach could in part be contrasted with the idea of *mathematizing*. Mathematizing is the experiential understanding or approach of constructing a math problem from the ground up. This could also be expressed as an experience of the formation of data. In the TWD approach, students need not to rely on their direct experience.

According to Vahey et al. (Vahey et al., 2012), an important aspect of data literacy is the understanding of proportionality, which they argue is greatly lacking in the general population, both students and teachers alike.

In their cross-disciplinary approach, they argue for the importance of compound measures when students are learning data literacy, as they are fundamental to both making predictions and comparisons as well as for the art of argumentation.

They specifically mention proportional transformation "of data from raw values (such as the total amount of water used by a set of countries and the populations of those countries) to a measure that combines two quantities" (Vahey et al., 2012), an example of a per capita measurement.

3.6 Concreteness Fading

The *concreteness fading* approach was developed within the fields of psychology and education and recommends beginning with concrete materials when teaching mathematics and science, before gradually moving toward more abstract forms.

According to Fyfe et al. (Fyfe et al., 2014), the concreteness fading theory states that mathematics education should focus on four areas:

1. Helping learners interpret ambiguous or opaque abstract symbols in terms of well-understood concrete objects.

2. Providing embodied perceptual and physical experiences that can ground abstract thinking.

3. Enabling learners to build up a store of memorable images that can be used when abstract symbols lose meaning.

4. Guiding learners to strip away extraneous concrete properties and distill the generic, generalizable properties.

The concreteness fading approach has its roots in *constructivism*, which has been a leading theory in the field of pedagogy since Piaget developed it in the 1950s (Piaget, 1970). A central point brought forward by Piaget was that children should be educated in a naturally progressive manner, building on and moving from knowledge about concrete objects and experiential content toward the more abstract.

A game focused on teaching elementary school students visual literacy was developed by Alper et al. (Alper et al., 2017). Their game was built on 2,600 visuals used in American schools grades K-4 in textbooks that adhered to the US Common Core Standards (See a synopsis in Figure 1).

They divided the material into five categories: tangible objects (or manipulatives), photographs of manipulative objects, illustrations of everyday objects, abstract shapes, and spatially organized notions. They found the concreteness fading approach applicable when developing their game. Even though they found that the concreteness fading approach was already utilized in the textbooks, they found that elements within their game design could enhance the connection between the different levels of abstraction, more specifically via animated transitions and brushing and linking, i.e., coordinating multiple representations of the same data to overcome the limitations of the single representation. We use the same feature in our first mini-game where the player helps a research organize research data simultaneously as the data is being transformed into a two graphs. Furthermore, Alper et al tried to mimic direct manipulation of the objects, allowing for variability in the level of input and the ability for customization of visuals and data. As they write, “The diversity in visuals and themes keeps children engaged, and helps them transfer their knowledge to different contexts” (Alper et al., 2017). It should be noted that no evidence is claimed for the game’s pedagogical utility. However, their methodology and theoretical focus were of use for us in the development of our game, primarily due to similarities in focus and basic ideas.

Figure 1: Figure from Alper et al. (Alper et al., 2017) detailing the synopsis of different forms of data presentation found in American textbooks for grades K-4.

As Garfield notes, students learn mathematics better by working on open-ended problems (Garfield, 1995). We had this in mind when developing the design of the game and mini-games. The student is an active participant in the construction of the solution in the first mini-game, and the created graph is then used in the second mini-game. In the second and third mini-games, the focus lies on what could be thought of as the secondary effects of statistical literacy, i.e., how information is communicated and collectively made sense of.

In later editions of the game, we see the possibility of this part of the game becoming more open-ended, where no definitive answer or feedback will be provided. Instead, the player will have to rely on their own judgement of what is correct, or what is the most effective strategy.

3.7 Engagement Theory

Engagement theory is an increasingly popular theory within the learning sciences that parts with earlier development within the field by taking a broader approach to students as situated within a learning environment. The reason for its widespread use is its effectiveness shown in addressing perennial problems in education such as low achievement rates and negative affects, like feelings of boredom (Fredricks, 2015). The research has further shown a positive correlation between heightened student engagement and higher grades and achievement test scores (Wang and Fredricks, 2014) (Fredricks et al., 2016). Engagement levels have also been found to be a strong predictor of math learning outcomes (Ayotola and Adedeji, 2009).

Engagement is a “multidimensional construct” (Christenson et al., 2012), meaning it focuses on the state of being engaged with regard to several dimensions of student interaction in both learning environments and environments outside of the school. The three most common dimensions of cognition are as follows (Fredricks et al., 2016):

1. *Cognitive engagement:* relates to deep learning strategies, self-regulation, and understanding.

2. *Emotional engagement:* relates to positive reactions to the learning environment, peers and teachers, sense of belonging, and interest.

3. *Behavioral Engagement:* participation, persistence, and positive conduct.

In general terms, an engaged student can be conceptualized as one that puts forth effort, persists, self-regulates their behavior toward goals, challenges themselves to exceed, and enjoys challenges and learning (Christenson et al., 2012).

A central component of engagement theory regards the practice of engaging students in such a way that they attend classes, follow teacher directions, complete assignments, and hold positive attitudes toward school and school work (Finn and Zimmer, 2012), i.e., active student behavior means an engaged student. Further, this focus usually entails studying student engagement as coupled with several aspects of the everyday life of the student, e.g., relation to family, peers, and teachers (Christenson et al., 2012), or other bioecological or sociocultural factors (Bond and Bedenlier, 2019) (Kahu and Nelson, 2018).

But for our purposes, we take interest in engagement theoretical constructs that deal with task-specific engagement, or learning activities, especially regarding mathematics, data literacy, and educational technology, through cognitive engagement and persistence.

An early definition within the field of engagement theory and EdTech defines engagement as “a category of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity and perceived user control" (O'Brien and Toms, 2008).

Previous research has found that game characteristics "like goals, rules, narrative, game levels, and the possibility of user-generated content" all have a positive effect on the engagement of players (Zagal and Bruckman, 2008) (Salen et al., 2004) to play the game in question. But when we design our educational game, we aim for the player to be engaged in the learning content.

4 Theoretical Background for Game Development

A recent report on the media habits of American children found that between the ages of eight and 11, around 81 percent spent at least 10 hours a week playing digital games (Ofcom, 2017). This ongoing change in habits and time allocation, together with technological development, has caught the attention of many researchers and enthusiasts of educational technology. One influential example is Marc Prensky, who coined the term *digital game-based learning* (Prensky, 2003) and defined it as the learning methodology that combines curricular content with digital gaming. The term was meant to capture the potentiality of gaming for educational purposes, as gaming plays a larger and larger role in the everyday lives of the youth of today. However, later developments within the field have come to be more inconclusive about the potential aspects of game-based learning (Linderoth, 2014).

4.1 Mini-games

A mini-game is defined as having a restricted set of rules and being restrained to one chapter or event. This has the effect of making the concept easy to grasp. It necessitates "no steep learning curve, [nor] does learning to play the game require an investment of many hours" (Jonker et al., 2009). Instead, the game, or a level of the game, generally can be finished within 10 to 15 minutes, which makes it easy to be implemented in the classroom environment.

We decided to incorporate mini-games into the larger game in part because they require less time to organize and develop, but also because they are easy to fit into a storyline. Furthermore, mini-games have been documented to be effective educational tools with motivational power for children learning mathematics (Alexopoulou et al., 2006) (Barmby et al., 2009).

4.2 Gamification

Another perspective comes from the concept of *gamification*, or "the use of game design elements in non-game contexts" (Deterding et al., 2011) (Deterding, 2012), and how it can be utilized to increase student engagement and the use and degree of task completion in children. We prefer to use the term gamification, even though we have created something that might more aptly be referred to as an education game, or game-based learning (Alsawaier, 2017). However, by using this term we can bring focus to the fact that we have designed specific incentives for students to keep playing and track their progress. We should also state that the definition of gamification might better be substituted by the term "pointification" as described by Hung (Hung, 2017). While it is arguable whether the gamification of an e-learning game seems tautological, it is used here to describe those design aspects that are put into place to motivate desired behaviors, specifically in an educational context. Studies (Brewer et al., 2013) (Alsawaier, 2017) have shown that creating such incentives can be more engaging for students than educational games or game-based learning, so some of these design ideas have been incorporated into the current project. However, the real incentives can thereby become lost, and this can hollow out the learning experience. For now, the only implemented feature is a progress bar that will count "characters cured," but other features such as avatar customization and gold star systems are in the design pipeline. The effects of such design implementations can have a great impact on challenge completion rates (Alsawaier, 2017). This way, we hope to incentivize the students to keep playing, and thereby learning, by reiterating.

4.3 Problematizing Gamification

The lure of using games in educational contexts is strong. If children willingly play games for hours, why wouldn't educators use games during instructional time? The last decade has seen much hope for the application of games in educational settings, but a more sober look at the field reveals much work to be done in order to understand the role of games in education. Though we present in this paper a game for providing students with foundational data literacy skills, we are aware that there are two sides to the coin of gamification. We present here some potential pitfalls and required future work for gamification and learning, as well as reflections on the implications for future work with the presented game.

Though it can appear that making progress in a game must imply some sort of learning, this is not necessarily the case. Linderoth (Linderoth, 2012) applied an ecological framework to the analysis of learning during gameplay to challenge the idea that games are designed with good learning principles. The approach considers the link between the perception and acting on of affordances within games. Affordances are the actions that the environment offers to a particular organism. Perceptual learning in this approach is becoming able to discover and act on available affordances. In the games that Linderoth analyzes, e.g., role-playing and massively multiplayer online games, so much support and structure are built in that players hardly need to develop skills or gain knowledge to progress and succeed. These types of games make certain exploratory actions obvious to the player by, for example, graphically highlighting their next moves, removing the need for exploratory work. Some games are structured on incrementally providing the player with better tools that allow progress, requiring only time from the player, not increased skill. Games like Mario Kart compensate players for performing poorly, equipping them with the tools to make it possible for them to better their position. In short, simply because a player progresses in a game does not mean that they are learning as the progress is built into the system. Linderoth proposes that it is perchance this false sense of progress and empowerment that make games so enjoyable for some. “Perhaps,” he writes, “some good video games offer a pleasure that comes from a continuous *illusion of learning* (Linderoth, 2012).

This issue of built-in progress is not an easy one to overcome. Learning is often not easy, and can come about in the presence of challenges. There is the desire on the part of the designer to allow players to progress so as not to frustrate them, which of course creates the possibility for players to simply click their way through the game simply to progress. This highlights the need for the mediating presence of the teacher, a subject Linderoth brings up in another article discussed below.

The last ten years have seen an abundance of research focusing on gamification, inside and outside of educational contexts. Dichev and Dicheva (Dichev and Dicheva, 2017) took a critical look at research from June of 2014 to December of 2015. Of the thousands of papers published on gamification, they looked at 41 empirical studies of gamification and learning that targeted affective, behavioral, and cognitive outcomes. Of those 41, 12 displayed positive learning outcomes, three negative learning outcomes, and 26 inconclusive results. The majority of papers whose results were deemed inconclusive by the authors were on the basis of inadequate research strategies or methods. These studies involved, for example, small population sizes, short study periods, or admittedly inconclusive results from the authors. Many of the studies did not discern which of the variables they were testing had which specific effect. For example, they included multiple variables such as points, badges, and leaderboards together instead of testing one at a time. Because of these inconclusive findings, the authors suggest the way forward to be a more scientific and systematic approach to the study of gamification and learning. Researchers need to be more specific if they want to come to any conclusive evidence. They should ask questions like ”whether game design elements G are effective for learners of type L participating in activity of type A“ (Dichev and Dicheva, 2017). The authors don’t mean to say that gamification will not prove successful in learning contexts in the future, just that it hasn’t yet been proven. ”Only continued theoretical and rigorous systematic empirical work in varying gamification settings and across contexts will enable us to establish a practical, comprehensive, and methodical understanding of the benefits of applying gamification in educational contexts“ (Dichev and Dicheva, 2017).

The call for such a systematic approach to studying the effects of gamification on learning will need to be taken to heart should the game we are presenting be studied in the future. One proposed idea would be to test the game with and without the progress bar to test for its contribution to the learning process. Gamified elements should be added one-by-one to the theoretical game in the future to be able to systematically describe their individual contributions to learning.

Linderoth is further critical of the idea that games are the magical cure that will make education fun and fruitful (Linderoth, 2014). He criticizes the presumption that games used in learning need to be digital. In fact, he proposes that analogue games can be more fruitful and suitable for the learning process. He makes the distinction between the rules and themes of a game, and makes the point that it is in rule-based games where the potential for learning lies. Thematically-based games have a tendency of trivializing the everyday meaning of the content presented, seldom leading to learning. The content is only seen in the context of the game and not generalized to knowledge about the world. Games with rule systems, on the other hand, can guide actions and provide experiences that can lead to learning. Linderoth argues for the use of analogue games in the classroom, for example board games and in-person role-playing situations. In these contexts, it is important for the teacher to have knowledge about the games being played, and to be present to mediate conflicts and guide learning. Also crucial is the need to avoid losing the real-world educational content of the game. This can be done with lessons and debriefings that accompany and reinforce the learning content over and above the game itself.

Implementing analogue games avoids the use of *commercial off the shelf* digital games that are often profit-driven and designed to be user-friendly and easy to progress. ”Interaction designers don’t like to create applications where people get stuck. However, from a learning perspective it is precisely when one gets stuck and with effort has to overcome obstacles that there are possibilities for learning“ (Linderoth, 2014) (my translation). This overcoming of obstacles can be facilitated by the teacher in the classroom who does not allow for the possibility of clicking one’s way through the game, but instead can provide support when difficult situations arise that, when overcome, can lead to learning.

Should the game be tested in a classroom setting with students in the future, the importance of teacher knowledge and presence should be emphasized. Teachers should play the game themselves to get a feel for it, and be present during testing to prevent students from simply clicking their way through the game. The game should not be presented in isolation but with supporting material about the underlying importance of data literacy in the real world.

5 Design guidelines

This section describes the design guidelines used in the project.

5.1 Fitts' Law

Fitts' Law, created by Paul Fitts, describes how quickly a user can select a target, for example a form field or a button in a user interface, depending on the size of the target and the distance to it (fit).

It is difficult to select a target that is small and distant. Targets which are often used should therefore be larger and closer to where the interaction is to occur (fit).

5.2 User Heuristics

User heuristics are a collection of general principles for interaction design (Nielsen, 1994). They can be used as a broad rule of thumb instead of as specific guidelines since not all of them necessarily apply to a given scenario. In total, there are ten guidelines. Some examples are:

- Match between system and the real world:

  "The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order" (Nielsen, 1994).

- User control and freedom:

  "Users often choose system functions by mistake and will need a clearly marked 'emergency exit' to leave the unwanted state without having to go through an extended dialogue. Support undo and redo" (Nielsen, 1994).

- Consistency and standards:

  "Users should not have to wonder whether different words, situations, or actions mean the same thing" (Nielsen, 1994).

- Recognition rather than recall:

  "Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate" (Nielsen, 1994).

- Error prevention:

  "Good error messages are important, but the best designs carefully prevent problems from occurring in the first place. Either eliminate error-prone conditions, or check for them and present users with a confirmation option before they commit to the action" (Nielsen, 1994).

5.3 Hick's Law

Hick's Law was named by the American psychologist team of William Edmund Hick and Ray Hyman in 1952 when they set out to describe the relationship between stimuli in an interface with user reaction time. As expected, more stimuli almost always leads to longer reaction time. This means that reducing the number of choices the user has to make in any given situation will make the interaction quicker (Soegaard, 2020).

The application of Hick's law can be found almost everywhere that information is presented in a designed way. An example is how Amazon does not list all product categories at once, but instead divides them into subgroups, so as to not overwhelm the user with choices (Soegaard, 2020).

5.4 Interaction Design Metaphors

Metaphors are used to make the learning process quicker by introducing new concepts in a way that users have likely encountered before in similar situations. If done correctly, they can reduce the cognitive load on the user and contribute to a more positive experience (Jiaqi Ma et al., 2009).

6 Process

The project was developed iteratively, which enabled a flexible working approach as small changes were implemented based on feedback. This section describes the process of the development phase and includes decisions related to *prototype 0 to prototype 8*. Each iteration is presented in chronological order in figure 2 and described in the sections below.

6.1 Idea Generation and Prototype 0

During this phase, different ideas for the project were discussed through both brainstorming and storyboarding. Based on the ideas generated, a low-fidelity prototype was created.

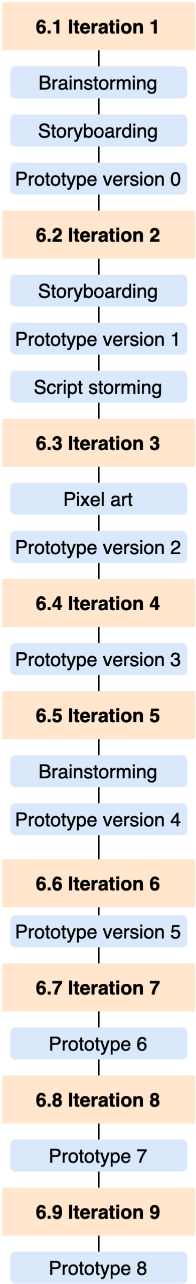


Figure 2: The process.

6.1.1 Brainstorming

Brainstorming is a method for developing ideas and concepts. During a brainstorming session, all ideas and thoughts should be taken into consideration without judgement. In this way, no ideas are excluded and the group is able to work creatively. A brainstorming web was created to visualize the concepts of the game and what content was to be presented (see figure 3).

6.1.2 First Iteration of Storyboarding

Storyboarding is useful for visualizing the context of a product. This method was used to understand how, where, and why the product would be used. During

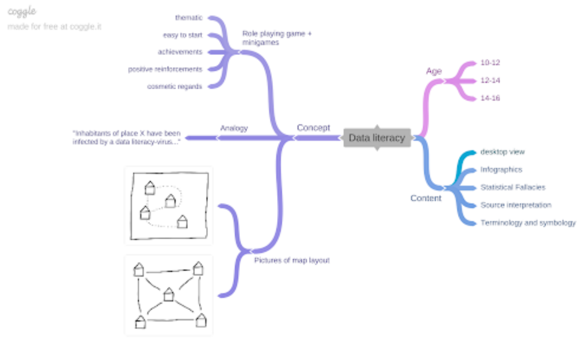


Figure 3: Brainstorming webs created during the first iteration.

the brainstorming session (see section 6.1.1), different options of the map view were discussed. The first iteration of the storyboard (see figure 4) made it clear that the game would be used in schools alongside traditional instruction to bolster learning outcomes for mathematics, statistical reasoning, and problem solving, and thereby provide students with a formative understanding of data literacy (201, 2019). The game could be played while students are learning related competencies that will form the basis of their data literacy skills.

The layout should make it possible for the user to play different parts of the game. The user should not be forced to complete easier levels to be able to play more difficult ones. In this way, the product can be adapted depending on the user. The storyboard also highlighted the importance of the idea that the product should be a role-playing game, so as to engage the user.

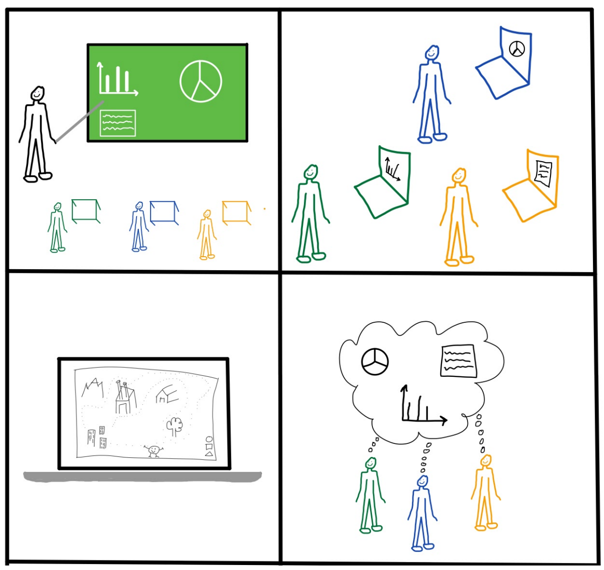


Figure 4: First iteration of the storyboard of the game.

6.1.3 Prototype 0

Prototype 0 (see figure 5) contains a dashboard view, a help view, and three different mini-games. A deeper explanation of mini-games is presented in section 4.1. The dashboard view contains a road map over the mini-games and visualizes the places the user can visit. Before the user enters a mini-game, they receive information about it from a character in the game. The character will be seen in each view to help the user during the game session. In this way, the game fulfills the requirements of the user heuristic *consistency and standards* (see section 5.2).

The goals of the three mini-games are to improve the user's knowledge about *visual literacy* (see section 2), *mathematics*, *natural sciences*, and *social science* (see section 3.3). Cross-curricular teaching (see section 3.2) is achieved as the three mini-games present mathematical concepts of data literacy within a sociocultural context. Numbers and graphical representations are presented in a quasi real-world situation, outside of the classroom.

In the first mini-game, the user learns aspects of *visual literacy* and *mathematics* by developing an understanding of how data is categorized and how it corresponds to a graphical representation. The user sorts items into categories and the sorting is simultaneously represented in an adjacent graph. The user is then able to develop an understanding of what the graphs represent. Allowing children to construct their own graphs might change their relationship to them. It might make the graphs more engaging and provide insight into what is behind them (see section 2). Folders are used to fulfill the user heuristic *match between system and the real world*. The folders are used as design metaphors. The users know how a folder is used on a computer which makes it easier for them to understand their mission.

In the second mini-game, the user develops their knowledge of the *natural sciences* by learning how graphical visualizations are used and presented. The user chooses a header for a newspaper article based on the graphs created in the first mini-game. The user needs to develop an understanding of what the graph is expressing and how to communicate this in a clear way, as well as the possibility of manipulating information.

The third mini-game focuses on the *social sciences*, aiming to teach the user about how information is interpreted in society. The user interacts with different characters in the game, discusses the data, and learns about how people can interpret information in different ways. The communication will take place via multiple choice options.

Each view will contain a small number of objects so the user can focus on the game and be as efficient as possible, in line with Hick's law described in section 5.3. Section 3.1 describes how students do not always pay attention to graphical representations as they can be overstimulated by information.

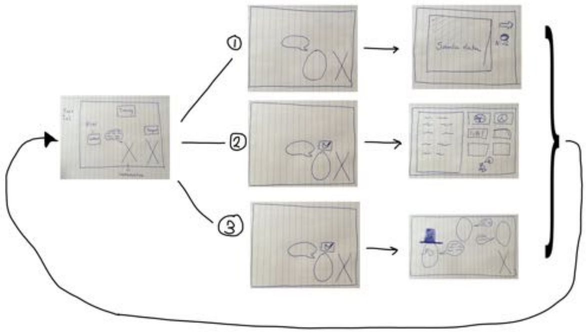


Figure 5: Prototype 0.

6.2 Storyboarding, Prototype 1, and Script

During this phase, a second storyboard was developed based on the decisions made during the previous phase. A high-fidelity prototype was also developed to improve the low-fidelity version, and a script was formulated.

6.2.1 Second Iteration of Storyboarding

The second iteration of the storyboard (see figure 6) changed the decision about the possibility of playing different parts of the game in a self-selected order. In other words, the order of the game became specific. The role-playing game could then be based on a story, and the user able to create an understanding of how data is collected, processed, and discussed. This was visualised in Prototype 0. It is still possible for the game to be adapted in certain ways to each user. For example, the same mini-game may be played many times, and text may be read aloud.

6.2.2 Prototype 1

For the first high-fidelity prototype, we set up the world that the player would enter. For this iteration, the player would have a *player character*, or *PC*, an avatar in the form of a frog. The PC would then interact with different *non-playing characters*, or *NPCs*. First, they would meet an in-game guide, an owl, and then go through the three different mini-games where they would meet other characters, i.e., a doctor and an assortment of citizens from the town (see figure 7). Using an owl as the sidekick was also an important factor to be able to help guide the player without breaking the sense of immersion. Owls often symbolize wisdom, which in the context of the game is a design metaphor. The back arrow allows the user to go back if they make a decision by mistake, which fulfills the user heuristic *user control and freedom*.

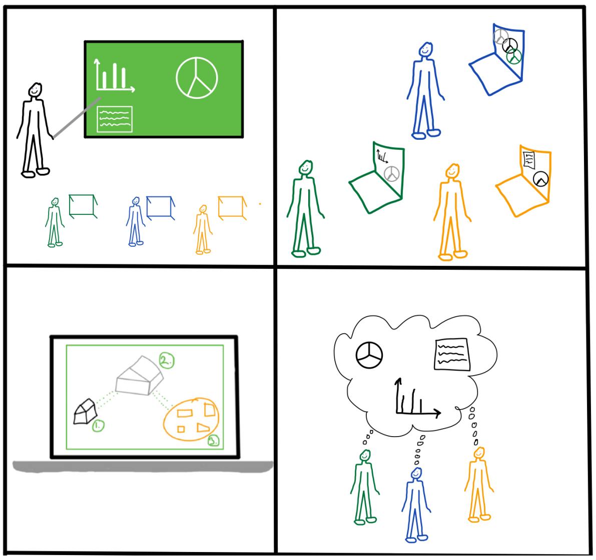


Figure 6: Second iteration of the storyboard of the game.

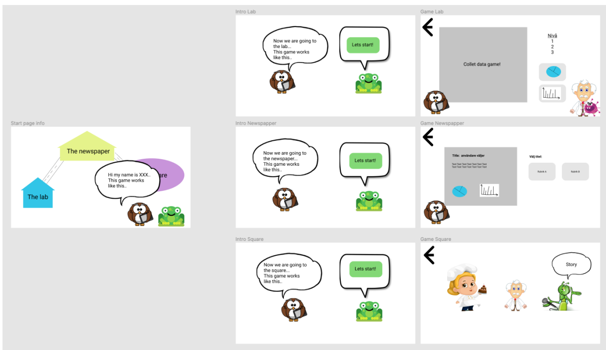


Figure 7: Prototype 1.

6.2.3 Script

During the writing process, we tried to develop the transitions between the different mini-games as well as introduce the player to the game world and its characters. This included a full story and plotline, with a clear beginning and end. During this time, previous inconsistencies were noted and changes to the overall story arc were made. One of the bigger changes was that we left the previous "open world" concept and made the game linear. This also served to make the story more engaging, to ensure that the games led into each other, and to give the game a goal that worked in parallel with the learning. Here we provided the *why* of the game, beyond the "because you're supposed to."

The script also gave us all the important plot moments and scenes necessary to outline "what happens when". The story is incredibly important for any game, and we wanted users not only to play it for learning, but also for pleasure.

We also decided to remove the avatar during the script and instead let the player be the PC. Though the ability was lost to make the PC customizable, the rationale was that the feeling of immersion would be achieved more easily and the effect stronger.

6.3 Pixel Art and Prototype 2

During this phase, pixel art pictures were drawn and prototype 2 was implemented.

6.3.1 Pixel art

The current graphic design in the game can be described stylistically as *pixel art*. This style was necessary when graphic cards had less power to generate pixels, but has lived on beyond the technical evolution (Techopedia).

During the development of the prototypes, pixel art was chosen as the preferred art style. It benefited the iterative way we worked where new additions could be added during a design meeting and feedback could be instantly given.

When every art piece is unique, issues with copyright are nullified. At the same time, it allows the prototype to have a consistent design profile.

6.3.2 Prototype 2

With a finished first draft of the script from the previous phase, we could build a working prototype. Based on the script, the frog was removed from the prototype and the PC became the student who was playing. For copyright reasons, sprites were drawn by hand (see section 6.3.1). The results can be seen in figure 8.

One change from the first prototype was the addition of a motivational greeting where the PC would be introduced to the situation and asked for help. The first mini-game could then be played, and the PC introduced to the scientist NPC. Because of limitations in the software used to create the prototype, all small interactions were excluded between the NPCs, though some were written down for future versions. The second mini-game was still missing certain aspects, as can be seen in figure 8. The third mini-game had multiple text interactions enabled, though some script mistakes were apparent, e.g., "FOOD X" instead of "pasta."

A progress bar was also added to the game, an example of *gamification* (see section 4.2). This serves as motivation for the player to continue playing and helping everyone in the village.

6.4 Prototype 3

A start page was added to prototype 3, with the name of the game: "The Rescue of Dataville" (See figure 9). This view enables the PC to continue playing after a break by using the button, "Load game" (see figure 9).

To make the game more attractive and easier to understand, backgrounds were added in some of the views. For example, a path was added to the map to

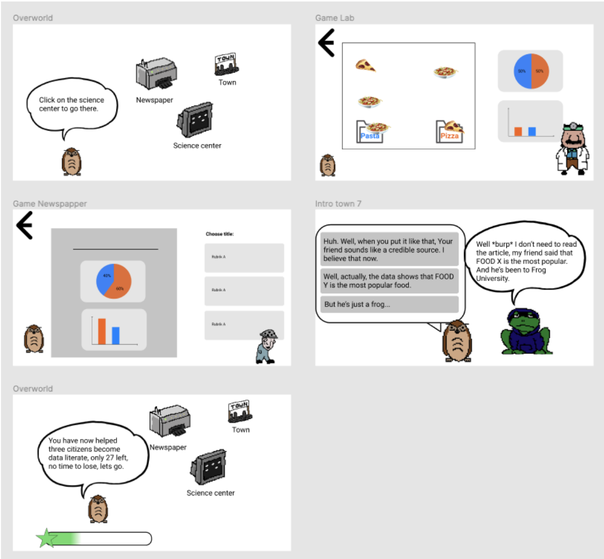


Figure 8: Prototype 2.

illustrate in which order the PC was supposed to play the game.

The newspaper design was added in mini-game 2 (See figure 9). The design shows in which context the title is supposed to be used and fulfills the user heuristic *match between system and the real world.* The graph and diagram from the previous mini-game are presented in the article. The PC does not need to remember what the graphs from the previous game looked like as the purpose is not to practice memory skills. Instead, the same graphs are used in the article and the user can recognize which fulfills the user heuristic *recognition rather than recall.* The multiple choice options in the mini-game are displayed as buttons. The buttons are larger than the text (see figure 9) because the PC should be able to quickly select an option based on the principles of Fitts' law (see section 5.1).

To fulfill the user heuristic *consistency*, the owl, the back and forward arrows, and the NPCs have the same position in all three mini-games. The owl is used to help the PC and provide instructions in the game. In prototype 2, the owl was talking in mini-game 3 as a part of the gameplay. In prototype 3, this was changed as it did not fulfill the user heuristic *consistency* and could confuse the PC as to what role the owl has in the game. "Choose your response" was added as a title to inform the PC of their task (See figure 9).

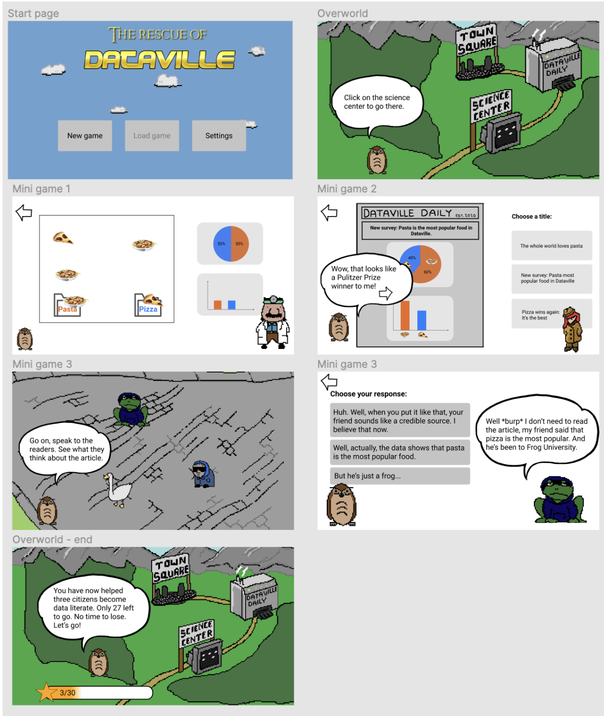


Figure 9: Prototype 3.

6.5 Idea Generation and Prototype 4

During this phase, different options for continuing to develop the game were discussed. Ideas were generated through brainstorming and implemented in prototype 4. Prototype 4 was a sketch and meant to be used as a communication tool to be further developed.

6.5.1 Second Brainstorming

Different options for continuing to develop the game were discussed during the second brainstorming session. It resulted in continuing to develop new games and using the same theme as before. The following part would therefore continue to unfold in Dataville and the PC would continue to save people from the data virus.

Two options for continuing to develop the game were discussed. We would either continue to develop new levels in the current mini-games or implement new mini-games. To be able to implement new learning moments, e.g., new types of graphs, we decided to implement new games. Ideas for new mini-games were discussed and resulted in three options:

- **Survey center:** The PC would generate its own survey based on information from people in Dataville.
- **Farm:** The PC would help a farmer in Dataville to buy a vaccine for its sick cows and motivate their decision using data.
- **Hospital:** The PC would practice repetitive exercises.

6.5.2 Prototype 4

Prototype 4 was implemented based on the second brainstorming session and was meant to be used as a communication tool. During the brainstorming session, three ideas were presented (see section 6.5.1). The farm and hospital were implemented in prototype 4. The survey center described in section 6.11.1 was not be implemented in the prototype because of the prototyping tool's lack of functionality. The two new areas, the hospital and farm, were added to the map view to illustrate that they are parts of the existing game (see figure 10).

Farm When the PC enters the farm, they are to help the farmer buy a vaccine for his sick cows, and their choice is to be based on the appropriate information. First, the user will calculate how many cows are sick with the different diseases and then choose the graph that represents the data which must be collected. In this way, the PC practices visualizing the data in different ways. The circles in the pictures illustrate cows (see figure 10).

In the second step, the user receives prior information on how many cows have been sick with different diseases. The user is to choose the diagram that represents the table. Based on the correct table, the user will read a value from the graph to practice understanding what the graph is saying. In the previous steps, the PC had only gone from table to graph and not from a graph to a specific value. The user will then discuss explanations as to why the values in the table have decreed. This section was added to be able to include real-world educational content and the possibility for the teacher to get the student to discuss the results, which is mentioned in sections 4.3 and 3.2. In this way, the game also includes questions which do not have a specific answer.

In the third step, the user gets information about different vaccines and how they have performed in healthcare studies. The user is to choose the vaccine which shows the best results in the studies. After that, the user are to choose a vaccine to buy for the farmer.

Hospital When the PC enters the hospital, they are able to practice short repetitive questions (See figure 11). The game checks if the PC is still healthy and not infected with the virus. During this session, the PC practices data literacy from previous mini-games and different subject areas.

The importance of practicing by using previous experience in new, but similar, situations is mentioned in section 3.2. This part of the game was developed to achieve that kind of learning situation.

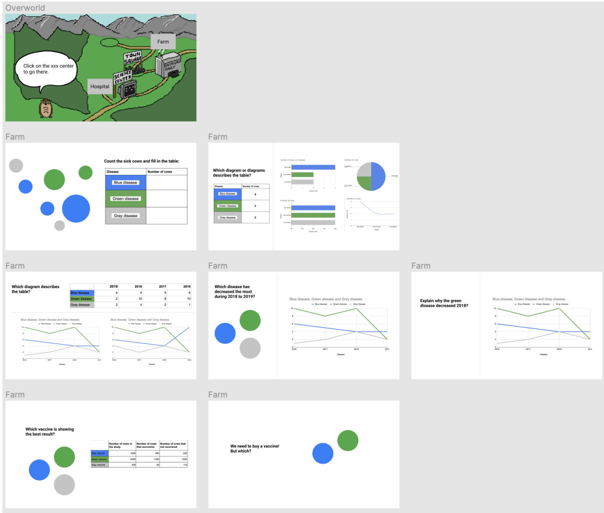


Figure 10: Farm in prototype 4.

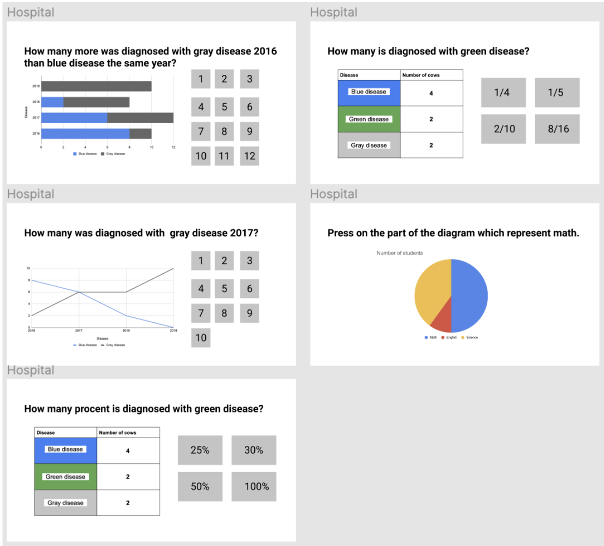


Figure 11: Hospital in prototype 4.

6.6 Prototype 5

Prototype 5 was an improvement upon the farm from prototype 4. In prototype 4, the farm contained three diseases, but it was decreased to two in prototype 5 to simplify the game (See figure 12). More diseases can be added in the future for future levels. The prototype illustrates the first level. Changes were also made in the task which showed the number of sick cows over time.

The number of cows in the table and the number of lines in the same graph were decreased. In prototype 4, the graphs contained multiple lines and did not clearly show which of the diseases increased or decreased. To be in line with the game's first level, they were decreased.

A discussion task was added after the PC had chosen a vaccine to buy. It will be similar to the discussion in the previous phase, described in section 6.5.2.

The formulation of the questions was changed in prototype 5. The PC is to give at least three explanations to the questions. In the earlier prototype, there were no requirements on the number of explanations. The PC is pushed to give more than one explanation, and students are of course allowed to find more than three explanations if they wish.

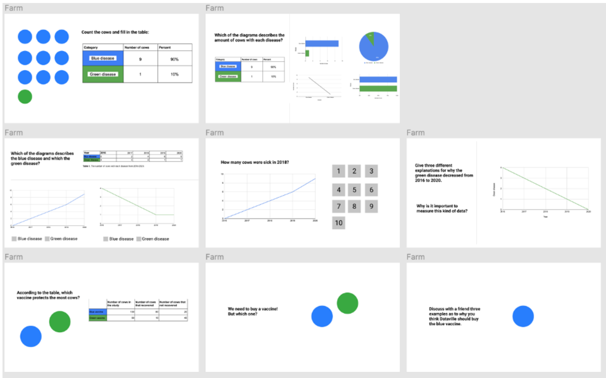


Figure 12: Farm in prototype 5.

6.7 Background Images: Mini-games one, two, and three

During this iteration, previous mini-games were updated. Mini-games one, two, and three were updated with background images (See figure 13). They were designed with pixel art to follow the previous design of the game. The update was done to make the game more appealing and to create a better understanding of the context.

A home button was added and the previous back arrow was removed to clarify its purpose (See figure 13). The home button is used when the PC wants to go back to the home screen with the header, “The Rescue of Dataville” (see figure 9). It has the same position in all views, which fulfills the requirements of the user heuristic, *consistency and standards* (See section 5.2).

Figure 13: Updates in prototype 6.

6.8 Prototype 7

In this iteration, prototypes 4 and 5 were further developed. The pixel art design was used to follow the theme of the game and new background images were added to improve the understanding of the context.

Farm The farm was updated with background images. Instead of visualizing cows and vaccines as circles, they were replaced with images (See figure 14). The owl and the farm were placed in the same positions they had in previous games. In this way, the game fulfills the requirements of the user heuristic, *consistency and standards* (See section 5.2).

A new function was added when the PC's task is to choose a vaccine. The PC is able to read a report of the results from the previous tasks performed on the farm (See figure 15). The graphs and figures are presented to help the user remember the previous data.

The layout was also updated in the discussion task so it can be used in other game situations. By using the same layout, the PC recognizes the images and understands what type of task they are expected to perform, which fulfills the user heuristic, *recognition rather than recall*.

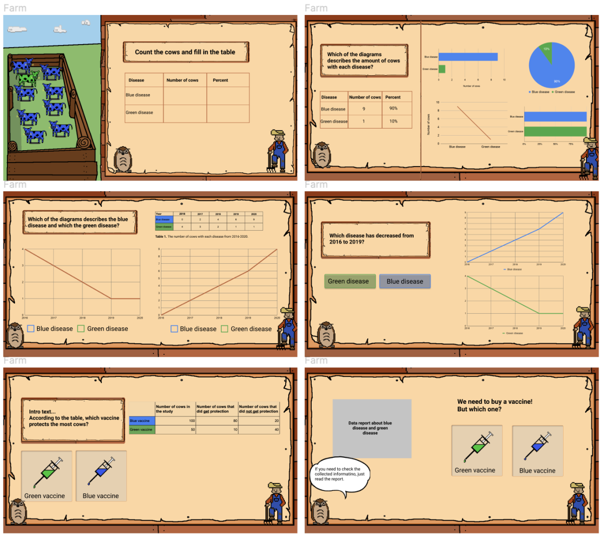


Figure 14: Farm in prototype 7.

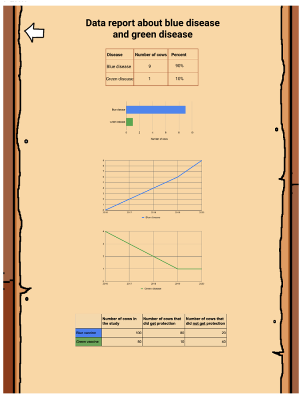


Figure 15: Report in prototype 7.

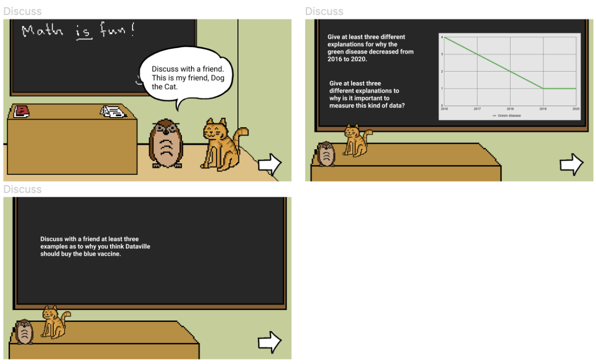


Figure 16: Talk to a friend in prototype 7.

Hospital The hospital game was also updated with background images. A start page where the PC can choose to take a test, practice, or read their medical history was added. The PC is able to practice as long as it wants during the practice task, but the functionality will not be implemented in this prototype. The same idea is presented in the implemented test phase, with the exception of the pre-determined number of questions. When the PC has completed the test, they will get a "diagnosis" which informs them if they have been infected with data illiteracy or not (See figure 17).

Figure 17: Hospital in prototype 7.

6.9 Prototype 8

When developing prototype 8, a play test, as seen in 7, had been conducted and many of the changes that were implemented in this stage were based on the feedback from section 7.5.

More explanatory texts were added in prototype 8 to increase the understanding of the story and to explain the PC's tasks (See figure 18).

The PC navigates to the next speech balloon by clicking on it. A hover effect was added as a signifier to demonstrate its function. When the PC is to navigate to another view, they use the white arrow (See figure 18). Prototype 6 was updated with a home button (See section 6.7), and to fulfill the user heuristic, *consistency and standards*, the home button was added to all of the other pages. When the PC clicks on the home button, an error text is shown (See figure 18). This was added in mini-games one and two to fulfill the user heuristic, *error prevention*. By adding the text *Are you sure you want to quit the game?*, the PC is given help to recover if they press the home button by mistake. In the future, this will be added to all the mini-games. Changes to pictures were also made in prototype 8. From testing (See 7.5), results showed that some of the pictures were perceived as unclear. The pictures were redesigned and added in prototype 8 (See figure 18). The map of Dataville was updated to visualize which places the PC should visit (See figure 18). The buildings were greyed out when they are to be played in the future. In this way, the PC is helped in understanding which order they are supposed to play the mini-games. To illustrate that the game is a draft and will be continued to be developed, a construction rat was added at the end of the game with the text, "Sorry, this is not built yet. Why don't you come back once I've had my coffee break?".

Figure 18: Prototype 8.

6.10 Future Additions and Features

While we have added some new aspects to our game, we've also started looking at future implementations and added features further down the pipeline. While there is probably no end to possible future additions, here we discuss some of the more developed and, to us, interesting improvements.

6.11 Graphics and User Heuristics

One key aspect that our prototypes found lacking, through playing and play testing, was graphical feedback. This includes effects such as interactive objects lighting up and non-interactive objects being greyed out, shaded, or see-through. As we don't want the user to be confused about where to press and what choices are available, creating a uniform system of highlights and gray-outs is an important implementation for future development. At the same time, we want to show that more material can be unlocked in the future, thereby raising interest in continuing to play so as to unlock these aspects.

Another important graphics related update that would be to better streamline the interface. As we mentioned earlier, a uniform interface creates ease of use and lowers frustration. In our case, this applies to the placement of NPCs on the screen, speech bubble design, and transitions between information chunks, mini-games, and scenery changes.

6.11.1 Survey center

Like our other mini-games, the survey center would provide a particular aspect of data literacy, i.e., to create and appraise different kinds of surveys and what they're used for.

In our first mini-game, data from a previous survey is organised by the player. The survey was a poll concerning what food was most popular in Dataville. With this mini-game, however, we want to expand on how a survey is created to answer specific questions, as well as provide discussion material for the class.

For example, the player is asked to investigate what pet is the most popular in Dataville. How should this be done? First, they need to decide on using a qualitative or quantitative method. In this case, a quantitative method would most likely be faster, easier, and cheaper. The next question would be the kind of quantitative method, e.g., survey, multiple choice, phone call, mail, standing at a street corner or on a square. What are their pros and cons? Say a survey is chosen with multiple choices that will be sent out to every citizen. Then the student has to create the survey. Do you write out the choices so people can fill in their preferred pet? Do you let them rank any number of choices? Do you let them write their preferred choice or ranked list themselves?

Taken all together, and used in different iterations, i.e., with different questions to be answered, this can

Figure 19: Model for Conspiracy theorist

provide a powerful tool for teaching players different methods of scientific studies and how to produce empirical data. It will also hopefully provide a better understanding for reading data in the future.

6.11.2 Returning citizens

In our current set up of the progress bar which includes helping all 30 citizens of Dataville, every "level" will introduce new citizens to be "cured" of their data illiteracy by interacting with the PC. For the first prototype, mini-game three, all the characters only needed a short response to right their error and change their opinion. While this is ideal, it's hardly realistic. Because of this, we wanted to introduce a "conspiracy theorist" (See figure 19. This figure was also previously a stand-in for the journalist in prototype 2 (See figure 8). He could potentially be a returning character and in every iteration there could be either a new error to correct, or an older one revisited. This way, the player would hopefully get a better idea for the long-term investment required for critical thinking and trust.

7 Play Testing

As we were progressing with the game and developing new aspects, mini-games, and functions, we also wanted to make sure that what we had already created was working as intended. For this purpose, it needed to be tested.

Since at the start of this part of the developmental process we had only just been able to finish our third prototype, the first to actually contain a somewhat coherent and interesting product, we knew we wouldn't be able to provide a beta, an early and playable, version of our game. Instead, we focused on testing other aspects including the graphics profile, "flow," story, and character content.

This section will present the purpose, test plan, focus, limitations, and feedback from our testing.

7.1 Purpose

As we were planning out the process from the beginning of the first course, we wanted a play test to be a part of the development as a tool for us to measure our creation against its purpose. Teachers could provide feedback into specific functions and students into playability, entertainment, and engagement. This feedback provides the developers with insight into where their ambitions and progress meet, as well as where the discrepancies are.

7.2 Test plan

Due to the physical constraints imposed by the current pandemic, we chose a version of "The play-test method" presented by (Davis et al., 2005), but with social distancing.

Our play-test subjects consisted of other students, mostly from InfoCom at LTH and cognitive science, as these were both easy to access and could provide valuable insight into things like design, interaction, game mechanics, and content. The intention was not to test learning of any kind.

The test leader (TL) set up a zoom meeting with each test subject (TS) and provided a link to the prototype. The TS was then asked to go to the provided page and share their screen over Zoom so that the TL could see what the TS was doing. When this was set up, a short introduction was provided (see appendix A) and then the Zoom record function was used so that the play-through could be further analyzed if needed.

Some TSs were providing their thoughts aloud as they were playing, and some were fully immersed in the game. In either case, a survey followed to collect information.

7.3 What was tested

As we had a functional prototype, prototype 3, and we knew we wouldn't have the time to both develop the prototype further *and* implement it into a playable beta version, we decided to play-test what we had to ascertain aspects about our progress. We could then use this information to revise our game and thereby provide a greater base for further additions.

7.4 Limitations

Since this was a play-test of an earlier prototype, it did not cover all parts of the final prototype. Also, since our play-test subjects didn't come from our target group, we instead focused the tests on aspects such as design, interaction, usability, story, and flow.

7.5 Feedback from Testing

The test subjects had a positive experience from playing through the prototype and wanted more when it ended. Most of the negative feedback was related to the technical limitations of the prototype tool. Features like drag and drop and sound feedback are not supported and therefore not in the prototype. However, a number of small improvements were proposed. The most important examples are as follows:

- Giving more characters a proper introduction.

- Not letting characters move around during conversation.
- Adding clearer instructions for the pizza/pasta sorting mini-game.
- Changing the back arrow appearance so that it is different from the forward arrow.
- Making it clearer when a conversation continues and when it is time to go to a new view.
- Changing some levels of pixelation so that the content is easier to interpret.
- Including an exit popup question so that the user does not leave the game by accident.

8 Conclusion

In this paper we have summarized the theories and research that were used when developing the concept of our educational game, *The Rescue of Dataville*, aimed at developing the data literacy skills of children in grades four through six. Our aim situates the game in a broader research framework, consisting of projects in areas of pedagogy, psychology, cognitive science, and educational technology with the collective aim of developing research, tools, and methods for increasing data literacy in the general population.

We started by reading up on the earlier literature of data literacy. We wanted to compare the different conceptualizations to identify areas within data literacy that we found to be applicable to the development of the idea of our game.

Recurrent and central aspects of data literacy are teaching how to reason, argue, and communicate with data, that understanding and interpreting data is for everyone, and that data literacy is an important aspect of our collective sense-making.

Assisting children in developing these critical skills is important to producing a more data literate adult populace in the future, making it easier for them to assimilate the everyday flow of information, and hence allowing them to become more data-literate citizens.

We continued developing our game collaboratively and iteratively, which made for a flexible working process. By combining our knowledge in areas of pedagogy, cognitive science, computing, information technology, and design, we quickly decided upon central elements of the game from which we developed a low-fidelity prototype. We then continued the development of the game based on research in educational technology and user heuristics, and using Figma, created a high-fidelity interactive prototype which better modelled the final product as well as enabling us to more easily test aspects of the game ourselves. *Link to Figma prototype* [1]

This is an ongoing project, and we will continue to develop the design and layout of our game as well as formulate our conceptions of how to best teach data literacy. This will include elaborating on our current theoretical framework by reading up on other relevant and established theories, such as engagement theory (Christenson et al., 2012). Future work will also include performing interviews with teachers and students as well as developing a method to test the usability of our high-fidelity prototype of the game in a classroom setting.

[1] https://www.figma.com/proto/q1pRhEjsUd7dAUhCyicDVS/Interaktion-1?node-id=124%3A0

References

Chantel Ridsdale, James Rothwell, Michael Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, and Bradley Wuetherick. Strategies and best practices for data literacy education: Knowledge synthesis report. 2015.

Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3):198–213, 2016.

OECD OECD. Future of education and skills 2030: Conceptual learning framework, 2019.

Tamara L Shreiner. Students' use of data visualizations in historical reasoning: A think-aloud investigation with elementary, middle, and high school students. *The Journal of Social Studies Research*, 43(4):389–404, 2019.

Fanny Chevalier, Nathalie Henry Riche, Basak Alper, Catherine Plaisant, Jeremy Boy, and Niklas Elmqvist. Observations and reflections on visualization literacy in elementary school. *IEEE computer graphics and applications*, 38(3):21–29, 2018.

Jacob Carlson, Michael Fosmire, CC Miller, and Megan Sapp Nelson. Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy*, 11(2): 629–657, 2011.

Oxford English Dictionary and WALTON Street. Oxford english dictionary. *Retrieved February*, 4:2019, 2019.

Jeremy Boy, Ronald A Rensink, Enrico Bertini, and Jean-Daniel Fekete. A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics*, 20(12):1963–1972, 2014.

Robert Gould. Data literacy is statistical literacy. *Statistics Education Research Journal*, 16(1):22–25, 2017.

Joan Garfield and Dani Ben-Zvi. How students learn statistics revisited: A current review of research on teaching and learning statistics. *International statistical review*, 75(3):372–396, 2007.

Programme for International Student Assessment and SourceOECD (Online service). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Publications de l'OCDE, 2006.

Hannele Niemi, Shuanghong Niu, Marianna Vivitsou, Baoping Li, et al. Digital storytelling for twenty-first-century competencies with math literacy and student engagement in china and finland. *Contemporary Educational Technology*, 2018.

Priti Shah, Richard E Mayer, and Mary Hegarty. Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of educational psychology*, 91 (4):690, 1999.

Carl Bereiter, Marlene Scardamalia, et al. From conversation to composition: The role of instruction in a developmental process. *Advances in instructional psychology*, 2(1-64), 1982.

Joan Garfield. How students learn statistics. *International Statistical Review/Revue Internationale de Statistique*, pages 25–34, 1995.

Catherine D'Ignazio. Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 23(1):6–18, 2017.

Jake Carlson and Lisa Johnston. *Data information literacy: Librarians, data, and the education of a new generation of researchers*. Purdue University Press, 2015.

Marcelino Caetano Luís. Increasing students' graphic awareness, from activities of converting a situation-context into their graphical representation. *The Educational Review, USA*, 2(6):330–338, 2018.

Trevor Kerry. Learning, cognition and cross-curricular teaching. *Cross-Curricular Teaching in the Primary School: Planning and facilitating imaginative lessons*, page 26, 2015.

Angela M Puccini, Marisa Puccini, and Angela Chang. Acquiring educational access for neurodiverse learners through multisensory design principles. In *Proceedings of the 12th International Conference on Interaction Design and Children*, pages 455–458, 2013.

Erika A Patall. Constructing motivation through choice, interest, and interestingness. *Journal of Educational Psychology*, 105(2):522, 2013.

Marina Papastergiou. Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & education*, 52(1):1–12, 2009.

Eric Jensen. Moving with the brain in mind. *Educational leadership*, 58(3):34–38, 2000.

James Paul Gee. Deep learning properties of good digital games: how far can they go? In *Serious games*, pages 89–104. Routledge, 2009.

Läroplan för grundskolan, förskoleklassen och fritidshemmet [elektronisk resurs] reviderad 2019, 2019.

Philip Vahey, Ken Rafanan, Charles Patton, Karen Swan, Mark van't Hooft, Annette Kratcoski, and Tina Stanford. A cross-disciplinary approach to teaching data literacy and proportionality. *Educational Studies in Mathematics*, 81(2):179–205, 2012.

Koeno Gravemeijer and Michiel Doorman. Context problems in realistic mathematics education: A calculus course as an example. *Educational studies in mathematics*, 39(1-3):111–129, 1999.

Emily R Fyfe, Nicole M McNeil, Ji Y Son, and Robert L Goldstone. Concreteness fading in mathematics and science instruction: A systematic review. *Educational psychology review*, 26(1):9–25, 2014.

Jean Piaget. Science of education and the psychology of the child. trans. d. coltman. 1970.

Basak Alper, Nathalie Henry Riche, Fanny Chevalier, Jeremy Boy, and Metin Sezgin. Visualization literacy at elementary school. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5485–5497, 2017.

Jennifer A Fredricks. Academic engagement in the inernational encyclopedia of social and behavioral sciences, vol. 2. 2015.

Ming-Te Wang and Jennifer A Fredricks. The reciprocal links between school engagement, youth problem behaviors, and school dropout during adolescence. *Child development*, 85(2):722–737, 2014.

Jennifer A Fredricks, Michael Filsecker, and Michael A Lawson. Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues, 2016.

Aremu Ayotola and Tella Adedeji. The relationship between mathematics self-efficacy and achievement in mathematics. *Procedia-Social and Behavioral Sciences*, 1(1):953–957, 2009.

Sandra L Christenson, Amy L Reschly, and Cathy Wylie. *Handbook of research on student engagement*. Springer Science & Business Media, 2012.

Jeremy D Finn and Kayla S Zimmer. Student engagement: What is it? why does it matter? In *Handbook of research on student engagement*, pages 97–131. Springer, 2012.

Melissa Bond and Svenja Bedenlier. Facilitating student engagement through educational technology: Towards a conceptual framework. *Journal of Interactive Media in Education*, 2019(1), 2019.

Ella R Kahu and Karen Nelson. Student engagement in the educational interface: understanding the mechanisms of student success. *Higher Education Research & Development*, 37(1):58–71, 2018.

Heather L O'Brien and Elaine G Toms. What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59(6):938–955, 2008.

José Zagal and Amy Bruckman. The game ontology project: Supporting learning while contributing authentically to game studies. 2008.

Katie Salen, Katie Salen Tekinbaş, and Eric Zimmerman. *Rules of play: Game design fundamentals*. MIT press, 2004.

U Ofcom. Children and parents: media use and attitudes report 2017. 2017.

Marc Prensky. Digital game-based learning. *Computers in Entertainment (CIE)*, 1(1):21–21, 2003.

J Linderoth. Spel i skolan: Det regelstyrda lärandets möjligheter. *Lärare i den uppkopplade skolan. Malmö: Gleerups*, 2014.

Vincent Jonker, Monica Wijers, and Frans Van Galen. The motivational power of mini-games for the learning of mathematics. In *Third European Conference on Gamebased Learning (ECGBL)*, pages 202–210, 2009.

E Alexopoulou, U Bennerstedt, M Childs, V Jonker, C Kynigos, D Pratt, and et al. Literature review on the use of games in mathematical learning. part ii: Deployment. report of the learning patterns for the design and deployment of mathematical games project. *Educational studies in mathematics*, 2006.

Patrick Barmby, Tony Harries, Steve Higgins, and Jennifer Suggate. The array representation and primary children's understanding and reasoning in multiplication. *Educational studies in mathematics*, 70(3):217–241, 2009.

Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: Defining gamification. volume 11, pages 9–15, 09 2011. doi: 10.1145/2181037.2181040.

Sebastian Deterding. Gamification: Designing for motivation. *interactions*, 19:14–17, 07 2012. doi: 10.1145/2212877.2212883.

Raed Alsawaier. The effect of gamification on motivation and engagement. *International Journal of Information and Learning Technology*, 35:00–00, 11 2017. doi: 10.1108/IJILT-02-2017-0009.

A.C.Y. Hung. A critique and defense of gamification. *Journal of Interactive Online Learning*, 15:57–72, 01 2017.

Robin Brewer, Lisa Anthony, Quincy Brown, Germaine Irwin, Jaye Nias, and Berthel Tate. Using gamification to motivate children to complete empirical studies in lab environments. 06 2013. doi: 10.1145/2485760.2485816.

Jonas Linderoth. Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming & Virtual Worlds*, 4(1): 45–62, 2012.

Christo Dichev and Darina Dicheva. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International journal of educational technology in higher education*, 14(1):9, 2017.

Jakob Nielsen. 10 usability heuristics for user interface design. 1994.

Mads Soegaard. Hick's law: Making the choice easier for users. 2020.

Jiaqi Ma, Xiaochun Wang, and Fei Wang. Research of interaction design method based on metaphor. In *2009 IEEE 10th International Conference on Computer-Aided Industrial Design Conceptual Design*, pages 142–145, 2009.

Techopedia. Pixel art. URL `https://www.techopedia.com/definition/8884/pixel-art`.

John Davis, Keith Steury, and Randy Pagulayan. A survey method for assessing perceptions of a game: The consumer playtest in game design. *Game Studies*, 5, 10 2005.

Appendix

A Test plan for play testing

A.1 Purpose

Testing usability and flow in the hi-fi prototype, “The rescue of dataville,” to find areas of improvement regarding design and interaction.

A.2 Questions at issue

- Does the user understand the purpose of the game and important game elements?
- Can the user interact with the game as the designers intended?
- What does the game miss from a user experience?

A.3 Implementation

A.3.1 Subjects

Initially the testing was supposed to be aimed at middle school students as this is the target group for the game. With the current situation this is not possible, so the subjects for testing will instead be other students and friends of the project group. The focus will not be on how much they can learn from the game but more on how they experience it and what they think can be improved in relation to its intended use. The goal is to be tested on between seven and 10 subjects.

A.3.2 Test environment

The testing will be conducted over Zoom where the subject will open the hi-fi prototype in their browser and share the screen during the test. The test will be recorded.

A.3.3 Pre-test information

Information about the purpose of game. Inform them that this is only a prototype and that only specific areas are clickable. Inform them that the test will be recorded and that the recordings will be deleted after the course has ended. Encourage the test subject to speak their mind during the test: every small thing, good or bad.

A.3.4 During test

When the test subject has opened the prototype and the recording has started, the test leader will take notes if the comments are made or any interesting interaction occurs. The test subject will be left alone to use the prototype as much as possible.

A.3.5 Post-test questions

- Name?
- Field of study?
- Earlier experience of gaming and e-learning games?
- What did you think of the prototype?
- How did the interaction and flow work?
- What did you think of the story and text content?
- How did you perceive the characters in the prototype?
- What did the prototype miss?
- Was anything confusing or frustrating?
- How do you like to learn?
- Anything else?

Variationsteori & historia för mellanstadiet

Emmy Edfors, Isak Aho Nyman, Amanda Eriksson & Benjamin Teir

em4146ed-s@student.lu.se; isak.aho.nyman@gmail.com; am4874er-s@student.lu.se; benjamin.teir@gmail.com

”Den glömska historieboken” är en prototyp av ett lärspel i historia för mellanstadiet som har utgått ifrån det teoretiska ramverk som erbjuds av variationsteorin för att ge ett nytt perspektiv på historieundervisningen. Rapporten nedan redogör för den teoretiska bakgrund som använts i utvecklingen av spelet, hur spelet i sig har formgetts och implementerats och slutligen även testats på en mindre grupp användare. Slutligen följer även en kortare diskussion där några av de centrala frågeställningarna för projektet tas upp och utvecklas. Projektgruppen ställer sig positiva till användningen av variationsteori inom historieämnet och menar att det lämpar sig bra för användning i en spelmiljö, men ser det som en viktig utveckling att genomföra mer genomgående tester på målgruppen för spelet (elever i årskurs 4-6), samt utvärdera hur man vill paketera de inlärningsmöjligheter som spelet erbjuder för att kunna ge en så utförlig förståelse av historieämnet som möjligt till de elever som spelar spelet.

1 Introduktion

Tänk er tillbaka till mellanstadiets historieklassrum, där allt man lärde sig hade sin plats i en viss ordning, med andra ord: en traditionell kronologiskt organiserad utlärning av ämnet historia. Den kronologiska ordningen i inlärningen av historia lever kvar än idag, men är det verkligen det bästa sättet att lära sig historia på?

Inom pedagogik finns det flera olika teorier om vad som är bästa sättet att lära ut diverse skolämnen på, en av dessa teorier är *variationsteori*. Variationsteori använder sig av kontraster, relevans och förhållande mellan aspekter och är beprövat för att lära ut skolämnen som exempelvis matematik. I denna rapport utmanar vi den traditionella synen på historieinlärning och undersöker huruvida det är möjligt att använda ett mer okonventionellt sätt att lära ut historia, med hjälp av ett variationsteoretiskt ramverk. Observera att alla infogade bilder finns bifogade som fullstora bilagor i slutet av rapporten.

Bakgrund

Den här rapporten är ett resultat av ett projekt som har utförts i kursen *Interaktion 1 - Neuromodellering, kognitiv robotik och agenter* samt kursen *Interaktion 2- Virtualitet och kognitiv modellering* vid Lunds Universitet. Processen startade med att vår projektgrupp samlades och fick ett projekt tilldelat, efter att ha individuellt fått rangordna olika typer av projekt. För att kunna arbeta fram ett genomförbart projekt som passade såväl uppgiften samt vår kompetens fick vi fria händer att skapa en egen infallsvinkel utifrån ett antal resurser. Några av dessa resurser var bland annat uppgiftsbeskrivningen där betoningen låg på att testa variationsteori samt ett redan existerande digitalt lärspel: *Historiens väktare*, som är utvecklat av forskningsgruppen ETG vid institutionen för kognitionsvetenskap på Lunds universitet.

Historiens väktare är ett spel för inlärning av historia där eleven, eller den användare som spelar, får utföra tidsresor tillbaka i tiden för att träffa kända vetenskapsmän. En rad uppgifter utförs hos dessa personer för att uppmana spelaren att lära sig mer om dessa inflytelserika människor. Eleven ska sedan ta med sig sin kunskap tillbaka till nutiden där hen får lära ut den nyfunna kunskapen till en så kallad *teachable agent* (se teoriavsnittet). Beroende på spelarens svar får hen antingen ta sig vidare i spelet eller återvända tillbaka i tiden för att hitta rätt svar.

Historiens väktare innehåller flera olika spelkomponenter och har utvecklingspotential. Vi fick därför redan i ett tidigt skede ta ett beslut huruvida projektet skulle utgå från spelet och om det ansågs vara möjligt att inkorporera det variationsteoretiska ramverket i det redan existerande digitala lärspelet, eller om spelet snarare skulle ses som en inspirationskälla och att målet istället skulle vara att ta fram ett nytt koncept för ett helt nytt spel.

Det stora fokuset i uppgiften har legat på att undersöka huruvida det är möjligt att inkorporera det variationsteoretiska ramverket i skolämnet historia, och i detta fall i form av en digital lärospelsmiljö. För att garantera att designbesluten baseras på variationsteori har detta projekt utförts såväl som en litteraturstudie samt som ett designprojekt.

Syfte

Att undersöka om ett variationsteoretiskt ramverk kan användas för att bygga ett läromedel inom historia för mellanstadieelever, samt utforska om detta kan göras i form av ett digitalt lärspel.

Mål - Delkurs 1

Att under första kursen *Interaktion 1 - Neuromodellering, kognitiv robotik och agenter* utveckla en konceptuell modell för hur ett potentiellt spel skulle kunna utformas, samt att få en bild över den teoretiska bakgrunden för inlärning med hjälp av variationsteori. Målet var även att en low-fidelity-prototyp (lo-fi) skulle utformas för att få en grafisk representation av den konceptuella modellen, som sedan skulle kunna utvecklas vidare i kurs 2.

Mål - Delkurs 2

Att under andra kursen *Interaktion 2- Virtualitet och kognitiv modellering* utveckla vidare prototypen från kurs 1 och slutligen nå fram till en slutgiltig high-fidelity (hi-fi) prototyp som är interaktiv och testbar, samt som kan användas för fortsatt utredning av syftet.

Avgränsningar

Då detta projekt utförs i en tidsbegränsad kurs har vi valt att inte utveckla en fullskaligt detaljerad produkt. Fokus kommer istället att ligga på att utforska det variationsteoretiska ramverkets möjligheter inom inlärning av historia, samt att utveckla en T-prototyp bestående av en överskådlig vy med ett eller några enstaka teman att kunna interagera med och testa.

Gruppen har även valt att begränsa testningen till enbart användbarhetstestning av prototypen, då test av inlärning hade krävt mer resurser i såväl tid för spelutveckling samt koordinering med en skolklass för att kunna kontrastera resultaten av spelet gentemot traditionell metodik.

2 Teori

Projektets syftet har som tidigare nämnt varit att undersöka variationsteorins applicerbarhet inom historieundervisning, mer specifikt till mellanstadieelever. Vad är då variationsteori? Som en inledande grund så är variationsteori en inlärningsteori om hur människor tar till sig information. I sin mest grundläggande form så menar variationsteorin att vi lär oss genom urskiljning. Enligt variationsteori bygger människor alltså upp unika profiler av diverse företeelser, objekt eller subjekt genom att se dem presenterade i samband med varandra och kontrastera skillnaderna emot varandra. Bussey, et.al. (2013) använder ett exempel där målet är att lära ut hur en mogen banan ser ut. Enkelt sagt så kan flera bananer presenteras samtidigt, där storlek och form är identiska medans färgen varierar mellan grön och mörk gul. I detta fall så ligger fokuset på den kritiska egenskapen färg.

Variationsteorins idéer om kontrasthöjande är enkla att integrera med uppfattningar om hur viss undervisning ska gå till, men inte all. Inom t.ex. matematikämnet kan teorin lätt omsättas i praktik då matematikämnet i mångt och mycket handlar om just att kontrastera olika angränsande begrepp och enheter genom att visa var de skiljer sig åt, kanske för att det är lättare att förklara dem i relation till varandra än i ett vakuum (Kullberg et.al., 2017). När det kommer till historia, som ofta lärs ut genom att presentera händelser som punkter i ett kausalt system blir det dock svårare. Vi har en vedertagen uppfattning om att historia och historien är uppbyggd inte så mycket efter enskilda enheter utan som efter antagandet att det finns naturliga brytpunkter i en händelsekedja, och att vi i dessa brytpunkter kan lägga fokus för att visa hur de *förhåller sig* till varandra, snarare än vilka skillnader eller likheter som finns mellan dem. När det kommer till historia verkar de vardagliga antagandena vi har antyder att det variationsteoretiska perspektivet lämpar sig dåligt för ämnet, men vi menar att detta handlar mindre om historieämnet i sig, och mer om sättet på vilket det konstitueras och formuleras för/av en elev eller lärare.

Variationsteori

Hittills har vi benämnt variationsteorin som en teori om kontrast och urskiljning, men det heter ju trots allt *variations*teori och inte *kontrast*teori. Så var kommer variationen in i det hela? Man menar att ett sätt att skapa den här höga kontrastsituationen är genom att hålla vissa parametrar i det som man vill att eleverna lär sig konstanta medan man just *varierar* andra. Ett exempel som tas upp i Kullberg et.al. (2017) är hur man undervisar begreppet om ekvationer med en variabel i skolan. Där pratar Kullberg et.al. om hur de i samråd med en lärare utvecklade en lektion som fokuserade på att visa de olika sätten att lösa en ekvation på och på så sätt demonstrera hur den fungerar. Man visade hur lösningsmetoderna skiljde sig om ekvationen innehöll multiplikation, addition, subtraktion etc. men höll resten av ekvationen konstant. Alla värden var alltså samma, men operatorn som ingick i uttrycket förändrades. Genom den här typen av undervisning kan man ändra vad det är man vill att eleverna ska lära sig, från att lägga fokus på en specifik lösning eller struktur till vad det är som skiljer olika strukturer eller lösningar från varandra.

Den aspekt av variationsteori som handlar om att ha en rörlig variabel medan andra hålls konstanta, är något som vi försökt implementera så mycket som möjligt och på ett så effektivt sätt som möjligt. Många av studierna som vi tagit del av i det här arbetet har varit inriktade på matematik, språk, eller annan icke-linjär undervisning, men inte i så stor utsträckning behandlat en mer processinriktad inlärningsform som t.ex. historia. På grund av den icke-diskreta strukturen som historieämnet har kan det vara svårt att urskilja facetter på det sätt som man kan göra inom matematiken. En initial lösning på detta har varit att strukturera upp historiematerialet efter mindre linjära och mer tematiska linjer. Istället för att ha en struktur där man lär sig historien i en följd har vi valt att presentera historien parallellt inom ett visst specificerat tema. Vi har använt oss av en tabell för att lättare kunnat konkretisera detta (se tab.1).
Tabellen är uppbyggd efter två axlar där y-axeln är den konstanta variabeln ”Tema” och x-axeln är den beorende variabeln ”Epok”. På det här sättet kan vi visa på skillnader i hur ett visst tema tar sig uttryck genom olika tidsepoker. Istället

	Epok (a)	**Epok (b)**	**Epok (c)**
Tema (A)	Aa	Ab	Ac
Tema (B)	Ba	Bb	Bc
Tema (C)	Ca	Cb	Cc

Tab.1 - Tabell för att illustrera variation över teman

för att gränsa av historien genom tidsperioder och tidslinjer har vi valt att gränsa av den baserat på hur ett godtyckligt valt tema tar sig uttryck under vissa perioder, något som utnyttjar den variationsteoretiska grundinställningen om att kontrast och variation bidrar till en ökad förståelse för det som ligger framför en, som man ska lära sig. I det “traditionella” sammanhanget, en historielektion som inte var organiserad utifrån variationsteorin, hade det antagligen sett annorlunda ut. Det är en lite haltande jämförelse eftersom historia generellt undervisas, framförallt i mellanstadiet, som ett flöde där epokerna hakar i och tar vid från varandra, men man skulle kunna säga att en traditionell historielektion hade varit organiserad tvärtom mot vad vi vill illustrera i tabell 1. Man låter eleverna studera t.ex. vikingatiden i sin helhet, genom strukturen ovan får man istället ett upplägg som varierar teman (hushåll, familjeliv, teknologi etc.) men håller tidsepoken (vikingatiden) konstant.

En kort utvikning behöver göras här för att förklara ett begrepp som kommer dyka upp nedan och som används av den större delen av den variationsteoretiska litteraturen, begreppet om ett “lärandeobjekt”. Enligt variationsteorin ses inlärning som en funktion av urskiljning, och urskiljning i sin tur som en funktion av variation. Lärarens roll i det variationsteoretiska klassrummet är i den meningen inte att överföra kunskap till eleven utan att tillgängliggöra något - lärandeobjektet - för denne (Kullberg et.al., 2012, Bussey et.al., 2017). För att kunna förstå variationsteorin såsom den är utformad i litteraturen måste man förstå att den gör vissa antaganden om hur inlärning fungerar och att en av dessa är att man ser inlärning som något som görs möjligt i en inlärningssituation, snarare än att en inlärningssituation är uppbyggd kring att en lärare “fyller” en “tom” elev med kunskap.

En teoretisk utmaning som projektet stod inför var hur vårt historiespel kan ackommodera en aspekt av variationsteorin som beskrivs i bl.a. Guo et.al. (2012), en aspekt som handlar om relevans. Guo et.al. (2012) presenterar hur det går att dra en skiljelinje i analysen av lärandeobjekt där det finns egenskaper hos lärandeobjekt som måste ses som relevanta eller irrelevanta för att uppnå ett visst mål eller lösa problem. Enligt Guo et.al (2012) är de här egenskaperna oftast uppdelade i två kategorier, vad som ligger på ytan och vad som ligger djupare i lärandeobjektet. En ytlig egenskap är i Guo et.al (2012)'s exempel antal eller namn, medan en djupare egenskap är de underliggande regler som styr en viss företeelse. När det kommer till ämnen som matematik blir detta lätt att koppla till ämnet. Att förstå en ekvation eller en formel har ingenting att göra med att veta värdena som ingår i ekvationen, lika lite som det är avgörande att vi kallar det för just en ekvation. Snarare är det viktiga - den djupgående egenskapen som låter oss nå vårt mål och till slut lösa ekvationen - vilka regler som gäller för manipulering och ekvivalens. Men när det gäller historia är detta inte lika entydigt. En händelses namn, eller antalet personer som var närvarande vid ett slag eller hur stora befolkningsströmmarna in i ett visst land var vid en viss tidpunkt kan inte sägas vara mindre relevanta för att förstå situationen och uppnå det mål eller lösa det problem man har framför sig.

Att identifiera vilka aspekter som är kritiska för att kunna använda sig av ett begrepp eller förstå hur ett skeende passar in i en större tidslinje är avgörande för att ett projekt som detta ska kunna vara användbart. Kullberg et.al. (2017) menar t.ex. rent konkret att olika saker är möjliga att lära sig i olika sammanhang, det går inte att lära sig vad som helst när som helst. Vilka begrepp och skeenden som presenteras samtidigt kommer påverka vilka aspekter av lärandeobjektet som görs tillgängliga att lära sig för den som ska lära sig dem. Läraren, eller materialet för den delen, kan inte ensam kontrollera för vad eleverna kommer att lära sig eller ta in från ett lektionstillfälle eller en viss bok; det är inte läraren som lägger informationen i eleven. Ska vi följa Kullberg et.al. (2017)'s linje är det snarare så att vi måste se processen som att eleven extraherar information och syntetiserar detta i sig själv. Läraren och materialet kan endast ge en förutsättning för detta att ske. Vårt mål med designen av spelet bör därför, om vi ska hålla oss så trogna som möjligt till den teoretiska bakgrunden, vara att möjliggöra inlärning hos eleven. Att identifiera exakt vilka detaljer som är relevanta eller inte för ett visst skeende är svårare när det kommer till historia än när det kommer till t.ex. matematik eller kemi, där man kan bestämma sig för att valens eller antal inte spelar roll i vissa kontexter. Att utforma ett lärspel i historia utefter ett variationsteoretiskt perspektiv kom att kräva inte bara en grundläggande insatthet i ämnet som ska läras ut, men också en förståelse för hur lärandeobjekten ska presenteras gentemot varandra just för att det inte går att bestämma sig för att det aldrig är relevant att diskutera antalet närvarande vid en viss händelse eller att det alltid är relevant att se till makroekonomiska aspekter i den direkta omgivningen.

För att göra en kort sammanfattning. Variationsteorin är enligt tidigare rapporter mest lämpad för användning på ämnen som saknar linjära förlopp och skeenden, i varje fall är den underutforskad inom fält som har den här linjära egenskapen. Variationsteorins centrala antagande är att allt lärande sker genom kontrast, och att egenskaper i ett lärandeobjekt bara kan göras tillgängliga för ett lärande subjekt genom att de ställs i kontrast mot något annat lärandeobjekt med liknande egenskaper men med någon central variation. På det sättet är variationsteorin en teori om urskiljning genom jämförelse. Vidare menar vissa forskare, bl.a. Guo et.al. (2012), att det finns en hierarki för varje lärandeobjekt där vissa egenskaper är mer centrala för förståelsen av objektet, och att sagda hierarki grundar sig på hur väl en viss egenskap bidrar till att ett lärandesubjekt kan använda sig av lärandeobjektet i konkret problemlösning eller för att uppnå vissa mål. Med allt detta givet borde variationsteorin vara väl lämpad för användning i ett lärspel, främst utifrån dess utforskande och interaktiva natur, men det kan finnas en utmaning i hur man ska tydliggöra vilka aspekter och egenskaper hos det icke-diskreta och kraftigt relativa historieämnet.

Tidsfördelad repetition

Tidsfördelad repetition har länge varit ett ämne som intresserat forskare inom inlärning. Det finns flera kognitiva teorier som ligger till grund till tidsfördelad repetition. Bland annat "encoding variability" teorin som syftar på att inlärning som sker med tidsmellanrum stödjer minnet för att informationen presenteras i olika kontext samt mentala tillstånd och därmed bidrar med ett mer mångsidigt perspektiv till kunskapen (Smolen, Zhang & Byrne, 2016). En annan kognitiv teori som stödjer tidsfördelad repetition är "Study-phase retrieval" teorin som framhäver att konsolideringen av minnet stärker sig ju oftare ett minnet återhämtas och aktiveras. Ifall kunskap testas flera gånger i ett och samma pass så kan de aktuella neuronerna inte återaktiveras då de redan är aktiverade (Smolen, Zhang & Byrne, 2016). Enligt denna teori är det alltså bättre ifall testning sker i enlighet med tidsfördelad repetition.

Teachable Agents

Utöver variationsteori och tidsfördelad repetition har vi också använt oss av begreppet "teachable agent". Idén bakom Teachable Agents (TA) är att en elev har lättare att ta till sig information om den inte bara lär sig för sig själv, utan också tydligt lär sig för någon annan (Blair et.al., 2007). Det absolut mest intuitiva sättet att åstadkomma det här är såklart att låta eleverna i klassen lära upp varandra. Ett problem som uppstår för den typen av undervisning är att eleverna på det sättet kan "skada" varandras inlärning. De som har en dålig lärare kommer få sämre förståelse av materialet än vad de som har en bra lärare får. En TA kan lösa det här problemet då en TA inte är en faktisk person. I Blair et.al. (2007) ges olika exempel på hur man kan utforma TAs, gemensamt för de alla är dock att det är TAn och inte eleven som examineras.

Blair et.al. drar upp fyra kärnprinciper som en TA bör ha för att vara effektiv:

1. Den ska använda sig av explicita och välstrukturerade visuella representationer.
2. Den ska kunna agera självständigt.
3. Den ska vara en modell av produktivt inlärarbeteende.
4. Den ska innefatta miljöer som underlättar lärandeinteraktioner (Blair et.al., 2007, s. 58-59, egen översättning).

Punkt 1) och 3) är båda två svåra att fånga utan en TA, i en elev t.ex. Vad gäller 1) kan en TA programmeras att hela tiden visa vad det är den har lärt sig, något som ger eleven konstant feedback av hur hens undervisning påverkar TAn. Vad gäller 3) kan man genom en TA konstruera den "ideala inläraren". Istället för att låta eleverna lära varandra och allt vad det innefattar av lässvårigheter, bristande kunskap, intresse osv. kan en TA bli en feedbackloop till eleven som lär upp den så att den själv kan utvärdera inte bara sitt lärarbeteende utan också sitt inlärarbeteende.

Punkt 4) är värd att förklara lite närmare då den formulering som Blair et.al. (2007) ger i sin rubrik är lite vag. Det som avses här är att TAn ska existera i en miljö som på olika sätt lämpar sig för inlärning. Exempelvis kan det vara att låta TAn göra ett prov eller på annat sätt testas i olika minispel inuti det större kontextet (Blair et.al., 2007, s.57, 60-61). Detta liknar effekten av 3): genom att förbereda TAn för t.ex. ett prov kommer eleven ta till sig informationen den använder bättre då den uppfattar sig själv vara i en omhändertagande och ansvarstagande relation gentemot sin TA (Blair et.al., 2007, s.58-59, Chase et.al., 2009). Vidare kommer eleven genom att observera TAns beteende och feedback kunna få en uppfattning om hur den själv ska bete sig i en liknande lärandesituation och därmed modellera sitt eget beteende efter TAn som den själv lär upp.

Genom att använda sig av en *teachable agent* kan man som läromedelsutvecklare skapa en situation där eleven fördjupar sin kunskap genom att delge den till någon annan, men också få feedback både om sin utlärningsförmåga och mer implicit feedback om sin egen position som elev och lärsubjekt.

3 Designprocessen

Historiens väktare

Projektet utgick som tidigare nämnt från lärspelet *Historiens Väktare* som är ett historiespel där användaren får resa tillbaka i tiden och besöka olika personer som har kommit på stora upptäckter, exempelvis Kopernikus, Galilei och Newton. Användaren ska resa i tiden och lära sig fakta om personerna för att sedan återvända och lära en alv, som fungerar som en teachable agent, den nya kunskapen som hen har lärt sig. För att kunna skapa ett lärspel med hjälp av variationsteori ansåg vi att det fanns för få tematiska aspekter att hämta i *Historiens Väktare.* Därför bestämde vi att ha spelet som inspiration och enbart behålla vissa koncept. Hela idén med att besöka personer från andra tidsepoker gav en större anknytning till nyckelkaraktärerna och därmed förståelse, vilket var något vi ville försöka behålla.

I spelet utförs dialogen med karaktärerna likt en envägskommunikation, men där användaren själv får navigera sig runt i miljön och trycka på de objekten som hen tycker verkar intressanta.

Alven i *Historiens Väktare* tillför mycket till spelet som en teachable agent, något som vi ville få med oss även till vårt spel. Själva testen i *Historiens Väktare* upplevde vi som kreativa och såg stora möjligheter att hämta inspiration från dessa och försöka att inkorporera variationsteori i dem. Variationen i spelen som presenteras bidrar till att upprätthålla intresset hos användaren.

Kurslitteratur för mellanstadiet

För att ha en bred grund började vi med att undersöka hur redan existerande historiespel är upplagda och om det fanns något spel som redan använde en tematisk uppbyggnad. Vi hittade dock inget konkret exempel där variationsteori har använts inom historiainlärning. I och med att vår uppgift gick ut på att undersöka variationsteori samt att utforma en prototyp och inte ämnet historia i sig, började vi titta på Skolverkets läroplan för att få information om vad mellanstadieelever ska lära sig om historia för att på så sätt hitta givande kurslitteratur. Källorna som vi främst har utgått ifrån är *Upptäck Historia* (Ljunggren & Frey-Skött, 2016) samt *Digilär Historia* (Körner et.al., 2018) som lär ut historia från 11 tidsepoker specifikt för mellanstadieelever via ett digitalt verktyg. För att utforma frågor inför att testa användaren på sin nyinlärda kunskap behövdes ytterligare information. Främst användes BBC, History Channel och sidan Kidsconnect.com. Vi ville se till att använda källor som både var relevanta och pålitliga, men syftet med informationssökningen var att illustrera sätt att formulera variationsteoretiskt relevanta *frågor* snarare än att utforma det större *ramverket*. Vi avgränsade materialet till ett fåtal teman för att fokusera mer på djupet. Ett krav vid valet av tema var att det måste finnas innehåll från alla tidsepoker om diverse teman i och med att vi just ville testa variationsteori. Med andra ord utformades teman utifrån materialet. De tidsepoker som vi kunde hitta i kurslitteraturen för mellanstadieelever som vi kände uppfyllde kravet var de som tillhör forntiden i Norden, alltså jägar- och samlarstenåldern, bondestenåldern, bronsåldern, järnåldern och vikingatiden.

Konceptuell modell

En konceptuell modell för lärspelet arbetades fram. Olika idéer diskuterades till en början och med inspiration från *Historiens Väktare* byggde vi vidare idén att spelaren skulle besöka de olika tidsepokerna och personerna från dessa. Vi ville att spelet skulle bestå av en värld där olika hus längs en väg skulle representera dessa olika teman. Den visuella idén som hela vår prototyp bygger på kommer från en mycket enkel skiss gjord i all hast, se bild 1.

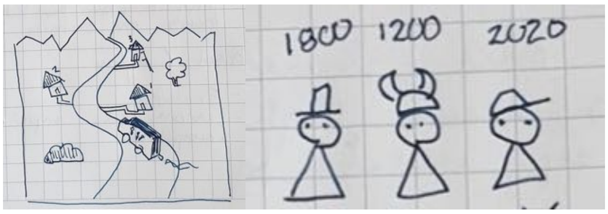


Bild 1. Första idéskissen av världen och karaktärerna

Många idéer har omarbetats under processens gång. Bland annat tänkte vi från början att världen skulle vara som en karta, där de husen som spelaren inte hade tillgång till skulle vara mörklagda. På så sätt skulle användaren få en översiktsbild över hur långt man har kommit i spelet, någonting som vi tyckte saknades i *Historiens Väktare.*

De olika tidsepokerna representeras av olika karaktärer likt bild 1. Karaktärerna föreställer barn som kommer från respektive tidsålder och urskiljer sig från varandra med hjälp av tidstypiska detaljer, kläder eller hattar. Exempelvis har karaktären som kommer från vikingatiden en stereotypisk vikingahatt. Varje hus representerar ett tema, och i varje hus bor en av karaktärerna. På så sätt får användaren hälsa på hos karaktärerna i något skede i spelet och där berättar de om delar ur deras liv. En annan bakomliggande tanke till att vi just ville använda oss av barn var för att mellanstadieeleverna lättare skulle kunna relatera till spelkaraktärerna med tanke på ålder.

I början av designprocessen hade vi även med en karaktär från nutiden, informationen som presenterades skulle då även kunna kontrasteras med nutiden, något som kändes relevant med tanke på variationsteori. Dock fick vi senare i processen exkludera den nutida karaktären på grund av praktiska skäl som vägde högre än det vi tyckte att den nutida personen tillförde. Vi ansåg bland annat att användaren kunde själv skapa samma kontrast till nutiden utan karaktären. Den nutida karaktären passade inte heller in i vår storyline som utvecklades sent i designprocessen, därav exkluderingen av denna.

I iteration två av den konceptuella modellen utvecklade vi skissen från bild 1 till en mer detaljerad modell, se bild 2. I mitten av kurs 1 ägde en mittredovisning rum där bild 2 presenterades som en konceptuell modell. Vi fick med oss värdefull feedback, bland annat uppmanades vi att fundera mer kring hur faciliteringen av inlärningen skulle ske. Vi fick även frågor kring om vi hade en storyline eller vad spelet skulle gå ut på, något som vi än inte hade kommit fram till.

Första skisser

Med bild 2 som grund skapades en enkel storyboard för att få en idé om hur spelet skulle vara uppbyggt samt vilka sidor som behövdes, se bild 3.

Grundidén var att användaren skulle få en presentation av karaktärerna på startsidan och att man sedan skulle gå in i det första huset där det första temat skulle presenteras. Inne i huset skulle varje karaktär berätta om sin historia och användaren skulle likt *Historiens Väktare* kunna välja olika alternativ för att styra hur dialogen skulle gå. Efter att varje karaktär berättat sin historia skulle huset avslutas med ett test, och därefter skulle användaren belönas med exempelvis stjärnor och sedan gå vidare till nästa hus med nästa tema.

Bild 4 visar en mer utvecklad storyboard och skiss över hur de första fyra sidorna av spelet hade kunnat se ut. Husen är uppritade längs en väg samt visas en idé om hur nutida karaktären ska få träffa på karaktärerna, i detta stadie har karaktärerna börjat kallas för *vänner.* En menyrad har även skissats upp.

Storyline

Storylinen till spelet kom till precis innan kurs 1 tog slut. För att ha en enhetlig storyline kände vi att det behövdes en koppling mellan inlärningen och varför användaren ska spela spelet. Vi ville även använda oss av en teachable agent såsom i *Historiens Väktare*. Vår TA är *Den glömska historieboken,* som också kom att bli spelets namn. *Den glömska historieboken* har som namnet tyder på, glömt all sin historia som han ska kunna och är därför mycket förtvivlad. Han ber användaren om hjälp att fylla i sina tomma sidor. Användaren kan då ta hjälp av vännerna från de olika tidsepokerna i historievärlden.

Till en början bodde *Den glömska historieboken* i ett litet bibliotek och målet var enbart att göra boken glad igen. Efter slutredovisningen av kurs 1 fick vi värdefulla idéer och tankar om hur en vidareutveckling av storylinen hade kunnat vara. Det behövdes mer motivation till att vilja fylla i *Den glömska historieboken* sidor. En teachable agent som spelaren skall lära upp har möjlighet att väcka känslor av ansvar och omhändertagande. Detta är något vi ville utnyttja i storylinen. Därav föddes idén till att *Den glömska historieboken* är förtvivlad för att han har blivit utkastad från världens mest kunskapsrikaste bibliotek, *Alexandria.* Han bor därför i en boklåda vid vägkanten och hans högsta dröm är nu att ta sig tillbaka in i biblioteket, men då måste självklart alla hans sidor vara ifyllda igen. Efterhand som sidorna fylls i av användaren blir *Den glömska historieboken* gladare och gladare och när han får lov att komma in i *Alexandria* igen är spelet avklarat och användaren har förhoppningsvis lärt sig mycket om historia på vägen.

Spelstruktur och dialog

Ett av de viktigaste stegen i designprocessen med tanke på projektets syfte var att komma fram till hur inlärningsprocessen skulle gå till. För att kunna börja skapa en prototyp av spelet krävdes en spelstruktur och en plan för hur dialogerna skulle genomföras. Varje hus i historievärlden innehåller som tidigare nämnt ett givet tema, och dessa teman lärs ut genom olika inlärningsspel i husen. Dessa spel är inspirerade av testen som finns i *Historiens Väktare.* Strukturen för de olika spelen i husen skapades i olika iterationer. Det första huset har temat vardagslivet, och därför började vi strukturera upp och välja ut information som kunde kopplas till just det tema. Spelet skulle gå ut på att dra information som presenteras i olika bubblor till rätt vän. Informationen strukturerades upp på det sätt som kan ses i bild 5.

Detta mynnade även ut i ett beslut om hur dialogen med karaktärerna skulle gå till. Som tidigare nämnt funderade vi på att lägga upp dialogen i spelet på ett lite annorlunda sätt än dialogen i *Historiens Väktare* där karaktärerna berättar sin historia och delger på så sätt information till spelaren. Vi ville istället delge information till användaren på ett mer interaktivt sätt och undvika envägskommunikation så mycket

Bild 2. Andra Idéskissen av världen

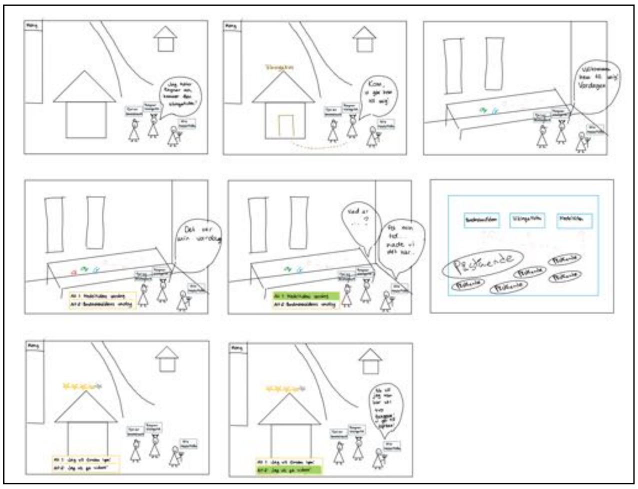


Bild 3. Storyboard

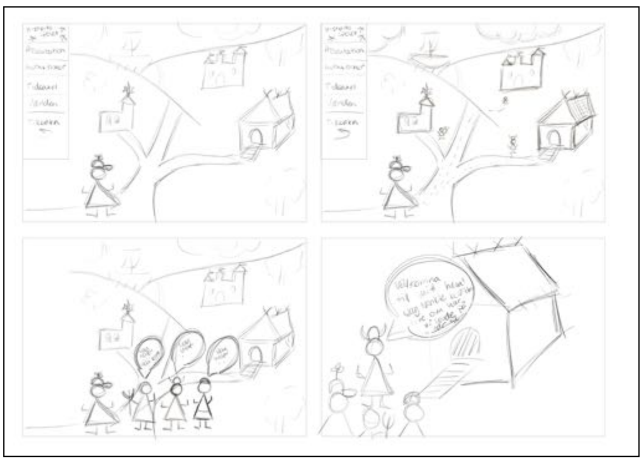


Bild 4. Skiss över de fyra första sidorna

som möjligt. I husen får användaren prova sig fram för att hitta rätt information och kan naturligtvis göra fel på vägen. Tanken är då att användaren får feedback från karaktärerna om vad som var fel eller rätt. En orsak till denna uppbyggnad är för att ta med variationsteorin i feedbacken genom att kontrastera de olika tidsåldrarna mot varandra. Till exempel kan feedbacken vara: *Nej detta har jag aldrig hört om* eller *detta var före min tid.* Variationsaspekten är i detta fall utvecklingen genom tiden.

Det första huset som kan besökas är vikingahuset, och det är här första temat - vardagslivet - presenteras. Det börjar med att alla karaktärerna samlas i huset och användaren ska sedan para ihop rätt tidsålder med rätt karaktär. Detta är en viktig grundpelare för att spelaren skall kunna koppla innehållet till de olika tidsepokerna och på så sätt förankra det. När karaktärerna är ihopparade med rätt tidsålder kommer ett antal olika påståenden som användaren på samma sätt ska dra till rätt tidsålder, drar hen fel kommer hen att få dynamisk feedback beroende på vilken tidsålder informationen blev kopplad till. Därefter presenteras ytterligare information inom temat. Detta är ett av flera testupplägg; tanken är att i varje hus ska olika inlärningsspel finnas för att få en variation i spelet. Övriga spelupplägg kom till senare i designprocessen, när vår hi-fi prototyp utvecklades.

En stor utmaning för spel inom historia är mängden information. Hur mycket innehåll kan presenteras på samma gång innan det testas? Miller (1956) menade att korttidsminnet kan behålla sju plus/minus två skilda enheter av innehåll på en och samma gång men vissa inom forskningsfältet anser att minnet är ännu mer begränsat, troligtvis mer så hos mellanstadieelever än fullvuxna individer. Med det i åtanke kommer spelarna sannolikt att behöva återvända till husen igen efter testet i historieboken för att hitta svar på de frågor som ställs i boken. Denna inlärningsprocess är i sig själv inte dålig, men spelarens självbild kan ta skada ifall hen upplever det konstanta behovet av att gå tillbaka som en indikation på hens egna okunskap. För att undvika denna frustration kan det vara en god idé att låta användaren ha tillgång till frågorna i förhand så att spelaren vet vad som hen borde leta efter.

Olika idéer lades fram angående testets upplägg. Bland annat att användaren skulle vara tvungen att göra ett riktigt test för att komma ut ur varje hus, alltså inte fick svara fel på någonting. Det uppstod dock en naturlig koppling mellan inlärningen och vår valda struktur på spelet när storylinen kom till. Efter varje hus testas inlärningen i *Den glömska historieboken* genom att användaren ska svara på frågor genom att trycka på rätt svarsalternativ. Varje tidsepok har ett eget kapitel och i varje kapitel finns alla teman representerade, se skiss över ett test i bild 6. Testen i *Den glömska historieboken* måste användaren svara rätt på, annars måste spelaren tillbaka till huset och göra om inlärningsspelet. I testen i *Den glömska historieboken* får användaren olika alternativ presenterade för sig, där enbart ett alternativ är korrekt, de övriga kan ha detaljer som är felaktiga. Det är användarens uppgift att urskilja det rätta svaret genom att lägga märke till detaljerna i svaren. Detta är ett designbeslut som uppmanar spelaren konsolidera och slå ihop den kunskap som hen lärt sig till ett enhetligt svar. Samma information som användaren nyss har fått lära sig i husen presenteras alltså i *Den glömska historieboken* men i omformulerad form.

Efter att ett tema är ifyllt i samtliga tidsepoker blir *Den glömska historieboken* gladare och ett nytt hus med ett nytt tema låses upp i historievärlden. Proceduren upprepas tills alla hus är upplåsta och tills alla sidor är ifyllda korrekt i *Den glömska historieboken.*

De två andra husen och dess teman kom till senare i processen i kurs 2 och under utvecklingen av hi-fi prototypen.

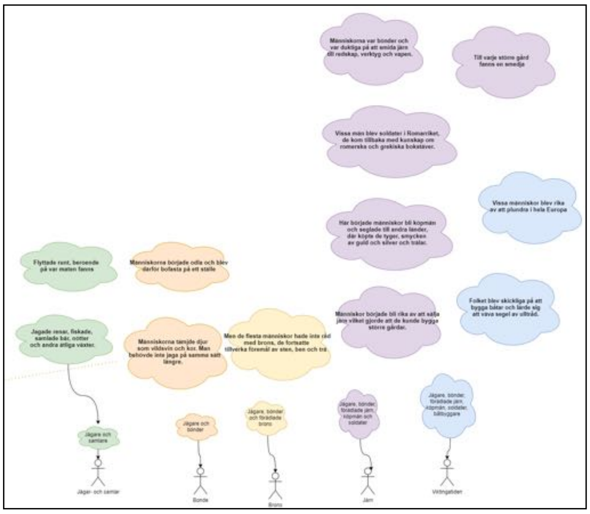


Bild 5. Information om de olika tidsåldrarna

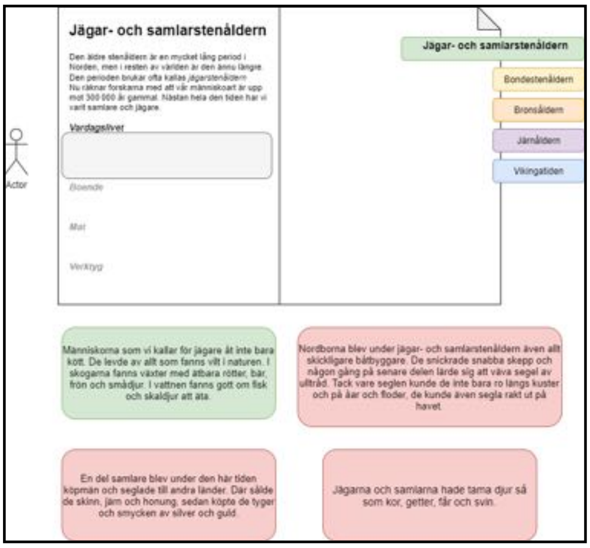


Bild 6. Skiss över test i historieboken

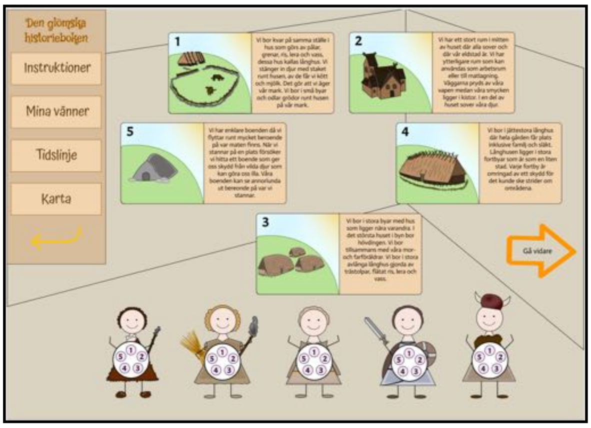


Bild 7. En annan typ av övning med visualiseringar i hi-fi prototyp

Det andra huset har temat boende och familj, där alltså användaren får lära sig om hur boende- och familjekonstellationerna skiljde sig åt under de olika tidsepokerna. Informationen presenteras på olika kort, med en bild av antingen familjen eller boendet i samband med en informativ text (se bild 7.) Vidare ska användaren para ihop rätt kort med rätt vän genom att trycka på rätt siffra som är presenterad framför vännen. Den bakomliggande tanken för att inkludera bilder

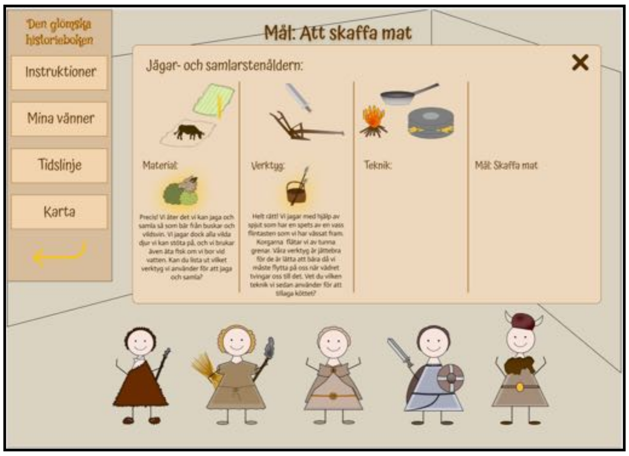


Bild 8. En övning för logik relaterat till verktyg i hi-fi prototyp

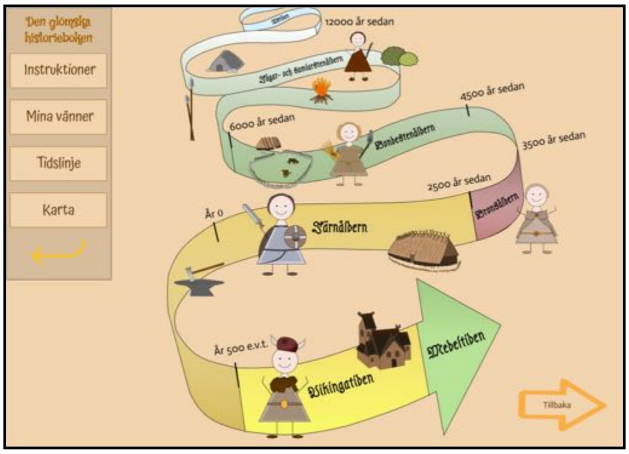


Bild 9. Tidslinje

var delvis för att göra innehållet mer intressant men även för att ge spelarna en visualisering som de kan förknippa innehållet med.

I hus nummer tre är temat verktyg och teknik. Användaren får olika mål presenterat för sig likt en matris med komponenter, se bild 8. De målen som i nuläget är implementerade i spelet är “Att skaffa mat” och “Att tillverka köksredskap”. Fler mål hade kunnat vara “Att bygga hus”, “Att skapa kläder”, “Att skapa verktyg och vapen” etc. Användaren ska för var och en av vännerna lista ut vilket material, verktyg och teknik man använde för att nå målet. I varje kategori får användaren välja mellan tre olika alternativ och väljer hen fel får hen feedback om varför det var fel.

Spelen i husen har utformats för att på olika sätt ta tillvara på någon av de variationsmönster som nämns i Bussey et.al. (2013, s.14-15). Spelet i hus tre har t.ex. utformats med mönstret *kontrast* i åtanke, där vi försökt skapa en situation där så lite som möjligt i den övergripande strukturen förändras för att på så sätt tydliggöra för eleven att det som förändras är just processen man måste gå igenom för att nå ett visst mål. Lärandeobjektet vi har försökt nå genom den här övningen är de förändrade förutsättningar som kommer med att tiderna förändras, och vi har valt att illustrera detta genom verktygen och processerna som krävdes för att ta sig igenom en vanlig dag med inskaffande och tillagande av mat.

Spelen i hus ett och två liknar varandra på det strukturella planet och fokuserar båda på att eleven lär sig skilja på olika lärandeobjekt genom att de presenteras samtidigt med varandra (Bussey et.al. 2013, s.11-12). När många lärandeobjekt görs tillgängliga samtidigt kan eleven använda sig av mångfalden av information för att upptäcka vad ett lärandeobjekt är genom att också identifiera vad det *inte* är. Eleven kan t.ex. lära sig om vad som definierade levnadssätten under jägar-samlarsamhället genom att se hur det skiljer sig från levnadssättet under t.ex. järnåldern.

Utformningen av spelen är ett av de ställen där expertis från yrkesverksamma lärare och/eller forskare hade kunnat underlätta utformningen och skapat ännu mer effektiva spel som tog ännu bättre tillvara på variationsteorin. Vi har utgått från en väldigt begränsad uppfattning om lärandeobjekten eftersom vi själva inte har någon större erfarenhet av varken historia eller undervisning. Hade man kopplat in en panel av lärare i åk.4-6 och intervjuat dem om vilka de anser vara de kritiska aspekterna för de lärandeobjekten som infattas i spelets mål.

Karaktärerna

Designen av karaktärerna har sin grund i bild 1. Vi valde att fortsätta designa dem som streckgubbar men med lite mer detaljer. Kläder och mindre detaljer har designats för att återspegla epoken som karaktären kommer ifrån, något som också blev ett sätt att urskilja dem från varandra. Karaktären från jägar- och samlarstenåldern har pälsliknande kläder och ett stenverktyg i handen. Bondestenåldern är lite mer utvecklad, och där ville vi försöka få in någon passande ledtråd vilket blev i form av en bunt med hö. Bronsåldern har fått sina kläder utsmyckade med bronssmycken, karaktären från järnåldern har fått ett svärd och en sköld för att ta upp utmärkande element från den tiden. Slutligen fick karaktären från vikingatiden som tidigare nämnt en vikingahatt och ett försök till vikingakläder.

Lo-fi prototyp

Lo-fi prototypen skapades i kurs 1 och består av skisser över de olika sidorna och den tänkta interaktionen. Prototypen är handritad användes främst för att konceptualisera våra idéer och för att ha något att få feedback på. På slutredovisningen i kurs 1 fick vi mycket positiv feedback och därav valde vi att fortsätta med koncepten som vi skapat i lo-fi prototypen och ta de vidare till en hi-fi prototyp. Lo-fi prototypen kan ses i bild 9 samt i fullstora skärmavbilder i bilaga 11-19.

Hi-fi prototyp

Prototypen är skapad i prototypverktyget Figma och saknar därmed all typ av logik som kan fås genom att hi-fi-implementering. På grund av begränsningar behövde prototypen anpassas efter dessa. Bland annat har delar av spelen behövt anpassas för att interaktionen överhuvudtaget skulle fungera. I detta stadie kan användaren enbart ta sig igenom prototypen med strikta instruktioner samt så är historieinformationen som presenteras begränsad. Fokus har istället legat på att visa konceptens funktionalitet, visualisera våra idéer och illustrera hur inlärningsteorier kan användas i historielärospel. Därför visas både vad som händer när användaren svarar rätt respektive fel i varje hus samt rätt eller fel i testen i *Den glömska historieboken.*

Spelet börjar med instruktioner om hur spelet går till och vad det går ut på och genom hela prototypen får användaren feedback och instruktioner. För att göra det extra tydligt vad användaren kan interagera med har *signifiers* tagits i bruk, t.ex. en illustration av en hand eller pilar för att rikta spelarens uppmärksamhet mot vart den ska klicka eller dra något. Dessa används enbart när någon ny funktionalitet presenteras för att framhäva nya *affordances* i spelet (Norman, 2013).

En av delarna som behövt anpassas har t.ex. varit TA-aspekten. Eftersom Figma saknar kodningsmöjligheter har den glömska historieboken betydligt färre av de egenskaper som beskrivs av bl.a Blair et.al (2007). En av de tydligaste av de här begränsningarna var att vi inte kunde göra historieboken särskilt aktiv själv och därmed tappade vi också en stor del av förtjänsterna med TAn i vår prototyp. Historieboken som den ser ut i prototypen ger feedback och säger t.ex. till spelaren när den har gjort fel och behöver spela om ett hus, men den interagerar inte med spelaren i någon högre instans, t.ex. en prov- eller testsituation där boken tar provet istället för eleven. Detta är något som med fördel kan tas vidare av en kommande grupp som arbetar vidare med projektet.

All visuell design är skapad med vektorgrafik. Utifrån våra skisser från lo-fi prototypen skapades historievärlden, *Den glömska historieboken*, alla vännerna och husen. Förutom att vara i historievärlden, inuti husen och inuti *Den glömska historieboken* finns det även en tidslinje som användaren kan besöka. Här presenteras vännerna i kronologisk ordning och med en tidsaxel för att användaren ska kunna sätta in tidsepokerna i ett annat perspektiv. Se tidslinjen i bild 10 (fullstora skärmavbilder av Hi-Fi-prototypen finns tillgängliga som bilaga 20-31). Detta är ett exempel på att presentera samma information i spelet på olika sätt vilket sannolikt bidrar med en berikad förståelse av innehållet.

Knappen i menyn som heter *Instruktioner* som bland annat kan ses i bild 8 är en funktion som gör att användaren kan få upp de instruktioner som hör till det test som hen håller på med om osäkerhet skulle uppstå för vad målet är med testet.

En sida som finns med i menyn men som inte har blivit implementerad är *Karta*. Här var tanken att användaren på ett enkelt sätt ska få en överblick över vilka hus som finns tillgängliga och därmed också vilka teman hen kommer att få lära sig. Detta känns speciellt viktigt om man utvecklar prototypen ännu mer och gör ett större spel med fler hus och teman.

Ytterligare en sida som vi ville implementera men som det inte fanns tid för var den som heter *Mina vänner*. Här var tanken att varje vän ska sparas och när ny information förekommer i spelet så skall denna information även synas under karaktärens profil. Detta för att användaren ska kunna gå tillbaka hit för att repetera vad hen har lärt sig och på så sätt underlätta inlärningen. En skiss av detta kan ses i bild 11.

Under utvecklingens gång har hi-fi prototypen testats med heuristisk utvärdering som testmetod i ett flertal iterationer. Fokus har legat på den feedback som presenteras, instruktioner och att det gick att ta sig från början till slut i prototypen utan problem. Detta för att upptäcka större fel, felformuleringar och uppenbara buggar innan vi testade användbarheten på personer utanför gruppen.

Som tidigare nämnt exkluderade vi den nutida karaktären på grund av praktiska skäl och även för att vi ansåg att användaren gör samma kontrast till nutiden med eller utan denna.

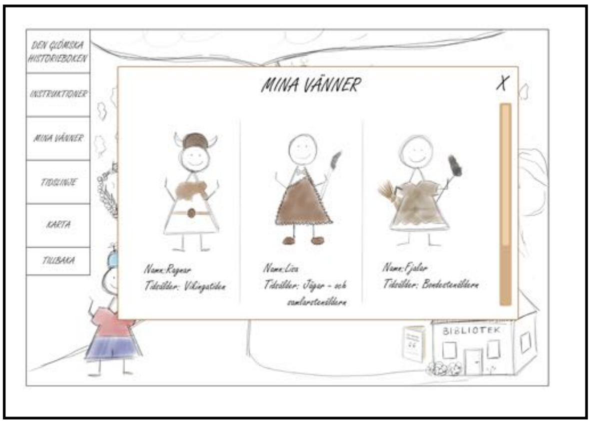


Bild 10. ”Mina Vänner”-sidan

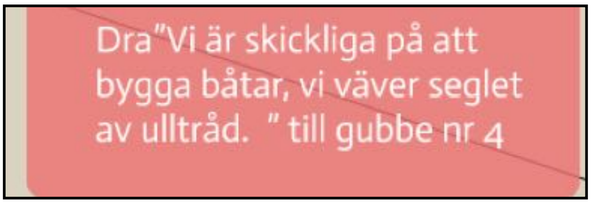


Bild 11. Exempel på instruktion

Bild 12. Exempel på instruktion

4 Testning

För att vidare undersöka hur prototypen tas emot av användare så testades prototypen. Testningen genomfördes efter att prototypen var färdigställd och syftade till att verka som grund för rekommendationer för prototypen och spelets framtida utveckling.

Då detta projekt varit tids- och resursbegränsat, delvis på grund av den rådande pandemin, valde vi att fokusera på att testa användbarheten i prototypen istället för att testa inlärningseffekten spelet hade haft på mellanstadieelever. Då projektet syftar till att testa variationsteorins möjligheter inom inlärning av historia så kan det upplevas som konstigt att vi inte i detta avsnittet har testning av inlärningen som huvudfokus. Detta motiverades med att testning av inlärning hade krävt både ett mer färdigutvecklat spel samt långvarig samordning med historielärare och deras klass. Dock följs denna rubrik av en diskussion där framtida möjligheter för testning av inlärningen diskuteras.

Syfte

Testa användbarheten i vårt utvecklade koncept för Den glömska historieboken, med hjälp av en klickbar high fidelity prototyp, för att kunna hitta potentiella förbättringsmöjligheter och rekommendationer för framtiden.

Frågeställningar

- Kan testpersonen navigera i gränssnittet på ett tillfredsställande sätt?
- Kan testpersonen ändamålsenligt följa det tänkta speluppläggett?
- Vilka förbättringsmöjligheter finns det i gränssnittet? (Saknas något?)
- Hur upplevs feedbacken som ges i de olika uppgifterna?

Urval av försökspersoner

Urvalet av försökspersoner fick också begränsas i omfattning, vilket resulterade i att vi dessvärre inte kunnat testa på den tänkta målgruppen, mellanstadieelever. Försökspersonerna har därför bestått av projektgruppens studiekamrater samt när-familj för att minska smittspridning under en rådande pandemi.

Metod

Testningen har genomförts enligt Wizard Of Oz-principen, där försökspersonen möts av vad den tror är en helt fungerande produkt, men i verkligheten är det en testledare, och i detta fallet ett digitalt verktyg, som styr interaktionerna (Dahlbäck et. al., 1993). Detta för att möjliggöra testning av de stegen som vi valt att implementera på ett bra sätt. Vi valde att bygga in testledaren i hi-fi-prototypen i form av tydliga instruktioner på vad försökspersonen skulle genomföra för steg i de olika spelen i totalt tre olika hus. Exempel på instruktioner som försökspersonen möttes av kan ses i två olika exempel i bild 12 och 13.

I samtliga hus möts försökspersonen av hur vi ser att feedback skulle kunna vara utformad vid såväl ett felaktigt som ett korrekt svar, exempel på dessa syns i bild 14 och 15.

När försökspersonen genomfört samtliga steg som prototypen innefattar och lyckats hjälpa Den glömska historieboken att bli en riktig historiebok så fick försökspersonen fylla i en enkät med följande uppbyggnad:

- Namn, ålder, sysselsättning, samt grad av tidigare erfarenhet av lärspel.

1. Vad är din generella åsikt om spelet?
2. Vad tyckte du om speluppläggett?
   2.1. Hur upplevde du Historieboken?
   2.2. Hur upplevde du vännerna?
   2.3. Hur upplevde du feedback:en i spelet?
3. Var det något du kände saknades?
4. Övriga åsikter.

Utformningen av enkätens frågor utgick ifrån ovan frågeställningar och syftade till att samla ihop försökspersonernas åsikter om prototypen.

Bild 13. Exempel på feedback vid ett felaktigt svar

Bild 14. Exempel på feedback vid ett korrekt svar

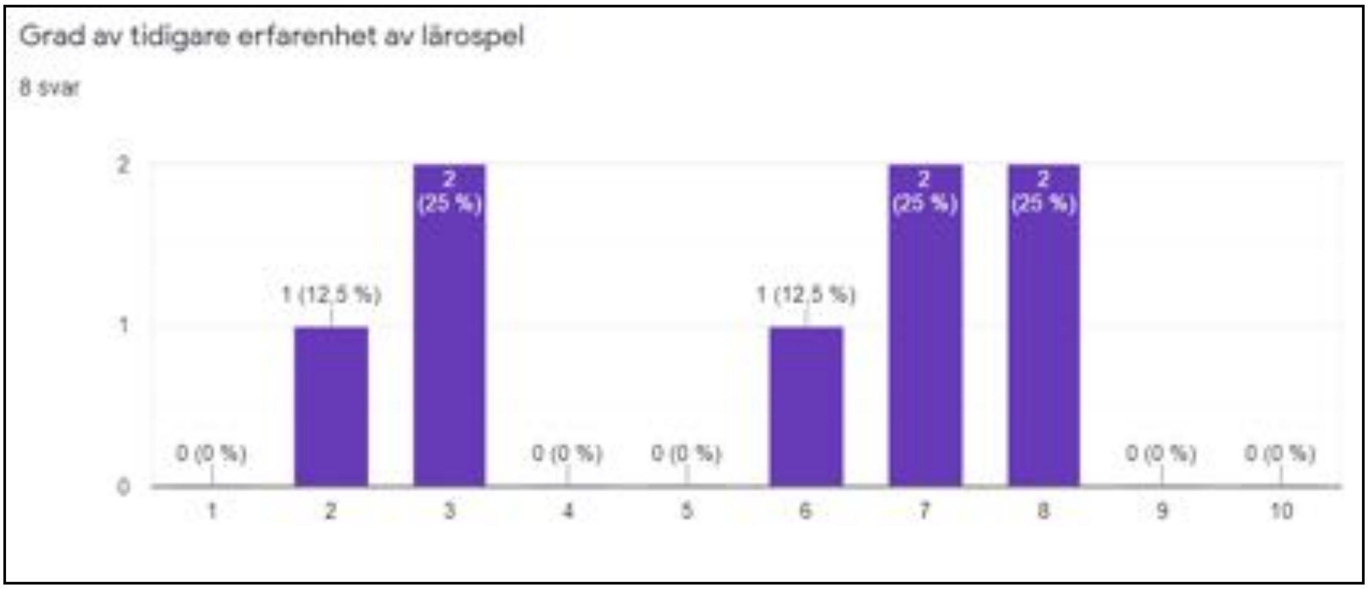


Fig.1. Grad av tidigare erfarenhet av lärspel

Resultat

Testningen genomfördes på 8 st försökspersoner, där 75 % av deltagarna var i åldern 24-26 år. Resterande försökspersoner va i medel- eller över medelålder. Graden av tidigare erfarenhet av lärspel var väldigt varierande, vilket skapade en omedveten, men bra, variation mellan försökspersonerna. Nedan följer resultatet av försökspersonernas åsikter, indelat efter kategorier.

1. Design

Den generella åsikten om spelets designelement var att allt var väldigt väldesignad och följsamt, med tydliga och fina illustreringar. Elementen upplevdes som lätta att manövrera, samt att de uppmanade till interaktion. En av försökspersonerna sammanfattade det som att ”Det var lockande layout som gjorde att man ville fortsätta spela och lära sig!”. En annan försöksperson påpekade att hen tyckte att vi arbetat med en bra symbolhantering och att det aldrig var tveksamheter kring vad som var vad.

Två av försökspersonerna tyckte att texten stundtals upplevdes som för liten, vilket kunde bidra till otydlighet i exempelvis momentet där man ska byta sida i historieboken. En annan påpekade att den hellre hade velat ha en ”okej”-knapp i instruktionssteget, istället för att som nu klicka på ”X”-knappen.

För att fortsätta på temat text så var det två försökspersoner som tyckte att de stundtals presenterades med för mycket text, med för mycket att ta in i varje steg, och att man därför lätt glömde saker. Deras lösning var att de gärna hade velat se att man som användare presenteras med mindre text i taget.

2. Spelupplägg

Samtliga försökspersoner tyckte att spelkonceptet var enkelt att förstå, där flera påpekade att det kändes informativt och lärorikt även för dem, trots att de inte tillhör den tänkta målgruppen. En återkommande åsikt var att spelet var roligt att genomföra och en försöksperson påpekade att om denna hade ”fått göra detta på skoltid hade jag blivit överlycklig.” En annan försöksperson sammanfattade den generella åsikten bra i ”Tycker själva berättelsen och historien är bra och rolig, vilket gör att man gärna vill fortsätta.”

Flera av försökspersonerna kommenterade att de gillade det tematiska konceptet, där jämförelsen mellan de olika tidsepokerna var något som sågs som väldigt positivt för att kunna se utvecklingen i historien. En kommentar kring detta var att det var extra bra att ” Lära sig att se tidslinjen i huvudet utan att lära sig det i kronologisk ordning.” Variationen mellan spelen i husen var också något som påpekades som positivt och motiverande för den spelande. En försöksperson såg potentialen att integrera spelet i ett lärandesammanhang, där hen påpekade att det framför allt är bra när elever vill jobba i sin egen takt.

Något som delade försökspersonerna i två åsiktsläger var det repetitiva upplägget. Några av försökspersonerna tyckte om att repetera i flera iterationer och uppskattade att man exempelvis var tvungen att gå tillbaka om man svarade fel i Historieboken. Dessa såg även potentialen i att lära sig genom repetitionsövningar, och en försöksperson framhävde att det var bra att man som spelare lärde sig genom att presentera den fakta man ”lekt” fram i husen för Historieboken. Av de som inte tyckte om det repetitiva upplägget, var det en som utvecklade sina åsikter ytterligare. Hen uttryckte sin förståelse för att man som spelare inte skulle kunna arbeta sig framåt i berättelsen genom att exempelvis gissa väldigt många gånger, och se mönster i det istället för att verkligen lära sig, men hade istället föredragit att ha en annan typ av lösning i form av ”i andan av "fail forward" mer än att repetera om”.

Åsikterna gällande historieboken var genomgående positiva. Flera påpekade att det gav en motiverade knuff att ha ett tydligt mål och skapade en empatisk känsla genom att man ville hjälpa historieboken att bli gladare och gladare. En av försökspersonerna påpekade att Historieboken var ett bra val av karaktär då den inte bara motiverade spelaren utan också kändes naturlig när text skulle presenteras.

Även vännernas roll i spelet gav alltigenom positiva reaktioner. Flera påpekade att det var ett bra interaktivt upplägg att ha en vän för varje tidsålder som sedan följde med en igenom spelet och att det gjorde att man som spelare relaterade den fakta man lärde sig till den respektive vännen. De flesta tyckte att utformningen illustrativt var bra genomförd och att det var enkelt att urskilja de olika vännerna, medan en försöksperson tyckte att de behövde vara mer distinkta i relation till sin respektive tidsålder för att enklare gå att urskilja. Personlighetsmässigt fick vännerna kommentarer om att de upplevdes som ”glada och informativa” och en försöksperson såg potentialen i att utveckla personligheterna ännu mer i framtiden.

Vad gäller feedback, mottogs även denna av försökspersonerna som välutformad och tydlig. Flera försökspersoner påpekade att det var bra att man som testare bemöttes av hur vi tänkt med såväl korrekta som felaktiga svar, och att det i svaren gavs mer information om vännen i fråga. Även här betonades den positiva aspekten i att kunna jämföra mellan vännerna. Tonen på feedbacken sågs som bra och positiv och två försökspersoner påpekade att tonen kändes uppmuntrande att försöka igen vid ett felaktigt svar. En försöksperson tyckte att svaren kunde varit lite mindre positiva vid ett felaktigt svar, men såg även poängen med att ha dem i en snällare ton för den tänkta målgruppen.

3. Utvecklingsmöjligheter

Den övergripande åsikten bland försökspersonerna för framtida variationer av spelet var att försökspersonerna såg utvecklingspotentialen och hade tyckt det varit kul att se fler delar av spelet, såsom andra hus, vad som händer i biblioteket, vänner från andra tidsepoker etc. En av försökspersonerna påpekade att hen ” Hade kunnat se det användas på antingen typ förskola eller ett museum!”. Några konkreta utvecklingsförslag var att det hade setts som positivt att ha någon form av ytterligare belöning utöver att göra Historieboken gladare och att ett välkommet inslag i spelet hade varit en förloppsindikator för att få en bättre överblick i hur långt man som spelare har kvar att klara av i exempelvis ett specifikt hus.

En återkommande åsikt var att det hade varit bra med ljudeffekter i spelet. Detta i form av exempelvis bakgrundsmusik eller mer naturliga bakgrundsljud för att skapa rätt atmosfär. Ett annat välkommet inslag hade varit att karaktärerna, såsom vännerna och historieboken, hade kunnat presenteras i både tal och skrift för att lägga till ytterligare en dimension där.

Slutsats

Sammanfattningsvis har återkopplingen från försökspersonerna genomsyrats av positivitet för vår prototyp samt vårt koncept och dess utvecklingsmöjligheter, trots deras variation av tidigare grad av erfarenhet av lärspel. Resultatet lever väl upp till och besvarar de frågeställningar vi ställde oss tidigare i detta avsnitt och när det gäller användbarheten kan den sammanfattningsvis ses som hög. Det som saknas i testningen är, som tidigare nämnt, testning av inlärningen, där vår tankar om potentiella utformningar presenteras i diskussionen nedan.

5 Diskussion

Det är intressant hur tidsaspekten förekommer så tydligt i t.ex. tidslinjen eller genom feedbacken från testningen i ett historiespel som är uppbyggt kring variationsteori. Man verkar kunna ha både ett "ickelinjärt" tidsperspektiv som grundar sig i variationsteori samtidigt som man har det mer traditionellt linjära perspektivet närvarande. Med det sagt så är variationsteori inom historia ett perspektiv som har stor potential. Användningen av variationsteori i historia är ett nytt fenomen som vi påbörjat att undersöka i detta projekt. Nedan kommer en diskussion om variationsteori och även tankar om dess framtid i *Den glömska historieboken.*

Variationsteori i historia, För- och nackdelar

För att variationsteori ska fungera inom historia bör innehållet först struktureras upp tematiskt för att identifiera händelser som kan kontrasteras mot varandra. Händelser som är lika i sin natur och tematik, men som varierar beroende på plats och tid då händelsen utspelade sig. En fördel med variationsteori inom historia är att det finns en ändlig mängd teman gentemot en oändlig mängd faktum. Därför hade det varit enklare för elever att få en överblick av det som de redan har lärt sig, och det som de inte än kan ifall historia presenteras ett tema åt gången. Detta är i kontrast till det klassiska kronologiska upplägget av historia som ibland kan upplevas som lösryckta historiska händelser förknippade med datum och namn. Detta tar oss till en annan fördel med variationsteori inom historia vilket är att historiekunskap blir bundet till teman i minnet istället för personer och tider. Det här är någonting som har potential att främja historiekunskap i vardagen hos gemene man då människor tenderar att snarare vara i tankebanor kring teman än historiska årtal eller faktum. Tänk samtalsämnen såsom vardagsliv, mat, kläder, verktyg, skolsystem etc.; dessa diskussioner hade troligtvis berikats av en historieundervisning som är uppbyggd i enlighet med dessa teman. På så sätt blir historia även mer av ett antropologiskt verktyg där kunskap från flera tidsåldrar hämtas för att sedan kunna bidra med ett rikare svar om frågor inom diverse teman.

Nackdelen för variationsteori är först och främst resurserna som hade krävts att sålla och sortera historia tematiskt. Men även ifall det hade gjorts är det inte sagt att alla faktum hade hittat ett fack eller tema som det passar under, eller kanske faktum hade passat under flera olika kategorier. Vad bör man göra i sådana fall? Detta är kanske något som man kan hitta en lösning eller kompromiss för ifall man tagit sig an projektet att strukturera upp historia tematiskt.

Historieämnets omformulering

Som tidigare nämnt behövs en grundlig omformulering av historieämnet äga rum för att kunna applicera det variationsteoretiska perspektivet ordentligt, både i en undervisningssituation och för spelet såsom vi utformat det. I grunden handlar den här omformuleringen om vilka lärandeobjekt man vill göra tillgängliga för eleverna genom spelet. Om historieundervisningens fokus ligger på att eleverna ska lära sig faktabaserad kunskap om när vissa händelser ägde rum eller liknande så kan det blir klurigt att få in det variationsteoretiska perspektivet.

För att lyckas med ett lärspel som ska fungera som ett hjälpmedel i historieundervisningen krävs en djupgående kunskap om ämnet, något som vi i gruppen alla saknade. Vi har kunnat utforma ett välfungerande ramverk med stora utvecklingsmöjligheter, men för att spelet och uppgiften ska kunna lösas på bästa sätt behöver det finnas en medvetenhet om lärandeobjektet redan innan man börjar formulera sina frågor.

Som exempel kan vi ta vårt minispel om att göra verktyg och skaffa mat, det här spelet hade kunnat se helt annorlunda ut beroende på vilken typ av lärandeobjekt man vill nå med hjälp av spelet. Vi har utgått från att försöka belysa ytliga och relativt tydliga skillnader mellan tidsepokerna då dessa på bästa sätt genererar kontraster och fungerar som effektiva illustrationer av våra idéer, men det är möjligt att de inte är de som är allra mest intressanta för varken lärare eller elever. Man skulle kunna uttrycka det som att man behöver ändra historieämnet inifrån för att kunna åstadkomma ett spel, eller annan undervisningsform för den delen, som har ett tydligt variationsteoretiskt perspektiv.

Just ett spel kan vara en applikationsform som lämpar sig bra för just variationsteoretisk applikation då det tillåter eleverna att få tydligt paketerade kontraster presenterade för sig, något som kan tillgängliggöra lärandeobjekt som annars hade varit svåra att presentera samtidigt. I en spelkontext är det mindre oväntat eller konstigt att få information presenterat för sig på ett mer fragmenterat eller icke-kronologiskt sätt, framförallt om spelet utspelar sig i en öppen spelmiljö där spelaren får göra egna val, något som är till *Den glömska historiebokens* fördel då det sänker tröskeln för eleverna att acceptera ett nytt och kanske bitvis icke-intuitivt perspektiv på ett ämne som både de och antagligen också deras lärare ser som tydligt och klart linjärt med en grund i en "objektiv" tidslinje.

En sista reflektion innan vi vänder blicken framåt mot framtida utvecklingsmöjligheter för *Den glömska historieboken* bör göras med avseende på möjliga teoretiska överlappningar. Om det skulle visa sig, t.ex. vid mer utförlig testning, att elever eller lärare inte anser att det variationsteoretiska perspektivet är tillräckligt eller tillfredsställande för att ta till sig den informationen som står i läroplanen skulle det kunna kombineras med begreppet om *threshold concepts.* Threshold concepts har en snarlik position inom ämnen som anses svåra att kombinera med variationsteori, t.ex. samhällsvetenskap och historia. Threshold concepts är begrepp som kan fungera som en portal eller en tröskel in till en djupare förståelse för andra begrepp (Meyer et.al., 2008).

En möjlig utvecklingsmöjlighet hade varit att försöka kombinera variationsteorin och threshold concepts i ett spel såsom *Den Glömska Historieboken* för att på så sätt kunna täcka upp så mycket av materialet som möjligt och presentera det på ett sätt som maximerar möjligheterna för eleverna att förstå de centrala begreppen och lärandeobjekten inblandade i ämnet.

Framtiden

Om vårt projekt hade fortlöpt under en längre tidsperiod finns det vissa områden som vi hade velat utveckla mer. För det första hade vi utökat antalet hus och därmed haft med fler teman inom tidsepokerna. En följd av detta hade varit att utveckla ytterligare spelstrukturer och lärspelet i sig hade visat en större bredd. Vi hade även implementerat de sidor som presenterades under designprocessen men som det ej fanns tid för.

Efter vår användbarhetstestning fick vi som tidigare nämnt värdefull feedback och konstruktiv kritik som omvandlades till utvecklingsmöjligheter och som vi hade velat implementera i prototypen om tid hade funnits. Bland annat hade en utveckling av storylinen gjorts för att få in ytterligare belöning för användaren än att Den glömska historieboken blir gladare. Ett alternativ hade varit att främja lärandet genom att visualisera och demonstrera den kunskap som eleverna redan har lärt sig genom spelets gång via olika narrativ som knuffar på att de skall vara stolta över deras framgångar. Som ett resultat av testningen hade vi även implementerat ljud och musik i spelet då det ansågs önskvärt av flertalet av våra testpersoner. Att utnyttja ett till sinne såsom hörsel ses fördelaktigt enligt multimediaprincipen (Mayer, 2014). En annan fördel kan enkelt nog vara att ljud och musik kan främja en känsla av lugn och därmed minska stress.

En aspekt av TA-strukturen som *Historiens väktare* använde men som vi inte hann designa var en "lärare" som rättar svaren som spelaren sätter in i historieboken. Denna karaktärs funktion hade varit att ge feedback indirekt till spelaren genom TA'n. På så sätt minskar chansen att eleven bemöter feedbacken defensivt eller som ett hot mot hens självkänsla. Detta är något som det bör läggas vikt vid i framtida utveckling.

På grund av resursbrister i form av tid för spelutveckling gjordes enbart en användbarhetstestning. För att verkligen testa inlärningen och för att se kontrasterna av att använda variationsteori gentemot traditionell metodik hade givetvis ett test av inlärning behövts göras. Därför behöver lärspelet utvecklas mer och sedan genomgå ett större test med exempelvis en skolklass. Något vi även hade som mål men som inte hanns med var att kontakta lärare inom historia för att få deras infallsvinkel på vårt koncept och vår idé. Med facit i hand var detta något vi gärna sett att vi hade gjort från början för att tidigt få stöd i utformningen av vårt lärspel.

Vidare hade det inför framtida utveckling krävts mer grundläggande och gedigen kunskap om vilken typ av historia som mellanstadieelever ska lära sig samt noga gå igenom Skolverkets läroplan för att anpassa innehållet mer exakt efter vad eleverna ska lära sig. I det här projektet är det inte det som har varit i fokus då vi mer har fokuserat på den teoretiska aspekten och designprocessen.

Tidsfördelad repetition var även något vi helst hade velat implementera mer i självaste spelet men som vi inte hann med. Vi ansåg det viktigt med tanke på att spelet är väldigt informations tungt, och då det hade varit ett ypperligt sätt att följa upp hur mycket eleverna har lärt sig och faktiskt kommer ihåg. Tanken var att beroende på hur bra eleverna svarade i själva historieboken så skulle kunskapen återigen hämtas upp mer eller mindre i sporadiska intervall under spelets gång. Till exempel, efter att spelaren klarat av de första tre husen så skulle spelaren få möjlighet att återigen besvara några av frågorna från det första huset. Med detta upplägg hade eleverna fått metaförståelse över vad de faktiskt kommer ihåg. Vi hade kunnat mäta variationsteorins inverkan på elevernas kunskap och även lärarna hade fått ett verktyg för att se vad eleverna kan eller har svårt med.

6 Slutsats

Utifrån vårt arbete med *Den glömska historieboken* har flera framsteg för att testa variationsteori i historieundervisning gjorts. Inom tidsramen för detta projekt hann inte konceptet utvärderas på målgruppen mellanstadieelever. Däremot så har effekten av variationsteori diskuterats och demonstrerats genom prototypen och i denna uppsats. I processen av att utveckla spelet har vi stött på motstånd men även flera insikter till möjligheterna som variationsteori har att erbjuda till historieundervisning. Teman inom historia går klart att identifiera och urskilja, någonting som inte var självklart när vi började projektet. Den tematiska uppdelningen har potential att främja gemene mans användning av historia i vardagliga diskussioner. I sin helhet tror vi att variationsteori kan bli ett viktigt komplement till traditionell undervisning. Vidare forskning behövs för att mäta variationsteorins effekt på inlärning inom historia.

Det stora problemet som framtida arbete med *Den glömska historieboken*, och i förlängningen även annat arbete med variationsteori inom historia står inför är hur man ska formulera ämnets lärandeobjekt. Detta måste ske på ett sätt som gör dem hanterbara i en lärandesituation och som hjälper till att skapa en djupare förståelse för historieämnet för de elever som använder perspektivet och materialet. Detta kan med största sannolikhet hjälpas genom både praktisk testning i skolmiljöer, men också genom samråd med lärare och andra ämneskunniga för att på så sätt få en titt “under huven” på historieämnet.

Referenser

Blair, K., Schwartz, D. L., Biswas, G., Leelawong, K. (2007) Pedagogical Agents for Learning by Teaching: Teachable Agents. *Educational Technology, Vol.47, No.1,* 56-61

Bussey, T., Orgill, M. and Crippen, K., (2013). Variation theory: A theory of learning and a useful theoretical framework for chemical education research. *Chem. Educ. Res. Pract., 14(1)*,.9-22.

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,* 334-352.

Dahlbäck, N. Jönsson, A. & Ahrenberg, L. (1993) "Wizard of Oz Studies - Why and How", *Proc. 1st Int'l Conf. Intelligent User Interfaces,* pp. 193-200.

Guo, J-P., Pang, M. F., Yang, L-Y., (2012). Learning from Comparing Multiple Examples: On the Dilemma of “Similar” or “Different”. *Educational Psychological Review (24)*, 251-269

Kullberg, A., Runesson Kempe, U., Marton, F. (2017) What is made possible to learn when using the variation theory of learning in teaching mathematics?. *Mathematics Education (49),* 559-569.

Körner, G., Lindberg, P., Götlind A (10 Augusti 2018), *Digilär Historia för årskurs 4-6,* hämtad från https://app.digilar.se/

Ljunggren, P., Frey-Skött, F., (2016). *Upptäck Historia.* Stockholm: Liber, pp. 4-108.

Mayer, R. (2014). Cognitive Theory of Multimedia Learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Cambridge Handbooks in Psychology, pp. 43-71). Cambridge: Cambridge University Press.

Meyer, J.H.F., Land, R. and Davies, P. (2008) Threshold concepts and troublesome knowledge (4): Issues of variation and variability, in: Land, R., Meyer, J.H.F. and Smith, J., (eds), *Threshold Concepts within the Disciplines*. Sense Publishers, Rotterdam, 59-74,

Miller , G. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review , 63 ,* 81–97 .

Norman, D. A. (2013). T*he Design of everyday things,* New York: Basic Books.

Smolen, P., Zhang, Y., & Byrne, J. H. (2016). The right time to learn: mechanisms and optimization of spaced learning. *Nature reviews. Neuroscience, 17(2),* 77–88.

Ternblad, E. M., Haake, M., Anderberg, E., & Gulz, A. (2018). Do preschoolers ‘Game the system’? : a case study of children's intelligent (Mis)use of a teachable agent based play-&-learn game in mathematics. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10947 LNAI,* 557–569. https://doi.org/10.1007/978-3-319-93843-1_41

Panoptis

Future UI for Pilot Training With Eye Tracking

Jesper Brodin Johannes Person Melina Tsapos Mergim Rruka

vov15jbr@student.lu.se bas15jpe@student.lu.se fte15mts@student.lu.se me7076rr-s@student.lu.se

Measuring eye movement has proven to be a very powerful tool in human problem solving. That is why there is great potential for future pilots and their instructors to benefit from implementing eye tracking into their aviation training. Our task has been to find a solution to how best make use of the eye tracking data, and to optimize its use in the simulator cockpit. With design principles in mind, as well as an insight into the instructors needs, we have developed a program called Panoptis. This report describes the steps in the developing process.

1 Introduction

It has long been know that there is a relationship between eye movements and problem-solving. It reflects what we are thinking. But recent research has not only shown that eye movements actively affect our cognitive processes, it may also influence how we think (). By measuring and analysing where people look as they perform tasks, we can gain insight into the cognitive processes involved. More than that, it is also believed that by directing eye movements of test subjects there is potential to affect the participants' ability to solve problems (eye, 2021). Utilising this research coupled with modern eye tracking devices, could be very powerful in many learning tasks and human problem solving situations. One such environment is in the simulator cockpit of an aircraft at aviation training facilities. Gaining insight into eye movement and gaze behaviours from the expert pilots and pass on the knowledge in the training of future pilots could potentially be a very useful tool for the instructors during training sessions and, as a result, improve flight safety for everyone.

Using eye tracking during pilot training sessions could give a great insight into pilots visual behaviour (Niehorster et al., 2020). The instructor is usually seated slightly behind and between the two pilots in training. From there he or she is keeping a watchful eye on the various movements and interactions the pilots have in the cockpit. However, from this vantage point the instructor can only assume that the student pilots have checked whatever was necessary before call-outs and actions taken. One challenge for the instructor, not being able to see the eye movements of the pilots, is to assess the decisions made (or not made). Did the pilot fail a particular task because he or she did not monitor the instruments correctly? If eye tracking was used in pilot training and performance assessment tasks the instructor could gain great insight into the pilot's work in the cockpit. Potentially, it would be useful as a real time aid while in the simulator cockpit, in flight, but also in the debriefing phase.

Our team's task has been to develop a user friendly interface in order for the instructor to be able to use the information from the eye tracker in an efficient and precise way. This of course has a number of challenges. Simply using eye-tracking data would obviously be much too time consuming and not very useful during a training session for the instructor, who already has a great cognitive load keeping track of everything that is going on. In other words, the availability of eye-tracking data does not necessarily lead to useful applications for the data. Our goal has therefore been to turn eye-tracking data into information that is useful for the instructor, in order to improve the pilot's performance. And thus realize the potential of incorporating eye-tracking as a support tool in pilot training.

2 Theory

One of the, if not the most essential (and obvious) principles in good UI and UX is to keep the user in mind. That's why we have adapted an approach that puts the human needs, capabilities and behaviour first so that our project can be a powerful tool presenting complex information in a design with good communication from technology to person. A human-centered design should start with a good understanding of human cognition and the needs that the product intends to meet. However, the theoretical ideas behind this approach have developed from two schools of thought, symbol system and situated actions (Lave, 1988; Norman, 1993; Newell, 1972), that are perhaps best seen as complementary rather than opposing; at least for our purposes in designing the interface for future pilot training.

2.1The symbolic and situated actions approach

The traditional symbolic approach to the study of human cognition and intelligent systems is the view that a physical symbol system interacts with its external environment by sensory stimuli that it converts into symbolic structures in memory; and it acts upon the environment in ways determined by symbol structures that it produces. An information system can take a symbol token as input and use it to gain access to a referenced object in order to affect it or be affected by it in some way. Symbols may designate other symbols, which is the case for example of a cockpit dashboard and or design interface using symbols symbolizing them. This is a process of encoding sensory stimuli into internal symbols. We will discuss a few examples where we have inferred an inner symbol system keeping conceptual models in mind.

Symbol systems are structured in memory representations of external situations. And the representations can be manipulated in a way of planed action. For example, the internal representation of the instructor may be highly incomplete or even inaccurate compared to the actual situation in the world. With the symbol system in mind, when developing interfaces a focus on how people use them and not just on how they think or expect the program to work is also important. This is usually the so-called "situated action" view.

This view criticizes the symbol system approach for focusing too narrowly on what is in the head without concern for the relation between the intelligent system (the instructor) and its surrounding (being in a simulator cockpit). In short they don't think internal states have causal effect on behavioural output (Suchman, 1987). So, a lot of emphasis is put on the plasticity and adaptability of humans as problem solvers. Rather, the behaviour is fully determined by the contextual situation in which it is elicited. To describe situated action simply, Gedamer does it rather well:

> "The very idea of a situation means that we are not standing outside it and never are unable to have any objective knowledge of it. We are always within the situation" (Gadamer, as quoted in Vera and Simon (1993), p. 12)

Winograd and F. (1986) argued that the most significant challenge facing interface design is to discover the true ontology of human beings with respect to computers. And so when designing, the study and understanding what they are used for and what problems are encountered in their use is what makes good design. Transparency of interaction is one way to help the user, and as such feedback is an essential part of it. The design should ideally allow the instructor to work immediately at a high level without needing to know anything about the details of the program.

Having considered both these approaches, we find it more useful for us to consider them complementary at least for our purpose trying to optimise the humans experience working together with the program as a tool. Intelligence could be considered fundamentally a property of appropriately programmed symbol systems, the symbol system operating adeptly in real time in a complex environment, such as a cockpit. However, it is crucial that it be studied with the human in situated actions. We believe our design process has greatly benefited incorporating ideas from both of these approaches.

3 Design guidelines

Equipped with the theoretical ideas, we closely studied the guidelines presented by Niehorster et al. (2020). It is argued that in order to best visualize eye-tracking data and meet the needs of the instructor simple gaze replays will not be especially useful in most training and assessment applications. They are simply not a good way to present the information to the instructor in a commercial pilot simulator. There are a number of reasons.

For example a problem with direct gaze displays in real time is that gaze location change is hard to keep up with (Niehorster et al., 2020). The flight instructor must make the judgment whether a pilot has looked enough, and this in turn strains an already high memory load. The instructor would need to aggregate all the fixations (three or more per second) into a tally of gaze per cockpit element, interpret the task and derive insight into the pilot's behaviour. Not many instructors would be convinced this is very useful if that much extra work and attention was required of them.

There have been efforts made by other researchers to develop a way to utilize eye tracking in pilot training assessments. One such effort is by Emirates so-called "Candy-bar". It provides a very intuitive way to show what the pilots are looking at in real time and one can make an instant comparison between the pilot flying (PF) and the pilot monitoring (PM) gaze behaviour. However, some crucial things are missing in order to be an optimal tool for the instructor. To mention only one example where a design principle of good conceptual mapping would improve the system, consider that they have chosen to color code the different cockpit elements. This indeed makes it a very cheerful display with all the colours of the rainbow, but lacking the fundamental idea of conceptual mapping, the colors are completely arbitrary, and other than the blue which represents the wind shield (because the sky is blue?), the others just need to be memorized to know what they mean. The colors do not project information in any intuitive fashion.

3.1Conceptual mapping

Conceptual mapping should follow a natural mapping (Norman, 2013). As such, the design should take advantage of spatial analogies which would lead to an immediate understanding and doesn't require extra learning for the instructor. The conceptual models in people's minds represent their understanding of how things work and what the signifiers indicate. Using icons is one way that would be more appropriate and be more effective mapping, for the instructor to easier keep track of the different elements represented in a cockpit. Conceptual mapping is also essential to enhance and evaluate the result. For us it meant that we had to think about what is natural for the "culture" or the industry (i.e aviation) specific practice when choosing the color coding to communicate the information to the instructor.

This was also a concern raised by Niehorster et al. (2020) stating that "the level of uncertainty about actual gaze position (and inference about attention location) [should be] obvious to avoid misinterpretation" (p. 21). So, for example, if we had chosen the color red as a signifier when the eye tracker did not record a fixation, perhaps that would have implied a fatal error for an aviation trained instructor; when in fact the pilot might have checked an alternative element or perhaps the data was obtained due to the inherent inaccuracy and imprecision of the equipment used. In cases like these red could be a much too alarming indication and does not leave the interpretation of the situation open for the instructor to make the final assessment.

3.2Signifiers and Feedback

Although one should expect some error and some confusion, we have nevertheless taken extra pains to design the interface such that these kinds of errors would be as cost-free as possible. We have strived to minimize the chance of inappropriate actions in the first place by the use of signifiers, good mapping and constraints that hopefully will make sense even to the first time user.

This requires intelligent feedback from the program to the user. The feedback must be immediate, since a delay would be extremely annoying and uninformative. Too many announcements would perhaps cause the instructor to ignore them, which means that critical and important information is apt to be missed. The user/instructor should at all times understand what has happened and what state the program is in, and what the appropriate action to take is. Then he or she can undertake his or her task more efficiently, and it will truly be a tool for the instructor, and not an obstruction.

With the philosophy of human cognition and the design principles in mind we aimed to design a system that could represent the eye tracking data at a higher level of abstraction in an intuitive and clear way for the instructor, in order for him or her to incorporate the information and make Panopti a part of every future pilot training session.

3.3The 10 Usability Heuristics for interaction design

Because of the time limit of the project and the restrictions of COVID-19, the usability testing was not performed. The usability testing would have needed Panoptis to be tested in the training cockpit and such thing would not have been possible to be done due to COVID-19. The usability testing was not done with the implemented version of Panoptis. To determine whether the implementation would be useful in a real case scenario, a full useability testing needs to be performed and analysed. However, the usability test could still be performed by using the test plan described in section 5.2. Since the useability testing could not be performed we can choose a different approach on evaluating the design of Panoptis. We chose to analyzing the 10 principles of Jakob Nielsen, as mentioned in section 3.3.

Jakob Nielsen has come up with 10 principles for designing user interaction and interfaces. The principles are not specific guidelines but give a general idea on how interaction design could be done. Because of the principles broadness, they overlap with some of the guidelines already mentioned in the sections above. The principles are (Nielsen, 2020):

1. Visibility of system status - The design should always allow the user to know what is happening in the system. This can be achieved through well designed feedback.

2. Match between system and the real world - terms, icons and concepts being used in the design should be familiar for the user. The user should not have to learn how to use the system.

3. User control and freedom - when user makes a decision by mistake, they should be able to return or undo without much effort.
4. Consistency and standards - Industry conventions should be followed so that the user does not have to wonder what different things could mean.
5. Error prevention - Prevent the user from making mistakes, this can be done by introducing constrains in different forms.
6. Recognition rather than recall - Do not force the user to remember information, they should recognize in instead. Offer rather help than forcing them to remember how to use the product.
7. Flexibility and efficiency of use - the product needs to be flexible enough to offer personalising for the experienced users.
8. Aesthetic and minimalist design - Only relevant information should be presented, otherwise it will compete with the information that does matter and thus distract the user.
9. Help users recognize, diagnose, and recover from errors - the users should be able to understand error messages.
10. Help and documentation - even though it is not desired to have documentation on how to use a product, if there is need for documentation, it should be easily accessible for the user.

3.4Brainstorming

Brainstorming is a method used to come up with new and different ideas or concepts to target a problem or a task. This method is used by groups of designers and while using this method there are different rules to have in mind, such as: "quantity over quality", "weird ideas are welcome", "build on each other's ideas" and more of the same kind. These rules act as a guideline when coming up with ideas and solutions, they promote creative thinking and sharing of ideas (B and B, 2012a).

3.5Storyboard

Storyboards are a series of illustrations that show how a user might use the product that is being developed. The illustrations might show different parts of the product in use, this technique also helps to come up with new and different designs from the beginning of the development (B and B, 2012b).

4 Process

This part of the report explains the process of the project, what was done, in what order and the prototypes that were produced.

4.1Iteration 1

4.1.1 Brainstorming session 1

The first session of brainstorming was combined with storyboards to come up with scenarios of how the dashboard would be used by the supervisors. Different use cases were created, the idea was that the supervisor starts with selecting the training session that the pilot student has to complete. Making sure that the training session contains all the different scenarios, fill in information about the student and proceed to start the training session, see figure **??**.

The next step would be the training session itself that would show all the different tasks of the training session and their status. Meaning that if the student has completed them, they will be marked in some way. A task contains different activities, however theses are not displayed until a student has failed to complete them. A timeline of the different parts of the training session is shown below the tasks list. See figure **??**. To clarify the hierarchy of a training session see figure **??**.

At any time the supervisor would be able to see why the task was marked as it was, by clicking on it and a video would show when that decision was made. Figure **??** shows the player that the supervisor would see when clicking on an task.

4.1.2 Lo-Fi prototype 1

The following prototype was created after the first brainstorming and storyboard session. Figure **??** shows the start screen that the user encounters when starting the application and the exercise selection screen when choosing to start an exercise. Figure **??** shows the screen before the training session and during the training session.

4.1.3 Learnings from prototype 1

From this iteration we learned that our first prototype lacks the ability to quickly show the supervisor what has gone wrong when something has gone wrong. The color marking would only indicate that something has happened but the supervisor would be forced to click on the task to get more information. This was the main point that was taken over to the next session of brainstorming.

4.2Iteration 2

4.2.1 Brainstorming session 2

In this session we focused on the learnings from the previous iteration. Moreover, the implementation of the application was taken into consideration and discussed as well. Since there can be many aspects to the implementation we tried to come up with ideas on how it might work and what would be needed.

While discussing the problem from prototype 1 we also started to look into the Standard Operating Procedures (SOP) and how the tasks might look like. A SOP contains a lot of small tasks (activities) that have to be fulfilled, meaning that the list in the dashboard with the different tasks will be long. To solve this, we decided to chunk the tasks into slightly bigger tasks that will be shown in the list. To give an overview of the status of the whole list, an scrollbar was added that also act as an indicator. Meaning that the scrollbar will be green as long as everything has been done correctly. If something has been done wrong an indicator will show up in the scrollbar and when click on it will take the supervisor to the task that was done incorrectly. Icons next to the task will be used to quickly indicate what the student has missed to do. See figure **??**.

4.2.2 Lo-Fi prototype 2

Figure **??** shows the second iteration of the prototype. This was based on the brainstorming session mentioned earlier. The biggest changes were the addition of the scrollbar/status indicator and the replacement of the timeline of the training session.

4.2.3 Learnings from prototype 2

From this iteration learned that we the movement of the tasks will catch the attention of the supervisor and we have to be careful with what becomes highlighted. If a task has been passed it means that it does not need attention than the task should not be marked in any way. Only the tasks that need attention need to be highlighted.

After showcasing this prototype for a supervisor at the School of Aviation in Ljunbyhed we realized that the bright colors of the prototype do not fit in the environment where the dashboard would be used. The cockpit for the training session is dark and dark version of the dashboard would be a better fit. From this meeting we also understood the importance of being able to restart a scenario. If a student has recurring problems with a part of training session and that a debriefing feature should exist. The debriefing tool would be used to give feedback to the students and help them improve till their next training session.

4.2.4 Lo-Fi prototype 2.1

As mentioned earlier this version was created based from what we learned from our visit at the School of Aviation in Ljundbyhed. See figure **??** for the updated version of our prototype. The colors were based on what are used in the cockpit. Black was used to match the dark cockpit, orange in the cockpit represents that something is happening or something is not okey and that is why it was used to indicate when the student missed something. The green is used in the cockpit to indicate that everything is okey and that is why we used it to check off everything that the student has done right in the scenario. In this way there is a mapping between reality and the software for the supervisor.

Testing of Lo-Fi prototype 2.1 The testing of Lo-Fi prototpye 2.1 was done according to section 5.1 and after testing it gained a lot of new insights and got very useful feedback from the test subjects (the instructors at the aviation school in Ljungbyhed). One major thing that we soon realized needed a slight update and improving was the conceptualization of the timeline at the top of the screen. Some of the participants mapped this onto the current task displayed. So, in order to make it more clear that the timeline in fact tracks the whole training session, not only displays the past stage, present, and the next upcoming stage in the session, we had to go back to the drawing-board and brainstorm. We then made a decision to display all the sessions that have passed. In other words, all the past events will always remain visible and therefore tractable from the activity screen. However, only the first coming session will be visible at any one time.

In general Panoptis was very well received by all test persons. The most common reactions we received from the test persons was surprise of how simple the tool was and how useful it could be for them.

All the feedback made us realize that a requirement specification will be needed before starting to implement Panoptis. The specification can be found in appendix B Disclaimer: The requirement specification was primary used at the start of the implementation phase and should not be seen a complete requirement specification for the entire system or future implementation.

5 Testing

This section will cover the testing processes for Panoptis and what the testing will be used for.

5.1Initial testing

The main goal of this testing was to get feedback on the interactions in Synpotis since the solution contained interactions and feedbacks in unusual ways and therefore needed to be further tested before being implemented.

5.1.1 Introduction and permission to record

The script below will be read to the test persons before the testing begins.

"Hi, we are a group of students from LTH doing a project on how eye-tracking could be used during pilot training session. The purpose of the testing is to observe how you interact with our latest prototype of our application - Panoptis. We are going to record the testing, however, it is only going to be used for our own purpose and not be published anywhere. The recording will be deleted after the project is done. Is this okey for you?

As mentioned earlier, we are only testing the prototype of the application, that means that all the elements in the screen are not clickable. To make it easier, we have prepared 3 tasks that we are going to read to you step by step. Remember that there is no right or wrong, we are only testing the interaction. Afterwards we would like to ask a few more questions about the interaction."

5.1.2 Test persons

The test persons for this test are supervisors from the School of Aviation in Ljungbyhed. There were four adult supervisors, three male and one female. The average time that each participant spent doing the tasks was 3 minutes and 2 seconds, with the longest at 5 minutes and the shortest at 1 minute and 50 seconds.

5.1.3 Tasks

- Task A: Starting training session.
  - Step 1. You want to log in. There is no need to fill in the fields.
  - Step 2. You want to start an exercise
  - Step 3. You are interested in flight to London", you want to check which items/scenarios this training session contains.
  - Step 4. You want to start the exercise.
- Task B: Finding activity with warning. Besides the icon that ha appeared next to the activity, there is another icon which is not visible on the screen:
  - Step 1. You want to find this item/icon.
  - Step 2. Continue the session without making any amendments.
- Task C: The training session is done.
  - Step 1. Go to debriefing.

.

5.1.4 Follow-up questions

- Was there something that you did not understand?
- Do you have any input on how the interaction could be made clearer?
- Did you find the presented information useful?
- Could you describe what the icons mean?

5.1.5 Procedure

The testing is done via Zoom where the test person has the ability to click in the prototype by themselves. The test persons will be greeted with the introduction mentioned earlier. The tasks will be read out loud to them and after the tasks are completed the interview questions from above will be asked.

5.2Usability testing

In this section the usability test plan for Panoptis will be described.

5.2.1 Purpose

The purpose of the usability testing is to investigate the design intuitiveness, the functionality of the application when used in the intended environment and get real input on how the user uses the application. The result of the testing will be used to improve the usability of future versions of Panoptis.

5.2.2 Problem

The following questions should be answered by the testing:

- What obstacles does the test person encounter when performing the predefined test-scenarios?
- How intuitive is the interaction for the test person considering the meaning of icons, colors and other elements of the application?
- Is there a difference between a first time user and a user already familiar with Panoptis?
- Comparing them to each other to the the following questions:
  - The intuitiveness of the application? (Meaning of icons and what is clickable)
  - The success rate of each test-scenario?
  - The time to complete the test-scenarios?
- Can the test person complete a test-scenario without help?
- What is the positive feedback from the test person regarding their use of the application?
- Does the test person find Panoptis responsive?

5.2.3 Method

Procedure
The following procedure will be used during the testing.

- Pretesting
  - Subprocedure
    - Receive test person
    - Inform the test person about the test.
    - Verbal consent from the test person.
    - Noting if the test person has previous knowledge of Panoptis.
  - Material
    - Script for information to be read to the test person.
    - Observation protocol.
  - Time
    - 10 minutes.
- During testing
  - Subprocedure
    - Test person performs the test-scenario.
  - Material
    - Test-scenarios
    - Observation protocol
  - Time
    - 25 minutes.
- Debriefing
  - Subprocedure
    - Test person answer the interview questions.
    - Test person fills the survey.
  - Material
    - Survey.
    - Interview questions.
  - Time
    - 15 minutes.

Test person
The test persons are instructors from Aviations School in Ljungbyhed that are used to being in a cockpit and they are familiar with the procedures and terminology. They currently use similar software devices and we therefore assume that they have some preexisting knowledge of how to interact with applications on touch screens.

Tasks
During testing there should be at least 2 people present. One of them should guide and be responsible for the test person during the testing and the other person should document, take notes and complete the observation protocol based on the test persons behaviour, their verbal reports and comments. The second person would also be responsible for the technical aspects of the testing, such as starting the testing and the video/audio recording.

Script to test person
Hi xxx,

Welcome to this usability testing. Thanks for participating. I'm xxx and this is my colleuge xxx. We will together assist you and make sure that everything works smoothly during the testing. I will give you a short introduction to what the product we will test today is. Panoptis is a tool that uses eye-tracking to track the eye movements of pilot students during training. This tool will be used by supervisors and help the supervisors notice if the students have missed to look at different parts of the panel that they should have looked at during the training.

You will be presented with a written test scenario which you will be able to read before starting the test, but will also be able to consult during the testing. After you have finished the test scenario we would appreciate it if you wanted to answer some brief questions that we have for you as well as answering a very short survey grading your experience from 1 to 5. The whole test will take about 50 minutes. During the testing we would like to do a recording of the screen and audio. This will be used only for our own internal reference and only if needed. You will remain anonymous and your participation will not be published anywhere. So, before we start and if you feel this is ok, we would like to confirm your consent. Finally we want to reassure you that the purpose of the testing today is solely to test the product and not your cognitive abilities, and if you wish to terminate the testing you can do so at any time.

Test scenario

You want to start a new training session for the student called Carl Larsson and his email is Carl.Larsson@student.lu.se. Your email is supervisor@lu.se. Before you start you want to do the following:

- T1.1 Make sure to add the student to the exercise.
- T1.2 Add yourself as the supervisor for the training session.

Since you did not have training sessions with Carl in the past, you want to do the following:

- T2.1 Use a preexisting training session that includes take-off, engine failure and landing for this training session.
- T2.2 While in training you want to see and understand what part of the panel in the cockpit did Carl miss to look at.
- T2.3 After landing you want to redo the the landing.
- T2.4 When you are done you want to go through the debriefing with the student.
- T2.5 During debriefing you want to see where it was decided that the student missed to look at the panel.
- T2.6 After the debriefing you want to send the results to the student.

Equipment and test environment

We will need a tablet with a touch screen that is able to run Panoptis, and a software to record the screen and audio.

In order to ensure that the tests are performed in the most realistic environment, the testing will be in a simulator cockpit at the Ljungbyhed Aviation School.

Clues for the test scenario

If the test person does not understand the timeline:

- C1.1 What do you think the bar to the right indicates?
- C1.2 What do you think the green color and the orange color stand for?
- C1.3 Do you think you can press them?

If the test person does not understand the icons:

- C2.1 What do you think the icon looks like?
- C2.2 Does it resemble anything from the cockpit?

If the test person does not understand how to find where in the students actions the decision was made that they missed something:

- C3.1 Where would you want to press to see the where the decision was made?
- C3.2 Do you think the icons are clickable?

Documentation

The testing will be documented with help of recordings, the tables below and Google forms for the survey. The different tests will be documented with help of the recordings. All the surveys will be done digitally with help of Google Forms. Table 1 will be used to take notes during the testing, the observer will be able to note the task, the amount of leads that were given, lead id, observations, time to perform the test and other notes. Table 2 will be used for note taking during the interview, the observer will be able to note the questions, answers, the follow up questions and other notes. Everything will be stored in Google drive and be deleted after projects end.

See Appendix A for the documentation tables that will be used during the usability testing. Table 1 will be used during the execution of the test and table 2 will be used during the interview.

Interview Questions

The interview will be done after the testing. It will be semi structured where there will be predefined questions but there will also be room for spontaneous followup questions for further discussion if needed.

- Q1 How happy are you with your performance and the interaction in general?
- Q2 What are your thoughts about the program and your interaction with it?
- Q3 Is there anything you would change with the application?

SUS-survey
A SUS survey will be used to measure the usability of Panoptis and it will be a complement to the interview questions mentioned above. See tabel 3 in Appendix A for the SUS survey.

6 Results

Because of the time limit of the project and the restrictions of COVID-19, the usability testing was not performed. The usability testing would have needed Panoptis to be tested in the training cockpit and such thing would not have been possible to be done due to COVID-19. However, the usability test could still be performed by using the test plan described in section 5.2.

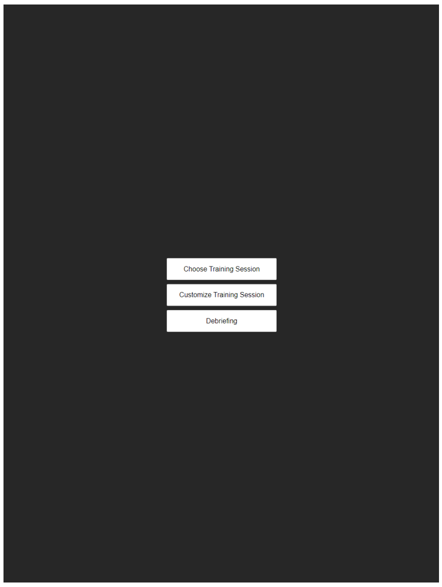
Since the useability testing could not be performed we can choose a different approach on evaluating the design of Panoptis. We chose to analyzing the 10 principles of Jakob Nielsen, as mentioned in section 3.3. *Principle 1* was fulfilled by highlighting in which part of the exercise the user is in at the given moment.

The icons and language used in Panoptis were such that the user can recognize from the real world. Icons that were used represent real parts of the cockpit, these aspects fulfill *principle 2*. *Principle 3* was not achieved since there are different parts of Panoptis where the user might not be able to navigate back, this should be looked into since the user should be able to navigate as they please without having to wait for the process to be finished before being able to return in the previous state.

Principle 4 was not fulfilled since there is not really a standard on how an application such as Panoptis should work. However, looking into the elements used, such as buttons and so on, they follow the standards. One element that might cause confusion among users is the scrollbar that indicates the status of the exercise and does not follow the standards on how a scrollbar is usually used. This might need explanation for the first time users. The *5th principle* was not fully fulfilled since there is a lack of confirmation steps before e.g. starting a session. More constrains should be introduced, such as not redoing an exercise before conforming it, not closing an exercise before conforming it and so on. Aspects of *principle 6* are mostly covered since most of the information is available for the use at all time, for example, when choosing an exercise they are always able to see the parts that are included in the exercise. However, they are not always able to see the name of the exercise itself, this should be looked into. *Principle 7* was fulfilled by allowing the advanced users create custom exercises if they want to and the novice users can use the pre-existing ones. *Principle 8* is one principle that was most worked on in this project. The whole goal was to only present the data that is relevant for the user and nothing else to disturb them. The aspects of *Principle 9* were not included as much in the design of Panoptis. Since we tried to minimize the errors, the design is not prepared for what happens if an error occurs. *Principle 10* is not fulfilled since there is no documentation on how to use Panoptis. This however would be beneficial to create because of some of the unique interactions that differ in Panoptis.

6.1Prototype result

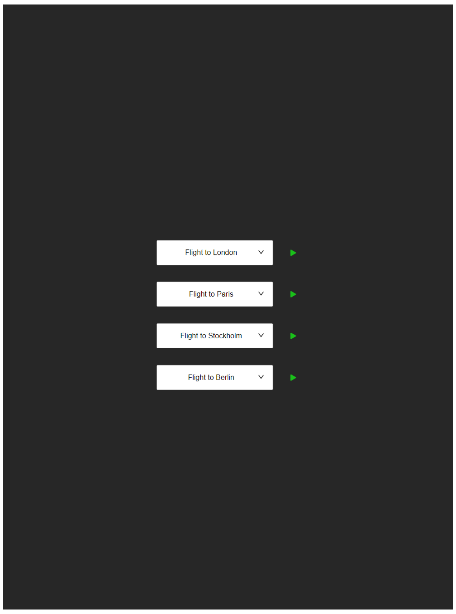
The project resulted in the following implementation.



Figur 10: Panoptis start screen.

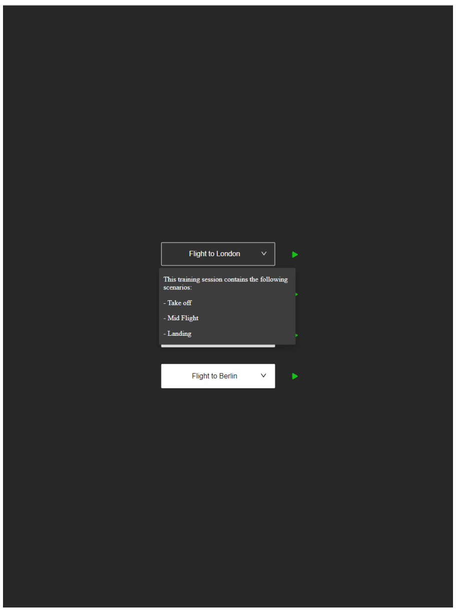
Figure 10 shows the start screen of Panoptis where the supervisor can choose what they want to do. They can start a training session, customize an training session or go to the debriefing.

When they choose to start a training session they will be asked to choose which training session they want to start, this can be seen in figure 11. By clicking on the arrow in a training session the supervisor can see the different exercises that are included in that that training session, see figure 12.

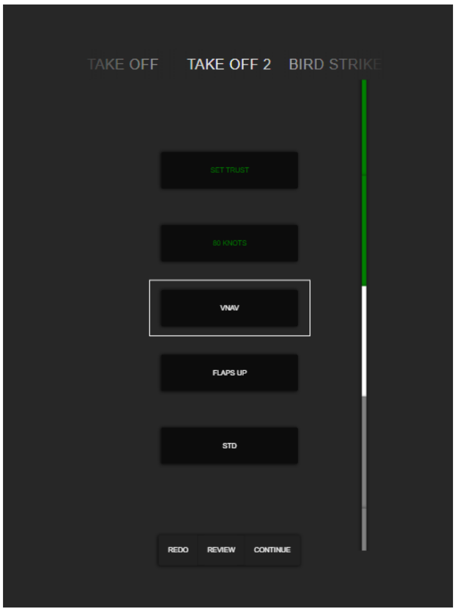


Figur 11: Screen for selecting training session.

When starting the training session they will be introduced to the screen seen in figure 13. Here is where the supervisor will be able to easily get information about the session and what the students have missed to look at. At the top, the bar will let the supervisor know at what exercise they are at right now and at the bottom there are different buttons to help the supervisor navigate. On the right side is the scrollbar placed that also indicated what has been missed and what has been passed. The scrollbar is clickable and can take the supervisor the the corresponding activity when clicked on.



Figur 12: More information shown about the training session.



Figur 13: Panoptis during the training session.

7 Future work

As mentioned earlier, the usability testing was not done with the implemented version of Panoptis. To determine whether the implementation would be useful in a real case scenario, a full useability testing needs to be performed and analysed. Depending on the results of the testing, further development of Panoptis can be done. Furthermore, the the 10 Usability Heuristics by Jakob Nielsen can be further investigated and how they can improve the design of Panoptis. *Principle 3* need to be worked on since right now the user is not always able to return if they have made an accidental click, this can be done by adding back buttons. *Principle 4* should be looked into since the scrollbar does not work in the traditional way and thus would need explanation. The explanation could be presented in the documentation/manual of Panoptis which would also satisfy the *10th principle*. Panoptis needs in general more confirmation windows e.g. before starting a training session, before finishing one etc. By adding confirmation windows the *10th principle* would be fulfilled. Error messages and how to handle when something goes wrong should be looked into and implemented, this would fulfill the *9th principle*.

There are some features that were not implemented and could be done in the future. For example, the debriefing has been mentioned as something very important and should therefore be looked into. For the supervisor it should be possible to log in and also register which student the training session is connected to. Moreover, the ability to see where Panoptis made the decision that something has been missed should also be implemented.

There are a lot of other different approaches one could take with Panoptis and even the other tools that the supervisors use. Such an example is AR, with a simple redesign of Panoptis it would allow the supervisor to see what the student has missed directly in the device when it happens and the supervisor would not even need to move their head to see that something has happened. Furthermore, AR can be a solution for the future where all the tools and screens that are used by the supervisors would be replaced with AR. This would allow the supervisor to choose what to see and when to see them. What to focus on at what time and without becoming a distraction for the students. With AR the student would not notice if the supervisor is checking anything and thus not be distracted by it. Moreover, when using AR, there is no need to have room for screens and other tools. This means that the cockpit for training would need less place to be installed in.

8 Discussion and Conclusion

It is a difficult challenge to design a usable application such as Panoptis since there are many different aspects to. Besides being able to show all the important information it also needs to be simple and not crave too much of the users attention. Based on the initial testing that was performed, the feedback we got was very positive and greatly appreciated by the instructors at the Aviation School in Ljungbyhed which leads us to believe that there is a good foundation for this application to be build upon.

Panoptis is not designed to replace any already existing tools that supervisors use and thus it will become an addition to the other tools. One might argue that supervisors might not need more tools to keep tack of while in training. That we do agree with, however, the use of Panoptis will hopefully lead to the supervisors not needing to look at the students all the time and be able to trust that Panoptis will notify them when the students have missed to look at something. This will hopefully reduced the stress for the supervisor and at the same time allow them to focus on other important aspects of the training instead.

Referenser

(2021). Eye Movement Can Affect Problem-solving, Cognition. publisher: University of Illinois at Urbana-Champaign.

B, M. and B, H. (2012a). Brainstorm graphic organizers. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*, page 22.

B, M. and B, H. (2012b). Storyboards. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*, page 170.

Lave, J. (1988). Cognition in practice: Mind, mathematics, and culture in everyday life.

Liversedge, S. and Findlay, J. Saccadic eye movements and cognition.

Newell, A. . S. H. (1972). Human problem solving.

Niehorster, D. C., Hildebrandt, M., Smoker, A., Jarodzka, H., and Dahlström, N. (2020). Towards eye tracking as a support tool for pilot training and assessment.

Nielsen, J. (2020). 10 usability heuristics for user interface design.

Norman, D. (2013). The deign of everyday things.

Norman, D. A. (1993). Cognition in the head and in the world: An introduction to the special issue on situated action.

Suchman, L. (1987). Plans and situated action: The problem of human-machine communication.

Vera, A. H. and Simon, H. A. (1993). Situated action: A symbolic interpretation.

Winograd, T. and F., F. (1986). Understanding computers and cognition: A new foundation for design.

Appendices

A Documentations

Tabell 1: Documentation protocol during testing.

Task ID	Amount of leads	Lead ID	Observation	Other

Tabell 2: Documentation protocol during interview.

Question ID	Answer	Follow up question	Other

Tabell 3: SUS survey

	Strongly disagree				**Strongly agree**
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

B Requirements specification

Panoptis

System Requirements

V 1.1

Jesper Brodin, Johannes Persson, Mergim Rruka, Melina Tsapos

Table of Contents

1. Introduction

1.1 Background

Panoptis is an application that aims to make it easier for flight instructors to identify why pilot students make errors during flight simulation training, while reducing the cognitive load for the flight instructor. This will be done by telling the instructor if students have looked at the correct flight instruments before they take a decision.

The following technologies are needed to create this application.

- ***Eye tracking:*** To track where the students look.
- ***Voice to text:*** To identify what the students are saying, this to make it possible for the application to tell what decisions are taken, thereby telling our system that a task is done.

- ***Tablet/Computer:*** To process the eye tracking, voice to text and provide a screen for the instructor to interact with.

2. Vocabulary

Training: A training refers to when a pilot student is training in a flight simulator.
Instructor: A teacher who leads the pilot students simulation training.
Student**:** A pilot student who uses the eye tracking equipment and performs the different exercise during different training sessions.
Standard Operating Procedure (SOP): Step-by-step instructions of how to perform certain operations for different situations.
Training Session: A set of different exercises that the student performs during training in the flight simulator.
Exercise: A specific scenario that the student performs during a training session, for example TAKE OFF. An exercise is based on several tasks.
Task: A specific task is either passed or failed by the student, depending on the eye tracking data. A task is based on one or more activities.
Activity: A specific activity where the eye tracking system checks whether the student has looked upon a required flight instrument or not. Based on SOP.
Pass:
Fail:

3. System description

The system contains a pilot student, an instructor, and our application. How they interact is described in Figure 1 below.

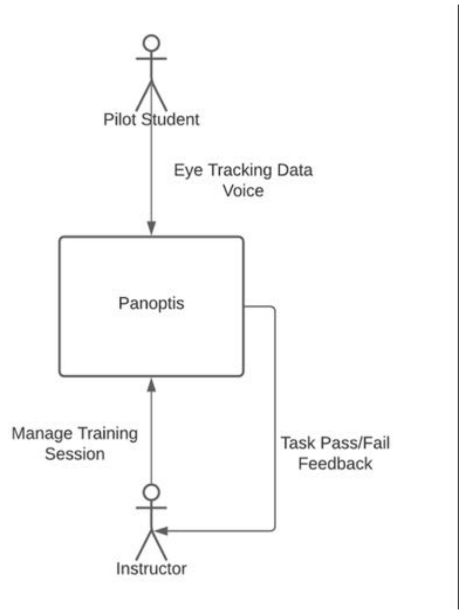


Figure 1: Context diagram of the Panoptis system.

4. Requirements

4.1 Goal-level Requirements

The goal-level requirements are listed in the table below.

Requirement	**Description**
R-Goal-1	The application shall reduce the cognitive load for the instructor during training.
R-Goal-2	The application shall make it easier for an instructor to debrief the students after a training.
R-Goal-3	The application shall enable useful & quick feedback during training.

4.2 Feature Requirements

The different feature-level requirements are described in the following respective subsections.

4.2.1 Training Requirements

Feature: Training Session

Gist	See section 2. Vocabulary.

Training-TrainingSession	**Training Session**
Purpose	
Pre-condition	1. Student information is entered in the system
Sub-tasks	1. The instructor starts the training by pressing start. 2. The system starts the first exercise in the training session. 3. When an exercise is done, the system asks if the instructor wants to review, redo, or continue to the next exercise. 4. When the instructor presses continue, the next exercise shall start. 5. When training is done, the debriefing screen shall be shown.
Variants	1. Sub-task 3: When the training session is on the last exercise the continue button shall read “finish” instead. 2. Sub-task 4: When the instructor presses redo, the exercise just completed shall be initialized. 3. Sub-task 4: When the instructor presses review the review screen shall be shown.

Feature: Exercise

Gist	See section 2. Vocabulary.

Training-Exercise	**Exercise**
Purpose	A training shall consist of multiple exercises.
Pre-condition	1. The training session has started an exercise.
Sub-tasks	1. The session starts the first task and shows it as the current. 2. When a task is done, the session marks the next task as the current task, and scrolls the tasks forward. 3. When the last task is done, the exercise tells the training session that this exercise is done and moves to the next exercise.
Variants	

Feature: Task

Gist	See section 2. Vocabulary.

Training-Task	**Task**
Purpose	To give feedback whether the pilot student has looked upon required flight instruments for a certain task.
Pre-condition	1. An exercise is active
Sub-tasks	1. The tasks registers if it’s activities have been looked upon or not. 2. If all activities have been looked upon, the text of the task is marked green.

	3. If one or more activities haven't been looked upon, the box of the task is marked orange. 4. Icon corresponding to the activity that has been missed is displayed to the left of the corresponding task.
Variants	

Feature: Activity

Gist	See section 2. Vocabulary.

Training-Exercise	**Activity**
Purpose	To give feedback whether the pilot student has looked upon a certain flight instrument.
Pre-condition	1. An exercise is active and it's tasks have all it's corresponding activities configured.
Sub-tasks	1. The system registers if the eye-tracking system outputs data that suggests if the pilot student has looked upon a certain predefined area. 2. If the system has registered that a certain instrument has been looked upon, the corresponding activity is marked as pass. 3. If the system has registered that a certain instrument has not been looked upon, the corresponding activity is marked as fail.
Variants	

5. Release plan

5.1 Description

This section specifies in what order we want to achieve certain features and requirements to end up with a MVP as early as possible, to be able to do testing.

5.2 Releases

Release	**Features**
Release 1	**Home View** **Training Session** (**Sub-task 1 & 2**) **Exercise** (**Sub-task 1**)
Release 2	**Training Session** **Exercise** **Task** **Activity**
Release 3	**Eye tracking** **Speech-to-text**

VR Environment

Anna Andersson Ewerman, Gabrielle Carlsson, Sam Bagra & William Rosenberg

The following report summarizes our work in the development of a Virtual Reality (VR) game with eco-conscience themes. It contains an account of our progress during the autumn semester of 2020, from initial brainstorm to final conceptualization, including sketches, underlying theory and as well as a description of the pilot study conducted in our VR environment with results.

By visualising the man-made changes that threaten the Scandinavian Mountains, (Swe: Fjällen), our project aimed to raise awareness in how an individual can make a difference. Furthermore, we studied the influence of eco-conscience feedback on users' in-game behaviour and our results shows an effect on the players' in-game priorities, however, due to COVID 19 health regulations, our sample was not large enough to declare this effect as statistically significant nor to explore what might drive it.

1 Introduktion

Flora och fauna i den svenska fjällen är känslig för stigande temperaturer. Det är därför viktigt att sprida kunskap om dess skörhet och hur den kan skyddas.

Under den här projektkuren har vi, tre ingenjörer inom informations- och kommunikationsteknik samt en kognitionsvetare, tagit oss an utmaningen att skydda en nationalskatt med hjälp av virtuell verklighet, även kallat virtual reality (VR). VR-teknik sägs komma att förkroppsliga nästa kapitel i skärmtidens historia och med museibesökande mellanstadie- och högstadieklasser som målgrupp är detta medium välanpassat för att nå ut med ett starkt budskap till yngre generationer. Utöver kursens tidsplan fanns inga fler avgränsningar, varför vi bröt ner uppgiften och satte upp egna mål. Tillsammans kom vi överrens om att skapa en high-fidelity protoyp, med ett inbyggt experiment, som går att testa med utomstående så snart som möjligt - vilket ska nu presenteras i denna rapport.

Rapporten redogör för vårt arbete och består utav sex kapitel**:** *Förarbetet, Miljöutveckling, Underliggande teori, Experimentsupplägg, Resultat, Diskussion av resultat, Diskussion av utveckling och Slutsats.*

2 Förarbetet

Gruppen började med gemensam brainstorming kring olika spel och delspel. Vi började sedan att undersöka vad som ingår i läroplanen för mellanstadie- och högstadieelever.

Med hjälp av det kunde vi banta ner det till fyra möjliga idéer. Dessa idéer var: *Plast i havet, Ekosystem i obalans, H_2O i olika tillstånd* och *Energiförsörjning.*

Varje gruppmedlem blev tilldelad en idé och gjorde då en storyboard för hur denna idé skulle kunna se ut och utföras på bästa sätt. Det fanns både för- och nackdelar med alla idéerna, vilka vi diskuterade och övervägde noggrant.

Plast i havet

Bild 1. Storyboard av *Plats i havet*

En fördel som vi kunde se med konceptet *Plast i havet* är att det är ett känt problem som behöver belysas, det är även naturligt avgränsat. Problematiken som vi såg med detta är dock att just detta fenomen med plast i havet och nedskräpning på stranden nästan återges bättre i verkligheten än vad det skulle kunna göras i en VR miljö. Det är dessutom en problematik som är lätt för eleverna att förstå och relatera till. Detta innebär även att det redan finns stora etablerade initiativ som arbetar aktivt med att utbilda skolelever kring dessa miljöproblem, till exempel Håll Sverige Rent (Håll Sverige Rent, 2020). Detta leder till att vi i detta fall inte bidrar med något nytt.

Ytterligare en aspekt med *Plast i havet* var att vi hade kunnat skapa tydliga konsekvensscener där man hade kunnat använda elevernas empati för att skapa förståelse. Idén vi hade om ett delspel var att spelaren först fick plocka och sortera skräp på en strand och sedan ta sig vidare till nästa del av spelet där de skulle vara en sköldpadda som simmar i havet och målet är att ta sig från en punkt till en annan utan att fasta i skräpet som finns i havet. Mängden skräp som sköldpaddan behöver undvika varierar beroende på hur bra det gick i det första spelet. Har man lyckats rensa bort allt skräp på stranden så är det enkelt att ta sig fram i havet, har man misslyckats och lämnat kvar skräp så kommer det vara svårt för sköldpaddan att ta sig fram utan att fastna.

Att utforma en konsekvens av det första spelet där spelaren får fortsätta spela och där det första resultatet påverkar svårighetsgraden på det andra tror vi hade varit ett intressant sätt att skapa en djupare förståelse hos eleverna. Ett problem med denna idé är dock att det hade blivit svårt att implementera med de kunskaper vi besitter inom VR-programmering.

Ekosystem i obalans

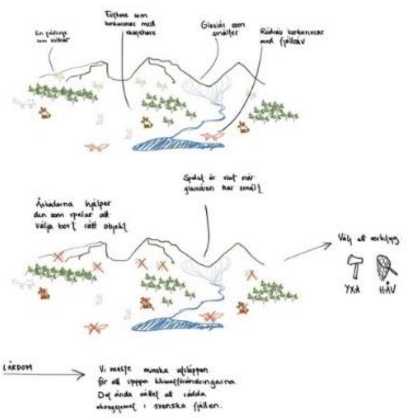


Bild 2. Storyboard av *Ekosystem i obalans*

Med konceptet *Ekosystem i obalans* såg vi en stor fördel i att man kan välja miljöer som inte är tillgänglig för alla och på det sättet bjuda in till något som må vara nytt för vissa. Sett till kunskapsinnehåll kan *Ekosystem i obalans* vara svårt att ta till sig och förstå med enbart teoretisk undervisning, så här såg vi chansen att utveckla något som kompletterar klassrumsundervisningen. Något som kan ses som både en för och en nackdel är att detta är ett mer komplext miljöproblem. En fördel är att det ger oss en möjlighet att skapa ett koncept som hjälper spelarna att förstå ett komplext problem. Ett problem är att det krävs en hel del eftertanke för att skapa ett spel som leder till större kunskap om ett miljöproblem av denna komplexitet. Det kräver även en hel del implementering för att vi ska få det att fungera, ett exempel på svårigheterna går att förstå genom vårt första utkast för *Ekosystem i obalans*.

Detta spelkoncept gick ut på att spelaren "styrde" över ett ekosystem. Vi visade detta genom att man befann sig i en fjällmiljö med en glaciär som smälte. Glaciären blir då en form av tidtagare för spelet som kommer att pågå tills dess att hela glaciären har smält.

För att hindra glaciären från att smälta skulle spelaren då se till att hålla kvar trädnivån på den ursprungliga nivån samt se till att rensa bort de djur som inte hör hemma i fjällmiljön från början men som med klimatförändringar klarar av att leva i denna miljö och då konkurrerar ut de arter som levt där från början. När trädnivån höjs skapar det sämre förutsättningar för de ursprungliga arterna att leva där och bättre för de arter som utgör ett hot. På grund av detta bör trädnivån hållas på dess ursprungliga nivå.

H_2O i olika tillstånd

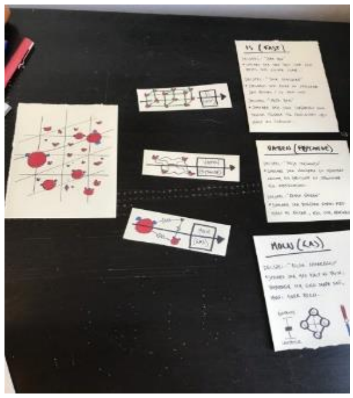


Bild 3. Storyboard av H_2O i olika tillstånd.

När det gällde konceptet om *H_2O i olika tillstånd* är det en unik konceptuell idé där spelet utgår från en miljö med flera H_2O-molekyler som spelaren ska få att interagera med varandra på olika sätt för att ta sig vidare till de olika delspelen. Ett delspel bygger på is, ett på vatten och ett på vattenånga. För att ta sig till spelet med vattenånga är tanken att spelaren ska få H_2O-molekylerna att röra sig fort i relation till varandra. Detta för att skapa förståelse för hur is, vatten och vattenånga byggs upp med H_2O molekyler. Idéerna på delspel var kopplade till is, vatten och vattenånga.

Nackdelen med denna unika idé var dock att det hade blivit tekniskt svårt för oss att illustrera hur molekylerna är ånga eller is. Det är även svårt att koppla konceptet till ett tydligt miljöproblem vilket innebär att det kanske snarare är ett koncept som kan vara ett komplement till klassrumsundervisningen i kemi snarare än i miljöproblem.

Energiförsörjning

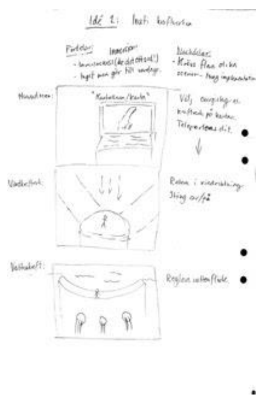


Bild 4. Storyboard av *Energiförsörjning*.

Med konceptet om att utveckla ett spel om energiförsörjning såg vi potentialen med att låta spelarna uppleva en miljö som är svårtillgänglig och riskfylld. Vårt spelkoncept hade även potential att involvera de elever som inte har på sig VR-setet i spelet. Detta koncept tog inspiration från spelet *Keep Talking and Nobody Explodes (Bomb Manual, 2020).* Konceptet gick ut på olika delspel på olika energikraftverk där något har gått sönder som spelaren måste laga.

Samarbetet med de som är utanför spelet är då att det är de som har lösningen på hur det ska lagas. Nackdelarna med detta koncept är att det är svårt att skapa en interaktiv konsekvensscen eftersom dem blir svåra att relatera till den enskilde individen, vilket innebär att det snarare handlar en överblick över vilka konsekvenser det skapar för samhället i stort.

Val av koncept och miljö

Efter mycket brainstorming och två omgångar med utveckling samt vidareutveckling av storyboards, valde vi konceptet med *Ekosystem i obalans*. Detta eftersom vi såg potentialen i att utveckla något som det inte redan finns lösningar för, men även för att det kan hjälpa till där den teoretiska undervisningen inte är tillräcklig för att skapa en djupare förståelse. Att vi valde att placera spelkonceptet i just en nordisk fjällmiljö beror bland annat på att det är en miljö som påverkas mycket när det sker temperaturökningar till följd av klimatförändringarna.

Höjningen av temperaturen förväntas också bli större i fjälltrakterna än i övriga delar av Sverige. Fjällen är samtidigt ett väldigt folkkärt naturområde för svenskar, och hem för en unik flora och fauna som numera är hotad. Vidare är fjällmiljön ett kulturarv som svenska skolelever förväntas ta del av och kan därför ses som en miljö med potential att skapa engagemang, mer empati och en vilja att förändra beteenden för att bevara den.

Trots att fjällmiljön är nära är det långt ifrån alla svenska skolelever som har möjlighet att få uppleva den och på detta sätt kan alla få ta del av den.

Utvecklingen av spelkonceptet och delspelen

När vi valt ett koncept att utforma spelet kring var det dags att hitta ett fungerande spelkoncept samt flera idéer på delspel. Vi började därför med att brainstorma i grupp för att kunna kategorisera olika idéer som skulle ligga till grund för ett koncept. När vi gjort detta fick alla gruppmedlemmar skapa storyboards för hur de såg ett spelkoncept eller ett delspel framför sig. Vi utvecklade sedan dessa till ett helhetskoncept som vi alla var eniga om att det fungerade och som vi trodde skulle vara möjligt för oss att implementera.

Tanken med spelkonceptet är att spelaren utgår från en typisk fjällmiljö och startar vid en fjällstuga. Utformningen bygger på att spelaren inne i stugan plockar upp olika föremål som representerar ett delspel och tar med sig detta ut för att starta delspelet.

När vi brainstormade uppkom flera idéer till delspel och de tre vi tyckte bäst om redovisas i nästkommande stycken.

Hjälp de utrotningshotade arterna

Målet med detta delspelet är att visa spelaren vad som händer när klimatförändringar gör att miljön förändras och blir beboelig för arter som inte hör hemma där egentligen. Här var det upp till spelaren att fånga in icke-hemmahörande rödrävar som sedan kommer att flyttas till en miljö där de inte utgör ett hot för fjällräven.
En idé vi har med detta delspelet är att spelaren inte har information om vilka av rävarna som ska fångas in utan att det är information som de utanför spelet har. På detta sätt kan de då samarbeta för att fånga in och frakta bort rätt rävart.

Se till att trädgränsen inte flyttar sig

Detta delspelet kan ses som lite kontraproduktivt då idén är att spelaren ska plocka bort träd som börjar växa över den ursprungliga trädgränsen. I och med att det är en kontroversiell handling skulle det dock kunna öka chansen att spelaren lär sig och kommer ihåg kunskapen. Anledningen till att man inte vill att trädgränsen ska höjas är för att arter som ursprungligen lever i fjällen missgynnas av en höjd trädgräns. Dessa arter löper då en större risk att utkonkurreras av andra arter och fjällens unika fauna riskerar att minskas drastiskt.

Plocka skräp i fjällen

Med det läge vi har i världen idag, med *Covid-19*, så är det fler än normalt som turistar uppe i fjällen. Detta har lett till mycket nedskräpning vilket är negativt för fjällmiljön.

Vår plan är därför att utveckla ett spel som går ut på att spelaren rensa och sortera skräp från naturen. För att spelaren ska få en positiv respons på att sortera skräpet kommer det finnas två mätare som visar hur mycket som samlats in, dessa mätare kommer visa på hur god väg man är att återvinna plast som räcker till att tillverka en kåsa och metall som räcker till att tillverka en termos.

Beroende på hur duktig spelaren är på att samla upp allt skräpet kommer spelaren dessutom få interagera med en fjällräv. När spelaren har gjort ett bra jobb så kommer en glad fjällräv med sina ungar att synas och om spelaren inte har varit tillräckligt snabb eller samlat upp allt så kommer den att få se en fjällräv som fastnar med nosen i en konservburk. Detta för att skapa en djupare förståelse för vad det får för konsekvenser när vi skräpar ner i vår natur.

3 Miljöutveckling

Inför delkurs två så modellerades en fjällmiljö i *Blender* - se bild 1. Miljön var byggd för att påminna om en nordisk fjällmiljö. Ett detaljerat narrativ skrevs och låg till grund för designbeslut både vad det gäller implementering och modellering.

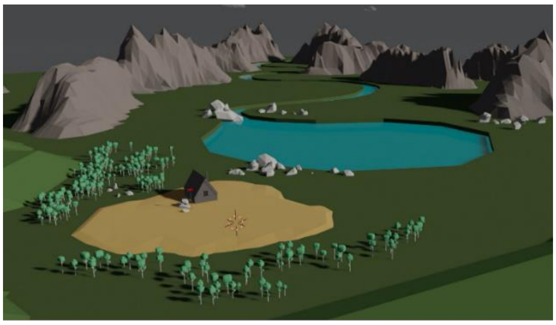


Bild 1: Modellerad fjällmiljö när delkurs två påbörjas.

Spelyta

Den modellerade miljön var stor och vidsträckt. För att lättare visa användaren vilka objekt som var av betydelse och vart narrativet utspelades så skapades bestämda ytor där användaren skulle kunna vistas. Det gjordes med specifika öar i en ljusare sandfärg - se bild 2. En ö där stugan står och en ö för tältlägret.

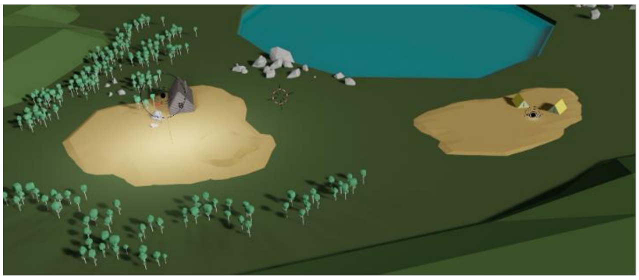


Bild 2: Öar av sand som visar spelytor.

Narrativ

För att användaren ska dras med in i VR-världen på andra sätt än det visuella behövdes ett narrativ. Följande beskrivning togs fram av gruppen – notera att *du* hänvisar till användaren ifråga: ”Du vaknar upp i en säng. När du ser dig om så upptäcker du att du befinner dig i en timmerstuga. En eld är tänd i spisen och lyser upp stugan med ett gult ljus. Plötsligt ringer en telefon. Du plockar upp telefonen för att svara och en naturvårdare berättar för dig att det är fullt med turister på fjället som skräpar ner en massa i tältlägret. Det gör att djuren far illa. Personen som talar berättar att den hört att en fjällräv till och med ska ha fastnat med huvudet i en konservburk. Det är hemskt och därför frågar hen om din hjälp! Du går därför ut ur stugan.
Utanför finns en skylt som visar vägen till tältlägret, den går att använda för att teleportera sig dit. När du kommer till tältlägret ligger det några skräpobjekt på marken och plockar upp ett så startar spelet.”

Skräp dyker då upp på marken inom ett begränsat spelområde, men även naturliga objekt som svampar och stenar dyker upp. Spelaren får poäng för att plocka skräp och minuspoäng för att plocka naturliga objekt. Om skräp blir liggande för länge på marken börjar de blinka för att sedan försvinna.

Konsekvenserna varierar, istället för att försvinna kan objekten bytas ut mot djur som far illa, ledsna växter, förorenad jord eller liknande. Användaren exponeras för positiv feedback när den plockar skräp. En poängräknare som går upp, ett tillfredsställande ljud, fina växter som blommar upp där skräpet var och djur som skuttar omkring. En timer avgör hur länge man plockar skräp. När all tid gått har även hela dagen gått. Himlen har blivit mörk och full med norrsken och det är dags att gå hem. När man anländer till stugan är det dags att sortera allt skräp. Det ligger i en stor hög mitt i rummet. Olika tunnor för plast, metall och kartong används. De insamlade soporna återvinns till nya grejer som du kan interagera med efteråt.

Skyltar

För att ge användaren möjlighet att lätt navigera mellan öarna så användes skyltar för teleportering. För att underlätta översättning mellan olika språk så användes inte skriven text på skyltarna utan istället symboler. Ett tält för tältläger och en stuga för att komma tillbaka. Fler skyltar diskuterades för att skapa en känsla av att miljön kunde vara större, men det antogs snarare leda till förvirring om det inte fanns en plan på att faktiskt implementera dessa, därför skrotades den planen.

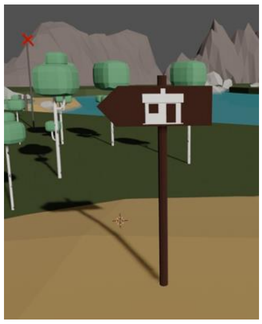
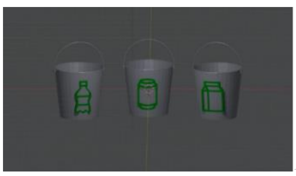


Bild 3: Skylt för teleportering till tältlägret. Bild 4: Skylt för teleportering till stugan.

Skräphinkar

På grund av begränsad tidstillgång skrotades planen på att sortera soporna i stugan i efterhand då det ansågs vara ett större moment att implementera än att direkt sortera soporna vid insamling. Kärl för skräpinsamling var nödvändiga i någon form. Sopsäckar, skottkärror och soptunnor har diskuterats som möjliga alternativ. Slutligen beslutades det att hinkar var det bästa alternativet - se bild 5 för skiss. Dessa förseddes med symboler för plast, kartongpapper och metall under modelleringen - se bild 6. För att ge användarna feedback på hur mycket skräp som samlats in skulle en siffra placeras ovanför respektive hink.



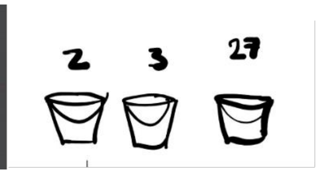


Bild 5: Skiss för skräphinkar. Bild 6: Modellerade skräphinkar med symboler för material.

Skräp och konsekvenser

Skräpet som skulle plockas upp efterliknade objekt som är typiskt förknippade med kartongpapper, plast och metall. Därför modellerades tre typer av objekt som tillhörde dessa kategorier. För att ge spelaren en förståelse för konsekvenser av deras handlande, modellerades även ett tillhörande objekt som skulle föreställa en förening mellan skräptyp och naturen ifall de misslyckades med att plocka upp ett skräpobjekt i tid. Som om naturen fallit offer för dåligt beteende bland fjällturisterna. Konsekvensen för kartongen kom aldrig att implementeras i spelet på grund av experimentets utformning - se bild 12.

Konsekvensen för konservburken som visas i bild 8 är en fjällräv. Den är modellerad från grunden samt riggad med en armatur för att sedan kunna animeras. Animeringen visar en fjällräv som uppenbarar sig vid konservburken för att sedan sticka in huvudet och fastna. Det gör att fjällräven börjar ruska huvudet fram och tillbaka för att försöka göra sig av med burken. En alternativ konsekvens för konservburken i bild 7 kom även att adderas likt den ihopknycklade plastflaskan i bild 10.

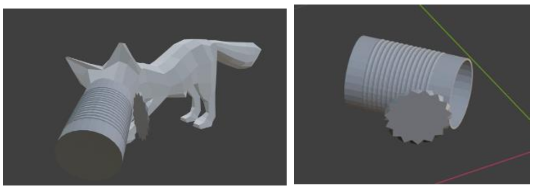


Bild 7: Konservburk. Bild 8: Konsekvens med konservburk och fjällräv.

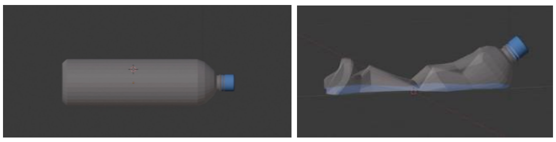


Bild 9: Plastflaska. Bild 10: Konsekvens med ihopknycklad plastflaska.

Bild 11: Kartong. Bild 12: Konsekvens med kartong.

Återvinningsmaskinen

För att användaren skulle kunna se en ytterligare positiv konsekvens av att samla in skräp uppkom idéen om återvinning. Själva återvinningsprocessen är i verkligheten komplex men förenklades i spelet i form av en återvinningsmaskin som kunde placeras i stugan - se bild 13. Användaren skulle manas till att trycka på en knapp för att fyllda hinkar, som i bild 14, skulle omvandlas till nya objekt.

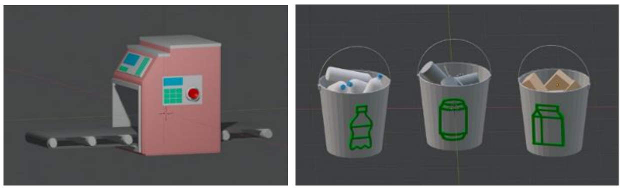


Bild 13: Återvinningsmaskinen. Bild 14: Fyllda skräphinkar.

Återvunna objekt

Objekt i kartong omvandlas till pappersrullen i bild 16, objekt i plast omvandlas till kåsan i bild 17 och objekt i metall till termosen i bild 15. Samtliga nya objekt var tänkta att upplevas som användbara i en fjällmiljö.

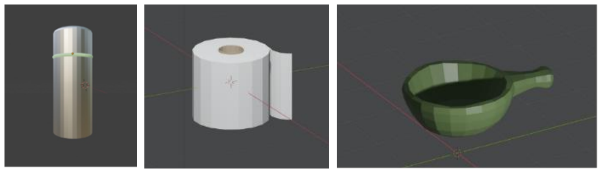


Bild 15: Återvunnen termos. Bild 16: Återvunnen pappersrulle. Bild 17: Återvunnen kåsa.

Begränsning av spelyta & flygande hinkar

Under våra första speltester förstod vi att spelytan för skräpplockningen var alldeles för stor för att kunna överblicka. Därför uppkom idéen om en naturlig barriär till spelytan som skulle ha mindre yta än sandön som tältlägret var placerat på. De naturliga objekt som ansågs passande i mijlön adderades i en cirkel runt spelytan. Vid upplockning av skräp så var det även svårt att se hur räknaren ovanför sorteringshinkarna ökade. Då uppkom idéen om att dessa skulle höjas ovanför marken.

Bild 18: Skiss av naturlig barriär till spelyta. Bild 19: Naturlig barriär till spelyta.

Övrig vidareutveckling

Efter provspelningar och tester förfinades spelmomentet ännu mer. Följande ändringar ansågs nödvändiga: För att användaren skulle kunna överblicka spelplanen så justerades teleporteringen till tältlägret, se position på bild 20. Detta gjorde så att vyn riktades direkt mot hinkarna och konservburken som startar timern - se bild 21. Konservburken som startar timern ansågs behöva en något som signalerar för att säkerställa att användaren förstår hur man startar timern - se bild 21.

Bild 20: Ny position vid teleportering. Bild 21: Vy över spelplan med signalerande konservburk.

När timern gått ut saknades feedback för att bekräfta att spelet slutförts. Därför planerades bekräftelse i form av ljud och fyrverkerier - se bild 22. Slutligen behövdes även något som signalerar att användaren ska använda skylten för att teleportera tillbaka till tältlägret då spelmomentet avslutats - se bild 23.

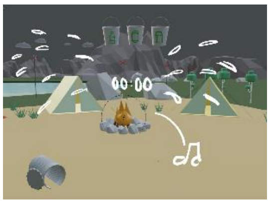
 

Bild 22: Skiss av naturlig barriär till spelyta. Bild 23: Naturlig barriär till spelyta.

Implementering i Unity

Spelets utveckling har bestått av flera olika designbeslut, både av större och mindre karaktär. De beslut som har varit mindre har tagits av utvecklarna i takt med utvecklingen medan större beslut krävde diskussion med hela gruppen. Exempel på mindre beslut är bland annat storleken på de olika objekten samt deras placering. Till de beslut som har krävt en diskussion har utvecklarna testat olika lösningar i *Unity* och sedan presenterat de lösningar som är möjliga att implementera.

För att få spelaren att förstå vad som är nästa steg i spelet beslutade vi att alla interagerbara objekt blinkar. Detta för att göra dem enkla att se samt för att kunna leda spelaren vidare på ett naturligt sätt. För att få blinkningen så naturlig som möjligt använde vi oss av så kallade *point lights*, dessa kan liknas vid en strålkastare som blinkar på ett objekt. Detta ansågs se mer verklighetstroget ut än att själva objektet blinkade genom att förändra färgen på det.

När spelaren skulle plocka upp objekt programmerades spelet så att objektet ansågs vara uppplockat då det nådde en specifik höjd. Detta kan designmässigt anses vara svårt för spelaren att förstå men var det som ansågs vara bäst ur ett implementeringsperspektiv.

När tidtagaren och räknaren i delspelet skulle implementeras diskuterades det om de skulle vara inom spelarens synfält oberoende av var spelaren tittade eller om de skulle vara placerade som objekt i miljö. Att ha den synlig överallt ansågs bidra till en “spelkänsla” och ta bort från upplevelsen av att vara i en värld. Därför beslutades det att placeras i miljön.

Avståndet mellan stugmiljön och tältlägret var för långt för att spelaren skulle kunna ta sig mellan dem med hjälp av *teleport areas*. Därför användes istället skyltar som fungerade som *teleport points*, genom att röra skylten teleporterades då spelaren till en förutbestämd plats i världen. För att få skyltarna att fungera har en *canvas* placerats framför skyltobjektet och det är när spelare interagerar med denna *canvas* som teleportfunktionen aktiveras. Både stugmiljön och tältlägret har även *teleport areas* för att spelaren på ett enklare sätt ska kunna ta sig runt i miljön.

4 Underliggande teori

Projektets experimentella syfte var att undersöka hur ett empati-drivet feedbacksstimuli kan påverka spelbeslut i en VR-mijö. Följande avsnitt ska redogöra för grunderna inom VR-teknik, tidigare empati-relaterad forskning relaterad till VR, VR som läromedel samt feedback och lärospel.

Närvaro och interaktion

För den som är ny till virtuella upplevelser är Tham, Duin Hill, Gee, Ernst, Abdelqader och McGrath (2018) en bra början. Däri beskrivs vanliga begrepp inom VR som forskningsfält, såsom *närvaro*, *interaktion, embodiment* och *immersion* samt några exempel på VR-utrustning.

Likt Tham et al. (2018) menar Cummings och Bailenson (2016) att användarens närvaro bygger på två huvudsakliga premiser: 1) Deras uppfattning av VR-miljön och 2) självlokalisering. Detta kan göras på olika sätt men oftast genom att synkronisera användarens rörelser med en avatar inom VR-miljön. När användaren förstått att de till viss grad befinner sig i en virtuell värld och att de kan påverka simuleringen antas det skapa en känsla av närvaro.

Detta samstämmer med forskning som visar att användare ger sin VR-upplevelse bättre omdömen ifall den underliggande sensomotoriska spårningstekniken har hög uppdateringshastighet och tillåter fler frihetsgrader sett till rörelse, jämfört med andra tekniska aspekter som hade mindre effekt på omdömen så som, synfältsstorlek, skärmupplösning eller ljudkvalité (Cummings & Bailenson, 2016).

En annan viktig poäng är att undvika vad som kallas *cybersickness* hos användare. I sin genomgång av fenomenet betonar Weech, Kenny och Barnett-Cowan (2019) att överansträngda ögon, yrsel och allmänt illamående är grundläggande problem inom VR som främst orsakas av fördröjningar mellan den virtuella och verkliga världen. I deras slutsatser understyrks vikten av att minska gapet mellan världar genom att använda utrustning med avancerad sensomotorisk spårningsteknik och att utveckla VR-miljöer som är intuitiva sett till navigering samt interaktion (Weech et al., 2019).

Vidare menar Biocca (1997) att användarens känsla av närvaro förhöjs om de får interagera med den virtuella världen. Med hjälp av Gibsons (1979) *affordance* begrepp, definierar han interaktion i VR som användarens nivå av sensomotoriska engagemang givet vad som är möjligt att göra i den virtuella miljön (dvs., affordances). Biocca (1997) problematiserar även begreppet interaktion då användare ofta kliver in i en VR-värld med förväntningar från verkligheten. Han menar att användarens uppmärksamhet bör styras aktivt för att undgå att världen upplevs som begränsad.

Bailenson (2019) lyfter fram en liknande poäng och menar att den virtuella världens narrativ kan utnyttjas för att leda användarens uppmärksamhet genom VR-upplevelsen – vilket är en teknik som användes i fjällmiljön.

Embodiment och immersion

När en VR-upplevelse inger användaren en känsla av närvaro, och bjuder in till interaktion, är det relevant att tala om embodiment och immersion. De båda sistnämnda begreppen är konceptuellt besläktade och används för att på olika sätt beskriva användarens förhållande till VR-miljön.

Till exempel talar Slater, Spanlang, Sanchez-Vives och Blanke (2010) om embodiment i termer av *body transfer*, dvs. att VR-simuleringen lyckas ge användaren en realistisk illusion av en virtuell kropp. Deras studie undersökte hur deltagare upplevde sin egen kropp när den virtuella kroppen utsattes för olika stimuli och fann att de kunde leva sig in i den virtuella kroppen trots begränsade interaktionsmöjligheter (Slater et al., 2010).

Immersion är, enligt Ditton och Lombard (1997), inom VR nära besläktat med närvaro. De menar att begreppen flyter samman då det är samspelet mellan VR-tekniken och personens inlevelseförmåga som avgör huruvida det går att leva sig in i VR-simuleringen. De illustrerar med ett exempel där en användare plöstligt märker en störning i simuleringen som stundtals distraherar uppmärksamheten och i förlängningen deras närvaro. I detta exempel är det alltså både användarens känslighet för distraktioner och VR-miljön som i samspel avgör hur enkelt en person kan återgå och leva sig in i upplevelsen igen samt att se förbi mediumet (Ditton & Lombard, 1997). En annan användare hade kanske reagerat på något annat, eller inte alls.

VR och empati

Vetskapen att VR-användare kan leva sig in i en simulering till olika grader är viktig för ett projekt som har målsättningen att studera hur feedback påverkar användarens spelbeslut. Frågan är vad som driver denna benägenhet. Tidigare studier på beteendeförändring, inlärning och liknande inom VR menar att empati är en viktig kanal.

Exempelvis djupdök Ahn, Minh Tran Le och Bailenson (2013) i VRs potential att öka människors benägenhet till perspektivstagande och hjälpande beteende.

I deras studie fick en grupp deltagare testa på att vara färgblinda med hjälp av VR-simuleringar och en annan instruerades till att föreställa sig färgblindhet i samma VR-miljö. Resultaten visade att den förstnämnda metoden var mer effektiv för perspektivtagande bland individer med lägre kapacitet för empati samt att deltagare med förkroppsligade erfarenheter, dvs. simulerad färgblindhet visade större vilja att hjälpa färgblinda utanför laboratoriet efter experimentet (Ahn et al., 2013).

I en studie om medkänsla för djur, visade Ahn, Bostick, Ogle, Nowak, McGillicuddy och Bailenson (2016) att deras deltagare uttryckte en starkare koppling till djur efter VR-upplevelser jämfört med traditionella filmupplevelser av samma material. Resultatet verkar lovande men forskarna hade inte uteslutit vissa underliggande faktorer, exempelvis användarens tidigare kapacitet för empati (Ahn et al., 2016).

Efter en undersökning om hur experimentsdeltagares empatiska kapacitet är kopplat till deras upplevese av virtuella miljöer kontra videoupplevelser, fann Shin (2017) att givet samma narrativ hade deltagarnas begängenhet till empati större inverkan på deras upplevelse än mediumet i sig. Vilket, sammantaget, innebär att termer som närvaro, interaktion, embodiment och immersion uppenbarligen påverkas av användarens inlevelseförmåga - samt, i en bredare bemärkelse, även deras kapacitet för empati.

VR som läromedel

Oavsett tillämpning bygger VR-tekniken på samma logik. Tanken är att simulerad närvaro, interaktion samt en hög grad av immersion och embodiment tillsammans skapar starkare minnen - vilket i förlängningen leder till inlärning och beteendeförändring. Men frågan om inlärning och beteendeförändringar är bestående eller om kunskaperna ifråga kan överföras till andra sammanhang kvarstår.

Ahn, Bailenson och Park (2014) tog sig an nämnda fråga i ett experiment om miljömedvetenhet. Genom att bryta ner en annars komplex och tidsfördröjd konsekvenskedja, instruerades testpersoner att såga ner ett träd i den virtuella miljön för att sedan studera hur denna upplevelse påverkade deras pappersanvändning efter experimentet.

Förväntningen att direkt interaktion med negativa miljöutfall skulle ha högre inverkan på deltagarnas beteenden än andra medium styrktes; en vecka senare självrapporterades även förändringar i deras generella inställning till miljön (Ahn et al., 2014).

Om detta exempel verkligen är ett sätt att skapa bestående beteendeförändringar och överförbarhet till andra områden i livet kan diskuteras, men låt det vara sagt att liknande utmaningar är ett genomgående problem för alla tekniska läromedel - inte bara för VR specifikt.

Bland fördelarna med VR i skolan listas visualiseringar som förhöjer känslan av närvaro, att förvalta möjligheten till interaktion (Ohlsson, Moher & Johnson, 2000), samt möjligheten att elever kan finna sig i situationer som annars är otillgängliga eller alltför riskfyllda.

Smith och Ericson (2009) utnyttjade den sistnämnda fördelen till fullo genom att utveckla ett träningsprogram för barn att öva på brandsäkerhet i virtuella miljöer. Deras resultat visade på högre engagemang och små förbättringar i kunskapsutfall jämfört med traditionella medier – dock testades endast enkla faktakunskaper eftersom det fanns en risk att ett alltför trovärdigt träningsprogram skulle tramuatisera barnen som deltog i experimentet.

Därtill undersökte Detyna och Kadiri (2020) huruvida VR kan användas inom högre utbildning. De kom fram till att vissa ämnen, i deras fall geografi, lämpar sig kursmaterialet bättre för VR-stöd än andra studerade ämnen (humaniora, främst digital design). Deras studie visade också att VR-stöd har positiva effekter överlag, men effekterna för engagemang var också här högre än förbättrade studieresultat.

Större engagemang och aningen bättre resultat räcker dessvärre inte för att rättfärdiga VR som pedagogiskt verktyg. Inte heller lyfter VR-förespråkare upp det som inom social robotik kallas *the novelty effect* (Leite, 2013), dvs. den förväntade beteendeförändringen hos användare när tekniken förlorat sitt nymodighetsglimmer. Det må vara en utmaning för framtiden men måste förr eller senare beaktas. Särskilt inom pedagogik, då barn lätt kan tröttna.

Bland utmaningarna är den mest grundläggande frågan hur VR-miljöer ska inkorporeras i pedagogiska moment; ska det vara ett komplement eller alternativ? Och vad innebär detta för eleven, lärare, klassrummet, osv? Inte heller är det tydligt vilka lärdomar, färdigheter, egenskaper eller – mer generellt - kunskapsutfall som är bäst lämplade för VR-simulering, för att inte tala om hur detta ska bedömas av läraren eller pedagoger (Fowler, 2015).

Från ett mer tekniskt perspektiv nämner Fowler (2015) att det även är oklart ifall en VR-miljö ska vara elev- eller lärarstyrd, hur tekniken används i grupp, vad som är rimliga nivåer på realism och interaktivitet samt hur kvalitén på referensmaterial (typ av källa, validitet och relevans) hanteras, vem som utvecklar materialet? – detta vägas mot kostnadseffektiva läromedel som redan styrks av beprövad erfarenhet.

Andra vanliga missuppfattningar berör förhållandet mellan fängslande innehåll och grafik, samt myten om inlärningsstilar (Sjödén, 2014). I det första fallet presenterar Sjödén (2014) Minecraft och Tetris som exempel på hur enkel grafik inte behöver stå i vägen för långvarigt fängslande innehåll. Han varnar för läromedel som till ytan ger ett gott intryck men – av olika anledningar - inte lyckas engagera elever över tid, vilket är en fälla även för VR ovanpå risken att längre spelsessioner leder till cybersickness, som diskuterades tidigare.

Vad gäller myten om inlärningsstilar har Pashler, McDaniel, Rohrer och Bjork (2009) tydligt visat att det inte finns något empiriskt stöd för fenomenet. Detta är en relevant utmaning för VR-tekniken eftersom preferenser på inlärningsmetod, deras önskan att bli sedda och deras olika fallenhetsnivåer felaktigt (eller i varje fall tveksamt) tillskrivs elevens påstådda inlärningsstil, vilket, bland annat, leder till att behovet för individanpassat studiematerial överdrivs (Pashler et al, 2009).

VR som läromedel verkar vara en svår nöt att knäcka. Det kräver att utvecklare, i samarbete med pedagoger, isolerar specifika kunskapsutfall som ska överföras, skapar en miljö och ett narrativ där färdigheten, lärdomen eller kunskapen hamnar i fokus, samt undersöka hur väl miljön överför det önskade utfallet och utvärdera ifall användaren kan bära det med sig till andra sammanhang.

Läromedel och feedback

Inom pedagogik är det givet att elevers inlärning påverkas av feedback. Hanna (1976) var bland de första som visade att enkla notiser om ifall ett inmatat svar är (in)korrekt kan påverka elevernas prestation till det bättre.

Mycket har hänt sen dess. I sin genomgång på området sammanfattar Hattie och Timperley (2007) olika typer av feedback och diskuterar deras för- och nackdelar. Bland annat beskrivs hur feedback kan syfta på flera saker, exempelvis kan feedback rikta sig till uppgiften, på eleven prestation ifråga, deras studieteknik, motivation samt elevens förmåga att reflektera över sina egna processer. De menar att det är, trots gott med empiriskt stöd för dess positiva effekter, inte självklart hur feedback inkorperas i pedagogiska sammanhang – vilket försämrar dess kraft. Deras modell utgår därför i ett klassrumssammanhang och har som syfte att para ihop rätt feedbackstyp med elevernas utvecklingsstadie, mål och mottaglighet (Hattie & Timperley, 2007).

Detta står i kontrast till Tärning (2018) som undersöker hur feedback används i pedagogiska lärospel. I sin kartläggning utgår hon från lärospel som används av lärare inom den svenska grundskolan och finner att de flesta lärospelen inte tar hänsyn till elevens utveckling inom spelet. Inte heller tar spelutvecklarna chansen att ge mer ingående feedback utöver så kallad *verifikationsfeedback* - där spelare enbart får veta om deras inmatade svar, eller valda alternativ, var (in)korrekt trots dess tveksamma effekter på inlärningen.

Därtill finns *korrektionsfeedback*, där det rätta svaret meddelas vid inkorrekta inmatningar; *uppmuntrande feedback*, som fokuserar på personen ifråga snarare än uppgiften samt *förklarande feedback*, vilket är ett samlingsbegrepp för alla typer av respons som förser användare med härledningar och alternativt beskrivningar på varför ett svar är (in)korrekt (Tärning, 2018).

Huruvida feedback är individanpassad eller generisk behöver dock inte alltid spela roll. Till exempel visade Johnson, Gardner och Wiles (2004) att deltagare som fick veta att deras uppmuntrande feedback var personanpassad presterade inte nödvändigvis bättre än den experimentsgrupp som fick höra att samma feedback var allmän och opersonlig. Deras studie visade också signifikanta samband mellan datorvana och inställning till uppmuntrande feedback, där mer erfarenhet ledde till lägre mottaglighet för smicker (Johnson et al., 2004).

Intressant nog registrerades inga större skillnader i deltagarnas omdöme av spelet, inte heller sett till hur länge spelaren var villig att interagera med spelet – vilket går emot slutsatserna i Higgins, Hartle och Skelton (2001) som menar att feedback bör vara personanpassad och utförlig för att elever ska kunna ta till sig feedback huvudtaget.

Oavsett typ eller personanpassning är det lätt att anta att elever faktiskt tar till sig feedback utan förbehåll. Tärning, Lee, Andersson, Månsson, Gulz och Haake (2020) undersöker *feedback neglect*, vilket, mycket kortfattat, är ett fenomen som syfter på elevers tendens att inte ta till sig feedback. Med deras modell kan interaktionen mellan feedback och elev följas steg för steg, från första ögonkastet tack vare eye tracking, läsning, bearbetning, till handling och elevens beslut att antingen ändra sitt efterföljande beteende eller inte – tack vare logdata från experimentsmiljön.

Deras resultat tyder på att elever till varierande grad ignorerar feedback under alla steg och att få klarar sig genom hela processen (Tärning et al., 2020).

Sammanfattningvis är feedback ett verktyg som kan användas för beteendeförändring och inlärning, VR likaså. Det är dock inte helt oproblematiskt att använda feedback i läromedel, inte heller VR som läromedel. Det är med denna bakgrund som vi vill undersöka hur feedback kan användas inom VR.

5 Experimentsupplägg

Givet kursens målsättning, tidsbegränsning, projektets komplexitet och våra implementeringsmöjligheter, avgränsas experimentet till en studie på några få feedbackstyper. Främst undersöks huruvida visuell feedback som anspelar på experimentdeltagarnas empatiska kapacitet kan påverka deras beteende i en VR-miljö samt till vilken grad det påverkar deltagarens VR-upplevelse överlag.

5.1 Val av feedbackstyp

Experimentmiljön är förhållandevis enkel: Spelaren instrueras till att samla ihop skräp från en virtuell campingplats på tid. Skräpobjekt valdes på två grunder: 1) vad har en fjällvandrare rimligtvis har med sig på en vandring? Och 2) vad skulle rimligtvis väcka spelarens empati och uppmuntra beteendeförändring?

Med dessa två kriterier modellerades tre sorters skräpobjekt med tillhörande skräpkonsekvenser:

1. *Plastflaska* → Ihopkrynklad plastflaska
2. *Kartong* → Kartong försvinner helt (ingen skräpkonsekvens)
3. *Metallburk* → Ihopkrynklad burk (Scenario A) eller räv som fastnar i burken (Scenario B)

Grundtanken är att visualisera vad som kan hända i en fjällmiljö ifall skräp inte städas upp med konsekvenser.

Att den skräpkonsekvens som skulle väcka empati hos spelaren blev en fjällräv som fastnade i en metallburk var i grunden ett designbeslut då det lika gärna kunde ha varit en fjällhare som fastnat i kartong. Effekten av feedbacksstimuli diskuteras även senare i rapporten.

I termer av Tärnings (2018) terminologi skulle skärpkonsekvenserna båda kunna beskrivas som förklarande feedback*,* då spelaren får en uppfattning om varför ett (icke)val av skräp var felaktigt eller otillräckligt, eller verifikationsfeedback eftersom skräpkonsekvenser inte går att plocka upp och att spelaren får besked att det är försent att plocka upp detta objekt och bör därför sikta in sig på andra. Det finns naturligtvis en risk att spelaren inte bearbetar skärpkonsekvenser som relevant information under spelets tidspress, vilket skulle räknas som feedback neglect enligt Tärning et al. (2020). För att kontrollera för feedback neglect ombeds spelarna att besvara på frågor som indikerade ifall de lagt märke till att vissa skräpobjekt omvandlas under spelet.

Sammanfattningsvis innehåller spelet tre typer av skärpkonsekvenser: en kontroll som är neutral (2), inanimata (1 och 3A), animala (3B), där det senare anspelar på användarens empati och koppling till djur i naturen. Skräpkonsekvenserna agerar oberoende variabel, och kommer jämföras med varandra för att se huruvida feedbacken ifråga har någon effekt på antalet insamlade objekt – den beroende variabeln.

Hypoteser och förväntat resultat

Målet med undersökningen är att förstå hur dessa konsekvenser påverkar testpersonens val av skräp. Först undersöks huruvida spelarens val av skräpobjekt kan påverkas med synliga konsekvenser (H1). Förväntningen är att det finns en positiv relation mellan skärpkonsekvens som feedback (oberoende variabel) och antal insamlade skräpobjekt (beroende variabel) då spelaren ser vad som händer ifall de inte väljer motsvarande objekt.

Därtill undersöks ifall empatidriven eller neutral feedback har mer eller mindre påverkan på spelarens val av skärp eftersom vissa konsekvenser antas väcka större empati (H2).

Här förväntas de animala konsekvenserna (3B) ha störst inverkan på spelarnas val av objekt och därigenom bli deras mest insamlade skärpobjekt medan inanimata skräpobjekt (1 och 3A) förväntas ha lägre genomslag än 3B, men högre än inga konsekvenser alls (2). Vidare antas denna effekt vara starkast bland människor som tror sig ha en starkare koppling till naturen och anser sig vara mer empatiska (H3).

Det är också intressant att undersöka ifall förekomsten av empatidrivna skärpkonsekvenser kan påverka spelarens VR-upplevelse i sin helhet. Det är inte tydligt vilken riktning skräpkonsekvenserna kommer påverka spelarens generella uppfattning men viss effekt förväntas (H4). Särskilt bland spelare med starkare empatisk kapacitet (H5) eftersom de anses vara mer mottaglig för den effekt skärpkonsekvenserna antas ha.

Hypotes 1, 2 och 3 kommer besvaras med jämförelser av antal ihopsamlade skärpobjekt av varje sort samt spelarens enkätsvar om deras inställning till djur, natur och andra människor innan experimentet genomförts (Enkät finns i bilaga 1). Hypotes 4 och 5 besvaras med enkätsvar, dvs. deltagarnas inställning till djur, natur och andra människor samt deras enkätsvar om VR-upplevelsen efter experimentet genomförts (bilaga 1).

Experimentsprotokoll

I VR-miljön ska användare spela ett spel som går ut på att samla ihop i fjällen vanligt förekommande objekt. Ifall testpersonen inte hinner samla ihop skärp inom utsatt tid förvandlas det till en så-kallad skräpkonsekvens. Under varje spelomgång spawnas lika många av varje objekt i olika takt. Innan deltagarna delas in i experimentsgrupper kommer de, som sagt, svara på en samtyckesblanket och därefter en enkät som screenar för tidigare VR-erfarenhet samt andra nödvändiga kontrollvariabler som kön, ålder och deras inställning till naturen och andra människor – vilket kommer sammanvägas för att genera ett approximerat mått på personens kapacitet för empati, även kallat *empatimått* i resultatdelen. Deltagarna kommer därefter att slumpmässigt delas in en av två feedbacksgrupper, se nedan, testa miljön samt samla ihop objekt i två spelomgångar.

Grupp 1: Utan räv som skräpkonsekvens (scenario A)
Grupp 2: Med räv som skräpkonsekvens (scenario B)

Efter sista spelomgången kommer deltagarna svara på andra delen i enkäten där de svarar på frågar om deras upplevelse av VR-miljön. På så sätt kan jämförelser mellan deras upplevelseomdömen och feedbackstyp genomföras.

6 Resultat

Här följer en presentation av resultat. På grund av Folkhälsomyndighetens restriktioner kring Covid-19 kunde inte den tänkta målgruppen studeras. Istället för mellan- och högstadie musiebesökande skolelever rekryterades 15 personer till ett pilotexperiment i VR-labbet på Ingvar Kamprads Design Centrum.

All analys genomfördes i Microsoft Excel™ med hjälp av Analysis ToolPak på 5% signifikansnivå och tabeller samt regressionsutskrifter finns bland bilagorna.

Testgruppen

Testgruppen bestod av 7 kvinnor och 8 män med en median ålder på 25 år. Tidigare erfarenhet med VR varierade bland deltagarna: Några hade inte några tidigare erfarenheter överhuvudtaget (5 av 15); vissa hade testat på VR fler än en gång tidigare (4 av 15), andra deltagarna bara hört talas om VR men aldrig testat (3 av 15), samt ett par som testat VR en gång tidigare (2 av 15). Endast en person hade testat VR flera gånger och läst på om ämnet.

Tabell 1 visar fördelningen av män och kvinnor i respektive scenario med tillhörande genomsnittliga empatimått (median). Notera att empatinivån är lägre i scenario B – vilket diskuteras längre fram.

	Närvarande räv (scenario B)	**Ingen räv (scenario A)**	**Empatimått (median)**
Antal män	5	3	22,5
Antal kvinnor	2	5	22
Alla deltagare	7	8	22
Empatimått (median)	19	23	

Tabell 1: Fördelning av kön och empati mellan scenarion

Val av skräpobjekt och synliga konsekvenser

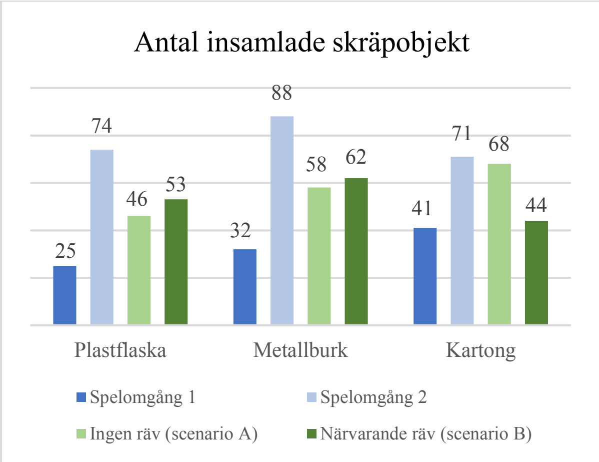
Det övergripande målet med studien var att undersöka ifall skräpkonsekvenser som feedback påverkar spelarens val av skräp inom VR-miljön (H1). Om så vore fallet hade det återspeglats med signifikanta skillnader i antal insamlade objekt.

Tabell 3 ger en indikation på att så inte är fallet då antal insamlade skärpobjekt är nästan jämt fördelade mellan de tre skräpkonsekvenserna i respektive spelomgång. Ensidiga ANOVA-test mellan spelomgångar fann inte heller några signifikanta skillnader då F-värden för spelomgång 1 landade på 0,91 och 0,59 för omgång 2. Vilket är lägre än tröskelvärdet $F(2, 44) = 3{,}23$ på 5% nivån. Fullständig ANOVA utskrift finns i bilaga 2.

Vidare är det viktigt att poängtera skillnaden mellan antal insamlade objekt för män respektive kvinnor – tabell 2 nedan. Män verkar samla in betydligt fler objekt.

	Spelomgång 1	**Spelomgång 2**	**Ingen räv (scenario A)**	**Närvarande räv (scenario B)**
Män	69	153	91	131
Kvinnor	29	80	81	28

Tabell 2: Antal insamlade objekt mellan kön och scenario



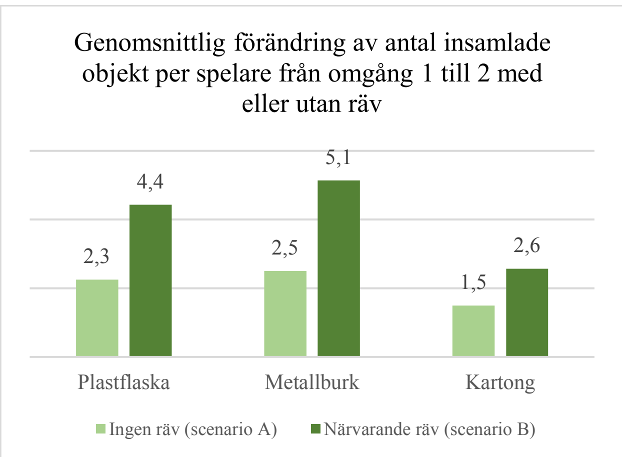
Tabell 3: Antal insamlade skräpobjekt

Intressant nog var kartong det mest ihopsamlade objektet under spelomgång 1, fast det minst samlade objektet under omgång 2. Det är också nämnvärt att spelarna lyckades samla ihop nästan dubbelt så många skräpobjekt i den andra omgången – vilket tyder på någon form av beteendeförändring mellan omgångar. Frågan om vad som driver dessa skillnader undersöks härnäst.

Olika feedbackstyper och deras effekter på insamling

Det enda som skiljer scenario A och B var skärpkonsekvensen för metallburken. I scenario A förvandlas den till en ihopkrynklad variant, medan i scenario B fastnar en räv med nosen i metalburken. Plast och kartong (neutral) hade samma skräpkonsekvens i både scenario A och B.

Tabell 4 visar den genomsnittliga skillnaden på antal insamlade skärpobjekt per spelare mellan omgångar.
Oavsett typ av skärp samlades det ihop fler skärpobjekt i scenario B jämfört med scenario A. Den största ökningen finns i antalet insamlade metallburkar, där varje spelare i genomsnitt samlade ihop fem fler metallburkar när räven användes som skärpkonsekvens jämfört med en ihopkrymplad metallburk.



Tabell 4: Genomsnittliga förändringen av antal insamlade skräpobjekt per spelare mellan omgångar med eller utan räv

Vid första anblick ger tabell 3 stöd för H2: feedbackstypena verkar påverka spelarens val av objekt olika mycket. Metallburkar var även, enligt förväntning, det mest insamlade objektet. Därefter plastflaskor, och sist, kartong – se tabell 3**.** Dock ska det noteras att alla skärpobjekt verkar ha gynnats av rävens närvaro under spelet enligt tabell 4 – den genomsnittliga insamlingen av plastflaskor fördubblades nästan mellan spelomgång 1 och 2 samt rävens närvaro är enbart signifikant vid insamlingen av plastflaskor på 5% nivån enligt tabell 5 (Notera att interceptet är bortaget). Alltså finns det inte stöd för H2, vilket väcker frågan ifall andra faktorer driver förändringen mellan omgångar.

H3 undersöker om det är deltagare som tror sig ha en starkare koppling till naturen och anser sig vara empatiska som driver skillnaden mellan omgångar. Tabell 4 visar icke-signifikanta resultat för empatimåttet, dvs. frågorna som kontrollerade för deltagarnas inställning till djur, natur och andra människor - vilket troligtvis beror på det låga antalet observationer då majoriteten av dem andra kontrollvariablerna inte heller visade signifikans, så som kön. (Fullständig regressionutskrift finns i bilaga 3).

Beroende variable: **Genomsnittlig förändring (Plastflaska)**	**Koefficient**	**Medelfel**	**P-värde**
Empatimått	0,38	0,21	0,10
Rävens närvaro (Närvarande = 1, Frånvarande = 0)	3,14	1,39	0,05*
Deltagarens kön (Kvinna = 1, Man = 0)	0,13	1,13	0,93
Tidigare erfarenhet	-0,07	0,46	0,88
Genomsnittlig förändring (Metallburk)	**Koefficient**	**Medelfel**	**P-värde**
Empatimått	-0,04	0,24	0,88
Rävens närvaro (Närvarande = 1, Frånvarande = 0)	2,04	1,57	0,22
Deltagarens kön (Kvinna = 1, Man = 0)	-0,83	1,46	0,58
Tidigare erfarenhet	0,71	0,52	0,20
Genomsnittlig förändring (Kartong)	**Koefficient**	**Medelfel**	**P-värde**
Empatimått	0,02	0,17	0,90
Rävens närvaro (Närvarande = 1, Frånvarande = 0)	0,65	1,15	0,59
Deltagarens kön (Kvinna = 1, Man = 0)	-1,24	1,08	0,28
Tidigare erfarenhet	0,17	0,38	0,66

Tabell 5: Regressionsutskrift på undersökta variabler

Olika feedbackstyper och deltagarnas VR-upplevelse

Med H4 förväntades förekomsten av en empatidriven skräpkonsekvens ha viss effekt på spelarnas inställning till deras VR-upplevelse överlag. Enkäten som användes till underlag finns i bilaga 1. Majoriteten av frågorna besvarades med fritext (markeras med *). Svaren finns sammanfattade i tabell 6 nedan. Alla deltagare la märkte till att vissa skärpobjekt förvandlades (fråga 14), vilket var en grund förutsättning för att ingå i analysen. Vidare är det viktigt att poängtera att endast 7 av 15 deltagare tog del av räven som skärpkonsekvens och deras svar inkluderas i parantes för att enklare återkoppla till H4.

Ifall det fanns stöd för H4, oavsett riktning på effekten, skulle andelen i respektive fråga och tillhörande svarsalternativ visa en tydlig över- alternativt underrepresentation. Detta verkar inte vara fallet då deltagarna som tog del av räven som skräpkonsekvens motsvarar ungefär hälften av alla inrapporterade svar utan någon tydlig riktning på deras innehåll eller värdering. Ensidiga ANOVA testa bekräftar ovanstående indikation med ett F-värde på 0,71, vilket är lägre än tröskelvärdet F(1, 24) = 4,26 på 5% nivån – ANOVA utskrift i bilaga 4.

Trots att det inte finns stöd för H4 kan effekten rimligtvis vara starkare bland deltagare med större kapacitet för empati, dvs, H5. Bland de sju deltagare som tog del av räven som skräpkonsekvens kommenterade endast två dess existens – vilket inte tyder på anmärkningsvärd effekt bland dem som tror sig ha högre empati. Givetvis kan det inte uteslutas att räven ändå påverkat deltagarna, måhända omedvetet, men underlaget visar inga sådana tecken.

11. Hur skulle du beskriva den här VR-upplevelsen? *	**Antal fall**
Talade positivt om VR-miljön ("kul", "häftigt", "spännande", "grym", osv.)	(7) 15
Påpekade att spelförståelsen blev bättre (efter "andra gången", "övning", "testomgång", osv.)	(3) 6
Uttryckte förvirring, svårigheter eller eftersökte tydligare instruktioner, osv.	(2) 6
12. Vad tänkte/kände du när du spelade första omgången? *	
Förvirrad, svårt att förstå navigering, vad som skulle göras osv.	(4) 11
"Wow", "roligt", "kul", "verkligt"	(1) 3
Jag blev exalterad och kände mig inne i spelet	(2) 3
Fråga 13. Vad tänkte/kände du när du spelade andra omgången? (Var det någon skillnad?) *	
Jag förstod bättre, kände mig inte förvirrad längre	(4) 8
Ja, det var skillnad	(3) 7
Fråga 15. Var det något som påverkade vilka skräpobjekt/typer du valde att plocka upp (i första resp. andra omgången)? *	
Prioriterade vissa objekt ("närmast", "såg dem först", "tyckte synd om räven", osv)	(4) 8
Nej…	(1) 4
Hade lättare att plocka upp vissa objekt…	(2) 3
Om objektet blinkade…	(0) 1
Fråga 16. Kan du tänka dig testa spelet igen?	
Ja!	(7) 15

Tabell 6: Sammanfattning på fritext svar

7 Diskussion av resultat

Följande avsnitt behandlar resultat som redovistats i föregående kapitel. Det finns som sagt indikationer för hypotes 1 och 2, men inte signifikant stöd. För hypotes 3, 4 och 5 finns inget stöd alls. Ifall detta beror på antalet respondenter kan diskuteras då fler expertimentsdeltagare skulle troligen stabiliserat resultaten - oavsett om det handlar om fler respondenter i rätt målgrupp eller inte. Det låga antalet respondenter orsakar skevheter.

Jämför exempelvis det genomsnittliga antalet ihopsamlade objekt mellan könen och faktumet att fler män var representerade i scenario B. Ett annat exempel är spridningen av empati mellan testgrupperna, vilket inte var jämnt fördeltat och kan ha förvrängt eventuella korrelationer till deltagarnas empatiska förmågor.

Vidare skulle fler deltagare öppna upp för att testa fler stimulikombinationer. Till exempel vore det intressant att variera kontrollobjektet, dvs. det objekt som inte hade en skärpkonsekvens, för att minimera sannorlikheten att underliggande variabler så som storlek, synlighet eller något annat skulle påverka utfallet.

Fler respondenter skulle även möjliggöra ännu fler typer av feedbackstimuli. Exempelvis hade det varit givande att studera huruvida liknande effekter i tabell 2 och 3 även stöds om fjällräven skulle bytas ut mot en fjällhare alternativt formge animala stimulin som visar på mildare och grövre konsekvenser. Även olika spawningsintervall hade varit givande att studera med fler deltagare då det kan tänkas påverka resultaten. Tänk, till exempel, om antalet objekt vore lågt och utspritt eller högt och frekvent. Detta skulle säkerligen medföra märkbara skillnader i spelarnas samlingsbeteende.

Log data

Det hade varit användbart att inkludera andra variabler i dataunderlaget, så som skräpobjektets initiala avstånd till spelaren, tiden det tar från att skärpobjektet spawnas till att det plockas upp samt objektens storlek. Någon form av reaktionstid. Dessa variabler skulle kontrollera för eventuella skevheter då, exempelvis, kartongobjekten var väsentligt större än plast och metallobjekten vilket innebär att de troligen var enklare att se samt att plocka upp (vissa deltagare poängterade detta). Alternativt åsynliggöra ifall spelaren utvecklar någon form av prioritetsordning under spelets gång, vilket verkar vara fallet enligt tabell 1 eftersom skillnaderna mellan omgångar är markant för objekten med tillhörande skräpkonsekvenser. Detta borde rimligtvis kunna brytas ner i detalj med hjälp av avstånds- och tidsvariabeler. Vidare hade en tidsvariabel gjort det möjligt att kontrollera för objektens spawningsfrekvens som eventuell felkälla.

Målgruppen

På grund av rådande pandemi var det inte möjligt att genomföra experimentet med den tänkta målgruppen. Detta lär ha haft en effekt på resultaten. Främst av två anledningar: 1) Barn antas ta till sig ny teknik snabbare än vuxna, och 2) Deras kapacitet för empati är fortfarande under utveckling. Detta är särskilt relevant när spelomgång 1 och 2 jämförs då en del av förbättringen måste tillskrivas spelarnas inlärning. Det vore alltså intressant att veta om och isåfall hur mycket av förändringen mellan omgångar drivs av teknisk ovana - vilket, under antagandet att barn lär sig snabbare, skulle minimera skillnaderna i tabell 2 och 3.

Sammanslaget med barns outvecklade kapacitet för empati kan det vara så att effekten av skräpkonsekvenser förminskas eller försvinner helt.

Förslag till framtida studier

Framtida forskning bör undersöka hur hypotes 1 och 2 står sig med fler respondenter. Det vore en bra början. Med fler respondenter kan även andra stimulikombinationer undersökas – vilket också lär ge olika typer av utslag. Viktigast vore att studera den eftersökta målgruppen samt balansera ut könsfördelningen bland deltagarna.

8 Diskussion av utveckling

När spelet testades uppmärksammades några faktorer som var svårare för försökspersonerna än vad som var väntat. Det hade varit fördelaktigt att förändra följande punkter i en ny version av spelet.

Navigering

En faktor som upplevdes svår för försökspersonerna var att förstå hur de skulle teleportera sig via stolparna, detta trots att de först fick träna på att göra det i en testmiljö innan de började spela. För att utveckla spelet vidare hade det kanske gått att komma på en bättre lösning för att teleportera sig från ett område till ett annat. Ett exempel skulle kunna vara att placera områdena närmare varandra så att de hade gått att nå med den "vanliga" funktionen för teleportering.

För att kunna leda spelaren genom spelets olika moment på ett smidigt sätt så blinkade de objekt som spelaren skulle interagera med härnäst. Detta fungerade på många sätt bra, men vid vissa tillfällen hade spelarna svårt att upptäcka blinkningarna. De objekt där problemet uppstod hade kunnat göras tydligare med, till exempel, ljudeffekter.

Synlighet

Ytterligare en svårighet i spelet var att försökspersonerna hade svårt att se tidtagningen och poängräkningen när de spelade spelet. Här hade andra lösningar diskuterats, till exempel att fästa dem i spelarens vy så att de kunde se tid och poäng oavsett var de tittade. Men det ansågs ge för mycket spelkänsla och ta bort från känslan av att man gått in i en värld.

Till en utveckling av spelet hade de kanske kunnat visualiseras på ett tydligare sätt. I momentet där spelarna uppmuntras att plocka skräp var det svårt för dem att förstå när de hade plockat upp ett objekt.

Onaturlig interaktion

Spelet är implementerat så att när man lyft ett objekt till en viss höjd så försvinner det och registreras som upplockat. Tanken med det var att spelaren ska uppleva det som att det läggs i en ryggsäck på ryggen, vi ville undvika att spelet blev fokuserat på att till exempel försöka kasta skräpet i behållaren.

Tanken var att fokus istället skulle läggas på konsekvenserna som kom då spelaren inte hann plocka upp en viss skräptyp. En annan implementeringslösning, alternativt en introduktion till spelet hade kunnat lösa problemet. Mobiltelefonens fysik som upplevdes svår att implementera på ett bra sätt, gjorde att försökspersonerna hade svårt att lyfta upp den. Spelupplevelsen hade blivit bättre om mobiltelefonen fungerade som den skulle gjort i verkligheten.

Snabb återkoppling

Överlag hade utvecklingsprocessen behövt fler tillfällen för testning. Genom att oftare få återkoppling från testpersoner hade det varit lättare att förstå behovet av ett narrativ för att guida spelarna genom de olika momenten. Detta var en insikt som kom till gruppen, men det hade varit fördelaktigt om det skett tidigare. Narrativet fyllde sin funktion, men hade kunnat vara mer utvecklat för att skapa en starkare koppling till budskapet som spelet ämnade att förmedla, att uppmuntra till att värna om känsliga miljöområden.

Feedback

Till viss del implementerades feedback och ljudeffekter, men det hade kunnat utvecklas ytterligare. Det saknades särskilt i det sista spelmomentet, när användaren trycker på knappen till återvinningsmaskinen. Då uppenbarar sig återvunna objekt på baksidan av maskinen, men ofta stod spelarna så att maskinen skymde sikten. Spelarna trodde ofta att de inte lyckades trycka på knappen eller att den inte fungerade. Detta hade kunnat undvikas genom enkel ljudfeedback.

9 Slutsats

Projektet hade som övergripande mål att skapa en high-fidelity protoyp, med ett inbyggt experiment som testas på försökspersoner så fort som möjligt - vilket är uppnått.

Experimentsresultatet tyder på att feedback inom VR kan påverka spelbeslut, men att olika feedbackstyper effekt inte kan fastställas. Vad som driver denna skillnad behöver studeras vidare. Det gick inte att koppla spelarnas förbättringar mellan omgångar till specifika feedbackstimuli, inte heller till deras empatiska förmåga. Därtill verkade spelarnas generella uppfattning av VR-upplevelsen vara oberörd av feedbackstimulins utforming.

Framtida forskning får sikta sig in på fler respondenter för att stabilsera resultatet. Slutligen har gruppen fått med sig många lärdomar som kommer vara användbara i framtiden.

10 Referenser

Ahn, S.J., Bailenson, J.N. och Park D. (2014). Short- and long-term effects of embodied experiences in immersive virtual environments on environmental locus of control and behavior. *Computers in Human Behaviour, 39, pg :235-245*

Ahn, S.J., Bostick, J., Ogle, E., Nowak, K.L., McGillicuddy, K.T. och Bailenson J.N. (2016). Experiencing Nature: Embodying Animals in Immersive Virtual Environments Increases Inclusion of Nature in Self and Involvement With Nature. *Journal of Computer-Mediated Communication, 21, pg. 399-419*

Ahn, S.J, Minh Tran Le, A. och Bailenson J. (2013) The Effect of Embodied Experiences on Self-Other Merging, Attitude and Helping Behaviour. *Media Psychology, 16, pg. 7-38*

Bailenson, J.N. (2019). Experience on Demand: what virtual reality is, how it works, and what it can do. *W.W. Norton & Company*

Biocca, F. (1997). The cyborg's dilemma: Embodiment in virtual environments. In Marsh, J.P., Nehaniy, C.L., and Gorayska, B. (Eds.), *Cognitive Technology: Humanizing the Information Age.* Second International Conference on (pp. 12–26). Stoughton, WI: IEEE

Cummings, J.J och Bailenson, J.N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology, vol 19, pg. 272–309*

Detyna, M och Kadiri, M. (2020). Virtual Reality in the HE classroom: feasibility, and the potential to embed in the curriculum. *Geography in Higher Learning, vol 44, pg. 474-485*

Ditton, T och Lombard, M. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *vol 3(2)*

Fowler, C. (2015). Virtual Reality and learning: Where is the pedagogy? *British Journal of Edu. Technology, 46, pg. 412-422*

Gibson, J.J. (1979). The Ecological Approach to Visual Perception. *Boston: Houghton-Mifflin*

Hattie, J. och Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, vol. 77(1), pg. 81-112*

Hanna, G.S. (1976). Effects of Total and Partial Feedback in Multiple-Choice Testing Upon Learning. *Journal of Edu. Research, 69(5), pg. 202-205*

Higgins, R., Hartley, P. och Skelton, A. (2001). Getting the Message Across: the problem of communicating assessment feedback. *Teaching in Higher Education, 6(2)*

Leite, I., Martinho, C. och Paiva, A. (2013). Social Robots for Long-term Interaction: A Survey. *Internt. Journal of Social Robotics, vol.5, pg. 291–308*

Ohlsson, S., Moher, T.G. och Johnson A. (2000). Deep Learning in Virtual Reality: How to Teach Children That the Earth is Round. *IEEE Computer Society, NW Washington DC, hämtad från: ww.evl.uic.eduroundearth*

Shin, D. (2017). Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior January, 2018:78:64-73*

Slater, M., Spanlang, B., Sanchez-Vives, M.V. och Blanke, O. (2010). First Person Experience of Body Transfer in Virtual Reality. *PLoS ONE 5(5)*

Sjödén, B. (2014). Vad är ett bra digital läromedel?. Från Persson, A. och Johansson, R. (Eds.), *Vetenskapliga perspektiv på lärande, undervisning och utbildning i olika institutionella sammanhang (pp. 65-78).* Studentlitteratur AB.

Smith, S. och Ericson, E. (2009). Using immersive game-based virtual reality to teach fire-safety skills to children. *Virtual Reality, vol 13, pg. 87-99*

Tham, J., Duin Hill, A., Gee, L., Ernst, N., Abdelqader, B. och McGrath, M. (2018). Understanding Virtual Reality: Presence, Embodiment, and Professional Practice. *IEEE Transactions on Professional Communication, vol 61, n.2*

Tärning, B. (2018). Review of Feedback in Digital Applications – Does the Feedback They Provide Support Learning? *Journal of Information Technology Education: Research, vol 17, pg. 247-283*

Tärning, B., Lee, Y.J., Andersson, R., Månsson, K., Gulz, A. och Haake, M. (2020). Assessing the black box of feedback neglect in a digital educational game for elementary school. *Journal of the Learning Sciences*, *DOI:10.1080/10508406.2020.1770092*

Weech, S., Kenny, S. och Barnett-Cowan, M. (2019). Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Front. Psychol. 10:158*

Hemsidor:

Håll Sverige Rent (2020): Om Håll Sverige Rent. Hämtad från https://hsr.se/om-hall-sverige-rent. (01-11-2020)

Keep Talking and Nobody Explodes (2020): Bomb Defusal Manual. Hämtad från https://www.bombmanual.com/. (01-11-2020).

Bilagor

Bilaga 1: Experimentsmanus och enkät
Hej!

Tack för att du vill vara med i vårt experiment.

Du kommer strax få testa ett VR-spel som vi utvecklat från grunden. Innan du går in i världen kommer du få bekanta dig med rörelserna och VR-miljön i några minuter. Viktigt att tänka på är att vi främst är intresserade av att utvärdera din interaktion med spelet, inte nödvändigtvis din spelprestation.

Ditt deltagande är anonymt och du får närsomhelst avbryta experimentet utan att behöva förklara dig. All data kommer att användas för vår forskning och inget annat. Vår data består främst av observationer under spelets gång och enkätsvar ifall du samtycker. Du får gärna tala högt under sessionen om du vill.

Hänvisa deltagaren till enkäten
Först vill ställa dig några enkla frågor om din bakgrund och några frågor om din relation till naturen och andra människor.

1. Är du Man, Kvinna eller av flytande könstillhörighet?
2. Hur gammal är du?
    - Under 18
    - 18-24
    - 25-34
    - 35-44
    - 45-54
    - 55-64
    - 64+
3. Hur skulle du beskriva din tidigare erfarenheter med VR? (Kryssa i ditt svar)
    - Jag har inga tidigare erfarenheter med VR
    - Jag har hört talas om VR men inte testat det personligen
    - Jag har testat en gång tidigare
    - Jag har testat fler än en gång tidigare
    - Jag har testat flera gånger och läst på om VR

Nu kommer några frågor om relation till naturen och andra människor

4. Generellt, hur ofta försöker du förstå andra personers perspektiv?
    - Nästan aldrig
    - Lite då och då
    - Ibland
    - Ofta
    - Nästan alltid
    - Jag vet inte
5. Generellt, hur ofta försöker du förstå hur andra ser på sitationen?
    - Nästan aldrig
    - Lite då och då
    - Ibland
    - Ofta
    - Nästan alltid
    - Jag vet inte
6. Jag känner en stark koppling till naturen omkring mig
    - Jag håller med påståendet
    - Jag håller inte med påståendet
    - Jag är tveksam inför påståendet
    - Jag vet inte
7. Jag tror att alla levande varelser har känslor
    - Jag håller med påståendet
    - Jag håller inte med påståendet
    - Jag är tveksam inför påståendet
    - Jag vet inte
8. Jag föreställer mig att mitt liv är en del av en större cyklisk process
    - Jag håller med påståendet
    - Jag håller inte med påståendet
    - Jag är tveksam inför påståendet
    - Jag vet inte
9. Jag tror mig ha en djup förståelse för hur mina handlingar påverkar omgivningen
    - Jag håller med påståendet
    - Jag håller inte med påståendet
    - Jag är tveksam inför påståendet
    - Jag vet inte
10. Jag känner att mitt välmående är beroende av naturen
    - Jag håller med påståendet
    - Jag håller inte med påståendet
    - Jag är tveksam inför påståendet
    - Jag vet inte

Innan eller när deltagaren i Tutorial miljön
Syftet med den här genomgången är att du ska bekanta dig med vissa rörelser och mekanismer.

1. Du ska lära dig greppa ett objekt (burken eller bollen)
2. Du ska lära dig teleportera dig med handkontrollen
3. Du ska lära dig teleportera dig med hjälp av skyltar

Innan deltagaren går in i vår VR-miljö, läs dessa tre punkter

1. Endast objekt som är kopplade till spelets narrativ får att interagera med.
2. Spelet kommer gå på tid, du får spela två gånger och din uppgift att samla ihop skräp.
3. Notera att skärpobjekten är lika värda och har en timer, vilket innebär att de inte kommer gå att plockas upp efter en viss tid.

Experimentet börjar
Ifall deltagaren inte förstår hur den ska gå vidare får vi uppmuntra dem att leta efter ledtrådar, så som blinkande objekt. Särskilt vid teleportstolparna, burken som startar spelet och knappen i maskinen.

Efter deltagaren tryckt på återvinningsmaskinen
Bra, det var första spelomgången. Nu vet du vad du kan förvänta dig av VR-miljön och därför vill vi att du ska spela igen. Har du några frågor innan vi sätter igång?

Efter testomgången
Nu var andra testomgången färdig. Vi har ytterliggare några frågor för dig att besvara på...

11. Hur skulle du beskriva din VR-upplevelsen?
    - Jag vet inte
    - Jag vill inte svara
    - *Fritext*

12. Kan du tänka dig testa spelet igen? (Kryssa i ditt svar)
    - Ja
    - Nej
13. Vad tänkte/kände du när du spelade första gången?
    - Jag vet inte
    - Jag vill inte svara
    - *Fritext*
14. Vad tänkte/kände du när du spelade andra gången? (Var det någon skillnad?)
    - Jag vet inte
    - Jag vill inte svara
    - *Fritext*
15. Märkte du att vissa skräptyper förvandlades? (Kryssa i ditt svar)
    - Ja
    - Nej
16. Vad tror du tror påverkade ditt val av skräpobjekt (i första respektive andra omgången(Kryssa dem ord som gäller för dig)?
    - Nej
    - Jag vet inte
    - Jag vill inte svara
    - *Fritext*

Bilaga 2: ANOVA utskrift

ANOVA: Spelomgång 1 utan räv

Groups	*Count*	*Sum*	*Average*	*Variance*	
Plastflaska	15	25	1,67	1,82	
Metallburk	15	32	2,13	3,98	
Kartong	15	41	2,73	7,40	
Source	*DF*	*SS*	*Mean Square*	*F-stat*	*P värde*
Groups (between groups)	2	8,58	4,29	0,91	0,41
Error (within groups)	42	198,00	4,71		
Total	44	206,58	4,69		

ANOVA: Spelomgång 2 utan räv

Groups	*Count*	*Sum*	*Average*	*Variance*	
Plastflaska	15	74	4,93	8,33	
Metallburk	15	88	5,87	6,12	
Kartong	15	71	4,73	11,40	
Source	**DF**	**Sum of Square**	**Mean Square**	**F Statistic**	**P-value**
Groups (between groups)	2	10,98	5,49	0,59	0,56
Error (within groups)	42	387,60	9,23		
Total	44	398,58	9,06		

Spelomgång 1 med Räv

Groups	*Count*	*Sum*	*Average*	*Variance*	
Plastflaska	7	11	1,62	1,93	
Metallburk	7	13	2,00	3,85	
Kartong	7	13	2,46	7,63	
Source	**DF**	**Sum of Square**	**Mean Square**	**F Statistic**	**P-value**
Groups (between groups)	2	0,38	0,19	0,07	0,93
Error (within groups)	18	47,43	2,63		
Total	20	47,81	2,39		

ANOVA: Spelomgång 2 med Räv

Groups	*Count*	*Sum*	*Average*	*Variance*	
Plastflaska	7	63	4,85	8,59	
Metallburk	7	72	5,54	4,86	
Kartong	7	58	4,46	12,25	
Source	**DF**	**Sum of Square**	**Mean Square**	**F Statistic**	**P-value**
Groups (between groups)	2	23,52	11,76	1,61	0,23
Error (within groups)	18	131,71	7,32		
Total	20	155,24	7,76		

Bilaga 3: Regressionsutskrifter

Genomsnittlig förändring (Plastflaska)

Regression Statistics	
Multiple R	0,63
R Square	0,40
Adjusted R Square	0,16
Standard Error	2,25
Observations	15

Plast	*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-6,07	4,77	-1,27	0,23	-16,71	4,57
Empati	0,38	0,21	1,79	0,10	-0,09	0,84
Räv	3,14	1,39	2,25	0,05	0,03	6,24
Kön	0,13	1,30	0,10	0,93	-2,77	3,02
Tidigare erfarenhet	-0,07	0,46	-0,16	0,88	-1,09	0,95

Genomsnittlig förändring (Metallburk)

Regression Statistics	
Multiple R	0,61
R Square	0,37
Adjusted R Square	0,12
Standard Error	2,54
Observations	15

Metall	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	2,12	5,38	0,39	0,70	-9,86	14,11
Empati	-0,04	0,24	-0,15	0,88	-0,56	0,49
Räv	2,04	1,57	1,30	0,22	-1,46	5,54
Kön	-0,83	1,46	-0,56	0,58	-4,09	2,44
Tidigare erfarenhet	0,71	0,52	1,37	0,20	-0,44	1,86

Genomsnittlig förändring (Kartong)

Regression Statistics	
Multiple R	0,46
R Square	0,21
Adjusted R Square	-0,11
Standard Error	1,87
Observations	15

Kartong	*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1,36	3,95	0,34	0,74	-7,45	10,16
Empati	0,02	0,17	0,13	0,90	-0,36	0,41
Räv	0,65	1,15	0,56	0,59	-1,92	3,21
Kön	-1,24	1,08	-1,15	0,28	-3,64	1,16
Tidigare erfarenhet	0,17	0,38	0,46	0,66	-0,67	1,02

Bilaga 4: ANOVA utskrift

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	13	40	3,076923	4,576923
Column 2	13	50	3,846154	6,141026

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3,846154	1	3,846154	0,717703	0,405267	4,259677
Within Groups	128,6154	24	5,358974			
Total	132,4615	25				

Crime Scene Investigation in Virtual Reality

Linnéa Gustavsson, Fredrik Magnusson, Emil Olsson, Kajsa Sikström
li8553gu-s@student.lu.se, fr2231ma-s@student.lu.se, em5456ol-s@student.lu.se, ka8265si-s@student.lu.se

The intent of this study is to investigate the gaze behavior of crime scene investigators using Virtual Reality (VR) and eye-tracking. The method was tested in a pilot study consisting of three different phases. In the first phase, the participants spent ten minutes inside a virtual crime scene built in Unity wearing a VR-headset with built-in eye-trackers. The second phase included a questionnaire with questions regarding the participants' thoughts of what had happened at the crime scene. Finally, during the third and final phase, the participants watched a recording of themselves while answering questions in an interview. The result of the study showed that if participants spent longer time watching relevant objects, for example, objects that potentially could be tied to the crime, they were more able to conclude what had happened at the crime scene. The performed study is still at an early stage and requires more test subjects in order to be fully representable.

1 Introduction

Visiting a crime scene is vital for understanding the causes leading up to a crime as well as the course of action. Crime scene investigators visit the location where the crime happened to collect evidence, clues and genetic material. An important part of searching a crime scene is to know where to look and differentiate between what is an actual clue from what is just an ordinary object. In order to become better at scanning a crime scene for important evidence, one needs proper training.

Purpose

The goal of this project is to create a training tool for crime scene investigators to perfect their gaze behavior. It is of utter importance that crime scene investigators have good knowledge of how to visually scan a crime scene. By focusing on the gaze behavior at an early stage, newly recruited crime scene investigators will hopefully become even better at collecting evidence and not missing key objects that might be crucial for the investigation. However, in order to become a successful crime scene investigator one needs sufficient practical experience that might not be available during initial training. It would therefore be helpful and easily accessible to construct a virtual environment where the police could train. In this project, VR will be used in order to create a virtual crime scene where the gaze behavior of the users can be studied with the help of eye-tracking technology. The novices can immediately be thrown into a virtual crime scene and act as if it would have been a real-life case. The purpose of the tool is to educate future crime scene investigators to better visually scan a crime scene in order to not miss any vital evidence.

VR and Eye-Tracking

VR is a concept for simulating an environment that places the user in an engaging, new version of reality. Most VR-technology uses stereoscopic lenses that create a feeling of depth for the user. In the VR-environment the user is able to interact with the surroundings, which can include picking up objects and open doors to name a few. The movement (locomotion) can be done by either physical walking, teleportation or the use of controls such as joysticks. By combining many different sensors to track the movement of the user as well as utilizing sound and vibrations, it is possible to encapsulate the user in the virtual world.

Eye-tracking is a technology that is widely used in research today. It is an exceptional way of understanding human cognition, for example how people think and make decisions. Eye-trackers come in different shapes and forms. Most common are specific eye-tracking goggles with regular glass that a user can wear in public while interacting with the real world. However, eye-tracking can also be utilized together with VR. Today's VR-headsets, or Head-Mounted Displays (HMDs) sometimes come with built-in eye-trackers. One example is HTC Vive Pro Eye. With a combination of VR and eye-tracking technology, it is possible to place a user in a virtual environment and also study gaze behavior. This opens up a lot of opportunities for different use cases and scenarios. For example, in a study made in 2019, the sense of direction and remembrance was tested by building an entire city in VR. The test subjects wore a VR-headset with built-in eye-trackers, and the researchers could therefore compile the gaze data to see where each participant was looking at while navigating in the virtual city (Clay et al., 2019). That is just an example of the potential and scale that a combination of VR and eye-tracking can have when conducting tests.

2 Theory and Background

Combining new technology in order to aid forensic investigation has always been a keen interest for the police. Cutting-edge technology in all fields, such as the discovery of DNA profiling in the 1980s (Håkansson, 2009) as well as being able to track movements via phone-masts has helped the police solve cases. The resurrection of interest for VR and the progress made in the field has aroused curiosity in the police, both international as well as in Sweden.

Previous and Related Work

The Swedish police have previously been involved in projects to obtain a teaching device in VR. For example, a master thesis was written in 2017 in collaboration with the Swedish National Forensic Centre (NFC). The study, which included both a literary study and interviews with experts in crime scene investigation resulted in a prototype of a VR-tool in which novices could practice and interact with the environment. Novices as well as experts in the field tried and reviewed the final product and the end result led to the conclusion that VR is a tool that will aid and support learning behavior at a crime scene. The application has however not been in use at the Swedish police since it needed further development (Dath, 2017).

Crime scene investigation and VR have also been in the spotlight for international conferences such as the 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR) (Nelis et al., 2018). A contribution to the conference presented a case where they had scanned a room using a 3D laser scanner and created an environment in VR using a point cloud from the scanned data. This proved that the future for using VR in crime scene investigations is bright and only awaits further development and distribution of the 3D laser scanning technology to be in practice. A 3D laser scanner has the ability to quickly and precisely collect 3D data from real environments. 3D laser scanning has the potential to aid forensic work, where it can be used to create an understanding of the space and depth in the crime scene in a way that is harder to recreate with regular photos. Using 3D laser scanners is not a common practice yet, but Swedish police have started to implement scanning of crime scenes, such as in a case of hit and run in Linköping 2019 (Norbergh, 2019). The reason for it not being more in use yet is, according to J. Olsson (personal communication, September 22, 2020), that there are difficulties to transfer information from the 3D cloud to a picture that is comprehensible to a viewer. But hopefully, in the not too distant future, 3D laser scanning will be an essential part in documenting and storing not only evidence but the full crime scenes that can be re-experienced within the immersive perceptual interface offered by VR.

Furthermore, a study made in September 2020 in the UK confirmed that a virtual environment is an effective application for both teaching and learning crime scene investigation skills (Mayne & Green, 2020).

In light of previously mentioned studies and in order to contribute to the field, the project group wanted to explore the potential of virtual crime scenes as educational platforms. Originating from popular theories and tools in the cognitive science literature, with this project we turned to a new and promising methodological integration to support forensic science, namely the interrogation eye-tracking technology and VR.

Eye-Tracking

Eye-tracking can be described as a system that identifies and measures the movements of the eye. By using it you can, among other things, measure and determine gaze origin and gaze direction. Easily explained, this is done by using a camera in combination with infrared light that acts as a light source that creates a pattern. Cameras then take pictures at high speed of the users' eyes in the perspective of the pattern from the infrared light. Algorithms are then used to calculate the position of the users' gaze (Tobii, 2020).

There is a wide range of metrics that can be derived from eye-tracking such as fixation duration, the number of fixations and time to the first fixation to name a few. These metrics reveal valuable insight into behaviors that are often implicit. One area of cognitive research that is especially interested in these insights endowed by eye-tracking technology is the study of differences in visual search patterns between novices and experts with a focus ranging from surgery to chess playing (Ozger, 2016). Studies have shown that the way forensic investigators differ in their behavior depends on if they are novices or experts in the field (van den Eeden et al., 2019). Experts are spending more time gazing at objects with higher evidential value, henceforth, their visual search approach is more action-based. Meanwhile, novices apply a more feature-based visual search approach, spending more time exploring the scene and every object it includes. This indicates that experts informed by their previous experience of crime scene scenarios better consider what material could be useful for their peers at later stages in the investigation compared to novices (Baber & Butler, 2012). Another eye-tracking study found similar behavioral patterns of experts scanning sequences in a crime scene and goes on to call their strategy "hypothesis-based evidence interpretation" (Watalingam et al., 2017).

It is difficult to find work and empirical evidence for what eye-tracking data about gaze and fixation duration really tells as it can indicate many things. Assuming the observer is on task, longer fixations can, for example, indicate that they are processing what they are looking at in detail, or that they are confused by what they see but you as a researcher can not be certain.

Definitions

A few definitions are recurring in the report and these are described below in order to get an overview and understanding of the concepts. When talking about VR in general, presence is a factor that is often discussed. Presence is described as "... the phenomenon of behaving and feeling as if we are in the virtual world created by computer displays" (Sanchez-Vives & Slater, 2005). Cummings and Bailenson give a slightly different interpretation of the same concept and describe presence as a concept about "being there". As the name implies, it is a concept to evaluate how the user experiences presence in the virtual environment (Cummings & Bailenson, 2016). In other words, presence can be used to investigate how realistic the environment created in VR is, and how good the user experience is.

Another definition continuously used in the paper is affordances. The term, described by Norman, means that the properties of objects can indicate to the users which actions are possible to take, and how the object can be used (Norman, 1999). Affordances is thus a collective name for the possibilities an object or a product has. In this project, affordances refer to what the user can interact with inside the virtual environment.

Research Question

The purpose of this study is to investigate participants' search behavior when looking for objects that are relevant for solving the cause of death at a crime scene. The experiment is mainly designed to answer if gaze preference for objects significant for crime investigation predicts higher certainty about the cause of death.

3 Methods

The virtual crime scene has been created in Unity, a game engine platform for building graphical environments in 3D. Unity is mostly used for developing games and has support for development on many platforms, of which VR is one. The objects used in the crime scene were pre-existing 3D models available for download. The project's prototype is made with and for a HMD from HTC Vive with built-in eye-tracking technology from Tobii. In order to analyze the eye-tracking data and get access to tools, libraries and scripts, Tobii XR SDK was used. For more specific needs, separate scripts were written in the programming language C#.

Design Process

The following section describes how the work with the project has progressed, primarily in the form of which decisions have affected the design process and why, but also what the practical implementations looked like.

The project began with a start-up phase where the primary focus was brainstorming. In the beginning, expectations with the project were also discussed, as well as what ideas and thoughts each group member had. The group was then able to orally discuss different conceptual designs in order to arrive at a basic idea that everyone in the group agreed on. The first idea was to use 3D laser scanning to scan a current crime scene and then use the collected data to create a VR-environment. The idea was also based on investigating how effectively a crime scene could be searched through by having two people searching and looking for evidence in parallel. This was to be investigated by comparing eye-tracking data from the two users, looking for any overlap, suggesting that they both had spent time looking at the same object and thus had spent the time wastefully. The group also agreed on a division of roles and decided that two team members should have the major responsibility for the technical development, whilst the other two would focus more on research and writing the report. Additionally, one team member was appointed as project leader.

Design Iteration I - Initial Environment

In order to move forward in the process and further build on the basic idea, a lot of research was done. In addition to this, the Swedish police were also contacted for interviews in order to gain a better understanding of 3D laser scanning, forensics and how crime scene investigations are conducted. The first digital meeting was held during the second week of the project with the police Philip Engström, who is the responsible coordinator for research and technical development at NFC. During the meeting with Engström, the group got a lot of ideas and input. Among other things, Engström considered that the focus should be placed on training police officers instead of investigating a real crime scene. In addition, he considered that it was more important to spend time investigating how to track eye-movements in a good way than to spend a lot of time on the model itself. He gave the group the proposal to start from a simple apartment built in Unity, instead of using 3D laser scanners. He suggested that there should be signs of a fight, with a body placed in the living room and evidence scattered around in the apartment. Moreover, Engström told us that he would look into it if he was going to be able to provide us with information, both regarding a handbook in forensic science as well as ideas for creating a realistic crime scene. Another meeting was also held with the police a couple of weeks later. This time by phone with police Jesper Olsson who works with documenting crime scenes with the help of 3D laser scanners. Olsson spoke in general about the work with 3D laser scanners and confirmed what Engström earlier had said about difficulties using 3D laser scanners for the project.

As previously mentioned, the group had a fairly clear conceptual model of the project from the beginning. The idea was to scan a crime scene using a 3D laser scanner and then use the data for creating an environment in VR. However, after the meetings with Engström and Olsson, the group received new input and thus the basic idea was further reworked. The idea of 3D laser scanning was dismissed and the focus changed to create an apartment in Unity. Despite the idea of scanning being dropped, the group saw great potential in the project since previous research had proven that an environment in VR does not need to be photo-realistic in order to invoke emotions and a feeling of realism in the virtual place (Cummings & Bailenson, 2016).

Also, by having a self-made environment, it is possible to configure it gradually if something new would come up.

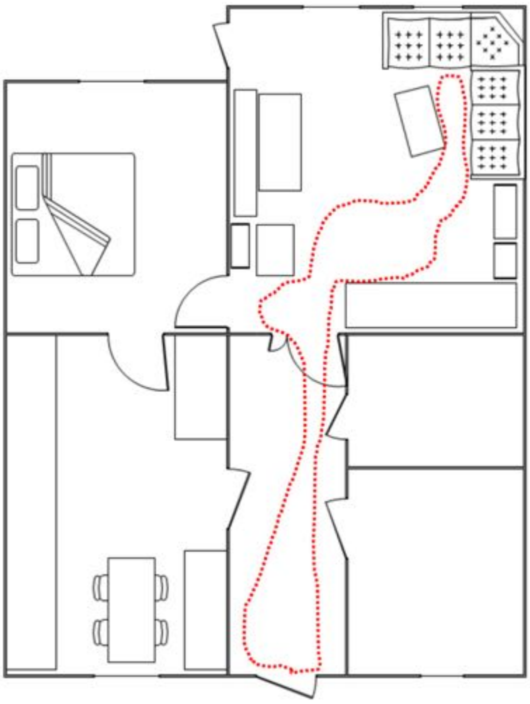


Figure 1. Schematic sketch of the floorplan from the real crime scene.

Thanks to Engström, the group gained access to a protocol from a real crime scene investigation (*Protokoll över brottsplatsundersökning,* 2020). The protocol was used as a basis for how furniture, evidence and random home goods should be placed in the apartment in order to make it realistic. The examined property was a three-roomed apartment with a kitchen and a small room for storage. A schematic sketch of the apartment from the real crime scene was found in the protocol and can be seen in Figure 1. The red dotted lines represent the most relevant area where the perpetrators conducted the crime, which in that case was an attempted murder and robbery.

The process of creating the specified apartment in Unity went easier than expected since the group had access to an environment that resembled the real crime scene. The environment was made in a VR-course held at LTH and was constructed in Unity (see Figure 2). In terms of the floor plan, the apartment was strikingly similar to the real apartment, which helped a lot. Therefore, not much had to be done in terms of the overall foundation. The existing model was a three-room apartment which consisted of one bedroom, one office/fitness room, one toilet as well as a combined kitchen and living room. In the real crime scene, the Area of Interest (AOI), which is represented by the red dotted lines in Figure 1, could be applied to the already existing model. The group, therefore, drew a similar sketch of the AOI in the virtual crime scene in order to compare the two models (see Figure 2). The similarities of the AOIs further made us certain that the already existing apartment was sufficient enough to use as a basis for the crime scene described in the protocol given to us by the police.

Figure 2. Schematic sketch of the floorplan from the virtual crime scene created in Unity.

As the focus of the crime had been in the kitchen and living room, according to the protocol, this is also where all the evidence and miss-evidence was placed. Thus, the other rooms did not contain any relevant objects. The group placed a knife in the kitchen and a hammer in the living room. Furthermore, a body was placed in the living room along with a trace of blood splatter by the sofa, on the living room-floor leading to the hallway and out. In addition, a phone, a pair of glasses, a couple of cans and other small everyday objects were placed around the apartment to make everything messier and more realistic. The idea was to make the apartment reasonably chaotic, to distract the user and see if he or she can discern which evidence was relevant.

All the furniture and pieces of evidence have been downloaded from various websites offering free 3D-models as well as Unity Asset Store. The body itself was made in a program called Autodesk Character Generator. With the program, the group had the opportunity to create a virtual character that could be animated and posed based on desires and needs. The decision to use pre-existing models and objects instead of modelling by ourselves was taken mainly due to lack of time and that the group considered that it was more important to work on the technical pieces than on the model itself. One constraint you can see with this, however, is that the selection is limited when it comes to finding free 3D-models. Thus, it was not easy to find objects that looked exactly like those at the crime scene. Above all, it was also difficult to find blood and footprints that would be perceived as real. On all objects in the apartment, the group has put something that in Unity is called colliders. This means that you can not move through the objects. Some relevant objects were also made interactable, which means that it is possible to pick them up, in order to further invoke a feeling of realism.

A comparison between the crime scene from the protocol and the apartment created in Unity can be seen in Figure 3.

Figure 3. Comparison between the crime scene created in Unity and the real crime scene.

Design Iteration II - Eye-Tracking

With a basic prototype of the crime scene finished in the first iteration, it was now time to work on the eye-tracking. Quite early on, the group saw limitations in examining eye-tracking from two or more people simultaneously on a single computer. This since it is technically difficult and demanding to track eye-movements from two persons on one computer. A study concluded that for two users to search together more efficiently, they had to be able to see where the other one was looking (Niehorster et al., 2019). This is not a possibility in the real world when searching. Hence, this led to that the idea of examining two users at the same time to see if the crime scene could be investigated more effectively was dismissed. Instead, the group decided to focus on the gaze behavior of one person only.

Since parts of the project involved eye-tracking, the next steps in the design process were to make this work for one person. The group started by experimenting with the Tobii XR SDK. After some attempts, the group managed to get the HMD to track where the person was looking at. The idea with the project is to be able to see where the person is looking when they enter the crime scene, which objects they are paying more attention to and what they deem as irrelevant. This is relevant and interesting since you want to be able to search the crime scene as efficiently as possible to find evidence in critical situations. The group discussed how the user's eye-movement should be illustrated and represented in the environment. Both a redpoint cloud of the viewed area as well as a single white circle were discussed as a way to illustrate the current viewpoint. In the end, the latter option was chosen due to it being the easiest manageable alternative for the data collection. Eye-movements are therefore shown with a small white circle, see Figure 4. Initially, the group experienced some issues with getting the eye-tracking to work in the environment. This was mainly due to the lack of previous experience as well as poor documentation. There were some example scenes for Tobii XR SDK in Unity, but they did not resemble the needs and functionality desired by the group. For example, the group wanted to collect different metrics regarding the gaze behavior in real-time, for example when the test user is in the actual environment. How this should be done could not be found on the internet, and therefore the group had to come up with a way themselves, which in the end worked out well.

An important aspect of eye-tracking is that all the users have different head shapes and interpupillary distance. It is, therefore, necessary to calibrate the eye-tracker for each test user so that the gaze trail tracks the eyes accordingly. This was however easier said than done. The documentation about eye-tracking calibration was sparse, and once again the group had to rely on their own. Luckily though, there was a suitable calibration scene already created in Tobii XR SDK. At first, the group tried to move the calibration functionality from that example scene into the crime scene. However, it was soon noticed that the calibration procedure needed space, time and concentration - criteria that could be jeopardized if done at a crime scene. Hence, it was decided that the example calibration scene included in the SDK should be used alone before anyone entering the crime scene. This is not only helpful for calibrating the eye-tracker, but also for letting the users be more familiar with the VR-headset and movement before the experiment is started.

Figure 4. The white circle represents where the user is looking.

Design Iteration III - Data Gathering

Once the eye-tracking worked as desired, the next step was to start saving and getting a representation of the gathered eye-tracking data. How this could be done was once again not found in any documentation or forums. The group, therefore, took the matter into their own hands and wrote a script for gathering data based on where and what a user is looking at. As mentioned earlier, each object in the scene contains a collider. When the gaze vector hits a collider, its object and the corresponding metrics are saved to a text file. At this point, the saved data contained information about which object you looked at, how many times you looked at the object and for how long (also called gaze length and is measured in seconds and milliseconds). A global counter in the script was utilized to keep track of how many times a specific object had been looked at. An example extract of the data can be seen in Figure 5. The data is only collected when starting a session in Unity and is therefore separated between two tests.

Later on, the group wanted to use this data to compare users to explore if it is possible to find a collective strategy for an effective search of a room at a crime scene. It was at this point not yet clear how the compilation and comparison of data would be analyzed.

Object:	Counter:	Gaze length (seconds):
shoerack_hallway	1	0.825
shoe_right_hallway	1	0.35
shoerack_hallway	2	2.349
shoe_right_hallway	2	0.645
shoe_left_hallway	1	0.449
shoerack_hallway	3	2.373
shoerack_hallway	4	3.6
Glasses_kitchen	1	0.11
Glasses_kitchen	2	0.23
phone_kitchen	1	1.614
Glasses_kitchen	3	2.497
Glasses_kitchen	4	3.33
Glasses_kitchen	5	4.617
phone_kitchen	2	1.638
phone_kitchen	3	3.304
blood_2_kitchen	1	0.14
phone_kitchen	4	3.468
phone_kitchen	5	4.273
blood_2_kitchen	2	0.205
coffee-cup2_kitchen	1	0.14
Ashtray_kitchen	1	0.547
Energy_drink_1_kitchen	1	0.289
coffee-cup2_kitchen	2	0.341
Energy_drink_2_kitchen	1	0.271
Energy_drink_1_kitchen	2	0.302
Vase_gym	1	1.864
TV_living_room	1	0.223
blood_under_corpse	1	0.171
RightUpLeg	1	0.31
Hips	1	0.194
Beer_floor_livingroom	1	0.127

Figure 5. Eye-tracking metrics output.

At the end of this design iteration, the group wanted to talk with the Swedish police about the design of the environment so far in order to see if they had any feedback. The meeting was held with Jesper Olsson, who we had been in contact with before, and Johan Ekman, who is an experienced crime scene investigator. They recommended that the corpse should be removed since it would take up too much of the attention. Instead, they wanted us to focus more on the objects in the apartment. In addition, they explained that a crime scene rarely looks like it does on TV and that we should focus on making the VR-environment as realistic as possible. As a result, we chose not to add any sounds such as sirens from police cars or pounding heartbeats to the crime scene. With that said, the group removed the body and added more clues and other miscellaneous objects in the apartment. The group also decided to change the lighting in the apartment in order to make it more realistic. A comparison between old and new lighting can be seen in Figure 3 and Figure 4, respectively.

Design Iteration IV - Environment Improvements

By this point, the group had a simple model of the crime scene, a working eye-tracking functionality as well as a script for collecting data. It was therefore time to create a plan and a design for the upcoming testing sessions. Our plan was to first conduct a pilot study with a small sample of test subjects without any previous experience, for example, fellow students taking the same course. The main purpose would be to test that the eye-tracker works well and that the data is saved and readable. Another reason for having pilot tests was to see if the users would feel enough presence in order to fully immerse themselves in the crime scene. The group also wrote down an informed consent that the test participants would have to sign before starting the experiment as well as a questionnaire that they had to fill in afterwards.

The initial thought was to, later on, test the crime scene on real investigators with a lot of experience together with newly hired investigators, in order to see any differences. The group had an ongoing discussion with the Swedish police about performing tests in the VR-lab located in the basement of Ingvar Kamprad Designcentrum (IKDC). They were interested but could not give any specific dates.

To fully prepare for the testing sessions, the team polished the virtual crime scene by adding additional items and placing them on locations similar to that of the crime scene protocol that was accessible. The group also decided on adding a few more gaze-related metrics such as time to first fixation and distance to the viewed object. The more data the better. Finally, since the test subjects probably were not familiar with movement in VR, the group decided to add a separate scene specifically for training on moving and grabbing objects. The sequence of the different rooms was to first place the test subjects in the movement scene, then the calibration room and finally the crime scene. The reason for performing the calibration after the movement was to make sure that the HMD had not been moved or fiddled with before the actual test was started.

Design Iteration V - Test Preparation

In the final design iteration, things took another turn. Due to the ongoing pandemic, the crime scene investigators that were planned to be tested could no longer participate. This was a big setback since the test procedure was more or less finished. Without any tests, it was impossible to conclude anything about the gaze behavior of experts versus novices. Therefore, the group had to improvise. When consulting the supervisors it was instead decided that the tests would be configured a bit and instead include other students that were more accessible. However, since the students were no experts in crime scene investigation the objective of the test had to be changed.

Instead of simply searching the crime scene for evidence, the test subjects were now to be instructed with the task of identifying if a murder or suicide had been conducted in the crime scene. The focus was instead shifted to finding relevant objects that could guide the test subjects to a conclusion. Some key objects were placed in the apartment. The relevant objects could then be tied to the two different scenarios of either a murder or a suicide. For example, an empty can of pills was located in the bathroom with the purpose of making the test subject believe it was a suicide that had happened (see Figure 6). Other suicide-related objects included: a letter placed on the TV table, another pillbox in the kitchen as well as bottles of alcohol scattered around the apartment.

Figure 6. An empty pillbox located in the bathroom.

Meanwhile, a bloody knife was placed in the kitchen, instead indicating that it was a murder (see Figure 7). Other relevant objects indicating a murder included a blood-stained hammer, blood on the floor in the hallway as well as the living room, a flipped chair and bloody glasses on the hallway floor. Some objects could be tied to both scenarios which were supposed to confuse the test subjects and make it more difficult to draw a conclusion.

All of the key objects were placed in either the kitchen, living room or hallway. The team tried to follow the crime protocol given by the police and in that case, the relevant objects (that were photographed in the protocol) were all located in those three rooms. In hindsight, it might have been better if the relevant objects were more scattered since many of the test subjects did not care to further investigate the rest of the rooms.

Figure 7. A blood-stained knife located in the kitchen.

Participants

As previously mentioned, the original experiment was based on comparing search behavior at a crime scene between experienced and novices in the field. This group, consisting of volunteers from the Swedish police in Malmö had volunteered themselves to partake in the experiment. Unfortunately, due to the Covid-19 situation in the region, they could no longer be available as participants. Due to this situation, both the experiment and the participant-group had to be rethought. After consideration, a decision was made that it would be interesting to use a group of participants that had no prior connotations to the work of the police, neither to crime scene investigation and that these trials could be seen as a pilot study for future experiments.

The participants were chosen via convenience selection and consisted of eight students, mainly from the KOGP10/MAMN15-course. The participants had little to no information about the project and the experiment prior to their sessions.

Procedure

The participants were scheduled to participate in one on one sessions. When arrived, the participants were asked to sign an informed consent before proceeding with the experiment.

The participants were asked to put on the VR-headset and to adjust it to their preferred fit. They were then put into a VR-environment where they were able to get comfortable with using VR, including movement via teleporting and how to interact with objects. The movement scene was made in order to reduce noise in the data later on due to inequality with previous experience with VR. The participants were then moved to another room for calibration for the eye-tracking device.

Before entering the crime scene, the participants were briefed about their task and the background information about the scenario they were to enter. The following information was read to them; *"A man has phoned the emergency hotline after he found his deceased wife in their home. The body has been taken away by the authorities. It is your task to examine the home and look for any objects that can have taken part in the death of the woman. Is this a crime scene where she was murdered or did she commit suicide? You have ten minutes to examine the rooms, interact with objects and draw a conclusion. You can exit the VR-environment whenever you want"*.

The participants were asked to examine the space and to draw a conclusion about the events. They were asked to search the rooms quietly without talking to the experiment leader about their experience while still in the VR-environment. The eye-tracking data and movement in the rooms were recorded on video.

In the next phase of the experiment, the participants were asked to fill in a questionnaire about their experience and what conclusions they had drawn about the proposed crime. The questionnaire can be seen in Appendix A. Afterwards, the participants were to look at the video of their eye-tracking movements and tell the experiment leader about their reasoning behind the conclusions they made. While watching, they were asked questions about what objects had stood out to them, what objects that they had found to support their conclusion and if they had changed their hypothesis about the situation after finding a new object or source of proof. A transcript of the responses from the participants can be found in Appendix B.

Lastly, there was a debriefing where the participants got the opportunity to ask anything about the experiment, get more information and to process the experience they had.

This approach was chosen in order to collect both qualitative and quantitative data. An overview of the test process can be seen in Figure 8.

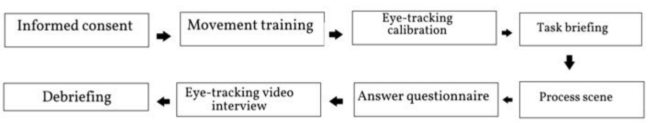


Figure 8. Experimental session scheme.

Variables

As mentioned, the task of the participants in the virtual crime scene was to conclude if a murder or suicide had occurred and look for objects relevant to draw this conclusion. To investigate if there were any differences in gaze behavior between objects relevant or irrelevant to the task, all objects placed in the scene were categorized into either high or low significance for crime investigation. The categorization was made partly on the basis of our intuition since we deliberately chose and planted murder associated-, suicide associated- and distracting objects, together with the participants' intuition as assessed by the post-experiment survey and interviews. Categorization can be seen in Figure 9-11. Optimally in future studies, you would want to have experienced crime scene investigators to inform which objects are relevant for concluding the nature of the crime.

As previously mentioned, in order to assess gaze behavior in the experiment, three basic eye-tracking metrics were used. These included: *counter* (further referred to as *fixations*), *gaze length* (further referred to as *fixation duration*) and *distance to object*. Based upon these metrics additional calculations were made to extract deeper insights brought by the eye-tracking technology. Following variables were created: *Time in VR-crime scene* (maximum 600 seconds); *Total fixations* (summed fixations); *Total fixation duration* (summed fixation durations); *Objects seen* (number of unique objects gazed at); *Duration per fixation* (*Total fixation duration* divided by *Total fixations*); *Fixation duration per object* (*Fixation duration* divided by *Objects seen*); *Distance to object per fixation* (*Total distance to objects* divided by *Total fixations*); *Fixation duration on high relevancy objects (%)* (*Total fixation duration on high relevancy objects* divided by *Total fixation duration*); *Time fixating at objects (%)* (*Total fixation duration divided* by *Time in VR-crime scene*).

In order to understand the participants' experience in the virtual environment, they were asked to fill in a short questionnaire. The questionnaire consisted of four questions of differing sorts. The first question asked the participants if they believed it was a murder or a suicide that had been committed and was to be answered through a Likert-scale of 1-9. The murder response was placed on one end and the suicide response on the other. The second question asked how certain the participants were of their conclusions by asking them to fill in a multiple-choice question-box with five choices between "very certain" to "very uncertain". The following question was a free-text paragraph where the participants were asked to resonate about their choice between the two scenarios. This question did not contain any minimum or maximum limit to the length of the text. Lastly, a question was asked about the amount of presence perceived. This question was answered with a Likert-scale of 1-9 with "very realistic" and "very unrealistic" as opposing extreme answers.

Data Analysis

Data were analyzed using simple linear regressions to evaluate if the number of *fixations*, *fixation duration*, and *distance to object* could be predicted by high versus low task relevance for objects. Furthermore, an additional simple linear regression was conducted to evaluate if the fixation duration ratio between high versus low task relevance for objects could be predicted by confidence in crime conclusion. All analyzes and graphs were made using R version 4.0.

4 Results

A summary of descriptive statistics for all gaze behavior data can be seen in Table 1.

Table 1. Descriptive statistics of gaze behavior.

Variable	n	mean	sd	median	min	max
Time in VR crime scene (s)	8.00	332.20	141.37	302.50	153.00	600.00
Objects seen	8.00	28.50	2.20	29.50	25.00	31.00
Total fixations	8.00	270.40	122.27	249.50	105.00	452.00
Total Fixation duration	8.00	118.80	59.38	114.20	40.90	210.30
Duration per fixation	8.00	0.40	0.12	0.40	0.30	0.60
Fixation duration per object	8.00	4.10	1.85	4.00	1.60	7.00
Distance to object per fixation	8.00	2.20	0.30	2.10	1.80	2.70
Fixation duration on high relevancy objects (%)	8.00	0.60	0.10	0.70	0.50	0.80
Time fixating at objects (%)	8.00	0.30	0.09	0.30	0.20	0.50

An overview of *fixations per object* can be seen in Figure 9. A simple linear regression was used to assess whether objects estimated significance for crime investigation predicted the number of fixations on objects. A significant regression equation was found, $R^2 = .173$, $F(1,29) = 6.08$, $p = .020$. Objects estimated significance for crime investigation significantly predicted the number of fixations on objects, $B = 3.19$, $t = 2.47$, $p = .020$.

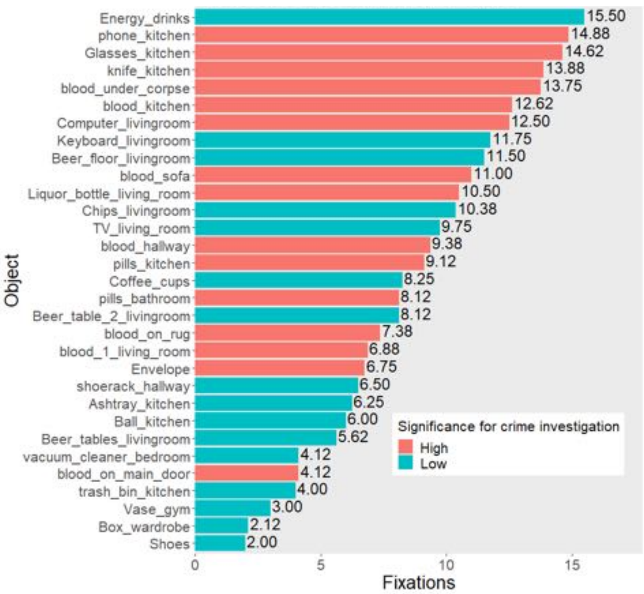


Figure 9. Plot of the average number of fixations per object.

An overview of *fixation durations* per object can be seen in Figure 10. A simple linear regression was used to assess whether objects estimated significance for crime investigation predicted fixation duration for objects. A significant regression equation was found, $R^2 = .314$, $F(1,29) = 13.3$, $p = .001$. Objects estimated significance for crime investigation significantly predicted fixation duration for objects, $B = 2.76$, $t = 3.65$, $p = .001$.

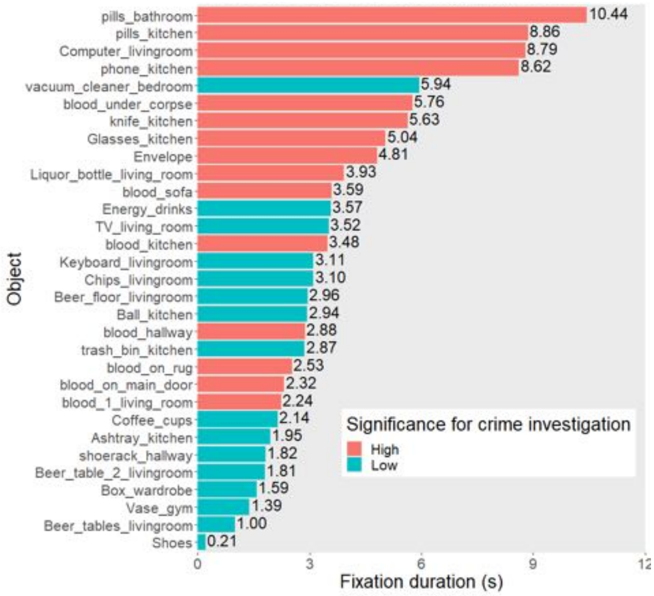


Figure 10. Plot of average fixation duration per object.

An overview of *distance to object* can be seen in Figure 11. A simple linear regression was used to assess whether objects estimated significance for crime investigation predicted distance to object gazed at. No significant regression equation was found, $R^2 = .040$, $F(1,29) = 1.20$, $p = .283$. Objects estimated significance for crime investigation did not significantly predict distance to object gazed at, $B = -.254$, $t = -1.09$, $p = .283$.

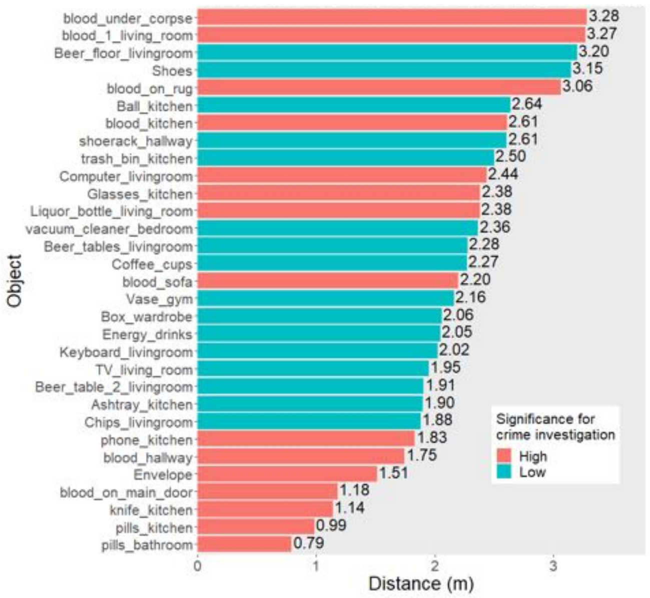


Figure 11. Plot of the average distance to object when gazed at.

A majority of the participants concluded that a murder had occurred, some were unsure, and no one concluded that suicide had occurred. Participants thought that the virtual crime scene was realistic, which indicates a high level of presence. See an overview of post-survey results in Figure 12. A simple linear regression was used to assess whether the ratio of fixation duration for objects with high compared to low estimated significance for crime investigation predicted confidence in crime conclusion. A significant regression equation was found, $R^2 = .703$, $F(1,6) = 14.17$, $p = .009$. The ratio of fixation duration for objects with high compared to low estimated significance for crime investigation significantly predicts confidence in crime conclusion, $B = 10.98$, $t = 3.764$, $p = .009$. In other words, the longer participants looked at objects highly relevant for determining if a murder or suicide had occurred, the more confident were they in their conclusion.

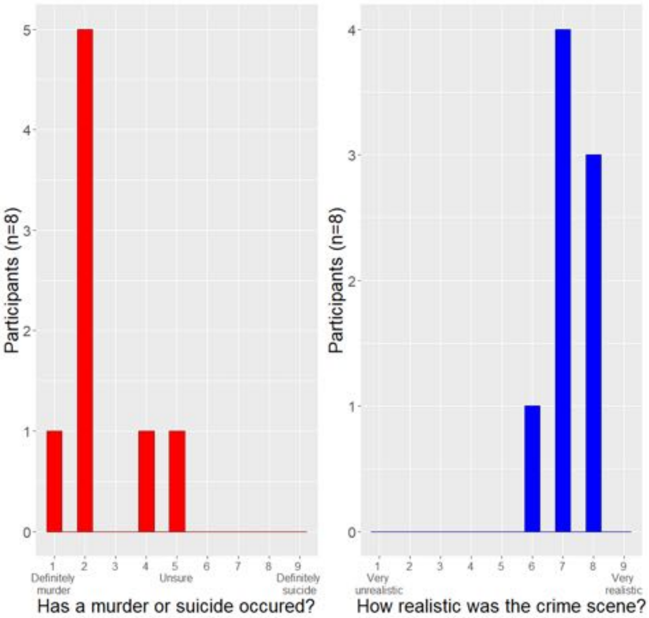


Figure 12. Post experiment questions about the confidence that a murder or suicide had occurred (left) and the virtual crime scenes' level of realism (right).

5 Discussion

Since this was a pilot study with a low number of participants (n=8) the interpretations of the results are as much focused on the reliability and validity of the measures as it is on the actual results. Almost all descriptive measures turned out to give some valuable information about behavior in the virtual crime scene in general and in relation to the crime investigative task. However, in some cases, validity can be questioned. For instance factors such as object size, location including proximity to other objects in the environment can clearly affect the number of fixations and fixation duration as well as the distance to object. For example, objects in the bathroom should generally be at a proximal distance from the gazer since it is required to open a door and enter the small room compared to the objects on the kitchen table which are visible from the living room. Likewise smaller objects might require longer fixation duration to comprehend their details. You might think on the other hand that larger objects should have more fixations and higher fixation duration since they must be randomly fixated at when scanning the scene. This is not necessarily the case because the threshold of 100 ms for counting a fixation compensates for random fixations when scanning the environment. The time limit for scanning the scene before participants had to make their conclusion was ten minutes. The fact that only one participant used the full time and the exit time was close to half of the total available time on average, in conjunction with an overwhelming majority leaning towards murder as their conclusion, one might doubt that some aspects in the scene made the conclusion obvious. Even though we never stated that there was a right answer, perhaps the crucial details that made suicide implausible to the participants was the blood drips spread out over the apartment. But as a professional crime scene investigator you must never omit the possibility of a staged crime scene. For example an open and empty jar of pills might look like a clue for suicide but can also be planted to make it look like a suicide. Correspondingly, glasses on the floor with blood stains might look like a clue for murder but can also be planted to make it look like a murder. Future addition of crime scene investigators as participants would be qualitatively prudent.

The two significant regression analysis of objects estimated significance for crime investigation as a predictor for fixations and fixation duration on objects tells us simply that 1) participant followed the task instructions and looked for objects important to the crime scene; 2) the object choice and categorization in lack of expertise made somewhat sense; 3) that participants had overlapping ideas of which objects are relevant for a crime scene investigation. One hypothesis was that the more relevant an object is the closer you will observe it, however, this was not significant but the previously mentioned problems of object size and location together with the small sample group might have impacted here. An alternative measure that we did not use could be how many times an object was interacted with. Another idea would be to ask participants afterward to associate every object with either a higher likelihood of murder or suicide or irrelevant and then categorize so that objects clue category could be used as a predictor. Conceivably the most interesting result was that the relative fixation duration for objects with high over low significance for crime investigation predicted higher confidence in the conclusion. The importance of what this tells us is twofold: 1) higher attention to the task which was to look for evidence, leads to greater confidence in conclusion which was if a murder or suicide had occurred; 2) one ground for hypothesizing crime scene investigator's who attends to relevant details in a crime scene are more likely to draw correct conclusions about the nature of the crime.

As mentioned in section *Definitions*, an affordance is a factor that indicates what a user can do with an object or a product. Most often, this design principle is seen as something positive, as it clearly indicates to the user how and what he or she can interact with. In this project, however, the concept of affordances has caused some problems due to limitations in the virtual environment, which might have affected the outcomes of the tests. Among other things, it is not possible to interact and open kitchen cabinets or wardrobes in the environment. Instead, it is only possible for the user to interact with objects attached to the crime scene by the project group itself. As a consequence of this, it becomes very clear for the user which objects that the project group thinks are relevant as evidence, which may have affected the results of the tests negatively.

Another thing that has been important to think about during the development of the virtual environment is presence. The whole idea of the project has been to try to make the environment as realistic as possible. However, previous research has shown that photo-realism is not necessary to create a good experience in VR. Instead, it is the consistency that counts in order to experience presence. As a result, the group tried to be as consistent as possible when creating the environment, making sure that the objects make sense together. The question *"How realistic did you feel that the apartment was?"* received a relatively high rating, indicating that the presence was more than acceptable.

As previously mentioned, murder-related and suicide-related objects were placed in the apartment in order to analyze if some objects influenced which thesis the test person chose in the end. One object that was seen to be neutral, and not related to either the murder thesis or suicide thesis was the stove. However, due to the fact that one of the stove plates was not turned off, many of the test persons saw this as very important evidence when trying to figure out what had happened in the apartment. The overall conclusion that the test persons took regarding the stove plate was that most likely, someone would not keep the stove on when committing suicide, and as a consequence of that, there were indications that a murder had taken place.

Even though the stove plate indicated that a murder had been committed, some test persons still were unsure whether it was a suicide or a murder. Many test persons commented that they were uncertain of their thesis due to the fact that they could not read the label on the pillbox.

Limitations

It was discovered after the tests were performed that some relevant objects did not have a collider. This includes a hammer located in the living room and the stove in the kitchen. When watching the recordings it was seen that these two objects drew much attention to the test subjects, and therefore probably would have topped the charts for the most looked-at objects.

Future Improvements

In the future, it would be interesting to study the gaze behavior of real crime scene investigators. However, during the pilot testing suite, it was not possible to include them. It would also be valuable to receive feedback from police officers who work with crime scene investigation if the virtual environment reflects a real crime scene in terms of manoeuvrability and object placement.

Even though the rating regarding presence was relatively high in the tests, there are always improvements that can be made when it comes to functionality. To achieve an even higher degree of presence, preferably more things should be made interactable, such as the kitchen cabinets and the wardrobe. While it would be preferable to make all objects interactable, there are some technical limitations that cannot be overlooked.

Additionally, it would be interesting to use heat maps to visualize the gaze behavior of the test participants. As of now, there is only numerical data which takes time to process when testing many subjects.

6 Conclusions

Applying virtual reality to crime scene investigations holds invaluable potential for the future. When forensic teams get equipped with and educated in laser scanning technology, visual representations of crime scenes no longer need to exist of large assemblages of photographs that are hard to systematically examine and revisit. Instead, full scenes can be saved in which not only crime investigators but suspects, witnesses and juries can enter or reenter. Adding eye-tracking technology to a virtual crime scene allows for an in-depth view into implicit behavior that can generate clues about the crime and improve investigative methodology. On the way to laser scanned scenes, virtual scenes can be created with 3D environment software with the benefit of easy object interaction and manipulation, features exceptionally admissible as an educational platform for forensic teams. We have created such a virtual crime scene inspired by a real-world crime scene and additionally built-in eye-tracking. The platform was tested by students at Lund University which considered the scene to be realistic. Through analysis of the testers' gaze behavior, it was revealed that looking longer at objects highly relevant for the crime investigation predicted higher certainty in answering the question if a murder or suicide had occurred. Summing up this prototype building and pilot testing we would like to end with a hypothesis for future studies: crime scene investigators who attend to relevant details in a crime scene are more likely to draw correct conclusions about the nature of the crime. Lastly and most importantly, the attention of the above type can through the aid of eye-tracking be evaluated and informed by a platform like the one we have built.

7 Acknowledgements

First of all, we would like to thank our supervisors Diederick C Niehorster and Mattias Wallergård. In addition, we would like to thank Philip Engström, Jesper Olsson and Johan Ekman from the Swedish Police for valuable feedback and input.

References

Baber, C., & Butler, M. (2012). Expertise in crime scene examination: comparing search strategies of expert and novice crime scene examiners in simulated crime scenes. *Human Factors*, *54*(3), 413–424.

Clay, V., König, P., & Koenig, S. U. (2019). Eye Tracking in Virtual Reality. *Journal of Eye Movement Research 12(1)*. https://doi.org/10.16910/jemr.12.1.3.

Cummings, J. J., & Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. In *Media Psychology* (Vol. 19, Issue 2, pp. 272–309). https://doi.org/10.1080/15213269.2015.1015740.

Dath, C. (2017). *Crime scenes in Virtual Reality: A user centered study* [diva-portal.org]. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1115566.

Håkansson, C. (2009). *DNA som bevis i brottmål*. http://lup.lub.lu.se/student-papers/record/1558437.

Mayne, R., & Green, H. (2020). Virtual reality for teaching and learning in crime scene investigation. *Science & Justice: Journal of the Forensic Science Society*, *60*(5), 466–472.

Nelis, J., Desmet, S., Wauters, J., Haelterman, R., Borgers, E., & Kun, D. (2018). Virtual Crime Scene. *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 165–168.

Niehorster, D. C., Cornelissen, T., Holmqvist, K., & Hooge, I. (2019). Searching with and against each other: Spatiotemporal coordination of visual search behavior in collaborative and competitive settings. *Attention, Perception & Psychophysics*, *81*(3), 666–683.

Norbergh, E. (2019-05-10). 3D-scanner ska hjälpa polisen i smitningsutredning. *SVT Nyheter Sörmland*. Retrieved December 22, 2020 from https://www.svt.se/nyheter/lokalt/sormland/3d-verktyg-ska-hjalpa-polisen-i-smitningsutredning.

Norman, D. A. (1999). Affordance, conventions, and design. In *Interactions* (Vol. 6, Issue 3, pp. 38–43). https://doi.org/10.1145/301153.301168.

Ozger, M. (2016). *Paying attention to the evidence: a comparison of perception and decision making processes in novice and experienced scene of crime officers using eye tracking in simulated crime scene scenarios* [Phd, University of Lincoln]. http://eprints.lincoln.ac.uk/id/eprint/26649/.

Protokoll över brottsplatsundersökning. (2020). Linköping: Region Öst, Forensiska sektionen.

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews. Neuroscience*, *6*(4), 332–339.

Tobii. (2020). *Världsledare inom eyetracking*. Retrieved December 22, 2020 from https://www.tobii.com/sv/.

van den Eeden, C. A. J., de Poot, C. J., & van Koppen, P. J. (2019). The Forensic Confirmation Bias: A Comparison Between Experts and Novices. *Journal of Forensic Sciences*, *64*(1), 120–126.

Watalingam, R. D., Richetelli, N., Pelz, J. B., & Speir, J. A. (2017). Eye tracking to evaluate evidence recognition in crime scene investigations. *Forensic Science International*, *280*, 64–80.

Appendix A: Enkät

Fråga 1: Tror du att ett mord eller självmord har begåtts? (Sätt svaret på 5 om du är osäker).

Mord 1 2 3 4 5 6 7 8 9 Självmord

Fråga 2: Hur säker är du på din tes?

- Mycket säker
- Säker
- Neutral
- Osäker
- Mycket osäker

Fråga 3: Vad var det som fick dig att dra din slutsats?

Fritextsvar.

Fråga 4: Hur realistisk upplevde du att lägenheten var?

Overklig 1 2 3 4 5 6 7 8 9 Verklig

Appendix B: Transkribering av intervjuer

Testperson 1

Objekt som testpersonen kommenterade att hen lagt märke till

- Blodet
  - På golvet
  - På handtaget på ytterdörren
- Brevet
- Glasögonen
- Hammaren
- Kniven
- Spisplattan
- Stolen

Kommentarer

- Kommenterade att det inte var möjligt att öppna skåpen i köket eller garderoberna. “Det var säkert inget viktigt där eftersom de gick inte att öppna”.
- Kommenterar att hen missade pillerburkarna och upptäckte dem först i videon.

Specifikt objekt som fick dig att känna dig övertygad om din tes?

Att hammaren hade blod på huvudet och inte på skaftet. Detta gjorde att hen trodde att det skett ett mord. Resonerar vidare att man antagligen inte använder en hammare om man vill begå självmord.

Kände du dig säker på din tes från början, om inte, vad var det som fick dig att ändra på din tes?

Hade ingen tes från början och fortfarande osäker. Lutar dock åt mord.

Testperson 2

Objekt som testpersonen kommenterade att hen lagt märke till

- Anteckningsblock
- Blodet
- Brevet
- Datorn
- Glasögonen
- Hammaren
- Kniven
- Mobiltelefonen
- Pillerburk
    - I köket
    - På toaletten
- Spisplattan
- Stolen
- Sängen

Kommentarer

- Kommenterar övergripande att det ser ut som att det har varit någon sorts kamp i lägenheten.
- Tittade på anteckningsblocket för att se om det stod något hemligt eller om någon hade lämnat en lapp.
- Tittar på brevet för att se om det kan vara ett självmordsbrev.
- Studerar hammaren noggrant för hen är osäker på om den har blod eller färg på sig. Jämför blodet på hammaren med blodet på golvet för att se om det är samma, eller om det kan vara färg på hammaren. Avgör att det är en annan färg och avfärdar således hammaren som bevis lite.
- Kommenterar att kniven har blod på sig.
- Tittar noggrant omkring sig vid pillerburken på toaletten. Detta för att se om någon har försökt spola ner pillerna i toaletten.
- Kommenterar att spisplattan är på.
- Kommenterar att hen vill undersöka stolen för att se om det finns någon snara på den som skulle indikera på självmord (hängning). Kommenterar vidare att hen tycker det är oklart varför stolen är vält.
- Undrar om den obäddade sängen kan betyda något.

Specifikt objekt som fick dig att känna dig övertygad om din tes?

Spisplattan i kombination med pillerburkarna. Troligaste hypotesen för självmord. Blod avfärdade självmordsidén.

Kände du dig säker på din tes från början, om inte, vad var det som fick dig att ändra på din tes?

Nej, höll båda teserna öppna under början. Letade efter saker som indikerade på självmord och efter saker som indikerade på mord. Menar att piller är mest typiska för självmord, men att det också finns en massa blod och att dessa två i kombination ej stämmer överens. Således borde det vara mord.

Testperson 3

Objekt som testpersonen kommenterade att hen lagt märke till

- Anteckningsblock
- Blodet
- Kniven
- Mobiltelefonen
- Pillerburk
    - I köket

Kommentarer

- Tycker att personen har rört sig lite väl mycket i lägenheten om det är ett självmord.
- Tycker det ser ut som att det har varit två personer i lägenheten. Testpersonen har svårt att förklara varför, menar att det mest är en känsla som indikerar på att det suttit två personer på två olika ställen.
- Menar på att mobiltelefonen kan indikera på att det är ett mord. Kan vara så att offret försökt ringa någon för att be om hjälp. Kan också vara så att mobilen indikerar på någon som ångrat ett självmord.
- Försökte läsa vad det stod på pillerburken, men var svårt att tyda och kunde därför inte dra en slutsats om vad det var för piller.
- Kommenterar att hen hela tiden lutar åt att det har skett ett mord, men att hen är väldigt osäker.

Specifikt objekt som fick dig att känna dig övertygad om din tes?
Kniven. Om det varit självmord kanske pillren hade räckt.

Kände du dig säker på din tes från början, om inte, vad var det som fick dig att ändra på din tes?
Osäker hela tiden. Lutar åt mord pga. telefonen och spåret av blod mot dörren.

Testperson 4

Objekt som testpersonen kommenterade att hen lagt märke till

- Blodet
    - I vardagsrummet
    - Som ledde ut i hallen
- Brevet
- Burkarna/flaskorna
- Dammsugaren
- Flyttlådorna
- Hammaren
- Mobiltelefonen
- Pillerburk
    - I köket
    - På toaletten
- Spisplattan
- Stolen

Kommentarer

- Tycker det ser ut som om det varit ett slagsmål och att något har hänt. Grundas i att stolen ligger vält.
- Försöker "följa blodet" när hen undersöker sin tes.
- Kommenterar att det står 5$ på brevet och undrar om det kan vara att någon är skyldig pengar och om det då kan vara orsaken till utfallen.
- Kommenterar att det känns som någon har varit iväg eftersom en dammsugare står framme och att det finns tomma flyttlådor.
- Kommenterar att hammaren känns konstig.

Specifikt objekt som fick dig att känna dig övertygad om din tes?
Blodet, de öppna pillerburkarna, spisplattan som var på, stolen som hade vält och att det fanns mycket alkohol.

Kände du dig säker på din tes från början, om inte, vad var det som fick dig att ändra på din tes?
Kände sig ganska säker när hen såg att spisplattan var på. Orealistiskt att den var på om det var ett självmord. Blodet som ledde ut i hallen stärkte tesen om mord.

Testperson 5

Objekt som testpersonen kommenterade att hen lagt märke till

- Blodet
    - På handtaget på ytterdörren
- Burkarna/flaskorna
- Glasögonen
- Hammaren
- Kiven
- Mobiltelefonen
- Pillerburk
    - I köket
    - På toaletten
- Spisplattan
- Stolen

Kommentarer

- Kommenterar att det är allmänt rörigt i lägenheten, t.ex. att stolen ligger ner.
- Känslan som testpersonen upplever i lägenheten är att det ser ut som någon som sitter och jobbar och som ska laga mat.
- Kommenterade att det fanns många flaskor med förmodad alkohol som var oöppnade, och att någon därför med största sannolikhet inte varit full.
- Tänkte att kniven skulle kunna vara mordvapnet, men kommenterar att det vore jättedumt att lämna den i köket. Undrar därför om kvinnan kan ha använt den för att försöka skydda sig.
- Ville kolla extra på mobiltelefonen så att det inte fanns missade samtal.
- Kommenterade att det var oklart vad det var för piller som fanns.
- "Man lämnar inte spisplattan på om man begår självmord".

Specifikt objekt som fick dig att känna dig övertygad om din tes?

Blodet på handtaget.

Kände du dig säker på din tes från början, om inte, vad var det som fick dig att ändra på din tes?

Nej, inte helt säker. Tänkte att det inte kunde vara ett självmord då hen såg att spisplattan var på. Blev dock osäker när hen sedan såg pillerburkarna. Kniven som bara låg i köket gjorde också att det var oklart. Drog sin slutsats när hen såg blodet på dörrhandtaget.

Testperson 6

Objekt som testpersonen kommenterade att hen lagt märke till

- Blodet
    - På golvet
    - På handtaget på ytterdörren
    - I vardagsrummet
    - Som ledde ut i hallen
- Brevet
- Burkarna/flaskorna
- Cigarettpaketet
- Flyttlådorna
- Glasögonen
- Hammaren
- Kniven
    - Båda knivarna i köket
- Mobiltelefonen
- Pillerburk
    - I köket
- Soptunnan
- Spisplattan

Kommentarer

- Undrade om det var ett brev som skulle skickas iväg, men såg inget frimärke.
- Förstod inte om flaskorna var tomma eller inte, samt att det var oklart om det i några utav dem var olivolja eller alkohol.
- Ansåg att hammaren var malplacerad. Kommenterar vid ett senare tillfälle att hen inte tittade ordentligt på hammaren.
- Kollar soptunnan då “skräpet kan säga mycket om en person”. Insåg dock att soptunnan var tom.
- Undrar om det är något som har hänt med tankte på att det står flyttkartonger framme.
- Förstod inte hur många som bodde i lägenheten, ställde frågan “har jag fått höra om de har djur eller barn?”.
- Menar att eftersom det finns blod på handtaget på ytterdörren så är det någon som har gått ut blodig. Testpersonen menar därför att det tyder på att det har varit flera personer i lägenheten.

Specifikt objekt som fick dig att känna dig övertygad om din tes?

Kniven fångade uppmärksamheten mest, men kommenterar att det är mycket frågetecken.

Kände du dig säker på din tes från början, om inte, vad var det som fick dig att ändra på din tes?

Osäker.

Testperson 7

Objekt som testpersonen kommenterade att hen lagt märke till

- Blodet
  - På golvet
  - Blodfläckar på soffan
- Burkarna/flaskorna
- Hammaren
- Kniven
- Mobiltelefonen
- Pillerburk
  - I köket
  - På toaletten
- Tidning
- Spisplattan
- Sängen

Kommentarer

- Kommenterar att hen trodde att det var mord redan när hen såg telefonen på golvet.
- Pillerburken får en att tro att det kan ha varit ett självmord, men kommenterar själv att hen tror att det är någon vanlig medicin. Kommenterar vidare att någon kan ha placerat ut pillerburkarna för att få det att se ut som ett mord, när det egentligen är ett självmord.

Specifikt objekt som fick dig att känna dig övertygad om din tes?

Blodfläckarna, hammaren och telefonen gjorde att hen trodde att det var mord. Att det är stökigt i köket och att stolen har trillat stärker tesen att det skulle varit ett mord enligt testpersonen.

Kände du dig säker på din tes från början, om inte, vad var det som fick dig att ändra på din tes?

Säker redan från början.

Testperson 8

Objekt som testpersonen kommenterade att hen lagt märke till

- Blodet
    - På golvet
    - Som ledde ut i hallen
- Hammaren
- Kniven
- Mobiltelefonen
- Soffan
- Stolen

Kommentarer

- Känsla av att det skett något pga. blodspåren.
- Kollade så att mobiltelefonen inte hade en sprucken baksida och att någon tappat den i farten.
- Oklart om hammaren hade blod på sig.
- Undersökte kuddarna i soffan för att se om det fanns blod på dem och att någon vänt på dem för att dölja bevis.
- Menar att den välta stolen är ett frågetecken, och att hen inte vet om man ska tolka den som att det har varit någon form av våldsamhet eller att personen har suttit där.

Specifikt objekt som fick dig att känna dig övertygad om din tes?
Kniven och blodet. Tankegång hos försöksperson: Känns inte som att personen har skurit sig i köket, för det är inget blod i köket. Blodet är istället i vardagsrummet vilket gör att kniven har hamnat i vardagsrummet i efterhand. Om det hade varit ett självmord så menar testpersonen att det är osannolikt att den skurit sig i soffan och därefter lagt kniven på köksbänken. Även tvärtom verkar osannolikt, dvs. att personen ska ha skurit sig i köket och sedan gått och satt sig i soffan. Slutsatsen är alltså att kniven har flyttats efter att personen har skurits, och att det inte är personen som har blivit skuren som själv har gjort det. Det är alltså en annan person som flyttat på kniven och därför är det ett självmord.

Kände du dig säker på din tes från början, om inte, vad var det som fick dig att ändra på din tes?
Höll sig till tesen om mord hela tiden.

Cell Phone distractions in a virtual classroom

Helena Sandquist[1], August Lindberg[2], Alexander Sölvestrand[3], Axel Ekström[4]

dic15hsa@student.lu.se [2]ppe14ali@student.lu.se [3]al8273so-s@student.lu.se [4]ax6217ek-s@student.lu.se

This report presents a project work with the aim of constructing a virtual reality (VR) experiment on the topic of attentional disruption in a virtual classroom. It follows the development of said experiment, from initial theoretical and practical considerations, to their implementations in the VR environment, emphasising on alternatives considered along the way. By providing participants with a dummy phone, we aim to investigate how the presence of a cell phone-type stimulus may be disruptive to attention in a virtual classroom environment. The goal of the present project was for a usable VR experiment, to be executed in the future.

1 Introduction

Background

In the courses MAMN10 and MAMN15, our group was tasked with developing a method of investigating stimuli distractive to attention in a virtual classroom setting. The project was a continuation and reconstruction of a prior pilot study with the similar purpose of examining distractions in a virtual reality (VR) classroom environment. This previous study included a virtual teaching agent, giving a presentation about biodiversity and ecosystems. Pre-study participants were to attend the lecture, which took place inside a virtual three-dimensional classroom. Afterwards, outside of the virtual environment, participants were quizzed on the material. During the lecture, participants were subjected to distractions in the form of a classmate at the bench next to the user scrolling through a Facebook-like news feed on a visual tablet device. In the present project we were also to investigate distractions in a VR classroom, but additionally tasked with reconstructing the study and its in-VR implementations to examine the topic from a different set of presuppositions.

Purpose

The purpose of this report is to detail the work on the project from initial idea to implementation. The end goal for the present project as a whole was a usable experiment, with ready pilot data and analyses. An additional goal was to uncover potential problems (and solutions to those problems) associated with constructing such an experiment.

Structure

Included in this report is relevant prior research on the topics of (1) attention disruption via cell phone and cell phone use, and (2) experiments conducted in VR learning environments, as well some relevant considerations on its implementation and realism. Multiple deliberations made in the course of the project's development are also described and discussed. The implemented prototype - a virtual classroom, with integrated in-VR quizzing and logging of multiple-choice answers - is described. Finally, the planned experimental procedure is discussed, and possible future directions considered.

2 Theoretical background

Phone usage as attentional disruption

In recent decades, everyday cell phone use has increased significantly. The percentage of Americans who owned a cell phone in 2018 was 96%, up from 62% in 2004 (PEW Research Center, 2019). Cell phone ownership has also apparently increased in younger age groups. According to the Common Sense Census on teen media use (Rideout & Robb, 2019) 69% of children aged 12 owned a smartphone. However, research has previously found that interacting with cell phones while concurrently engaged in other tasks may disrupt attention (e.g., Stothart, Mitchum, & Yehnert, 2015; Glass & Kang, 2019). This presents the possibility that cell phone presence and use in classroom settings may be distracting to students' learning. Indeed, previous research results suggest this. In an observational study on university students, greater cell phone use was found to be significantly and negatively related to academic achievement, defined as students' grade point average (Lepp, Barkley, & Karpinski, 2015). Also working with college students Uğur and Tuğba (2015) have also found support for a similar interpretation (but see also Dos, 2014).

Experimental setups have further indicated the distractive quality of sending instant text messages during lectures (Bowman, Levine, Waite, and Gendron, 2009; Fox, Rosen, and Crawford, 2009; Ellis, Daniels, and Jauregui, 2010; Mendoza, Pody, Lee, Kim, & McDonough, 2018) and cell-phone ringing in the classroom (e.g., End, Worthman, Mathews, & Wetterau, 2010). Additionally, Froese et al.,

(2012) found that students performed ~30% worse on a pop quiz after texting during a lecture, compared to controls. However, available research on attentional disruption by cell phone has largely been conducted using college-age subjects, neglecting younger students. But research on attention suggests that the attentional faculties of adults are generally more developed than those of young adults and especially children (see Plude, Enns, & Brodeur, 1994). This leaves the field of educational psychology with an important literature gap for understanding the implications and consequences of cell-phone presence and use in the classroom, and for learning in school and high-school age students.

The VR classroom

Modern technology allows for realistic simulation of classroom environments, enabling up-close scientific scrutiny. One such technology is virtual reality (VR). VR technology has previously been used successfully for training purposes, such as flight simulators (for an overview, see Blascovich & Bailenson, 2012). Since it has gained traction in behavioral science literature, several researchers have argued for its use in social psychology (see Blascovich et al., 2002; Pan and Hamilton, 2018). The true potential of VR in behavioral contexts lies in its high-end illusory capabilities. That is, users are never *convinced* that what they are seeing is true or real; yet, the system may still provide a strong enough level of presence, that they experience the illusion of being there (for a discussion, see Slater, 2018).

According to Slater (2009), participants tend to respond realistically even when the test environment is replaced to a VR environment. The reason for realistic behaviour is stated to be a combination of place illusion (PI) and plausibility (Psi). PI is often referred to as presence. PI is a sensation of actually being in a real place and is constrained by the sensorimotor contingencies of the VR system. Psi is about what is actually happening and how the participant reacts to it like if it was for real. When both PI and Psi is achieved the participant still knows that they are not actually "there" for real but the illusion is enough for them to act and respond like if they were.

While classroom applications of VR have long been discussed in the literature (e.g., Pantelidis, 1993), it has only begun to be implemented quite recently - but with promising results for education outcomes (e.g., Liou & Chang 2018). Additionally, available literature supports studying attention disruption in a VR classroom setting. For example, in one study by Adams et al. (2009), students with ADHD were shown to be more affected by distractions in a virtual reality environment than controls (see also Rizzo et al., 2000).

The goal of the present project was to investigate the distractiveness of cell phone use in a VR classroom environment. We aimed to develop a prototype of an education environment to this end, including integrated learning setting, testing, and logging of educational outcomes. The stated goal of this virtual environment was that it be used for further education-setting testing and experiments.

3 Methods

VR - prototype

The aim of the project was to further develop the given VR-prototype to a new version that suited the planned test scenario. The final product was intended to be used in the pilot study. The method for creating the prototype can be divided into three phases: a planning phase, the design phase and the implementation phase. For all phases, work was iterative.

Planning

Early planning was focused on adapting the previous pilot study to our needs. The present project was initially based on a previous pilot study conducted by the Educational Technology Group (ETG) at Lund University. In this earlier study, they designed a VR-classroom and conducted an experiment to investigate how distractions in a (virtual) classroom environment affect the learning performance of high-school students. The participants listened to a ~20 min lecture and then had to answer a questionnaire; these test questions were developed in conjunction with a high-school teacher. The questions were designed to test the general knowledge that the participant gleaned from the lecture. For a distraction stimulus, an animated avatar of another student close to the participant (inside the VR-environment) was shown scrolling through a social network page. However, the result from the pilot study showed no effect of the manipulation on learning outcomes.

The file size of this project was too extensive to be easily maneuverable. As we were also to use a common university computer for developing the project, we initially needed to remove non-essential components. The project was also constructed using an outdated version of Unity (Unity.org). These things considered, an early significant decision was made to start work with a new Unity project, rather than simply adapting the one provided. At this point, the group also worked closely together and discussed the expectations and goals of the project continuously. Tools used to simplify the communication within the group and to structure the project were Trello (Trello.com), Google Drive and the Facebook Messenger application.

Design

Having selected what was to be kept or discarded from the previous project, the group moved on to designing a new test scenario in the VR-environment. We used brainstorming, bodystorming and cognitive walkthrough. Next, the question arose if it was possible to create a stimuli that was distractive enough to affect learning in a VR-environment. We therefore designed the current project with the goal of identifying a candidate stimulus to this end.

Distractive stimuli

In discussing potential distracting stimuli, we agreed that there were likely some factors of the VR-environment that had diluted or mitigated the distraction in the original pilot study, such that the overall effect was null. We agreed on one of two possibilities: Either the VR-environment itself is so novel or distractive, that all participants are disturbed regardless of other distractive factors (a ceiling effect of the stimulus environment). Alternatively, it could be that the distracting stimuli competes with other factors in the VR-environment, which renders it inconsequential (a distraction floor effect). One hypothesis was that the stimuli used in the pilot study (the scrolling page of the social network), was unsuccessful due to its visual nature. Since VR is an inherently visual medium in itself, it could be that the intended effect of the stimulus was cancelled out in a thoroughly novel VR-classroom. Instead, we agreed to investigate sound-based stimuli with high social relevance. Social networks have significant relevance for the person using it. Therefore, a different, also highly salient and personally relevant stimuli was selected: cell phones.

We judged that a vibration similar to that of a cell phone receiving a text message should not compete with the visual experience of the VR-classroom, while simultaneously being a personally relevant stimulus to the participant. Indeed, several studies show detrimental effects on attention and study results by the usage of mobile phones (see Section 2.1. above). The general goal in implementing this novel stimulus, then, to investigate if it was able to create a noticeably distractive effect in the VR-classroom. If our stimulus were successful, it could show novel potential for future research using virtual classroom environments, in a controlled VR-experiment setting.

For the study proper, a 3D Classroom was built in the Unity engine (Unity.com). The new project was based on the latest version of Unity, version 2019.4 LTS. The project was also built to have support for SteamVR. The SteamVR was used to connect the HMD to Unity, making switching between different hardware simple. SteamVR also provides significant methodological benefits, such as the ability to simulate vibrotactile ("feeling sounds") feedback through the hand controls. Having settled on vibrotactile feedback (as simulated cell phone signals) for a distractive stimuli, we went on to investigate the material that we had taken over from the previous pilot study.

Implementation

Equipment for the present project was provided by the VR lab at Lund University Faculty of Engineering (LTH) where the chosen head-mounted display (HMD) was the HTC Vive, which both provides essential VR features, and a high-resolution display. It also provided the ability to provide sounds through the provided headphones.

First in the implementation procedure, the previous project was evaluated and tested; a new, empty project was then created, also in the Unity game engine. The old project was scaled down and the most important parts were copied to the new one, including a set of assets including but not limited to a classroom environment with a blackboard; student desks; a female teacher agent model; and four students - were deemed acceptable for present purposes (see Figure 1).

Initially, The VR equipment selected for the present study was the HTC Vive, which allowed for haptic feedback. Later in the development phase, the equipment was changed to a Valve Index (valvesoftware.com, see Figure 2). Both options were largely functional for our purposes, but Index controls made it possible to receive more feedback from finger movements, which, we argued, could be a distraction for the participants if lacking. Furthermore, the vibrotactile feedback from the HTC Vive controllers was deemed more similar to those of a cell phone than the controllers of the Valve Index.

Figure 1. The virtual reality classroom environment

Figure 2. The Valve Index HMD virtual reality headset (middle), cameras (mid-left, mid-right), and controllers (far left, far right)

Therefore, the HTC Vive was used in subsequent development and implementation.

Apparatus

The Valve Index provides a 1440x1600 LCD panel display for each eye, for a combined resolution of 2880x1600, with a refresh rate of 120 Hz and a 130 degree field of view. which is higher than the resolution provided by the HTC Vive setup (1080x1200 for each eye, for a combinated 2160x1200, with a refresh rate of 90 Hz and a 110 degree field of view). While the visual specifications of the Valve Index outmatch the HTC Vive, we deem the vibrotactile feedback of the controller as more important for the study's purposes. However, at this stage in the process we are using the Valve Index for simplicity's sake.

Pre-study

The stimulus chosen to evaluate attention in the previous study was a lecture on biodiversity (~19 min). For a question to be relevant to present our project participants, its answer had to be specific enough that participants could know the answer only by paying attention to the lecture. It was necessary to select against questions the answers to which participants knew beforehand. To control for university students' average knowledge of biodiversity, an online pre-study was conducted. Participants were asked to provide answers to 19 questions on the topic. Questions were derived from a set of lecture-relevant questions from prior pilot studies, and a questionnaire containing all potential questions sent out to university students (*n* = 20). For selection, any question where . less than .50% of . respondents answered . correctly was deemed acceptable. Results yielded a total of 4 questions that met the criteria. The pre-study was constructed using Google Forms.

Based on the result from the pre-study, the previous material was modified. The lecture was first divided into two shortened segments, ~5 min and ~3 min respectively. The segments were divided on the basis that they contained relevant information regarding a specific subject (so called "E-reasons" and ecosystem services). Based on the acceptable questions from the pre-study, 7 new questions were formulated, each of which tested for specific knowledge that was contained in a specific part of each segment of the lecture. All questions were designed as multiple choice.

Technical limitations

From the assignment of the project and moving forward, the group suffered many technical difficulties. For example, the size of the initially assigned project was ~18GB, likely because it included not only model assets, but comprehensive model animations also. These were originally intended to increase the immersion in the VR environment. However, many included files were deemed non-essential to our primary purpose. One example of non-essential files were scripts for the logging of eye tracking data (other VR headsets allow for integrated eye tracking, which had been used in previous pilot studies). These were removed to keep the project file size minimal, and also to keep the project manageable within the appropriate given time frame.

Further technical limitations arose around the question of data logging in VR. We preferred data to be logged directly in the VR-environment, to increase accuracy of response while also avoiding the risk of the subject losing immersion. One possibility was to let the subject answer questions vocally while the experimenter marked down the answer. However, this would require more work as (1) notes/recordings would have to be post-experimentally synchronized with the material and (2) the experimenter would have to separately monitor the subjects progress. We deemed that both of these factors would increase the risk of misinterpretation of the subject's answer. Finally, this option would also limit what types of questions could be asked.

We also discussed the possibility to let participants provide free-text answers. While not impossible, and significantly less error-prone than other solutions (e.g., speech-to-text), our experience was that keyboards in VR-environments felt uncomfortable and reduced immersion. Thus, we settled on using multiple-choice questions. Implementation was fairly natural and simple, and all data could be logged directly within the VR-environment.

Experimental design

In the earlier design phase, the group had decided on an experimental design for the project. The main issue was how to manipulate distractions in the experiment, to control for ordering effects. The group decided to synchronize distractions (i.e. cell phone notification signal vibrations) with the presentation stimulus, such that distractions would be presented at the same time as information that would be required to answer the later quiz questions correctly. However, there remained the issue of when distractions should be presented. Finally, the group settled on presenting distractions selectively in one of two parts of the presentation. For the first condition, distractions would occur in the first part; for the second, distractions would occur in the second part. This allowed us to control for what information was (potentially) to be cancelled out by distractions.

Covid-19 safety measures

The work described in this report took place during the Covid-19 global pandemic. Thus, it was critical that we also took into consideration measures to ensure the safety and health of both participants and experimenters during testing. The experiment took place in the Virtual Reality laboratory at the Ingvar Kamprad Center for Design, Lund University. The lab houses not only houses the equipment intended for use, but is spacious, ensuring that participants and experimenters can keep safe distance at all times during experimentation. At any one session, only one experimenter and one participant was present. Finally, the used HMD will be thoroughly cleaned between participants, using dry microfiber cloth.

Furthermore, halfway through the project the Swedish government issued stricter regulations regarding physical contact and presence. As a result, our group decided to shift all activity to digital means. All subsequent meetings were conducted online, and all work was done remotely when possible. As a result, the VR-environment was developed mostly without access to an HMD, as no group member had access to one at home. Thus, when doing the final implementations before testing only one group member with the shortest commuting distance to the VR-lab was actually present, while the rest of the group joined remotely.

4 Prototype

The prototype derived from a scaled-down version of the previous project, constructed in Unity. Animations, scripts and other sizable files were removed to reduce the size of the project from ~18GB to ~2GB. This helped make the work more accessible, both to collaboration among ourselves, and for developing new functionalities in the project environment itself. After reducing the size of the project files, updating the project files to the latest version of Unity was prioritized. This meant updating not only project files, but the VR integrated compatibility tools were also updated to the latest version of SteamVR. This tool makes it possible to experience VR on any hardware, making the project's development more flexible, so as to enable implementation of the resulting prototype on other systems.

The first functionality to be implemented was the triggering of vibrations to one of the controllers of the VR hardware. Early in the project, the controller was set to vibrate when a user pressed the "V" key on the computer keyboard. In subsequent versions, vibrations occurred at specific time markers linked to in-VR presentation (see above). Next, the capability to process user responses was implemented. In early iterations, five buttons, denoted by the letters A-E. These buttons were placed in-VR on the desk in front of the user point of view, and set to appear anew upon each new question. Subsequently, to control for the risk of double-clicking or mistakenly selecting a non-intended alternative, an additional "OK" button was implemented. To move on from one question to the next, users had to first select an alternative, then press "OK", at which point, the answer was logged. A logging function was later implemented to save the submitted answer key, including a timestamp signifying when the response was submitted. On the conceptual level, this allowed for accessible evaluation of participant performance. Critically, the percentage of *correct* answers should decrease where disruption has been introduced. In addition, the condition assigned to the participant was also logged. Finally, a set of questions were included in the debriefing procedure (see Appendix 1). These were designed to investigate the levels of immersion experienced in the virtual classroom environment (see Slater, 2018).

5 Test

The following details the procedure for project evaluation and future data collection.

Procedure

The participant was welcomed by the experimenter. They were informed that their data would be anonymized, and asked to sign an informed consent document. When signed, they were asked to sit at the experiment station, where the HMD and controllers were located. They were instructed that during the experiment, they will have a Vive Pro controller strapped to their outer thigh, and that the controller might vibrate during the experiment.

Preparation

Prior to the experiment, participants (n = 5) were asked to empty their pockets, so as to reveal in which pocket they commonly kept their cell phone. The VR controller was then strapped to the leg on the same side. After the controller was securely fastened, the participant was instructed to sit in a chair at the prepared area and put on the HMD. They were given the possibility to ask questions before putting on the headset, and helped to adjust it by the experimenter if necessary. Before the experiment began, participants were instructed to hold any questions that might arise during the VR-lecture either until the middle questionnaire or until the end, since the lecture cannot be easily paused. Participants were then randomized to one of two conditions, and the program procedure initiated. During the course of the experiment section in VR, the experimenter sat some distance away from the participant, observing the participant's behavior for evaluation purposes.

VR technology, while increasingly popular, is still largely a novel experience to many users. Indeed, most potential participants should be assumed to have little to no experience with the technology at all, expected to lead to a sense of novelty, and exploratory behaviors. Thus, prior to starting the lecture, the participant was asked to answer some basic questions about themselves (age, education, previous experience with VR, etc.), with the purpose of introducing participants to the VR-environment and its associated controls, as well as provide an opportunity to adjust the HMD for optimal fit. The preparation questions were thus designed to address the novelty of the VR simulation. The pre-experiment procedure was identical for both conditions.

Experiment

After preparations, the presentation was initiated. The experiment was divided into two sections of a presentation, with a quiz on its content following each of the two sections. The two sections lasted some ~5 minutes and ~3 minutes, respectively. After each section, participants were asked questions about the content of the lecture. Participants were able to complete data selection within the VR environment itself, by pressing the model button corresponding to the question-appropriate answer (see Figure 3). Participants were randomly assigned to one of two conditions, which decided for which section (first or second) vibrations would be implemented. Vibrations were timed to interfere with information needed to answer the questions at the end of the section. As there is a significant amount of sound leakage when using the provided headphones, such that outside noise is not cancelled out, experimental conditions were controlled, and the lab was kept silent during experiments.

After the conclusion of the experiment, participants were instructed to remove the VR equipment, exiting the virtual environment. They were then interviewed about their experience in the environment, in a semi-structured interview (see Appendix 1). Question 4 in the interview (see Appendix 1) was collected and translated from the Slater-Usoh-Steed

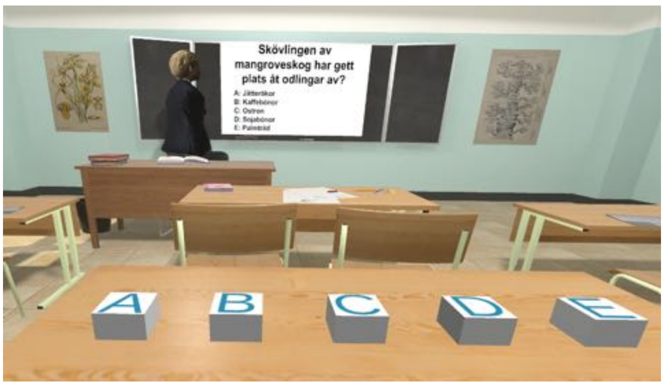


Figure 3. The prototype set-up for collecting participants' responses.

(SUS) presence questionnaire (Usoh, Catena, Arman & Slater, 2000). The purpose of including the question was to determine the amount of presence that the participant felt in the environment. The participant was then able to ask any questions about the setup of the experimenter, and summarily debriefed on the true nature and purpose of the experiment. It is worth mentioning that the experiment was not conducted without problems. In testing we discovered an issue with controller vibrations either not activating or activating for a shorter period of time than planned. Occasionally, these issues were sometimes present during the experiment itself. The on-site experimenter did not possess sufficient programming skills to solve this issue himself, but received assistance digitally from more proficient group members.

Results

Quiz data showed no significant difference in response accuracy between the vibration and control condition. This was expected, as the present sample size was very small. Nevertheless, qualitative data collected from observations of participant behavior in the VR-environment as well as from the semi-structured interview reveal some interesting findings.

Interview questions focused heavily on the experience of the participant's experience of the VR environment. Responses varied between 3 to 7 (on a 7-point Likert scale), indicating that the environment was not able to reliably induce a feeling of presence across participants. Interestingly, some responses were highly similar across all participants. Generally, all participants reasoned that their feeling of presence would increase if it was possible to interact physically with the environment. One example was the possibility to take notes, which is natural activity during presentation settings. This possibility was mentioned by three of the five participants. Another possibility mentioned in the interviews was the ability to rest one's hands on the in-VR table, as well as being able to see the rest of the body (presently, only the hands are visible).

Participants also requested more interactivity. One participant wished it was possible to interrupt the lecturer in order to ask questions. Another participant would have liked the whiteboard to be interactive, in the sense that the teacher agent writes on it live while speaking. Technical limitations of the setups were also mentioned. Three of the participants mentioned graphics as a factor limiting of the feeling of presence. Two of the participants also mentioned the lack of other people in the classroom as a factor that might have decreased their feeling of presence.

Finally, participants also commented on the vibration as a distracting stimulus and as a stand-in for phone notification. Only one participant responded that they thought the vibration was a phone notification. The other participants reported that they were surprised by the first vibration but ignored it for subsequent vibrations, or that they were confused about the purpose of the vibration. Two of the participants reported that the vibrations caused them to be more attentive. After debriefing one participant claimed he had not thought of the vibration as a phone notification since the controller had been strapped to the one leg, whereas he usually kept his phone in the other.

6 Discussion

Motivating testing in the virtual environment.

The pilot from which the present project was derived had participants test their knowledge after the conclusion of the lecture, and outside of the virtual environment itself. For our project, we sought to integrate testing inside the virtual classroom environment itself. The rationale was two-fold. First, it allowed for testing at multiple points during the lecture presentation. Second, it also - optimally - allows us as experiment designers to control for any effect of breaking the immersion inside the virtual environment.

Motivating a non-eye tracking paradigm

The original pilot project, from which we derived our files, used eye tracking measures to examine attention. In the work described in the present report, we elected not to make use of such measurements. This decision was not made based on considerations with regards to eye tracking data, but rather the nature of our chosen distractive stimulus, the cell phone. While cell phones also elicit visual attention, their visual qualities are not (primarily) what attracts a user's attention to the device. This stimulus is either audible (the signal elicited e.g., when receiving a text message) or tactile (the vibrations associated with the same). As such, eye tracking data was deemed unnecessary for investigating our chosen stimulus.

Moving forward

The stated end goal for the described project is a realistic classroom-environment test of attention disruption by the presence and use of cell phones. In this environment, we aim to integrate realistic learning opportunities, cell phone signaling, and testing. In this section, we provide suggestions for future work on this topic.

For a realistic cell phone sound stimulus, A 3D model of the target cell phone might also be considered. Future research on this topic should also seek to investigate the independent effects of either text audio signals, its associated vibrations, or combinations of the two. By implementing and selectively triggering an audio cue, such a relationship could easily be tested using the present project.

For the tactile vibration stimulus, possibilities outside the above described controller-vibrations should be considered. One such possibility is the use of an Android application, allowing experimenters to arbitrarily trigger cell phone vibrations of participants' own cell phones, for a significantly more realistic stimulus. So as to make the VR environment more realistic, future experimenters may also seek to implement realistic background noise (e.g., the sound of cars driving by on an outside road; student chatter in the corridor outside the classroom). This is potentially of high significance, as there is no way to say with certainty that the lack of immersion described by some participants resulted from the vibrations in the VR controller, or the virtual environment itself. Nonetheless, more closely simulating a real-life classroom environment would likely increase participants' sense of immersion in the virtual classroom.

It is also worth noting that many cognitive processes are involved in attention, not all of which are conscious - and none of which were controlled for in our pilot experiments. Implementing additional measures, such as simultaneous recording of electrocardiograms or skin conductance, as measures of bodily stress, may help paint a broader picture of the processes involved in classroom cell phone distractions. Finally, the tendency to be distracted is mediated by personality factors, including autism spectrum disorder (ASD, e.g., Adams & Jarrold, 2012) and ADHD (see Adams, Finn, Moes, Flannery, & Rizzo, 2009), as well as more general freedom from distractibility, as measured on the Wechsler Intelligence Scale for Children (WISC; e.g., see Dickerson Mayes, Calhoun, & Crowell, 1998), neither of which have been considered in the present project. In an experiment proper, experimenters should seek to control for such variables of personality on cell phone distractions in classroom settings. In the present work, we have provided the groundwork for such an experiment.

7 References

Adams, R., Finn, P., Moes, E., Flannery, K., & Rizzo, A. S. (2009). Distractibility in attention/deficit/hyperactivity disorder (ADHD): The virtual reality classroom. *Child Neuropsychology*, *15*(2), 120-135.

Adams, N. C., & Jarrold, C. (2012). Inhibition in autism: Children with autism have difficulty inhibiting irrelevant distractors but not prepotent responses. *Journal of autism and developmental disorders*, *42*(6), 1052-1063.

Blascovich, J., & Bailenson, J. (2012). *Infinite reality: The hidden blueprint of our virtual lives.* New york: William Morrow.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological inquiry*, *13*(2), 103-124.

Bowman, L. L., Levine, L. E., Waite, B. M., & Gendron, M. (2010). Can students really multitask? An experimental study of instant messaging while reading. *Computers & Education*, *54*(4), 927-931.

Dickerson Mayes, S., Calhoun, S. L., & Crowell, E. W. (1998). WISC-III freedom from distractibility as a measure of attention in children with and without attention deficit hyperactivity disorder. *Journal of Attention Disorders*, *2*(4), 217-227.

Dos, B. (2014). The Relationship between Mobile Phone Use, Metacognitive Awareness and Academic Achievement. European Journal of Educational Research, 3(4), 192-200.

Ellis, Y., Daniels, B., & Jauregui, A. (2010). The effect of multitasking on the grade performance of business students. *Research in Higher Education Journal*, *8*(1), 1-10.

End, C. M., Worthman, S., Mathews, M. B., & Wetterau, K. (2009). Costly cell phones: The impact of cell phone rings on academic performance. *Teaching of Psychology*, *37*(1), 55-57.

Fox, A. B., Rosen, J., & Crawford, M. (2009). Distractions, distractions: does instant messaging affect college students' performance on a concurrent reading comprehension task?. *CyberPsychology & Behavior*, *12*(1), 51-53.

Froese, A. D., Carpenter, C. N., Inman, D. A., Schooley, J. R., Barnes, R. B., Brecht, P. W., & Chacon, J. D. (2012). Effects of classroom cell phone use on expected and actual learning. *College Student Journal*, *46*(2), 323-332.

Glass, A. L., & Kang, M. (2019). Dividing attention in the classroom reduces exam performance. *Educational Psychology*, *39*(3), 395-408.

Lepp, A., Barkley, J. E., & Karpinski, A. C. (2015). The relationship between cell phone use and academic performance in a sample of US college students. *Sage Open*, *5*(1), 2158244015573169.

Liou, W. K., & Chang, C. Y. (2018, February). Virtual reality classroom applied to science education. In *2018 23rd International Scientific-Professional Conference on Information Technology (IT)* (pp. 1-4). IEEE.

Mendoza, J. S., Pody, B. C., Lee, S., Kim, M., & McDonough, I. M. (2018). The effect of cellphones on attention and learning: The influences of time, distraction, and nomophobia. *Computers in Human Behavior*, *86*, 52-60.

Pan, X., & Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395-417.

Pantelidis, V. S. (1993). Virtual reality in the classroom. *Educational Technology*, *33*(4), 23-27.

Pew Research Center. (2019, June 12). Demographics of Mobile Device Ownership and Adoption in the United States [Web page]. Retrieved from https://www.pewresearch.org/internet/fact-sheet/mobile/

Plude, D. J., Enns, J. T., Brodeur, D. (1994). The development of selective attention: A life-span overview. Acta Psychologica, 86, 227–272. doi:10.1016/0001-6918(94)90004-3

Rideout, V. J., & Robb, M. B. (2019). *The common sense census: Media use by tweens and teens*. Common Sense Media.

Rizzo, A. A., Buckwalter, J. G., Bowerly, T., Van Der Zaag, C., Humphrey, L., Neumann, U., ... & Sisemore, D. (2000). The virtual classroom: a virtual reality environment for the assessment and rehabilitation of attention deficits. *CyberPsychology & Behavior*, *3*(3), 483-499.

Slater, M. (2009). Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments. Philosophical Transactions: Biological Sciences, 364(1535), 3549-3557.

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, *109*(3), 431-433.

Stothart, C., Mitchum, A., & Yehnert, C. (2015). The attentional cost of receiving a cell phone notification. *Journal of experimental psychology: human perception and performance*, *41*(4), 893.

Uğur, N. G., & Tuğba, K. O. Ç. (2015). Mobile phones as distracting tools in the classroom: College students perspective. *Alphanumeric Journal*, *3*(2), 57-64.

Usoh, M., Catena, E., Arman, S. & Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence*, 9(5), 497-503.

Appendix 1. Interview questions.

No. of question	*Question*
(1)	*Vad tror du att experimentet gick ut på?* Eng. What did you think the experiment was about?
(2)	*Har du någon tidigare erfarenhet av virtual reality?* Eng. *Do you have any previous experience with virtual reality?*
(3)	*Hade du några strategier när du lyssnade på föreläsningen?* Eng. *When listening, did you use any strategies?*
(4)	*Vänligen rangordna din upplevelse av att faktiskt vara i den virtuella miljön på en skala mellan 1 till 7, där 7 representerar den normala upplevelsen av att vara på en plats.* Eng. *Please rate your experience of being in place in the virtual environment on a scale from 1 - 7, where 7 represents the experience of actually being in that place that the virtual environment was emulating.*
(5)	*Hur skulle du jämföra din upplevelse i 3D-miljön med en vanlig föreläsning? Hur naturligt kändes det?* Eng. *How would you compare your experience in the 3D environment with a regular presentation? How natural did it feel?*
(6)	*Finns det något som skulle gjort upplevelsen mer realistisk?* Eng. *Was there anything that would have made the experience more realistic?*
(7)	*Hur upplevde du vibrationerna i kontrollen? Kändes det som att det hörde till miljön eller som att det kom utifrån?* Eng. *How did you experience the vibrations in the controller? Did it feel like it was part of the environment, or like it came from outside the environment.*

Johansson, B., Gulz, A., Haake, M., Wallergård, M., Alce, G., Niehorster, D., Nirme, J., Ternblad, E. M., & Tärning, B. (Eds.) (2021). Intelligent, socially oriented technology V: Projects by teams of master level students in cognitive science and engineering. *Lund University Cognitive Studies, 180.*