Intelligent, socially oriented technology IV: Projects by teams of master level students in cognitive science and engineering

Wallergård, M., Haake, M., Balkenius, C., & Gulz, A. (Eds.)

Lund University Cognitive Studies, 175

Wallergård, M., Haake, M., Balkenius, C., & Gulz, A. (Eds.) (2019). Intelligent, socially oriented technology IV: Projects by teams of master level students in cognitive science and engineering. *Lund University Cognitive Studies*, *175*.

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

Table of Contents

AEGIS

HOLOGRAPHIC

Augmented Security

Philip Alm, Joakim Andersson, Axel Hellman, Henrik Josander, Andreas Månsson

Abstract

Guards in the manned security industry today are daily facing problems regarding workload, time efficiency and difficulties in localization. Modern technology has the potential to reduce these problems and help guards in their work life, thus reducing costs for security companies. One of these promising technologies is augmented reality. We have developed and tested a prototype for augmented reality to be used in the manned security industry. The report includes theoretical background, the process of building the prototype, user testing and suggestions for the future. Results from the user tests suggests that using waypoints as symbols in augmented reality guiding enhances situation awareness and performance compared to arrows. Even though our prototype shows the concept, we conclude that the technology is not yet ready to be used in this line of work in a real life scenario.

1 Introduction

Augmented Reality (AR) technology can work as a mean to enhance a person's interpretation and perception of their environment, and lets them interact with it by projecting virtual information with the help of hand gestures into the real world through computer displays (Liu & Seipel, 2018). This augmentation could provide an increase in specific task performances, a change in available useful information, increased situation awareness, and have therefore attracted a lot of research and potential application areas.

One of these industries is the security industry. Companies in the security industry have started to look into implementation areas and concepts, that could be improved by implementing new technology based around augmented reality (Peter Andreasson, personal communication, November 27, 2018). One area that has raised a lot of interest is revolving the work of a security guard, and how their work could be improved. Both when it comes to completing their tasks, raise their situation awareness, and improve the overall security revolving their work.

A guards' work involves a tough and high risk working environment, with difficult decisions and a demanding cognitive awareness. After conducting semi-structured interviews with guards we have identified a couple of key problem areas that characterize their work. There is simply too much and too dense information connected to the object, where the guard performs their patrol by walking around and inspecting the area. This information consists of everything from physical maps that are hard to interpret, with additional sub information such as different types of alarms and emergency areas, keys and pin-codes connected to different doors, to a task list with a vast amount of individual tasks that differs from object to object. All these problems boiled down to the same conclusion in all cases. There is simply too much to do, explore and keep track of, with too little time and means to do it. Their current technology consists of handheld devices in the form of a PDAs, radios, and physical objects such as maps. These tools are simply not able to provide them with the assistance they need. This is where the new augmented reality technology comes in.

One of the problems with PDAs, even though deemed convenient, is that they exist separated from their users in a much higher degree than wearable computers such as an augmented reality wearable display (Billinghurst & Starner, 1999). However, even if AR have been reported to be accurate and have positive impact on high demanding tasks, the final result will always be based on how well the intended user can realize the systems intended use (Liu & Seipel, 2018).

Our research goal has been to create an augmented reality prototype that aims to solve the major identified problems in a guards work listed above, and to present a possible solution for future implementations and research, that would further improve their work and awareness. The prototype have been created for the Microsoft HoloLens, and user tests for testing the prototype have been carried out and are displayed in the paper. All the theoretical research and the design process that the prototype is based on are listed and presented in the upcoming chapters.

2 Background

Situation awareness (SA) has been identified as one critical factor regarding performance in complex tasks (McKendrick et al., 2016). Lack of SA has also been named one of the major challenges in mobile collaborations in emergencies in the security domain (Reuter et al., 2014, cited in Lukosch et al., 2015). High situation awareness enhances performance, although it does not guarantee successful performance (McKendrick et al., 2016). Although a universally accepted definition of SA still is nonexistent, SA can be explained as a cognitive product of information-processing. It includes the understanding of one's environment and situation as context for one's actions (Lukosch et al., 2015). Endsley (1996) identifies three levels of SA. The first level is the perception of critical information, the second level involves comprehending the meaning of the critical information and the final level is the ability to project this information into the future. For someone to be completely aware of the situation it is thus not enough to simply see or perceive their environment, one must also understand the meaning of critical information in the environment in regard to one's own situation and future goals. Endsley presents a model of SA (See figure 1). Simplified, the model shows how SA (amongst other factors) leads to successful performance via decision making. The second and third steps in SA is heavily dependent on working memory to integrate perception of critical information with other information, compare the information with an existing goal state and project this information into the future. Other factors that come into play in this model are long term memory to help direct limited attention, expectations that affect how information is perceived and automaticity by reducing demands on limited cognitive resources. However, it is important to point out that the relation between SA and performance is a probabilistic one. Good SA does not automatically lead to good performance, but it increases the probability of good performance, just as poor SA does not automatically lead to poor performance.

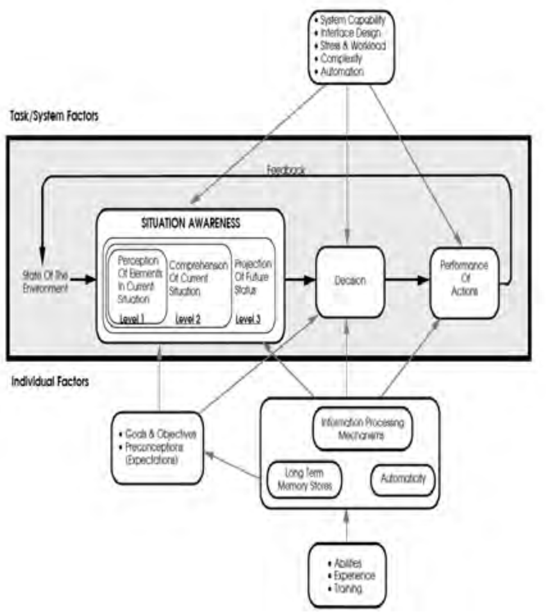


Figure 1. A model of situation awareness (Endsley, 1996).

SA is thus highly dependent on attention and working memory, much like mental workload. As an effect, augmented reality has the ability to both reduce and enhance SA depending on how it is used. Degradation of divided attention due to cognitive tunneling may be a risk factor in augmented reality. Optimal divided attention is accomplished through rapidly shifting or splitting attention between different inputs. Intense focus on a display when using an augmented reality headset may instead lead to degradation of divided attention by suboptimal allocation of focused attention to different inputs in the environment. As stated above, this is closely related to mental workload, which refers to the brains' limited capacity to process information. Augmented reality can ease mental workload by reducing information that needs to be held in working memory via holograms directly in a user's visual field, thus reducing time and distance between visual fixations (McKendrick et al., 2016).

McKendrick et al. (2016) set up an experiment to compare navigation, SA and divided attention between participants using either an augmented reality headset or a hand-held display. Participants were to navigate around a college campus, while experiment leaders assessed divided attention through an auditory n-back test as well as SA through questions regarding the surroundings. Pre-frontal cortex activity was measured using portable functional near infrared spectroscopy (fNIRS). The pre-frontal cortex was chosen due to its functional relationship with working memory, and former research has found fNIRS measurements to be an adequate index of mental workload as well as divided attention. Results from this study show that the augmented reality headset reduced mental workload for participants while navigating. Both participant-groups performed the task successfully, although the participants using augmented reality showed superior working memory recall. No difference in performance on the SA task was observed between the groups, however, hemodynamic differences were observed. The authors suggest that the augmented reality technology still can be perfected to enhance performance on SA tasks as well. One way in which this might be accomplished is to reduce cognitive tunneling by regulating the amount of unwanted attention capture.

Results from this study are in line with findings from another experimental study, investigating augmented reality as a user interface for the internet of things (IoT) (Alce et al., 2017). The study examined three different augmented reality interaction models for discovering and selecting IoT devices, using the Microsoft HoloLens. One model showed the world in miniature (WIM) and clustered icons together in real space. A second model showed a floating menu showing six different devices to select from on each page in a menu that is fixed in the field of vision of the user. The third and final model, the floating icons model, placed holograms in close proximity to the actual devices that the icon interacts with. Results show that participants completed the tasks faster and more easily in the condition using a lower hologram icon density level. This might be due to a high level of hologram icon density increasing the amount of unwanted attention capture, and thus increasing cognitive tunneling, as suggested by McKendrick et al. (2016). Further findings from the Alce et al. (2017) study suggests that with more holograms and more devices to interact with, ease of use diverge between different interaction models. The WIM model seemed to stand out as difficult and time-consuming to interact with when the level of hologram icon density is high. This might be due to the fact that with a lot of holograms to interact with, the WIM model, by its nature, assemble a lot of holograms in the same space, making selection and interaction harder. By contrast, the floating menu model only shows six devices at a time, no matter how many options there are. In this regard, most participants found the floating menu model to be easy to use. The floating icons model seemed to be the best model, from participants responses to a post-test questionnaire, in terms of situation awareness. Participants answered that they became more aware of the real world in both the high and low icon density levels when using the floating icons model in comparison with the other two models. These results taken together seem to suggest that not any one model is optimal in all situations, but a fully developed AR-system should combine different models for interaction depending on the situation (Alce et al., 2017).

It might not come as a big shock that augmented reality holds a lot of promise in providing assistance in tasks without capturing too much attention, we would argue that this is true only if the application and interactions are designed in the right way. Too many holograms and annoying interactions (Alce et al., 2017) might easily distract a user and make the task harder to complete instead of helping the user. Clark (2014) offers an interesting discussion about how the world, or at least portions of it, often is used as an extraneural memory store. To explain this claim, the author uses the example of the expert bartender who uses distinctively shaped glasses in specific arrays, as memory cues to deliver complex orders in a noisy environment with skill and accuracy. The bartender thus makes use of both internal and external resources to complete the task successfully. Without the use of these external cues, performance among expert bartenders have been shown to drop in experimental settings (Beach, 1988, cited in Clark, 2014). The domain of cognitive technology (as termed by Clark, 2014) involves not only brain and external objects, but also our body as situated in a social and cultural world full of props and aids. The idea is that the brain does not function in a vacuum. It links patterns of incoming sensory stimuli with associated information. We can thus transcend the limitations of our own mind by combining internal, neural operations with external props and aids that help reduce complex problems to ordered sets of simpler operations of the kind that the brain is more comfortable with. Clark (2014) expands this discussion by painting a picture of a kind of reciprocal evolution between technology and cognition. The first generation of brains and technology affects each other and cooperates to design the second generation of brains and technology, which in turns affects and cooperates to create the third generation and so on. If this is true, it gives us reason to be careful and considerate at every step of technological development to optimize the cognitive/technological evolution. With technology such as augmented reality at our doorstep, we want to make design choices that lets the technology reduce complex problems to ordered sets of simpler operations. The technology should provide cues and scaffolding and let the brain do what it is good at, such as recognizing patterns. It should strive to avoid providing the user with tasks that are inherently difficult for the biological brain, such as executions of long, arbitrary sequences of operations. To once again return to the analogy of the bartender, the technology should provide the bartender with distinctively shaped glasses, ordered in helpful arrays and not just uniform glassware that does not simplify the operations made by the biological brain.

2.1 Measuring SA

There have been several proposed ways of measuring SA. Some of these are indirect and some are more direct. McKendrick et al. (2016) used fNIRS as a way of measuring mental workload and divided attention. They also used post-test questions to assess how aware of their surroundings the participants were during the test. Endsley (1996) proposes eye-tracking as a way of measuring how attention is deployed in acquiring SA. These may be regarded as indirect ways of assessing SA. More direct ways to measure SA can be divided into subjective techniques and questionnaires (Endsley, 1996). Subjective techniques involve letting an experienced observer or the participants themselves rate their SA. Although these techniques are fairly easy to use in different settings, they can lack in validity due to their subjective nature. Other standardized ways of measuring SA directly have been developed (Endsley, 1996), such as a modified version of the Subjective Workload Dominance (SA-SWORD), or the Situational Awareness Rating Scale (SARS). These techniques, however, are still subjective (SA-SWORD), or developed primarily as an evaluation of air combat operations (SARS). The Situation Awareness Global Assessment Tool (SAGAT) has been developed as a global tool to assess all three levels of SA (see figure 1. above). The SAGAT has been used in several different domains and has been shown to have high validity and reliability in measuring SA. It involves queries regarding all levels of SA by randomly pausing the test-phase to administer the SAGAT. The pausing of the test-phase is done to overcome the difficulty of memory decay over time if a questionnaire is administered after the test is done. Although Endsley (1996) argues that this is the best technique for measuring SA, it does have its' downsides, the SAGAT has to be prepared for the specific domain in which it is to be used by a detailed analysis of SA requirements in the specific domain.

We have used the NASA Task Load Index (NASA-TLX) for measurement in our user study. The NASA-TLX is a multi-dimensional subjective rating procedure that gives a overall workload score (TLX-score) based on six subscales. These subscales are mental demand, physical demand, temporal demand, performance, effort and frustration. A second part of the NASA-TLX is available which entails pairwise comparisons between each subscale. The NASA-TLX has been widely used in numerous studies for measuring overall workload and situation awareness (Hart, 2006).

2.2 Current Limitations

One serious limitation with the current version of the Microsoft HoloLens is the users' field of view that is estimated to be about 30°×17.5° per eye, compared to the human visual field of about 200°×130° (Kreylos, 2015). This limitation is also mentioned by Alce et al. (2017) and it means that the user has a very small part of the screen where holograms can be seen and interacted with. When a user turns his head so that the hologram reaches the edge of the display, it is cut off and vanishes. This might interfere with a users' sense of object permanence as suggested by Kreylos (2015). In real life, we know that objects are permanent even when they are out of sight, placing an object on a table and then turning your head away does not mean that the object no longer exists. With the HoloLens, a hologram placed on a table vanishes when the table is still in the human field of view but out of the HoloLens field of view. The invisible boundary created by the HoloLens field of view thus interferes with the users' sense of object permanence and creates an unnatural sensation where the user has to turn his/her head so that the hologram is in the center of the visual field at all times not to vanish. The reason for this small field of view might be both a problem of processing power as well as an optical problem, either way it might be hard to solve in the foreseeable future without using diamond instead of glass if we are to believe Kreylos (2015). However, more recent head-mounted AR-devices seems to have already overcome this limitations. The Meta 2, for example has a field of view that is about three times the size of the HoloLens (Odom, 2017). One of the downsides with the Meta 2 is that it is tethered and must be connected to a computer via a cable while using it. Thus, it is not optimal for our purposes and we have instead chosen to use the HoloLens. However, this development in hardware leads us to believe that future devices will not suffer from the same limited field of view that the current HoloLens does.

Another limitation mentioned by Alce et al. (2017) is the amount of recognized hand gestures that can be used to interact with the HoloLens. The HoloLens we are using is an upgraded version but the possibilities of interaction are still limited to voice commands, a clicker device and two simple hand gestures. The two hand gestures are called bloom and tap. The bloom is used for opening menus and exits the current running application and is therefore not used within the prototype. The tap gesture is the gesture that can be used, it's performed by pointing with your index finger and then bending your finger in a tapping motion towards your thumb. The In our current prototype, we have focused on forms of interactions that avoid these limitations, such as voice commands and image recognition that is triggered simply by looking at an object.

3 Process

During the course of this project, the team has had a clear design process in order to develop as great of a product as possible. This process can be divided into three major phases; the planning phase, the design phase and the implementation phase. During early stages of development the group focused on generating a large amount of different ideas, whereas these were filtered and refined in later stages of the project.

3.1 Brainstorming

Starting the project off, the group set out to generate a vast amount of ideas. The focus at this point was not the quality of the ideas but rather the quantity, which is certainly of a quality in and of itself - at least in the beginning of a project. The "good" ideas, meaning the ideas that shows the most promise after extensive discussion, can be picked out and get refined in later stages of the project. The idea generation was done by using the specific method of brainstorming called "brainwriting", which is a method of generating a diverse amount of ideas that has been analysed and elaborated on by all the members of the group. The way it works is that all the group members, which in the case of the project group was five individuals, have five minutes to write down three different ideas. When the time is up, these ideas are then passed on to the person next to you and the timer is set for another five minutes. The participants can at this point either elaborate on the ideas given to them, or write down new ones. This is repeated until every participant has written on every paper. The major advantage for this method of brainstorming, which is especially useful for this course, is that it benefits from having participants from different backgrounds and past experiences, such as studying different programs. This advantage is of course that the method encourages the exchange of knowledge between participants as well as the elaboration of ideas that could otherwise have been useless without further examination by different thinking individuals. The following two images, figure 2 and figure 3, are examples of the brainwriting session:

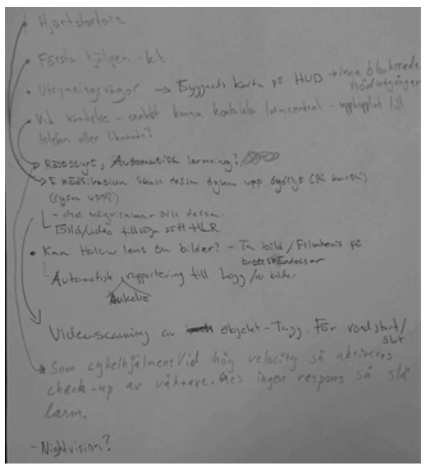


Figure 2. First paper from the brainwriting session

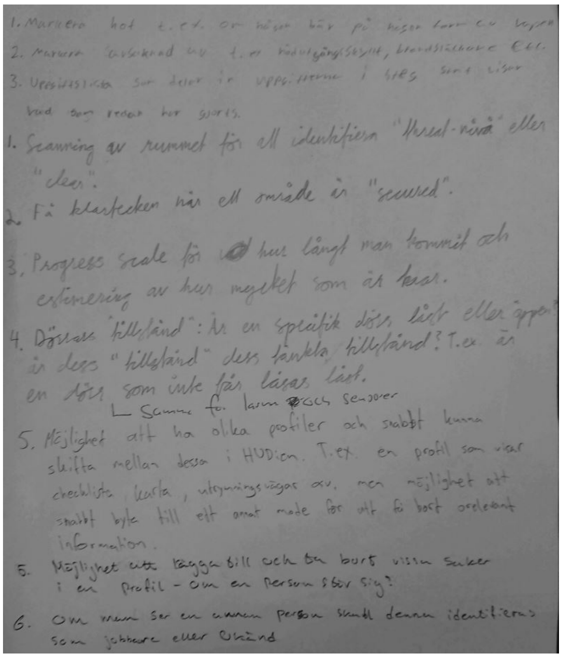

1. Markera hot t.ex. om någon bär på någon form av vapen
2. Markera avsaknad av t.ex nödutgångsskylt, brandsläckare etc.
3. Uppdragslista som delar in uppdragen i steg samt visar vad som redan har gjorts.
1. Scanning av rummet för att identifiera "threat-nivå" eller "clear".
2. Få klartecken när ett område är "secured".
3. Progress scale för hur långt man kommit och estimering av hur mycket som är kvar.
4. Dörrars "tillstånd": Är en specifik dörr låst eller öppen / är dess "tillstånd" dess [illegible] tillstånd? T.ex. är en dörr som inte får låsas låst.
└ Samma för larm och sensorer
5. Möjlighet att ha olika profiler och snabbt kunna skifta mellan dessa i HUDen. T.ex. en profil som visar checklista, karta, utrymningsvägar osv. men möjlighet att snabbt byta till ett annat mode för att få bort orelevant information.
5. Möjlighet att lägga till och ta bort vissa saker i en profil - om en person stör sig?
6. Om man ser en annan person skall denna identifieras som jobbare eller okänd

Figure 3.Second paper from the brainwriting session

The result of this brainwriting session was then categorized by use of an affinity diagram. This was done to make it easier for the group to analyse, combine and otherwise refine the generated ideas. The outcome of the affinity diagram can be summed up by the following categories:

1. Checklist
   Patrol checklist: A checklist that keeps track of a guards patrol and makes sure that all mandatory parts of the specified patrol is fulfilled.
   Checklist combination: Instead of having a pure checklist that is more or less just a hovering HUD, could it be integrated with anything else in the virtual space? For instance it could be combined with a map.
   Progress bar: Some indicator to show the user how far they have come and how much is left, in order to adjust the speed of the patrol accordingly.
   "Gamification": Take ideas from the game industry and apply that to the checklist, such as achievements and the like.
   Timer: Timer to tell how much time the guard has left to finish the patrol.

2. Map
   Patrol map: A map that shows objectives such as doors that needs locking, alarms and sensors. The map could be in 3D, a virtual display on the users arm or as a HUD in front of the user.
   GPS functionality: Guidance to objectives in the building and navigation while driving to a building, all done by showing the user around in augmented reality by arrows, lines or "waypoints" for example. In addition, this function could include the possibility to see the distance and the estimated time to the objective.
   Emergency mode: Show important information to the user in case of an emergency, such as first-aid kits or escape routes. This can be seen in the users field of view or on a map for example.
   Minimap: The ability to show a minimap of the surrounding area, not only as a way of showing the buildings different floors and its layout, but also as a way of showing where the guard should go next by integrating it with the GPS function.

3. Options
   Customization: Ability to add or remove functions that the user might deem useful or unnecessary. Choice of how much guidance in the form of arrows, lines or waypoints is wanted and how transparent these features, among other, should be.
   Modes: Depending on what kind of job the security guard is doing, differents customizations might be preferred. For example, a user might have a driving mode and a patrol mode.
   Profiles: Different users could use unique profiles that saves for example settings, customization options and saved modes.

4. Security and deviations
   Communication with headquarters: Ability to quickly connect real time with the central headquarters and communicate via speech or text, dependent on the urgency. Get auditory and visual feedback from the headquarters. Used for both informative reports and personal security.
   Deviation detection: See if a deviation has occurred, such as an unlocked door that should be locked. Ability to easily make a report on the deviation with voice recognition and pictures taken with the HoloLens, as well as audio and video recordings if needed.
   Highlight important objects: Objects and items of interest to security could be highlighted, such as exit signs, fire extinguishers and first-aid kits, thus making them hard to miss.
   "Threat-level": A comprehensive function that detects the level of threat of different scenarios, buildings, rooms and individuals. With buildings and rooms, the idea is to automatically surveil areas with security cameras and detect such things as abnormalities and movement in order to determine if a room might be more dangerous than usual and should be searched as such. This information would be shown before entering said building or room. With people, the headset could detect the individual and see if they equip some sort of weaponry or mask, as well as identify the person and make sure they are permitted in the area and rate their threat

level accordingly. This could be displayed above that person. A function like this would of course be hard, if not impossible, to implement with today's technology but maybe the technology of the future is good enough to make such function a possibility.
Video feed: The ability to take control of security cameras and see what the camera sees, in order to safely inspect and determine the safety of an area. The camera could be used as the users second pair of eyes, and would be controlled by moving your head and seeing the information almost as if it was virtual reality.
Panic/emergency button: A button or otherwise accessible way of triggering an "emergency mode" in which the headset starts recording both video and audio, automatically connects the guard with the headquarters, and tracks the location of the guard. Another way of activating this mode could be if the headset detects an abnormal amount of velocity that would occur if for example the guard falls over or gets hit in the head. The headset would in this case then automatically go in to the emergency mode within a short timeframe if this is not aborted by the guard as a miscalculation.

5. Information
Database of keys and codes: Show what kind of key or key code is necessary to open a lock. If a key is needed, it could be a 3D representation of said key that is shown in order to make it easier to identify and compare. A security guard today can have a key ring with a large amounts of keys and a log of different key codes, this would make the process of opening locks a lot more efficient.
Night/thermal vision: If a specialized headset was created specifically for security guards or similar personnel, it could include features such as night and thermal vision in order to heighten the situation awareness, and make it easier to find and identify living creatures, whether it is humans or other animals.

3.2 Personas

After our initial brainstorming we went ahead and created a single pair of personas, displayed in figure 4 and 5 below. The personas aims to give us some guidance regarding what a person with exaggerated personality features would think of our prototype. When creating our pair we therefore tried to get a wide spread when it came to all aspects of their personalities and life situation. By creating the fictive pair we could both get insight into what these people would need from our prototype and how they would regard it. It also served the purpose of making the intended segment of security guards more alive in the project, and worked as a guiding tool during the actual implementation of features and functionalities.

When creating the personas it's important to make them as alive and realistic as possible. By creating a fictive background and life situation, we could relate more strongly to the characters and have them in mind during the whole project. After our discussion we decided that the most important information to include in our personas would consist of their work and technology experience, behavior, age, and acceptance level of new technology. Their life situation and personal life was created in line with best practises but could be excluded. Another area of use for the future would be the possibility to create fictive scenarios based on our persona pair.

Bengt

Residence: House owner outside the city.

Family: A wife and two teenage girls along with the family dog.

Age: 52

Work Experience: Long experience, been around since everything was paper and analogue based.

Behavior: Very routine driven and precise in his analogue working method.

Level of motivation: Believes everything has always worked, but has gone through some changes with a positive attitude after the initial phase.

Technology experience: Has some basic IT experience but doesn't usually seek out new technology by own will.

Figure 4. The first persona Bengt

Adam

Residence: Recently moved out of his parents house into his first apartment in the inner city.

Family: An older sister and his two parents.

Age: 20

Work Experience: Short experience, recently graduated high school.

Behavior: Insecure in his own capabilities and tends to recheck his own work multiple times.

Level of motivation: Always open for new suggestions that could improve his work and everyday life.

Technology experience: Extensive technology experience since early childhood. Loves to try new technology but has never tried AR before.

Figure 5. Second persona Adam

3.3 Interviews

There were several purposes behind conducting interviews. The different work segments and persons we have been able to interview consist of two guards, the CEO and CTO of Blue Mobile Systems (a software development company for guards and security solutions), three employees at AXIS Communications and two of their AR developers. In addition to this, we have also had a quick discussion with three additional guards about the idea and our suggestions.

Security guards

First of all we needed a better understanding of the everyday work situation and routines of an actual security guard, and what scenarios they might encounter. Without a clear understanding of their work, we would have no validations behind our initial ideas and functionality suggestions. We would also easily have missed major problem areas we might have overlooked. We therefore held two separate semi structured interviews with two different security guards. The interviews were focused on getting to know their work, the problems that existed in it and where they thought AR or additional technology might be able to help. After these points we explained our ideas and suggestions to get to know their reactions and if they had any feedback or saw any problem with them. The order of the interview was very important to us. Just like we didn't want to have a too narrow mindset when we brainstormed our ideas, we ourselves didn't want to affect the guard or guide them towards our ideas initially. This could have narrowed the information we received and left out valuable lessons or information.

After the interviews it became clear that there were major problem areas in their everyday work that were in line with our suggestions, but three functionalities stood out as the most important ones to the guards. A guiding system inside the object, some way of keeping track of all their tasks in a more manageable way than memory or paper form, and a way to solve the problem regarding what key or pin code to use at a specific door. These three functions therefore became our top priorities.

Blue Mobile Systems

In addition to understand what the security guards do on a daily basis, we also needed a better understanding of their equipment and its problems if we are supposed to be able to improve it. We conducted an interview with the CTO and CEO of Blue Mobile Systems (BMS). The CEO brought the experience of 1200 plus hours with following guards and conducting structured studies regarding their problems and categorising them. Something that the company later analyzes and try solving with the help of applications.

In addition to the information about the work of a security guard and their equipment, he provided us with useful information about the leaps of effectiveness that has occured when guards switched from pure paper form to handheld devices. He was also a firm believer that AR could be the next leap, but that we are not there yet. A final note was the agreement that showing the way for a guard is the single most important feature to implement and solve. Besides the telephone interviews with BMS, we also visited the company's headquarters in Gothenburg to further discuss our prototype as well as getting hands on experience of their products.

AXIS Communications

A meeting was conducted with three employees at Axis Communications to discuss ideas regarding our prototype. They also put us in contact with two AR-developers that we conducted an interview with.

From this interview we got a host of implementation pointers and tricks that we will mention in the development chapters. They recommended different techniques and tools that we could use to help our development, things like the Mixed Reality Toolkit (MRTK) which is a Unity package containing premade functions and scripts specifically for the AR. How we could use the HoloLens for room scanning and tips for how we could match the AR worldspace to the real world one. The key takeaway from this interview was to focus on a few functionalities for our first prototype and leave the rest for future versions to come.

3.4 Conceptual model and Lo-Fi

After the brainstorming and interviews we had a clear idea of the three most vital functions to implement. But even though we knew the chosen features would be a guiding system, a key/code supporter and a checklist, we all still had different ideas what exactly these features meant, and how they could be designed. The next step was therefore to ensure we all had the same conceptual model in mind. According to Norman (1988), a conceptual model is a correlation between affordance, constraints and mapping of interaction with an object in mind. In our case the conceptual models will reflect how each of us in the group have thought to create this correlation with our application, the users perception of interaction and physical interface choices.

To describe our different conceptual models we both discussed and used lo-fi models to describe our individual concepts to each other. We could then merge all of our best ideas into one final concept that we were going to implement. We did this for all of the three main functions. It was a challenge to illustrate how things were going to work in augmented reality with a low-fi model, so the discussions around them were just as important.

We realized right away that we had different ideas on how to implement a guiding system for both of the guiding concepts. One concept consisted of guidance using a map and the other was to have something in the floor that showed you where to go. We realised that the second option utilises AR by guiding the guard without being the only focus in the field of view (FoV). We therefore believe that this solution will have a lower impact on the users workload memory and an improved SA. But we still had different ideas how to implement and design this. The three main options consisted of guiding arrows, a line, or a small circular object we have chosen to call waypoints. An advantage with arrows is that you can clearly show the direction. The idea with the lines could be good from a design perspective but we realized that it would be harder to implement in our prototype. The final idea is guiding with waypoints. The advantages with waypoints is that it is easier to place since we do not have to consider direction when placing them. We believe that another advantage might be that waypoints motivates the user placing them to do so more often than they would with arrows. However if that is the case is not tested. It can also be viewed from a steeper angle since it is a higher model than the arrow. This is useful when guiding a guard up a staircase. The advantage that arrows show direction could also be added to waypoints by animating them. Another reasoning behind our decision is inspired by the gaming industry. Their idea of guiding in a virtual environment is similar to ours in the real world. Since we don't have any concrete sources of what option would conclude in the best result, we therefore decided to implement both arrows and waypoints, to test the difference. Since the solution with line requires a completely different implementation testing this is out of this projects scope. With a common ground for how the first feature would look we moved on to the next.

The second main feature was the key and code support functionality for locked doors. Just like our previous feature we all had different understandings of how it could be done. From our interviews we learned that the patrolling guards have to learn which key goes to which door. This is also the case for larger buildings where there are over 25 different keys for all the doors. However, we also learned that none of the patrolling guards would like any large distractions during their whole patrol when they try to find the information. This is something very understandable since using any complex system would negatively impact the users situation awareness by having to work through menus or searching for which code or which key fits to what door. Therefore we decided to keep information close to where it is needed, and designed a hologram that would emerge from the door when the user looks at it. The idea is that when the guard arrives to a locked door, they shall only need to look at the door and a hologram emerges from it displaying information on how to get through it. What key is used, what code to use, or what keycard is needed (see figure 6).

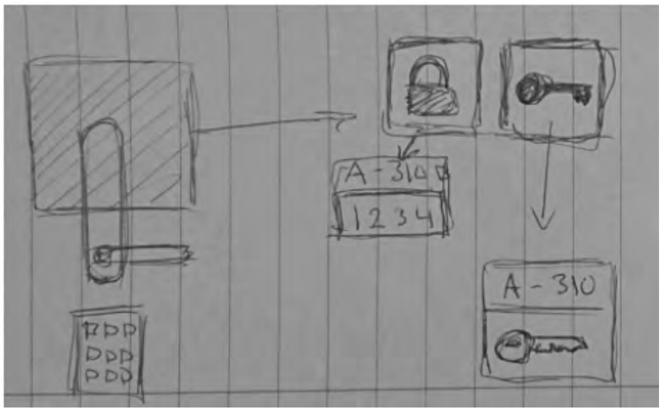


Figure 6. Lo-Fi on how key codes and keys can be displayed at a door.

The final major feature that is going to be implemented in our prototype is a checklist. With the information gathered from the interviews we started on a lo-fi for how the checklist could be implemented. However, we didn't want to commit to a solution before visiting BMS and checking out what their current checklists looks like. After visiting BMS we made some small tweaks to make the checklist more familiar for guards but still work in AR. The final design combined the second main feature and our first preliminary checklist design (see figure 7). This checklist would appear outside of rooms that contain new tasks on the guards route. It displays the current progress, a list of tasks, time and if needed to gain access to the room it displays what key or what code the door needs. We decided that we wanted to only show the task that the user currently needs. We do so by displaying the task right as you enter the room where said task are to be performed. This should in theory reduce the mental workload as discussed in section 2 and therefore enhance the guards situation awareness.

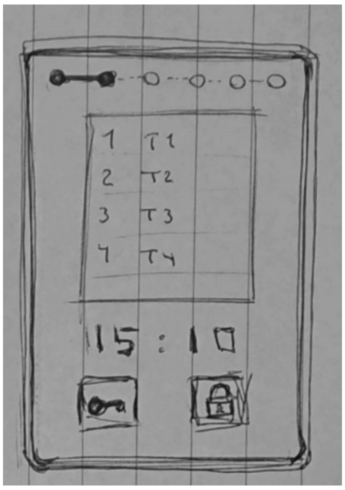


Figure 7: Lo-Fi on the checklist combined with key

3.5 Hi-Fi and Development

With the results from the brainstorming, interviews and Lo-Fi discussions the development could now begin. For starters the challenges were about developing for Microsoft HoloLens using Unity. The technology is still not mature and there is still low support for development compared to development for regular displays or head mounted virtual reality displays. The HoloLens runs on Windows 10 and a holographic processing unit. To deploy Unity projects to the HoloLens, some tools are necessary. First, a correctly setup visual studio is required. The installation of visual studio requires the right options checked, for example usb connectivity to be able to transfer files to the HoloLens. Furthermore the software development kit (SDK) that visual studio uses have to be one that is supported by windows mixed reality which is not always true for the latest SDK. Unity also needs to be setup with the right build and player settings to be able to run the application on a HoloLens. With the right version of Unity with correct settings and visual studio correctly setup we could now build and deploy a solution from Unity. To do so a couple of steps are necessary. First Unity needs to build a visual studio project. This project can then be opened with visual studio to be able to build an Appx which is an application package ready to be distributed. The Appx can then be deployed to the HoloLens which can install and run this format. This process then have to be repeated every time you want to test your current build. Luckily there is tools to make the process a little bit easier and one of them is MRTK. The same process still runs in the background but every step can now be made within Unity. MRTK also contains support for basic functionality such as voice speech input recognition, gaze control and gestures. This toolkit was highly recommended by the developers we got to interview through Axis. We therefore decided to use MRTK for our Hi-Fi prototype. A small difference in the last step to deploy the solution is that the MRTK uses the Microsoft device portal to connect with the HoloLens, this requires different credentials than the direct connection with usb and a pin. We had some initial problems with this but once the right credentials were set up this solution was much simpler than what needs to be done if you do not use MRTK.

Once everything was working we could start experimenting with our test environment. In our test environment we implemented the basic functions we knew our prototype was going to use later on. These functions were voice control, gaze control and gestures. To implement this using MRTK was fairly easy wich made using the tool an even better decision. In our test environment we could now switch colors on three different spheres by looking at the one to change and saying an added command like “red”, “green”, “blue” or “reset all” to reset all spheres back to white (see figure 8).

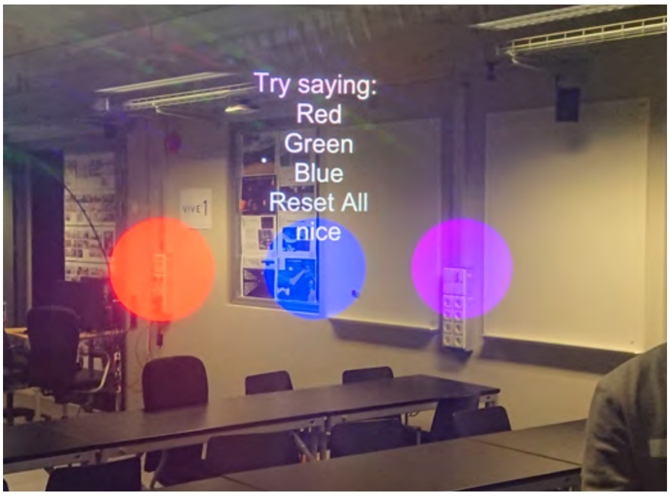


Figure 8. Voice command test environment in the HoloLens.

Another functionality we wanted to get working was image recognition. In the beginning we wanted to use this for most of our functions. We had the idea of placing pictures on the floor to trigger the guiding system but we later decided on a better solution that will be further described in the next paragraph. We also wanted image recognition to display information upon arrival to a new door. To get image recognition we used the tool Vuforia which supports development in Unity. Vuforia allows you to create a local database with images and the features of those images. With access to a camera Vuforia can then find these features and display a 3D object relative to the position of those features.

We finally decided on a much better way to implement the guiding system, by having the user first place the holograms that was later going to be followed. This is done by saying the word “place”. This places a hologram at the cursor which follows the users’ gaze. The hologram represents a 3D model of a waypoint that is one of our custom made 3D models used for the application and can be seen in figure 9. To make the placement easier we added tap to move functionality to these objects which allows the user to perform the tap gesture once to to start moving the hologram and then once again to stop moving it. We also added the command “remove” to give the user that places the guiding full control. The model can also be changed to be an arrow. The model is changed by using the voice commands “waypoint” or “arrow” (see figure 10). This solution turned out much better than the original idea to have images trigger guiding. The holograms are with this solution visible from every angle. With the picture triggering guiding this would not be possible. We still want to use Vuforia’s image recognition to display information upon reaching a door.

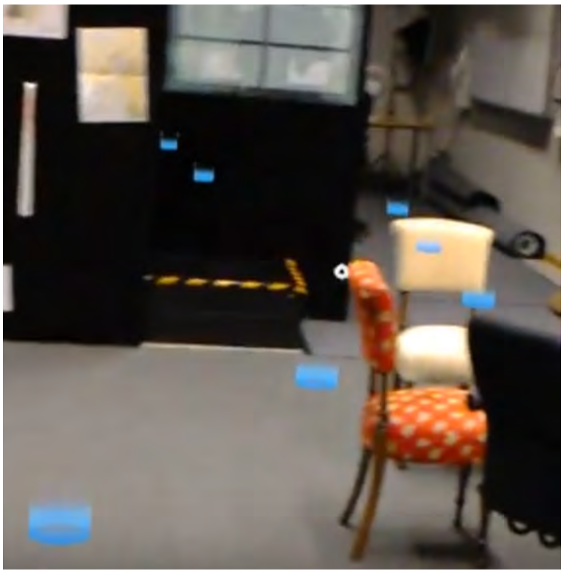


Figure 9. Waypoint guiding system in the HoloLens

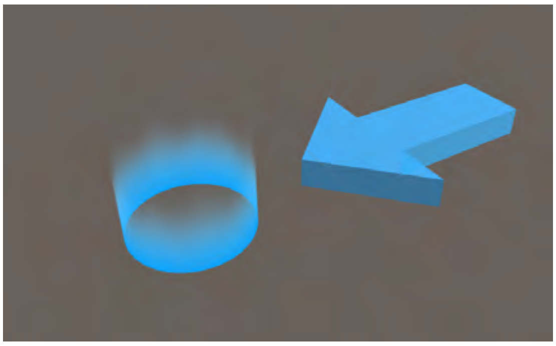


Figure 10. Model of a waypoint and an arrow.

Upon reaching a door that either needs to be unlocked or have tasks that for the user perform a GUI is shown. The GUI contains information on how to unlock the door and or what tasks that should be performed within the room the user is entering. The unlocking method shows whether the door should be unlocked with a key or a code. If a code is needed it will be shown in the GUI (see figure 11).

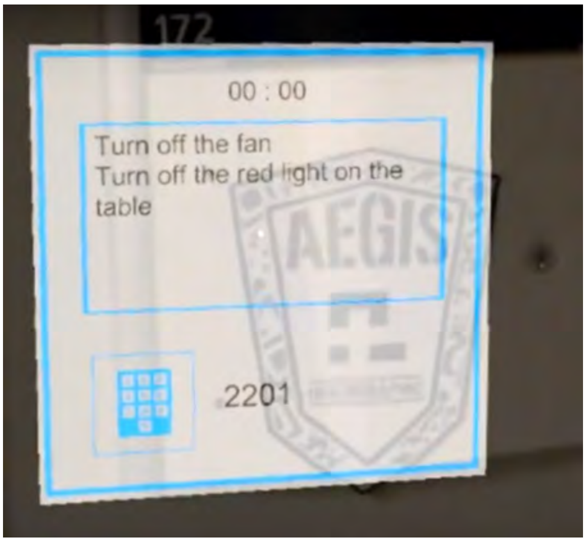


Figure 11: The GUI that is shown at a door.

During development testing we figured that there could be some distance between the door hologram containing the tasks and the physical tasks themself. That would mean that the guard could forget what tasks there is to complete in the room. However we didn't want to obstruct the users field of view by clotting it with information that the user might not want to see at all time. Therefore we created a solution so that the user can summon a smaller hologram containing the current task, we called this the gadget.

The gadget is a unique image that the user wears on their wrist and when detected by the HoloLens creates the task hologram (see figure 12). With this solution the guard can choose when to summon the task to receive information and when to not.

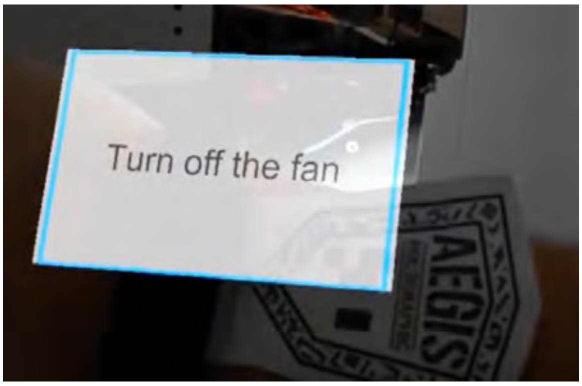


Figure 12. Wrist gadget.

The gadget only shows one task since it’s a small GUI and text is not always easy to read on a HoloLens. To go to the next task the voice command “next” is used. This shows the next task in the gadget and changes the text of the task to be green on the door GUI. It’s also possible to use the command “back” to go to the previous task.

Right before testing we found a big issue that needed to be solved for testing to be possible. When using the HoloLens, especially when switching user frequently, the

HoloLens lost tracking of the room and all holograms shifted in position. To solve this we added an anchor that all holograms are connected to. Therefore when this problem occurs, only the anchor needs to be set in place. This is done by initially deciding on an origin position and saying the voice command "set". When the HoloLens then loses tracking the position is reset to the origin position by going back to the chosen origin position and again using the voice command "set".

4 User Testing

This section describes the user tests that were conducted.

4.1 Participants

The participants consisted of 20 people, 8 women and 12 men, aged 24-60 (M = 29.85, SD = 11.15) from Sweden. 3 participants had tried head mounted augmented reality before, and 17 participants had never tried head mounted augmented reality before the test. The sample was selected through convenience sampling.

4.2 Material

The main hardware used in this study was the Microsoft HoloLens. The Microsoft HoloLens is a head mounted augmented reality device. The HoloLens is completely self-contained, meaning that there are no wires, external parts or need for a connection to a phone or PC (Microsoft, 2016). This creates a very mobile device that suits the needs of a working security guard. Besides the HoloLens and the displayed holograms we used three A4 papers with printed pictures of a red warning triangle symbol for participants to locate during the test, and two small papers, one on the arm displaying the current task (the gadget), and one placed outside the lab displaying all tasks. An electric fan and a small, red, LED-light was used as objectives in the tasks during the test. Participants signed an informed consent form on a paper and wrote down their participant ID, demographics and the amount of warning symbols they saw on another paper. A laptop was used for the participants to take the NASA-TLX test (Sharek, 2009).

4.3 Procedure

All user tests took place in the Virtual Reality lab at the Department of Design Sciences in Lund. All participants signed an informed consent form as well as a pre-test form for demographics, experience with head mounted augmented reality and their profession. We also controlled for use of glasses, all participants that normally would wear glasses did so during the test. After filling in the pre-test form, participants read the instructions for the test (see appendix 1). We employed a within subjects design, each participant first did a trial with either holographic arrows, or waypoints as guidance, and then a second trial with the remaining holographic symbol (see figure 10). We alternated the guidance symbols each participant began with between arrows and waypoints to control for training effects. When beginning the first trial, the participants stood in a predetermined spot outside the door of the virtual reality lab. They received a Microsoft HoloLens, a keycard to the door and a paper strapped to their wrist (the gadget) with a "AEGIS Holographic" logo (see appendix 2) that displayed a hologram of their current task during the trial. The participants were instructed to look at a "AEGIS Holographic" logo next to the door to receive a hologram of the code for the door as well as a list for all tasks they had during the trial. When entering the virtual reality lab, the participants were guided by either holographic arrows or holographic waypoints, depending on trial, through the virtual reality lab. All participants had two tasks during the trials, the first was to turn of a fan halfway through the virtual reality lab, after doing this, participants had been instructed to look at their current task hologram and say the voice command "next". This instructed the HoloLens that the first task had been completed and the current task hologram was updated to the next task. The second and final task in the trial was to turn of a small, red light, placed at the end of the route inside the virtual reality lab. After completing this final task, participants had been instructed to again use the voice command "next", upon which the current task list was updated and instead displayed "all tasks completed". This meant that the trial was completed and the participants could remove the HoloLens and return to the test leader. During the trials, three warning symbols printed on paper were placed around the reality lab. Participants were instructed to be vigilant and note how many of these they saw during the test. The warning symbols were constantly placed at the same spot during all the arrow trials and switched to another, constant spot during all the waypoint trials. The test leader timed each participant during each trial. Between trials, participants wrote down how many warning symbols they noted and took the first part of the NASA-TLX. After this, participants did the second trial in which placement of warning symbols were switched and the guidance symbols were switched from arrows to waypoints or vice versa. After completing both trials, participants wrote down how many warning symbols they saw in the second trial, took the first part of the NASA-TLX a second time, and answered semi structured interview questions (see appendix 3).

4.4 Hypotheses

Our first hypothesis is that participants will perform better in the waypoint trials than in the arrow trials, meaning that they will notice more warning symbols in waypoint trials than in the arrow trials. Our second hypothesis is that participants will prefer waypoints over arrows, meaning that they will have a lower NASA-TLX score in the waypoint trials than in the arrow trials and subjectively prefer the waypoints when being interviewed.

5 Results

In line with our first hypothesis, frequency of found warning symbols was significantly higher in the waypoint trials than in the arrow trials, exact binomial $p = .039$.

A paired-samples t-test showed no significant difference between time for the waypoint trials (M = 83.85 sec, SD = 25.03 sec) and time for the arrow trials (M = 93.05 sec, SD = 23.54 sec), $t\ (19) = 1.632$, $p = .119$.

Contradictory to our second hypothesis, a paired-samples t-test showed no significant difference between NASA-TLX scores for the waypoint trials (M = 19.63, SD = 10.59) and for the arrow trials (M = 22.50, SD = 16.03), $t\ (19) = .65$, $p = .523$.

Further, paired-samples t-tests for each NASA-TLX measurement showed no significant difference between the waypoint trials and arrow trials (values displayed in Table 1).

Measure	df	T-value	P-value
Mental demand	19	-.032	.975
Physical demand	19	.146	.885
Temporal demand	19	.459	.651
Performance	19	.765	.454
Effort	19	.248	.807
Frustration	19	1.160	.260

Table 1. Values for paired-samples t-tests between waypoint trials and arrow trials.

Figure 13 and 14 (see appendix 4 &5) below represents a boxplot of each NASA-TLX subscale score for the two different trials, where the lower and upper quartiles are represented by the box, and the whiskers represents the min and max values.

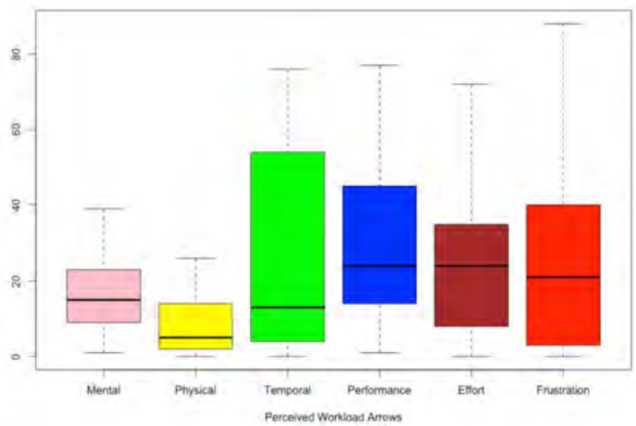


Figure 13. Boxplot of NASA-TLX Arrow trials subscales

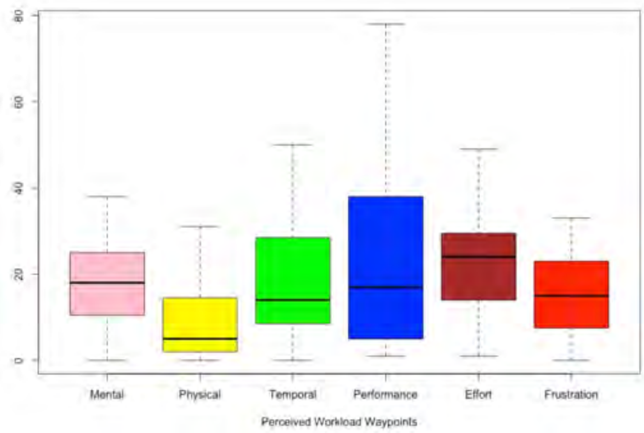


Figure 14. Boxplot of NASA-TLX Waypoint trials subscales

A total of 60 warning symbols could be found for the twenty participants during each trial. Figure 15 displays the total number of warning symbols found during the user test for each trial. Figure 16 & 17 displays how many participant found 0, 1, 2 or 3 warning symbols during each trial, where three symbols were the maximum.

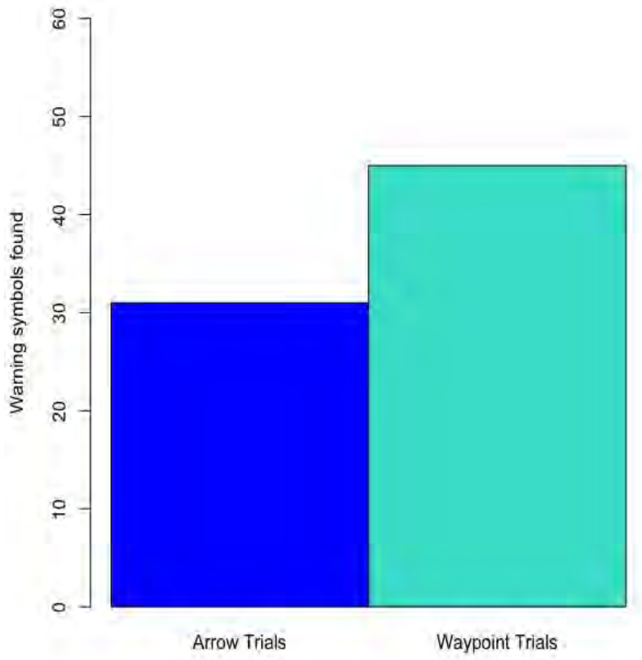


Figure 15. Total number of warning symbols the participants found

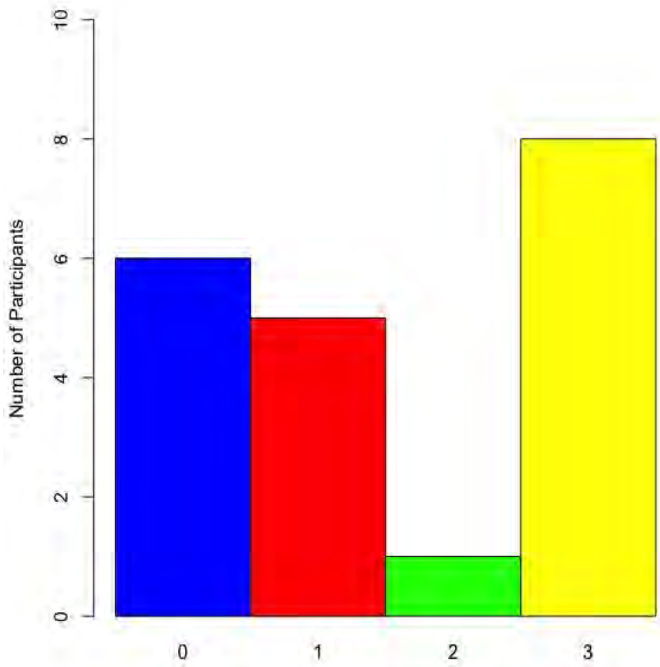


Figure 16. Number of warning symbols found during Arrow Trials

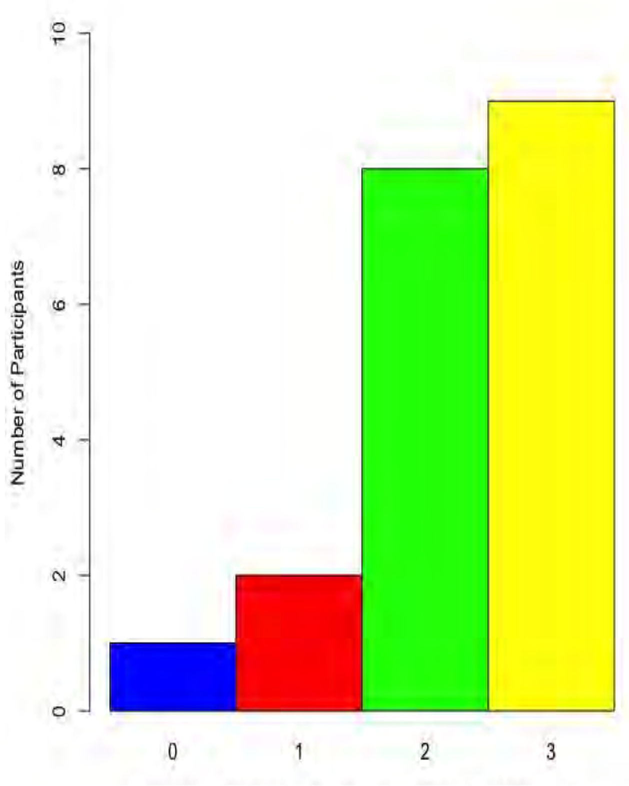


Figure 17. Number of warning symbols found during Waypoint Trials

The interviews resulted in the following subjective data. When asked about hardware and software problems the following problems were brought up: 8 participants found the headset uncomfortable, 14 stated they had problem with tracking, 5 mentioned low FoV, 7 experienced blurry text in the GUI, and 2 said they had problems with voice commands. When questioned if the holograms were distracting we found out that: 7 out of 20 participants found the holograms distracting, where 4 of them specified that it was worse with arrows. Everybody stated they understood both the GUI and the test leader. When asked if they prefered arrows or waypoints, 15 prefered arrows, 3 prefered waypoints, and two had no preference.

6 Discussion

The main goal for this project was to develop and test an augmented reality prototype for use in the manned security industry. Our user tests focused on investigating the difference between two different types of guidance symbols. The reason for choosing this as our focus in user tests was that initial interviews with guards working in the industry identified guiding as one of the biggest problem areas that augmented reality could help solve. Results from our user tests showed that 75% of participants preferred arrows as guidance symbols over waypoints in post-test interviews. This stands in contrast to our second hypothesis. Furthermore, we found no significant difference between NASA-TLX scores for waypoint trials and for arrow trials. Interestingly, in line with our first hypothesis, participants noticed significantly more warning symbols in the waypoint trials than in the arrow trials. We would argue that this has to do with the participants´ situation awareness. As discussed above, SA has three levels: perception, comprehension and projection (Endsley, 1996). As for the first level of SA, perception of the guidance symbols might not be any different between arrows and waypoints. For the second and third levels of SA however, comprehension and projection, it might be that participants comprehend the arrows as suggestions to simply move in the direction of the arrow. Projecting this information into the future could mean that the goal is to simply keep moving forward as long as there are arrows suggesting this action. This could create a sort of cognitive tunneling as discussed above (McKendrick et al., 2016). With the waypoints however, the information is more ambiguous, meaning that participants use the waypoints as guidance help when needed, but without losing track of their current tasks and SA. In short, the waypoints created less cognitive tunneling, allowing participants to perform better. This might happen completely implicitly, without the participants actively thinking about it which could explain why participants still preferred the arrows over the waypoints, even though they performed better (noticed more warning symbols) in the waypoint trials. We also timed all participants in all trials. Although we did not have any explicit hypothesis about time, we noticed in our user tests that participants seemed to complete trials faster in the waypoint trials. Although participants had a faster mean time in the waypoint trials, we found no significant difference in time between waypoint trials than arrow trials.

During user testing and development we ran into a lot of difficulties due to hardware restrictions. Some of the problems we were able to solve by adding functionality to our prototype, but some of the problems will remain until new better hardware comes out. The first thing we all noticed when using the HoloLens for the first time where the small FoV. This was something that was also brought up by 5 of the participants although all participants didn't realise that the hardware were the limitation. A hardware problem that we were able to solve with a feature was the loss of tracking. It's nowhere near completely fixed and in a real life scenario our fix is not sufficient. When the HoloLens loses tracking the user can reset anchor and fix all the holograms positions. This works when demonstrating a small demo area, even though it comes with some frustration since the user has to go back and restart. Luckily we implemented this feature before the testing which would not be possible otherwise. This feature had to be used a lot during testing and 14 out of 20 participants complained about the loss of tracking during the user test. If a guard would need to go back to reset the position even once during a patrol, the application would not be usable since the guard already is pressed for time. Another hardware problem is how poorly the HoloLens performs in low light. When the lighting is bad the hololens have troubles recognising the area and therefore loses tracking which leads to the holograms being in the wrong place. This is not that big of an issue when testing out a prototype to show a concept, but again, if it were to be used in live scenarios with an actual guard, this problem would be huge since a lot of the patrolling is done at night.

Problems with development tools were also present during the project. The major factors were getting the development environment correctly setup, which meant getting the correct Unity version and all the necessary dependencies according to Microsoft's HoloLens Guide (Zeller, 2018). For instance, it was very important to have the correct SDK for a certain Unity version or else it wouldn't compile. This is something we weren't completely aware of and working on a shared workstation meant that people would update Unity and add dependencies that fit their project, which meant that each time we returned to the workstation our project won't compile if someone updated Unity.

Even though we managed to solve and workaround many of the problems we faced during the development of the prototype, there were still some major issues that we couldn't figure out or didn't have time to investigate. Problems that could either be software or hardware related. The previously mentioned problem with loss of tracking might be possible to fix or at least minimize the occurrence by using different tools, or make your own code for scanning a room. Another problem is that the prototype runs slow and voice commands are therefore delayed. In some cases this leads to the user saying the voice command again triggering it twice. We therefore added a command "back" to go to the previous task. If this problem is due to too many voice commands, too many holograms, too many images to recognize, or slow hardware is not known. In reality it's probably a combination of everything listed, meaning that improvements can be made but the problem won't go away entirely until new hardware comes out. An option here would be to use the Meta 2 we briefly mentioned earlier. It's tethered to a PC and uses the PC's hardware for processing. This would give a much faster prototype but add the limitation of being forced to carry a computer.

When analysing the interview we noticed a problem that we haven't thought of before. Only two participants actually complained about it, but during the testing, most of the participants could not get the voice recognition to work. Some of them could not get it to work at all. This is highly dependent on the users accent. However, this was not explicitly tested since a user can perform all tasks without using voice commands. We believe this is the reason why so few mentioned the problem even though it was a problem for most.

Another point that was mentioned by the participants was that 8 out 20 participants thought that the headset was uncomfortable to wear. They thought it was heavy, bad weight displacement and didn't work well with glasses. On the contrary this was the first time most of the users tried on the headset and might be something a person could get used to after longer usage. Another problem that was mentioned in 7 test is that the text was hard to read. This might be linked to the fact that it was first time users. It might be that the headset is not properly on as this can lead to blurry holograms, or that we are using to small texts in our prototype. It would seem as its the first case since none of the users complained about the text size.

7 Future prototype versions

The Hi-Fi prototype as of today has focused on three main features, which has been described throughout this paper. These features being:

1. Navigation through waypoints
2. A checklist system that is shown in full detail on real life elements such as doors as well as in a scaled back form on the users arm
3. A system for showing the keys and key-codes for different locks

If more time were to be given to this project there is a whole lot of features we would want to see implemented in order to make the application more comprehensive, and viable for real life situations. In the section " 3.1 brainstorming" we described an array of different idéas that were possible implementations if time allowed it, and it was feasible in a technical sense. Though, in order to not get too attached to old idéas, we instead list what we believe the current problem areas of the prototype is, as to promote future brainstorming sessions around these areas. The following subsections describe problem areas we would like future versions to solve in descending magnitude of priority, the first one having the highest priority.

7.1 Users missing objectives and tasks

In the prototypes current state there is no visual indicator specifically showing the user where a task is to be performed, other than the waypoints leading up to said task. As a consequence, users may miss important objectives and tasks. There are plenty of solutions to this problem, one may be to alter the colour and size of the waypoint next to the objective so that the user know this position is of importance. Another solution could be to highlight the objective itself with a hologram, which could also be used in combination with the aforementioned example. This has a high priority due to how essential it is for a guard to be able to consistently solve tasks without major problems in order to not lose time or concentration. Furthermore we see this as a problem that would be easy to fix, and fast to implement, as it is just an extension of an already implemented feature.

7.2 Safety and deviation features

In the development of the current prototype we unfortunately had to forgo implementing features that would be essential for the real life work environments of a guard. Guard safety and deviation detection are problems we chose not to implement but which are very important. More specifically the problems include for example not being able to see where general but important objects are, such as fire extinguishers or first-aid kits, which in a future version might be highlighted and shown through walls. Furthermore we see a problem regarding the guards personal safety, if the user is being attacked or is otherwise feeling unsafe there should be a feature to alleviate this as much as possible, for example there could be an emergency or panic mode where the headset might start recording both video and audio, automatically connect the guard with a headquarter, and track the location of the guard. Regarding communication between the guard and a headquarter - not being able to connect with a central headquarters and communicate with them is also a problem, for example in order to get help or give informative reports. Lastly there is the problem of deviation detection, meaning we would like a feature that makes it easy for a guard to find things that are out of the ordinary so that this could be checked out.

7.3 Keeping track of time

Currently, if a user wants to keep track of the time they will have to look at a separate unit, such as a smartphone. This is something we instead would like to have implemented in the prototype, as looking at another unit will lower the users situation awareness. Though, just adding a clock that is permanently hovering in the users field of view will most definitely be distracting and would therefore defeat the purpose. Therefore an idea could be to show the time in combination with the checklist. Furthermore some sort of progress bar could also be implemented to be used in combination with the clock in order to improve time management. This is a concept that was nearly implemented in the current version of the prototype, but got postponed due to time restrictions. Therefore the concept art is already done, as can be seen in Fig. 7, which would make the actual implementation of a feature like this easier.

7.4 Different users/use of features

The last problem is not really a problem as such, but rather an improvement to the users experience. In order to create a better user experience, we would have ideally liked if there was some sort of customization option, where for example a user could specify which features is active as well as what colour or how visible certain UI elements are. An experienced guard might feel like some features are distracting, an inexperienced guard might want as much help as possible, and a colour blind guard might want to change the color of certain UI elements. This could be developed even further with the inclusion of “modes”, for example a driving mode and a work mode, where a certain set of customization options could be turned into a mode. Further, profiles could be added in order to create an even more personalized experience. In the personal profile all your current customization settings and modes could be stored.

8 Conclusion

With all the problems discussed, we’ve concluded that the hardware is not ready to be deployed to patrolling guards. Despite this we managed to implement our three stated prioritised functions that is guiding, checklist as well as key and code assistance. However, the prototype can be improved to demonstrate the concept better as well as making it more viable for a real world application by adding new features, such as safety and deviation features as well as customization options, and by making improvements to the currently existing features.

The results from our user tests suggests that guiding the user with waypoints instead of arrows enhances the user´s performance. Thus, we suggest that waypoints should be used as a means of navigation in future augmented reality guiding applications.

Acknowledgements

We would like to thank our supervisors Günther Alce and Diederick Niehorster, Peter Andreasson and Manne Engström at Blue Mobile Systems and Axis Communications.

References:

Alce, G., Roszko, M., Edlund, H., Olsson, S., Svedberg, J., & Wallergard, M. (2017, October). [POSTER] AR as a User Interface for The Internet of Things—Comparing Three Interaction Models. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)* (pp. 81-86). IEEE.

Billinghurst, M., & Starner, T. (1999). Wearable devices: new ways to manage information. Computer, 32(1), 57-64.

Clark, A. (2014). *Mindware: An introduction to the philosophy of cognitive science*. Oxford University Press.

Endsley, M. (1996). Situation awareness measurement in test and evaluation. *Handbook of human factors testing and evaluation*, 159-180.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Kreylos, O. (August 2015). “Hololens and Field of View in Augmented Reality” (PDF). Retrieved from http://doc-ok.org/?p=1274

Liu, F., & Seipel, S. (2018). Precision study on augmented reality-based visual guidance for facility management tasks. Automation in Construction, 90, 79-90.

Lukosch, S., Lukosch, H., Datcu, D., & Cidota, M. (2015). Providing information on the spot: Using augmented reality for situational awareness in the security domain. Computer Supported Cooperative Work (CSCW), 24(6).

McKendrick, R., Parasuraman, R., Murtza, R., Formwalt, A., Baccus, W., Paczynski, M., & Ayaz, H. (2016). *Into the wild: neuroergonomic differentiation of hand-held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy.* Frontiers in human neuroscience, 10, 216.

Microsoft. (2016). Microsoft Hololens webpage. Retrieved from https://www.microsoft.com/microsoft-hololens

Norman, D. (1988). *The Design of Everyday Things*, Doubleday, New York. ISBN 0465067107

Sharek, D. (2009). NASA-TLX Online Tool (Version 0.06) [Internet Application]. Research Triangle, NC. Retrieved from http://www.nasatlx.com

Odom, J. (2017). *What's the difference between Hololens, Meta 2 & Magic Leap?* Retrieved from https://next.reality.news/news/whats-difference-between-hololens-meta-2-magic-leap-0181804/

Zeller, M. (2018). *Install the tools*. Retrieved from https://docs.microsoft.com/en-us/windows/mixed-reality/install-the-tools

Appendix:

Appendix 1: Manus användartester

Hej,
Tack för att du vill ställa upp i detta användartestet.
Vi vill att du föreställer dig att du är en väktare som ska göra din rond inne i denna "byggnaden".

Du kommer nu att få ta på dig en Microsoft HoloLens. Du kommer också att få ett nyckelkort som du kan använda för att ta dig in genom dörren här med en kod. Du kommer även att få en *gadget* i form av ett papper som sitter fast på din arm som du kan titta på för att se din aktuella uppgift.
Genom att titta på symbolen vid dörren så får du upp en lista på alla uppgifter du har under testet samt en kod till dörren som är kopplad till nyckelkortet.

På din arm har du även en symbol som du kan titta på för att få upp en checklista där din aktuella uppgift står.
Genom att säga röstkommandot "next" så kan du markera att en uppgift är klar och på så sätt få nästa uppgift på din arm. För att gå tillbaka till föregående uppgift kan du använda dig av röstkommandot "back". När du är klar med alla uppgifter kommer det att stå "tasks completed", när detta sker är testet klart och vi ber dig då att komma ut till bordet och svara på en kort enkät.

Din uppgift är att följa vägbeskrivningen samt fullfölja de uppgifter som du får via symbolen på dörren.

Va uppmärksam på omgivningen, det kommer att finnas ett visst antal trekantiga varningssymboler i pappersform (visa) någonstans i din omgivning under testets gång. Vi vill att du kommer ihåg hur många av dessa varningssymboler du ser. Vi kommer att be dig att anteckna hur många varningssymboler du såg, en gång efter första testet och en gång efter det andra testet.

När du har genomfört uppgiften kommer du att få göra en liknande uppgift en gång till med skillnaden att vägvisningen kommer att se annorlunda ut.
När båda testomgångarna är klara kommer du att få svara på en kort enkät samt intervjufrågor.

Hela testat uppskattas ta ca 30 minuter. Har du några frågor?

Appendix 2: A.E.G.I.S Holographic loggan

Three variations of the Aegis holographic logotype

Appendix 3: Intervjufrågor

1. Upplevde du några problem med hårdvaran? D.v.s satt obekvämt på huvudet o.s.v?

2. Upplevde du några problem med mjukvaran? Buggar? Tappade tracking o.s.v?

3. Tycker du att hologrammen var distraherande?

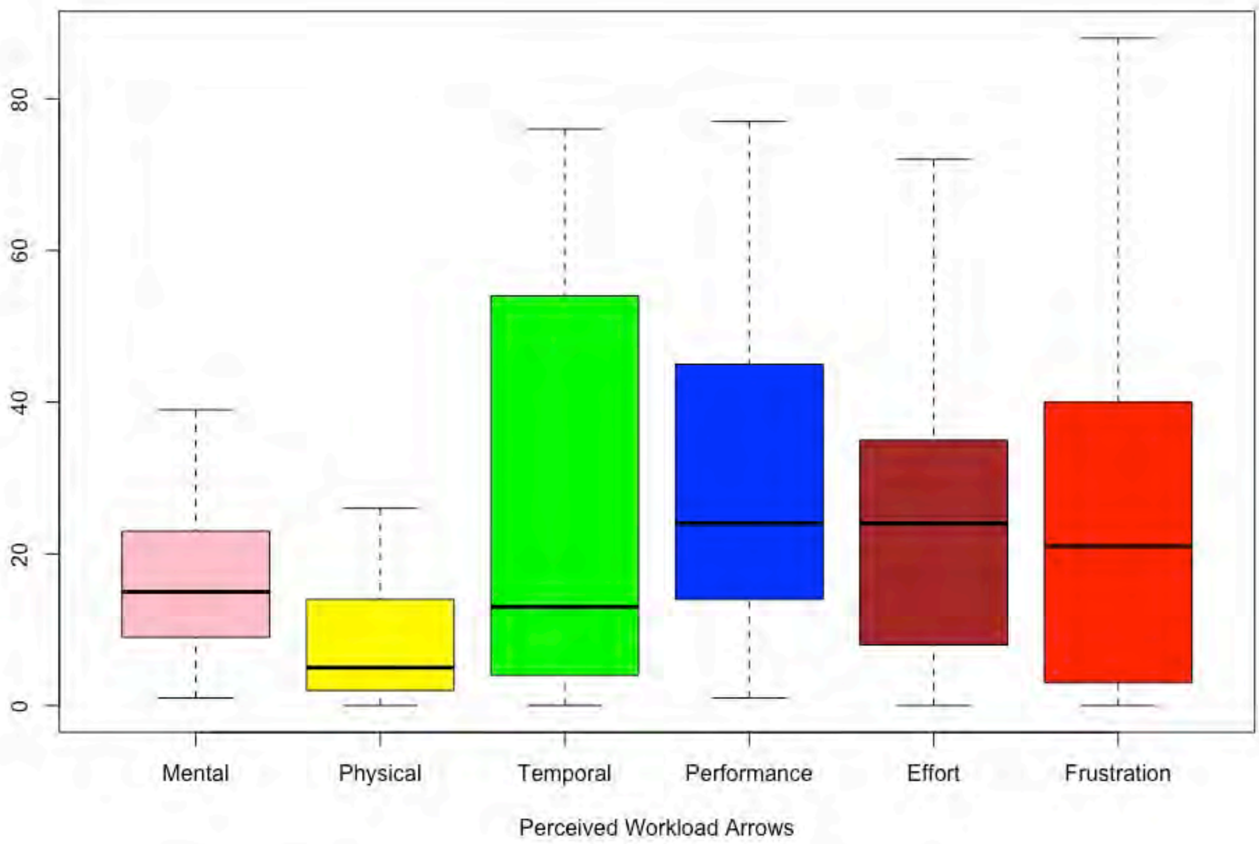
4. Förstod du uppgifterna som du fick av UI:n?

5. Förstod du instruktionerna från testledarna?

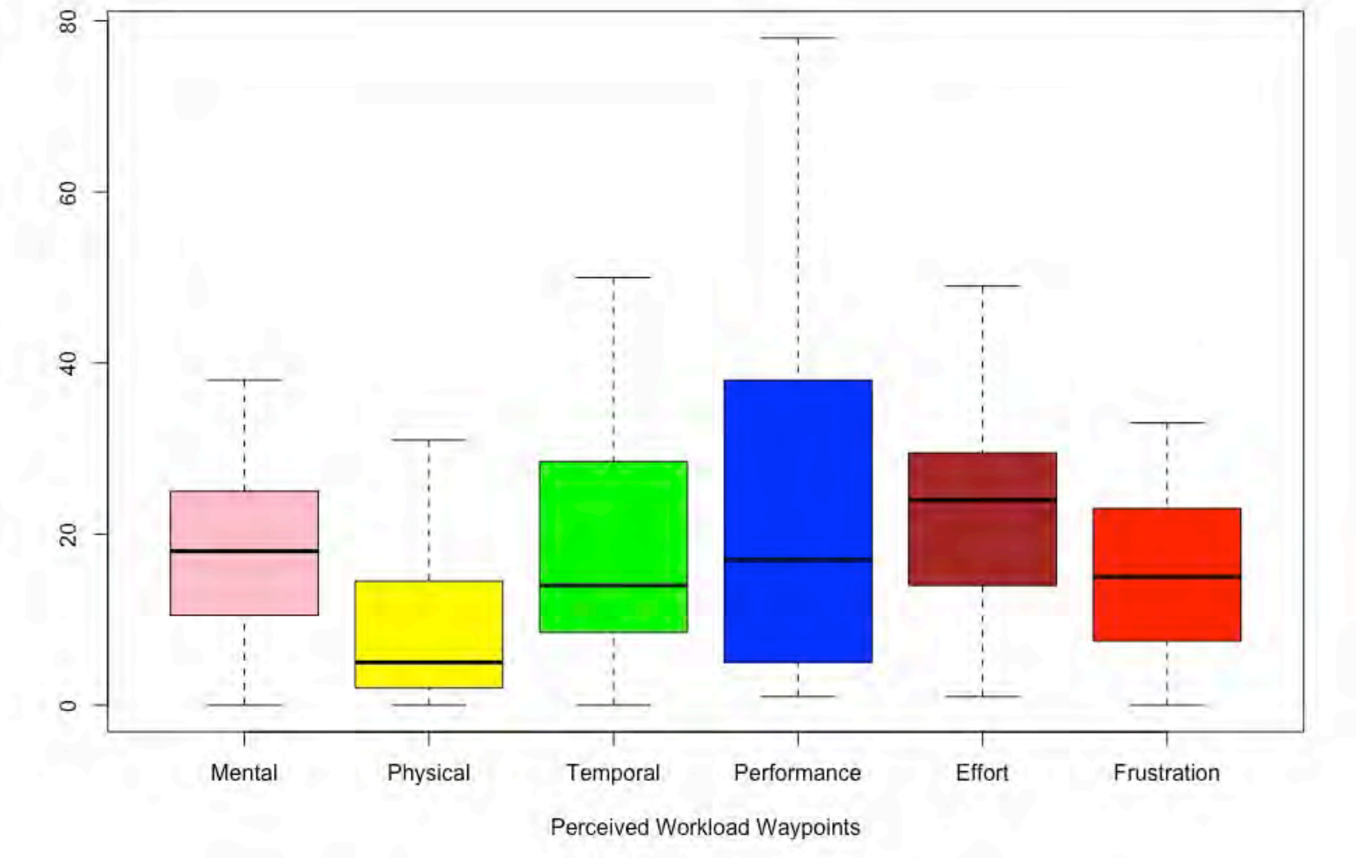
6. Föredrog du navigation via pilar eller waypoints? Var vänlig förklara varför.

7. Har du några andra kommentarer om prototypen som du vill tillägga?

Appendix 4: Perceived Workload Arrows



Appendix 5: Perceived Workload Waypoints



EnVRment

Axel Mulder, Felicia Carlsson, Jakob Håkansson,
Johannes Törnblom, Magnus Johansson

MAMN10/15 - 2018-12-23

The project EnVRment attempts to educate youths in the ages 10-15 in how one can counteract the environmental problems society is facing, in a fun and creative way - by utilizing the Virtual Reality-medium.

*EnVRment consists of snippets of everyday life where one can affect and help with the environmental issues on a personal level. The product contains different scenes; two **introductory rooms** explaining different interactions within the game, a **kitchen** where two mini-games are available for the user, a **hallway** connected to the kitchen, a **parking lot** outside the previous apartment, a **schoolyard**, and a **classroom**. Following the mini-games are different **"consequence scenes"** that are shown to the user, depending on their performance, displays the consequences caused by performing/not performing a certain environmentalfriendly action (like recycklig, saving energy and water or choosing the better option for transportation).*

All in all there are four different scenes where the user interacts with the environment (which can be seen as the yellow circles in Figure 1). These are:

- ***Trash Game**: The task of the user is to recycle different materials into different containers, gaining points displayed on a large scoreboard.*
- ***Energy Game**: The user has to turn off different kitchen appliances to save their battery (energy), trying to survive for 90 seconds.*
- ***Transport**: The user has a choice between riding a bike or traveling by car to get to school - and then experiencing the resulting air quality depending on their choice.*
- ***Classroom**: Three different small-scale models of environments are placed on tables in the classroom - where the user is tasked with constructing and learning about different ways of generating sustainable energy.*

This report also details many of the ideas behind the design choices and why or why not these have been made, accounting for both pre-existing theory about both learning and Virtual Reality, as well as describing specific testing conducted by the project group to establish a final product of good quality.

1 Produktbeskrivning

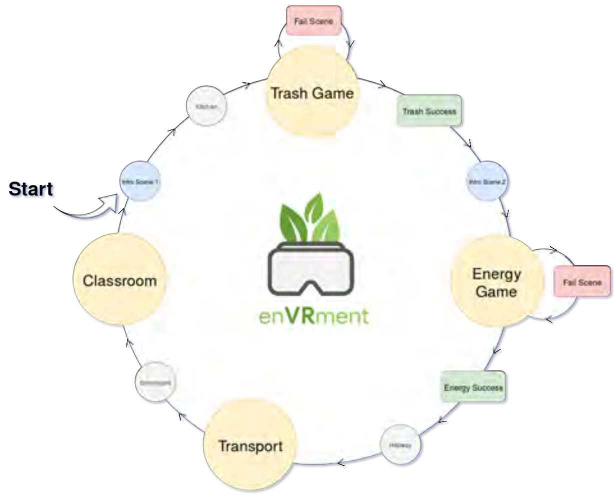


Figure 1: En förenklad visualisering av flödet i EnVRment

2 Bakgrund

Syfte

Projektet EnVRment har sina rötter i den nationella satsningen "Wisdome", som syftar till att skapa en mötesplats för visualiseringar, vetenskapskommunikation och interaktivt lärande. EnVRment-gruppen fick i uppgift att skapa en prototyp för en VR upplevelse, med målet att ge förslag på vad man skulle kunna ställa ut på Wisdome i Malmö.

Nyckelbegreppen i uppgiften var:

- Hållbar utveckling och miljö.
- Integration - då målgruppen har blandad etisk och socioekonomisk bakgrund.
- Inspiration - till att studera en teknisk utbildning i framtiden.

Ytterligare direktiv var att målgruppen främst skulle innefatta skolungdomar i åldrarna 10-15 och att projektet skulle utformas inom Virtual Reality (VR).

Teori

En bra återkoppling är av största vikt i design av alla system som syftar till att interagera med en mänsklig användare. I virtuella lärmedel är särskilt *typen av* återkoppling en viktig aspekt.

Betty Tärning gör en ingående analys av olika alternativ[1], och framhäver särskilt *corrective feedback (cf)* och *elaborated feedback (elf)*. Den förra innebär att man visar *vad man borde ha gjort istället* (när man t.ex. svarat 3 på frågan 1+1=?), medan den senare visar *varför* man borde ha gjort så (dvs hur de aritmetiska reglerna ser ut). *Encouraging feedback (enf)*, är ett medel som ofta används i spelsituationer, där användaren (genom t.ex. ett grafiskt ljusfyrverkeri som följer på ett korrekt svar) uppmuntras att fortsätta spela. Andra, mindre lyckade alternativ är *verification feedback (vf)*, som bara bekräftar om man gjort rätt eller fel, och *trial-and-error feedback (tef)*, som i fortsättningen planlöst bara låter användaren *testa* sig fram till det korrekta svaret. Både *vf* och *tef* behöver kompletteras med t.ex. *cf* för att inte rentav ha *negativa* inlärningseffekter: den *förståelse* som är nödvändig för inlärning, dvs steget från vad L. Vygotsky kallar den aktuella nivån till den potentiella, kan inte uppnås utan explicit vägledning. *Efter att denna givits* kommer eleven att oberoende kunna utvärdera sin prestation.

I vanliga spelscener översköljs användaren av möten med andra agenter (djur eller människor). Ett empiriskt skäl att faktiskt låta bli detta, åtminstone inte då ett huvudsyfte är *inlärning*, är att sådana, om de i själva verket uppfattas som avatarer (dvs kontrollerade av en människa), tenderar att kunna hämma användarens prestation. Att t.ex. observeras av en människa medan man försöker lösa en viss uppgift, innebär helt enkelt en större press[3]. Dessutom finns det studier som visar på att prestationen i just tävlingssituationer påverkas allra mest av närvaron av andra varelser, agenter eller avatarer[3]. I *förlängningen* talar dock forskningen för att man som designer, åtminstone i lärsituationer, genomgående bör sträva efter att den varelse användaren interagerar med ska uppfattas som *mänsklig*.

En bra *immersion* innebär att en miljö uppfattas som verklig, medan *presence* har att göra med hur väl en användare kan ”leva sig in” och *orientera sig i* situationen. I artikeln ”How Immersive is Enough” slår Cummings och Bailenson fast att ljud- och bildkvaliteten spelar roll för hur verklig omgivningen uppfattas som, men man påpekar också att just immersion inte är avgörande för vilken närvaro (presence) användaren upplever[4]. Därtill verkar det som att det är kvaliteten på hårdvaran (hur precis trackingen är, om synfältet är binokulärt och om synvinklarna i höjd- och sidled är verklighetstrogna) som är av störst vikt för vilken presence en VR-upplevelse kan erbjuda.

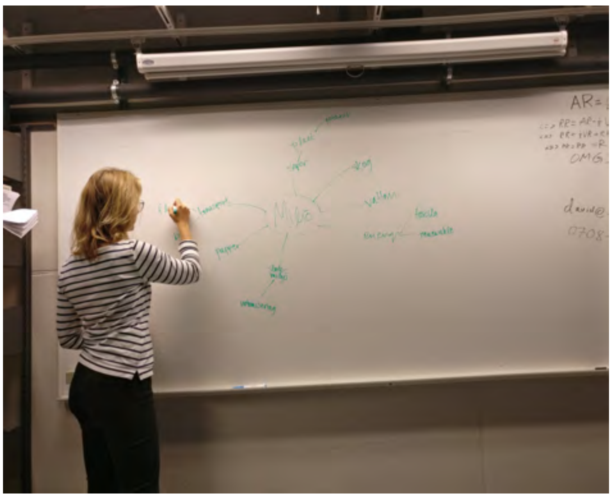


Figure 2: En tankekarta med fokus på Miljö

3 Metod

I detta avsnitt beskrivs hur arbetet har framskridit i projektet i form av dels designval och designprocess, dels hur det praktiska genomförandet sett ut och hur tester har gått till.

Designprocess

Startpunkten för arbetsprocessen var en brainstorming-session där tankekarta (se Figur 2) med fokus *Miljö* konstruerades. Idéerna som tog sin form här omarbetades och mynnade slutligen ut i tre koncept som presenterades för handledarna:

- **”A Day in your Life”:** Användaren får följa med på en resa genom en 10-15 årings dag och får möjligheten att uppleva vardagliga händelser, där man ställs inför val som är bra eller dåliga för miljön. Idén har tagit inspiration av Världsnaturfondens fem fokusområden som handlar om vad en vanlig person kan göra för klimatet [9].
- **Sopsorteringsspel:** Ett spel där användaren får sortera och återvinna olika typer av sopor. Spelet går på tid och ger plus-/minus-poäng när man sorterar rätt/fel.
- **Plasts kretslopp i vatten:** Användaren får följa med en mikroplastpartikel genom dess kretslopp, från och med att det hamnar i havet, och hur det påverkar olika organismer på dess väg.

De olika idéerna utvärderades utifrån deras respektive för- och nackdelar, och **”A day in your life”** bedömdes sticka ut som ett särskilt lovande koncept. Här sågs möjligheter att kunna visa individen ”vad just *jag* kan göra för miljön”. Det faller sig naturligt att en sådan vinkel är mer greppbar och enklare att relatera till, än att fokusera på åtgärder som berör samhället i sin storhet (som att minska mängden flygtransporter eller styra om elförsörjning från fossila till förnyelsebara kraftkällor).

Genom att peka på konkreta exempel på saker som ett barn kan påverka för att bekämpa miljö- och klimatproblem, fanns en förhoppning om att förändra dessas vardagliga attityder och handlingar mot att bli mer miljö- och klimatsmarta.

Konceptet **”Sopsorteringsspelet”** valdes först bort, på grund av en potentiell risk att användaren skulle bli distraherad från budskapet om miljö- och hållbarhet, utan istället lägga sitt fokus på själva spelmomentet. Denna nackdel kunde dock även vändas till en fördel, då en speladaption av sopsortering – ett ämne som kan uppfattas som ett tråkigt av gemene man - kunde verka intresseväckande och göra uppgiften roligare att utföra. Ur integrationssynpunkt kunde det vara en god introduktion till sopsortering för någon som aldrig har prövat på det tidigare - exempelvis den som är nyanländ från ett land där man möjligtvis inte har ett lika utbyggt återvinningssystem som vi har i Sverige. Därför valde man att inkorporera detta koncept som en subdel av ”A day in your life”.

Som handledarna på projektet påpekade under det första mötet, så har VR fördelen att man kan transportera användaren till platser som annars inte hade varit möjliga att uppleva. Det finns studier, genomförda på VHIL (Virtual Human Interaction Lab) vid Stanford University, som visat att en upplevelse i VR kan påverka en användares inställning till olika frågor i högre grad än exempelvis ett budskap förmedlat i textform

[6]. I konceptet **”Plasts kretslopp i vatten”** utnyttjades denna möjlighet på ett intressant vis. Upplevelsen skulle här kunna gestalta faktiska konsekvenser av människans konsumtion av plast för våra hav. Dock fann projektgruppen en problematik i att rent praktiskt göra upplevelsen interaktiv, vilket samtliga gruppmedlemmar var eniga om var en önskvärd egenskap hos produkten. Valet föll till slut på att inte gå vidare med denna idé, men tankarna kring att visa konsekvenser för miljön som följd av människors handlingar fick leva kvar - i form av konsekvensscenerna.

Huruvida språk i form av tal och/eller text skulle inkluderas som en del av VR-upplevelsen var ett diskussionsämne som fortgick under hela kursens gång. Fördelen med tal- och skrift är att de möjliggör förmedling av mer djupgående och komplexa koncept, än vad exempelvis bilder och symboler gör. Det fanns även en idé om att kunna använda VR-upplevelsen som ett verktyg för att lära sig svenska för nyanlända barn (och vuxna), exempelvis genom att förklara vad olika sorteringsmaterial kallas på svenska. Men detta ansågs vara utanför de tidigare satta avgränsningarna för projektet och avfärdades därmed. Att använda tal- och skrift som medium valdes även bort i slutändan, på grund av otillräckliga resurser.

Att målgruppen förväntas tala olika språk skulle nämligen leda till att skrift och tal inte bara skulle behöva finnas tillgängligt på svenska, utan även på en rad andra språk (främst t.ex. arabiska, kinesiska, spanska och engelska). Projektgruppen ansåg sig sakna kunskaper och kontakter för att samla in den mängd material (i form av text och tal på tillräckligt många språk) som krävts för att hela målgruppen skulle täckas in. Istället valdes bilder, ljud, gester och andra typer av effekter till huvudsakligt medium för att förmedla koncept och kunskap. Målet är att budskapet i EnVRmen ska vara universellt begripligt, det vill säga att det ska gå att förstå oavsett vilket/vilka språk användaren talar och vilken kulturell bakgrund hen har.

Nästa steg var att konkretisera de designval som dittills gjorts. Första scenen som användaren möts av förlades till hemmet, eller närmare bestämt köket. Köket sägs ofta vara hemmets hjärta och är en välbekant miljö för en 10-15-åring: det är här en vardag kan tänkas starta, för att sedan fortsätta vidare med någon form av transport till skolan och en dag i klassrummet. Dessutom sker sopsorteringen i ett hem som oftast i köket, där en stor del av hushållssoporna blir till (i samband med matlagning och dylikt). Sorteringskärlen befinner sig även vanligen i skåpet under diskhon, enligt projektmedlemmarnas erfarenheter. (Hur en användare skulle överföras till sopsorteringsspelet valdes därför just till att denne ska öppna köksluckorna under diskhon i ett senare skede av processen.) Kopplingen mellan kök och sopsortering var anledningen till att sopsorteringsspelet valdes till scen nummer två.

Men hur skulle den resterande delen av upplevelsen i VR organiseras? För att utforska möjliga metoder användes flödesschemat i Figur 3 som stöd i gruppdiskussioner. Schemat visar vilka scener en användare kan mötas av under upplevelsens gång.

Figure 3: Flödesschema över scener användaren möter

En central fråga som diskuterades under denna session var hur budskapet om varför användaren behöver sopsortera skulle förmedlas. En aspekt som kom på tal var att konsekvenserna av att återvinna sopor beror på vilket material soporna består av. Man valde att avgränsa sig till tre olika material:

- Metall
- Plast
- Papp

Dessa blev de olika material som användaren fick sortera i sopsorteringsspelet. På så vis blev spelet en metafor för att sopsortera i det verkliga livet, men den skillnaden att användarna får möjlighet att i kontrast till verkligheten upplleva direkta konsekvenser av att de personligen inte återvinner tillräckligt av plast, metall eller papp. Här föddes idén om de ovan nämnda s.k. ”konsekvensscenerna”; en per respektive material, med syfte att visualisera miljökonsekvenser av att inte återvinna det materialet.

En mer detaljerad beskrivning över respektive scens designprocess finns samlade under nästkommande subrubriker.

Introscen 1: Lära sig kontrollerna

För att användare skulle kunna navigera och interagera med vårt Virtual Reality-projekt - så behövde vi någon slags utgångspunt för vart en potentiell användares tidigare erfarenheter med Virtual Reality och dess headsets/kontroller, låg.

Efter initial testning med individer som både testat VR förut - och de som aldrig gjort det - beslöt vi oss för att utgå ifrån att användaren **aldrig** använt eller upplevt VR förut. Detta, då de som faktiskt testat/använt VR förut - ej heller direkt greppade kontrollschemat.

Därmed behövde vi ta fram ett enkelt sätt att lära ut hur kontrollerna och interaktionen fungerar - vilket ledde oss till att skapa en separat “introscen”. Det första användaren möts av i vårt projekt - är ett stort vitt rum, med en stor röd dörr med handtag [IMAGE HÄR]. Målet med scenen är att användaren skall lyckas öppna dörren - för att ta sig vidare till vår första “riktiga” scen; /textitköket.

För att kunna dra i handtaget och öppna dörren i scenen - behöver användaren lära sig att greppa objekt i VR. För att göra detta används den s.k. “trigger”-knappen på HTC Vive’s bägge handkontroller - detta för att simulera en hand som sluts runt och greppar ett objekt. Ifall användaren inte tittar ner på sina händer när de har på sig VR-headset:et - så kommer handkontrollerna efter ett par sekunder att vibrera för att tillkalla sig uppmärksamhet.

När användaren tittar ner på handkontrollerna så finns det symboler som pekar på handkontrollernas “trigger”-knapp. När användaren väl hittar knappen och trycker ner den - så visas tydlig feedback i form av att handkontroller-nas modell i spelvärlden, byts ut mot ett par händer. När användaren håller ner “trigger”-knappen på en handkontroll så knyts näven på den korresponderande handen i spelvärlden.

När användaren tagit fram båda händerna - så ges användaren ytterligare vägledning om dess uppdrag i scenen - genom att det börjar höras knack-ljud från dörren. När användaren öppnar dörren genom att greppa handtaget - så tas användaren vidare till nästa scen; *köket*.

Längre in i projekets process - så fick vi önskemål av våra handledare att lägga in ett stöd för val av språk i vår pro-dukt. Därmed så lades det in ett par flaggor (med tillhörande text) i scenen för respektive språk - som blir användarens första uppgift att interagera med. När en flagga och därmed språk har valts - så börjar knack-ljudet att spela som ovan beskrivet - och användaren kan sedan öppna dörren.

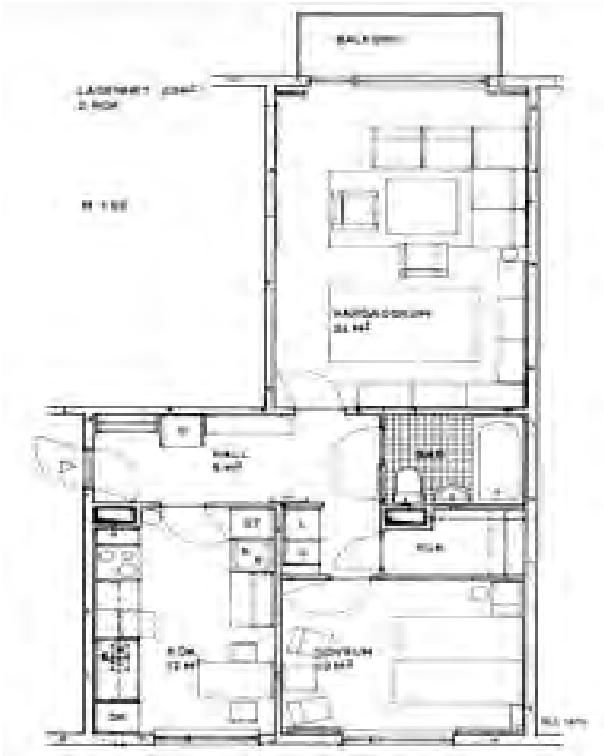


Figure 4: Planlösning miljonprogrammet i Kallhäll [10] där köket är placerat längst ner i vänstra hörnet på ritningen.

Köket

Som designer är det lätt att bygga in sina egna upplevelser och åsikter i det man skapar, inte minst när det kommer till en hemmiljö så välbekant som ett kök. Men då målet med köksscenen var att användaren skulle kunna knyta an till köksmiljön för att förstärka VR-upplevelsens immersion, ansågs det vara av stor vikt att köket hade en utformning som så många användare som möjligt kunde känna igen och relatera till. Att undvika fällan att lägga in element från projektgruppens egna barndomshem eller nuvarande studentlägenhet, som kanske inte alls kändes igen av slutanvändaren, hade därmed hög prioritet i designprocessen.

Därför valde man att ta inspiration från ett kök som en stor del av Sveriges befolkning faktiskt har en relation till, nämligen ett som återfinns i lägenhetshus som uppfördes under miljonprogrammet. Kökets planlösning går att finna i Figur 4, som är hämtad från lägenheter i Kallhäll. En annan bidragande orsak till att denna planlösning valdes, var att m˚anga nyanlända i dagsläget bor i bostäder som liknar miljonprogrammets. Att välja en köksmiljö som särskilt dessa individer skulle känna igen och relatera till, var därför ett sätt att jobba mot målet att VR-upplevelsen skall vara integrerande.

Inredningen i köket är inspirerad av diverse bilder från miljonprogrammens kök (vilka går att finna online). En av dessa bilder visas i Figur 5 och föreställer ett kök som är renoverat under ett upprustningsprojekt mellan 2011 och 2014.

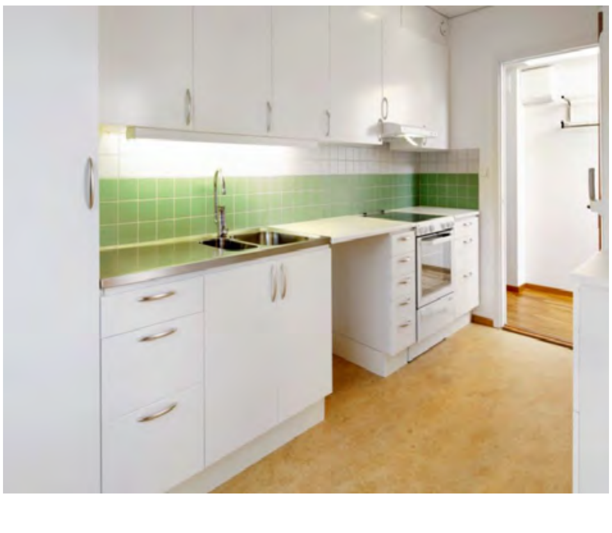


Figure 5: Inspiration till köksinredning [11]

För att göra köket i VR-upplevelsen verklighetstroget så gjordes samtliga skåp och lådor öppningsbara, och alla övriga objekt - med några få undantag - går att interagera med och flytta runt. Utanför fönstret syns en mängd andra höghus av den typ som lägenheten kan tänkas vara inrymd i, och att lägenhetsfönstret är placerat högt över marken är tänkt att förstärka detta intryck. Bakgrundsljuden som hörs i köket är utvalda för att passa in i miljön, med lekande barn utanför fönstret och musik av orientaliskt slag som spelas med låg volym på radion.

Steget som användaren får ta för att ta sig till sopsorteringsspelet sattes som tidigare nämnt till att denne får öppna köksskåpet under diskhon. Två metoder användes för att denna handling skulle ges tydliga *signifiers*[5]; den ena var att använda en färgstark kontur längs med köksluckornas kanter för att få dom att sticka ut bland de andra luckorna. Förhoppningen var att konturen skulle dra till sig uppmärksamhet såatt användaren skulle lockas att undersöka vad som fanns i skåpet. Den andra metoden var att en tavla placerades över köksbordet med en bild över de öppnade skåpsdörrarna, som en bildinstruktion till användaren.

Sopsorteringsspelet

Sopsorteringsspelet utspelar sig också i köket, men rummet i sig ändrar form. Det blir en övergång där själva köksväggen med spis, vask och skåp animeras iväg från spelaren och skapar en lång spelbana, samtidigt som de gränser (i form av spelyta) som HTC Vive begränsar till, behålls. Detta görs då genom att att själva ytan som spelaren kan röra sig på är den samma som innan, d.v.s köksarean, medan spelplanen finns *utanför* rörelseytan. Detta fungerar, eftersom att spelet går ut pa ˚ att kasta objekt ifrån en avgränsad yta, in i ”mål/sopkärl” som är utanför rörelsearean. För att få spelaren att naturligt stanna inom rörelsearean lades en låg avgränsningvägg till framför spelaren. De objekt (sopor) som ska kastas, tillkom-mer endast i själva rörelsearean.

Spelet går på tid och har ett poängsystem. Tiden är satt till 60 sekunder, då beskrivningen av projektet innebar att produkten skulle finnas på ett slags museum (där flertal personer kan tänkas vilja testa produkten) och då bör det hela alltså inte ta för lång tid. Tiden i sig visas på spelbanan som en ”3D-text” i formatet ”mm:ss”, där ”mm” är minuter och ”ss” sekunder som räknar nedåt. För att dra till sig uppmärksamhet inleds scenen med att tiden animeras fram och visas i stort format, innan det minskas till en mer passande storlek för spelets gång. När tiden sedan (under spelet) började ta slut, med 10 sekunder eller mindre kvar, så blir tiden färgad i rött istället för vitt (som varit standarden).

Spelets poängsystem fungerar enligt följande. Spelaren kastar tre olika typer av material i tre olika sopkärl. Materi-alen är viktade olika där metall ger 3 poäng, plast 2 poäng och papp 1 poäng. Poängen summeras och visas för spelaren under tiden i form av en 3D-text med formatet ”X points”, där X är antalet poäng. För att lära spelaren hur man ”kastar rätt” så ger spelet feedback på flertal sätt. När ett objekt prickas i ett sopkärl så undersöks om det var *rätt* typ av objekt. Om ja, spelas ett positivt ”korrekt-ljud” upp, samtidigt som poängen ökas och visas i grön färg över en kort tid. Om nej, spelas ett negativt ”fel-ljud” upp, samtidigt som spelaren bestraffas med en reduktion i poäng och totala poängen nu visas i röd färg.

Objekten; en kartonglåda, en metallburk och en plast-flaska, kom i första iterationen fallande ifrån taket, d.v.s. objekten föll ner mot spelaren och landade sedan på golvet. Väl när de slog i golvet avgav dem ett ljud som speglar deras material (t.e.x en metallbruk som klirrar till metalliskt). Detta för att locka uppmärksamhet till var objektet befinner sig.

Varje typ av objekt kom till i spelet efter en viss timer. Även i den första iterationen så tillkom objekten olika ofta. Papper tillkom oftare än metall och så sätt fanns det betydligt fler papperskartongerän metallburkar i spelet.

Efter testning och feedback kom projektgruppen fram till att detär förvirrande att objekten slängs ner ifrån taket samt att det är jobbigt att plocka upp objekt från marken. För att lösa detta problem ändrades logiken i nästa iteration, genom att en sorts behållare, i vilken objekten skapades, tillfördes. Detta löste b˚ ada problemen i den meningen att man slapp böja sig långt ner till marken eller behöva leta runt efter de olika objekten. I samma iteration löstes problemet att objekten kom olika ofta: nu tillkom ett objekt av varje material *samtidigt*, på en slumpmässig plats i behållaren. En ytterligare funktionalitet har tillförts, så att tiden mellan dåobjekten uppstår ökar linjärt. Detta innebär att det alltså blir fler objekt och mer stressigt ju längre tiden går. Detta lades till för att addera ett element i spelet och tillföra en sorts stressfaktor.

Övergångarna till och från sopsorteringsspelet har även haft en tilltänkt design. När man går från första köksscenen så öppnar man skåpet under vasken, varpå bilden bleknar till vitt för att sedan gå tillbaka från helvitt till sopsorteringsspelet. För att spelaren inte ska känna sig placerad annorlunda så har har scenen byggts upp på samma koordinater. På grund av spelrum så har dock skalan ändrats, genom att vyn bleknar till vitt och sedan tillbaka. Genom att få användaren att ”förlora” sina ut˚ angspunkter helt (då allt är vitt) och sedan föra dem tillbaka till en väldigt liknande scen, så kommer användaren inte att märka att små detaljer (t.ex. storleken hos olika detaljer i köket) modifierats. Samma övergång till konsekvensscenerna gör upplevelsen kontinuerlig, och en fördel är att den är mjuk och inte snabb eller plötslig.

Konsekvensscen: Papp

Den första konsekvensscenen som togs fram var för materialet papp. Inspirationen till denna scen togs till stor del från ett försök på universitetet i Stanford [6] som i korta drag visade pa ˚ att försökspersoner som fått uppleva ett kalhygge där träd avverkades i VR, i större grad blev påverkade att bry sig om pappersförbrukning än de försökspersoner som fick läsa en skriven text om skogsskövling. En liknande effekt var önskvärd hos användarna av EnVRment, och därför valdes ett kalhygge som bakgrund för denna konsekvensscen.

För att förhöja känslan av att befinna sig på en arbet-splats där träd avverkas, har ljudet av en motorsåg lagts in, vilket följs av att ett träd faller med tillhörande ljud. Sålångt saknades dock en koppling mellan att 1) användaren inte sopsorterat papp tillräckligt väl i det föregående spelet, och att man 2) sågade ner träd i skogen. Den första tanken var att detta budskap skulle förmedlas genom olika bilder som visades på en skylt - men bilder som bara dök upp ur tomma luften skulle sakna kontext, och härmed sågs en risk att sambandet skulle bli svårbegriplig.

Här föddes idén om att införa agenter (som nämns i Teori-avsnittet ovan), som med sin närvaro skulle tillföra kontext till historien om återvinning av exempelvis papp och skogsskövling. Ett djur som blir påverkat av skogsskövling är ekorrar (som bor i träd), och ett exemplar av dessa valdes ut till ”offer” i en historia där denna, på grund av att papp-kartongerna i spelet inte återvinns, blivit av med sitt hem och nu är väldigt ledsen (se bilder i Appendix C). Då människor ofta tillskriver djur känslor sågs potential i att låta ekorren få framkalla sympati hos användaren. Förhoppningen var att det skulle lägga till ytterligare en dimension till budskapet att˚ atervinning är viktigt. För att understryka att det är ekorrens historia som berättas, placerades den på en trästubbe i höjd med skylten.

Då VR till sin natur ger användarupplevelse en 360-graders vy, så krävs medel för att rikta användarens uppmärksamhet så att inte ekorrens historieberättande missas. Detta särskilt då bilder istället för ljud används som medium. Ekorren ger därför ifrån sig gälla och pipande ljud, vilka, för att tillkalla sig användarens fokus, hörs från just den position den befinner sig på. Ekorren har även animerats så att den rör sig med småhopp, som att den vore arg och ledsen.

Konsekvensscen: Plast

På senare tid har flera bilder på stora plastansamlingar i haven visats i medier, och på grund av dessa valdes dessa ut till att bli en del av denna konsekvensscen. Skräpet bildar öar som innebär både faror och försämrade livsvillkor för fåglar.

Agenten i scenen är en fiskmås, som stående på en av plastöarna berättar sin historia, och runtomkring den flyter b˚ ade plast och lik av djur. Ljuden som hörs är av skvalpande v˚ agor, plastobjekt som rör sig mot varandra och dämpade m˚ asskrin, allt för att göra situationens immersion så stark som möjligt.

Konsekvensscen: Metall

Den sista konsekvensscenen till metall var svårare att ge ett lika engagerande och tydligt koncept som de andra. Till en början övervägdes en gruva, i vilken användaren under en berg- och dalbanefärd skulle få se både den mörka gruvmiljön men samtidigt därtill miljökonsekvenserna. Efter att ha konfererat tillsammans och med handledarna beslöts dock att en soptipp vore bättre för att visualisera hur elektronik och utsläppta kemikalier påverkar miljön.

Agenten i skissen (som visas i Figur 6) är en strykarhund, men inför modelleringen ansågs att denna inte riktigt hörde hemma på skrottippen. Då inget annat djur heller kändes riktigt passande valdes istället en robot. Denna visar sitt budskap: att kemikalier från elektronik läcker ut i grundvatten och därmed dödar fiskar. Gester och känslouttryck har prioriterats och ägnats med tid för denna agent än de båda tidigare, inte minst för att få den att framstå som så mänsklig som möjligt.

Figure 6: Konceptuell bild för en konsekvensscen

Konsekvensscen: Användaren lyckas med sopsorteringen

Med alla konsekvensscener på plats - gällande ifall användaren misslyckades med sopsorteringsuppdraget - så var det även viktigt att det konstruerades en konsekvensscen som speglar ett positivt utfall av sopsorterandet. Vi hade många diskussioner inom projektgruppen om positiva och negativa konsekvenser i samband med lärande. Efter ytterligare diskussioner med våra handledare så valde vi att fortsätta på sp˚ aret att försöka kombinera de två; att både exponera användaren för negativa såväl som positiva konsekvenser gällande dess handlingar i hopp om att finna en balans mellan teorierna kring korrekt utlärande (se teori-delen ovan).

När användaren lyckats med sopsorteringsspelet (vilket efter mycket testning på våra målgrupper samt inom gruppen - valdes till att 4 stycken objekt av varje typ, skulle som minst sopsorteras) - så möts denne av en ljus, vacker och grönskande skogsglänta [IMAGE HERE]. Då de andra “negativa” konsekvensscenerna utspelade sig i miljöer med tillbakahållna färgpaletter och ljussättning - så var förhoppningen att detta skulle innebära en stor förändring mot vad användaren varit van vid - och på så sätt kommunicera budskapet i scenen ännu tydligare. Budskapet var i detta fallet enkelt att förmedla: “genom din goda prestation med att sopsortera - resulterar det i att vi har en natur och miljö som mår bättre”.

Vidare, valdes det att göra en stark återkoppling till de agenter som använts i övriga “negativa” konsekvensscener. När användaren möts av denna ovan beskrivna “positiva” konsekvensscen - så finns samtliga agenter på plats - och de uppvisar stor tacksamhet gentemot användaren och dess insatser (illustrerat genom uppspelandet av varje agents eget ljud och sprudlande hjärtan som emanerar från de olika agenterna).

Till sist så skapades det även ett plakat som agenterna var placerade runtom - som tydligt visar användaren hur mycket av varje typ av objekt som denne lyckats sopsortera. På så vis ges det även incitament till användaren (och dess vänner som kanske står och tittar på, medans användaren spelar) att vid ett återspel av sopsorteringen - att göra ännu bättre ifrån sig.

Introscen 2: Lära sig nya kontroller

När användaren lyckats med sopsorteringen - så tas användaren vidare till samma introscen som denne stötte på innan denne kom till köksscenen. I nästkommande del av upplevelsen behövde nämligen användaren lära sig ett nytt interaktionssätt.

Detta interaktionssätt involverade manipulation av spakar/reglage i scenen. Därmed möts nu användaren av en spak som skall dras i. När användaren drar i spaken - så dras ett galler framför en dörr upp i enlighet med hur långt användaren dragit spaken. Mappningen mellan spaken och gallret gör att användaren tydligt får en förståelse för att det är viktigt *hur långt* en spak eller reglage dras.

När användaren dragit i spaken tillräckligt långt - så dras gallret upp helt ovanför dörren - och användaren transporteras vidare till nästa scen.

Energispelet

Som tidigare nämnt, var köket som spelaren först möttes av - något stökig och dovt upplyst. Efter att ha lyckats med sopsorteringsspelet möts nu spelaren av ett rent, ljust och uppstädat kök - i egenskap av feedback för detta. Efter en liten stund så spelas det upp ett uppmärksammande ljud - och ett batteri med tillhörande klocka fälls upp ur köksbänken bredvid användaren.

Batteriet har tydliga indikatorer för batterinivån - och bredvid batteriet finns det en klocka med timer, där sekunder visas i stort. Efter att ha fällts upp ur köksbänken så börjar klockan bredvid batteriet att ticka ner. Det börjar nu spelas upp indikationsljud runt om i lägenhet - för att uppmärk-samma användaren på att saker försiggår.

Uppdraget för användaren är nu att spara på energi genom att stänga av diverse objekt runt om i köket som drar mycket ström/energi (genom att dra i/trycka på objektets korresponderande spak/reglage/knapp). I enlighet med våra övergripande designs kring köket och projektet - så är de objekt som sätts igång - objekt som vardagligen kan lämnas påsatta av misstag i ett vanligt kök; spisen, elementet, droppande från kranen, öppen kylskåpsdörr och lampor. Allt - för att ännu en gång - göra en stark återkoppling till vad en vanlig ungdom inom vår målgrupp kan göra i sin vardag för att förbättra miljön och hållbarhet.

När ett objekt är aktivt - så spelas det upp ett specifikt ljud för detta objekt; t.ex. hörs det ett droppande från kranen ifall denna är aktiv. Beroende på vilket objekt det är -så kommer även materialet eller utseendet på objektet att förändras för att indikera aktivitet. T.ex. så lyser lampskärmen upp ifall lampan är aktiv - och elementet lyser en varm, röd ifall det är aktivt. Vidare, när ett objekt är aktivt - så kommer batteriets batteriniva ˚ att sakta falla - vilket indikeras tydligt genom användning av färgade "bars" (som går från grönt till rött). Klockan bredvid batteriet kommer även ticka ner - och användaren blir varse om hur länge de måste se till att "hålla batteriet vid liv".

När tiden på klockan når 0, och det fortfarande finns batteriniva ˚ kvar - så har användaren lyckats. Omvänt så ifall batterinivån når 0, när det fortfarande finns tid kvar på klockan - så har användaren misslyckats.

Konsekvensscen: Användaren misslyckas med energispelet

För att upprätthålla en kontinuitet inom projektets lärande - beslöts det att även i fallet av vårt andra "spel" i vår produkt - att använda oss av konsekvensscener beroende på hur användaren presterat.

Ifall användaren misslyckades med energispelet så kommer dess syn att sakta svartna - för att sedan mötas av en stor, mörk, smogfylld scen. Mitt i scenen, framför användaren, tornar ett stort kolkraftverk upp - vars stora skorstenar spyr ut tjock, kolfylld smog. Ett allt större led av vagnar med kol kan synas rulla in i kolkraftverket. Även framför användaren, så syns de olika objekten som skulle stängas av, under energispelet. Från varje objekt, så leds det en tjock elkabel som går till det stora kolkraftverket. En efter en så tänds elkablarna mellan objekten och kolkraftverken - och objekten blir aktiva. Efter en stund svartnar det för ögonen på användaren igen - och man återvänder till energispelet för att ge det hela ett nytt försök.

Konceptet för denna scen var att tydligt illustrera för användaren konsekvenserna av att låta olika objekt stå och ständigt dra (ibland stora mängder) energi. Realiteten är den - att vi i många länder idag, får den större delen av vår energi från fossila bränslen. Därmed följer det - som ovan nämnt - att vår energikonsumtion har en betydande, reell kostnad för vår miljö. Ett kolkraftverk tyckte vi illustrerar detta väl - då man tydligt kan se hur processen i sig självt - ej är naturlig - och hur förbränning och dess resulterande smog även *ser* dramatisk och toxisk ut.

Konsekvensscen: Användaren lyckas med energispelet

Likt den positiva konsekvensscenen för sopsorteringsspelet - så behövde vi givetvis se till att vi hade en korresponderande scen, för energispelet.

Denna scenen utspelas på samma sätt som den "negativa" konsekvensscenen beskriven ovan - men med följande skillnad:

Efter en liten stund - så hörs det ett högt ljud - och en piedestal med en stor gul knapp (samma färg som "elen" i elkablarna i scenen) - reser sig upp ur marken framför användaren. När denne trycker på knappen, så hörs ett högt "avstängnings-ljud" - och alla vagnar med kol stannar, samtliga objekt som var aktiva stängs av, och kolkraftverkets många ljus släcks.

Sedan följer det ett par demolitions-ljudeffekter - och hela kolkraftverket faller samman. Efter att kolkraftverket brakat samman, så börjar små blommor att poppa upp ur marken runt om de olika objekten från energispelet. Sedan reser sig en hel grönskande miljö upp ur marken - och himlen går från svart till klart, färgsprakande blått. Solen tittar fram och fjärilar och fåglar kvittrar. Kvar i miljön står objekten från energispelet - och det nu raserade kolkraftverket.

Konceptet bakom denna scen var att tydligt illustrera att ifall vi stänger av våra energislukande produkter - så får det en tydlig effekt på vår energiproduktion också. I det här fallet så illustrerades det att då samtliga objekt från energispelet stängdes av - så behövdes det inte el från kolkraftverket längre. Genom att vi upphörde med vår “onödiga” energiförbrukning - så fick miljön och naturen tid till att återhämta sig och florera.

I kontrast till vår “positiva” konsekvensscen för sopsorteringsspelet - så beslöts det att inte visa någon form av scoreboard som visade användaren exakt hur bra det hade gått under spelets gång. Bara för att vi är duktiga med energibesparande så betyder det inte att de produkter vi använder försvinner - eller bör försvinna. Fokus lades istället på att kommunicera den fundamentala tanken kring att det viktiga är att vi försöker spara så gott vi kan - i den mån som är möjlig.

Hallen

När användaren avklarat energispelet - så återvänder användaren till lägenheten - men i detta fallet till hallen, bredvid köket. Denna “scen” var en viktig del av vårt koncept kring att vi skulle skapa en **“A day in your life”-narrative** inom ramen av vårt projekt.

På något sätt så skulle användaren nu behöva lämna lägenheten för att fortsätta med “sin dag” - vilket i detta fallet betydde att göra sig redo för att åka till skolan. Det bedömdes vara en viktig del i vårt arbete med *immersion* inom projektet - att inte bara plötsligt transportera användaren till nya scener. Istället får användaren uppleva att dess egna handlingar leder till att man kommer vidare/transporteras till nästa scen. Själva produktens narrativ ger stöd och ledtrådar till vilka handlingar som användaren bör genomföra för att nå sitt mål.

I hallen så får nu användaren i uppdrag att packa ner ett par skolböcker i en ryggsäck - för att sedan öppna ytterdörren till lägenheten. Som tidigare under projektets gång - såanvänder vi av oss identiska signifiers för att indikera att ett objekt bör interageras med - vilket illustreras genom att böckerna i denna scenen, t.ex. - skimrar och lyser en stark gul färg.

Genom dessa uppdrag av att återigen öppna specifika dörrar, och sedan transportera användaren till en ny scen - såfår man tydlig, konsekvent feedback som blir genomgående för hela projektet.

Transport

Efter att ha öppnat ytterdörren till lägenheten - så befinner sig nu användaren nere på innergården till bostadshuset som

lägenheten befunnit sig i. Mot bostadshusets vägg, ut mot vägen - så står det en skylt med stora bokstäver som säger “Skola”.

På motsatta sidor om användaren står en cykel och en bil. Cykelns ringklocka plingar till - och bilen tutar (för att dra till sig användaren uppmärksamhet), samtidigt som tjocka avgaser kommer ut ur dess avgasrör. Nu står användaren inför ett val: denne måste välja cykeln eller bilen för att ta sig till skolan. Genom användningen av samma signifier som använts konsekvent genom projektet - en skimrande, lysande gul färg över ett objekt - så får användaren en indikation om att den bör interagera med antingen cykeln, eller bilens dörr.

När användaren har gjort sitt val, så sätts användaren antingen pa ˚ cykeln, eller inuti bilen - och transportmedlet börjar köra ut pa ˚ vägen och iväg till skolan. Detta var även en scen som var viktig för vår **“A day in your life”-narrative** inom ramen för projektet - då det fortsatte flödet som började i lägenhetens kök och hall.

Konceptet bakom denna scenen var att användaren skulle få göra ett val i en situation som för personer i vår målgrupp tillhör vardagen. Våra val av transportmedel är en viktig del i miljöarbetet - och att kunna bidra med utlärning inom det, bedömdes viktigt. För att ändå ge någon slags indikation till att välja bilen som transportmedel inte är en god idé med tanke på miljön - så gavs ytterligare *affordance* till bilens utformning. Detta gjorde sig uttryck i de avgaser som pyr ut ur avgasröret till bilen. Bilar som åker förbi ute på vägen utanför innergården ger även de ifrån sig stora moln av avgaser.

En viktig del av designprocessen för denna “Transport”-scen - var att se till att begränsa mängden modellering som krävdes av projektets modellerare. Utomhusmiljöer - kräver i regel en hel del arbete för att inte verka “sterila”. Därmed lades mycket vikt på att försöka designa en innergård som skärmade av användaren så mycket som möjligt från den “yttre världen”. Således kunde mer resurser och tid läggas ner på att göra själva innergården desto mer verklighetstrogen.

Skolgården

Efter att ha valt sitt transportmedel - så anländer användaren till en skolgård. Beroende på vilket val av transportmedel som gjordes, kommer denna scen att se annorlunda ut.

Istället för att göra specifika konsekvensscener som tar användaren till nya miljöer - så ville vi se om vi kunde kombinera dessa, med den fortsatta resan genom användarens dag (**“A day in your life”**). Ifall användaren tar bilen till skolan - så kommer skolgården att vara fylld av smog, och bilen kör sakta bort samtidigt som mer avgaser spyr ut ur avgasröret. I kontrast till detta - ifall användaren tar cykeln till skolan - så är skolgården fylld av sol och fint väder, med fjärilar och fåglar som cirkulerar kring den grönskan som finns på skolgården.

Efter ett litet tag av att ha observerat sin miljö - så kom-mer skolklockan att ringa in till lektion. Då användaren står placerad vid en av skolbyggnadernas dörrar - så kommer handtaget här att börja lysa och skimra (likt tidigare signifiers). När användaren öppnar dörren till skolbyggnaden - såtas denne vidare till nästa scen.

Viktigt här för oss var att - likt kökets utformande - att se till att skapa en skolgårdsmiljö som kändes genuin - och som kunde kännas igen av ungdomar med olika bakgrunder - oavsett hur deras egen skola möjligen ser ut. Genom att studera flertalet skolgårdar runtom i Sverige - samt attåterkoppla till våra egna skolgårdar under våran uppväxt - såformgavs en miljö som innehåller detaljer från alla möjliga platser (basketbollplaner, gungor - flaggstångar osv.). Ett exempel på en skola som stod som inspiration för skolscenen är Sunnerby i Nynäshamns kommun, se Figur 7.

Figure 7: Sunnerbyskolan i Nynäshamns kommun.

Figure 8: Inspirationsbild till scen i kemisal.

Klassrummet

Nästa scen skulle komma att bli den sista scenen som vi skapade för projektet. I denna scenen befinner sig användaren på lektion i själva skolan.

För att återigen återkoppla till att när användaren öpp-nade dörren till skolbyggnaden och att man då transporterades in i själva skolbyggnaden, så kan man tydligt se detta genom att skolgården finns synlig utanför fönstren i lokalen.

Då det var tänkt att användaren skulle få laborera och pyssla lite i scenen - så föll det sig naturligt att modellera lokalen utifrån en kemisal. Utformningen i lokalen har tagit inspiration från kontemporära kemisalar i svenska skolor, se ett exempel i Figur 8. Under större delar av projektet - s˚ a har användaren varit i en mindre betydande maktposition, omslutna av stora miljöer - så i denna sista scen - ville vi nu undersöka och möjligen se en ändring på detta. Därmed framkom idén om att sätta ut tre stora bord med tre stycken miniatyrer av stora landskap/biomer.

Dessa tre distinkt olika biomer skulle alla på något sätt illustrera och lära ut grundläggande koncept inom förnybara energikällor; sol-, vatten- och vindkraft. Genom att just använda sig av miniatyrer - så kan nu användaren få en känsla av att ha en stor makt eller inflytande över de "små" biomerna och miljöerna som finns på borden. Ett utlärande av ett koncept, där *du* har kontroll.

På varje bord sattes det ut en viss typ av miljö:

- **Solkraft**: På en öken-miljö, fylld av stenar, kaktusar och en del kameler - skall man placera ut ett antal olika sol-celler - och sedan en stor energistation som knyter dem samman.
- **Vindkraft**: På en lantlig miljö (ej helt olik den skånska landsbygden), fylld av gräs, träd, med vinande vind, åkermark och en liten bondgård med tillhörande traktorer - skall man placera ut ett antal olika vindkraftverk. Men för att kunna placera ut dessa vindkraftverk så måste man först bygga ihop dem genom att använda sig av ett antal olika delar som ligger på bordet bredvid biomen.
- **Vattenkraft**: På en kuperad, bergig miljö (som skall försöka likna det svenska upplandet), fylld av kullar, bar-rträd, en stor sjö med tillhörande flod som rinner nerför berget - skall man placera ut ett vattenkraftverk i flo-dens mynning. Men för att kunna placera ut detta vat-tenkraftverk så måste man även här - först bygga ihop den genom att använda sig av ett antal olika delar som ligger på bordet bredvid biomen.

Ett viktigt koncept gällande denna scenen var också att vi genom hela projektet - ville försöka hitta någon slags balans mellan olika typer av interaktion. Tidigare i projektet så har de flesta interaktioner varit relativt fysiska och i vissa fall explosiva (kastande av föremål i sopsorteringsspelet, springa runt i köket för att stänga av olika objekt i energispelet o.s.v.). Här ville vi därmed introducera en annan typ av interaktioner, som riktar sig mot en målgrupp som kanske inte tycker det är lika roligt med "action"-fyllda spel, utan snarare föredrar att klura och lösa problem så som pussel och liknande.

Till varje bord så finns det en elkabel kopplad, som löper till varsin glödlampa, ståendes på katedern i salen. Allteftersom användaren sätter ut samtliga objekt på de olika biomerna - så kommer korresponderande glödlampa att tändas på katedern. När samtliga glödlampor har tänts - så rullar det ner en projektorduk och sedan projiceras det upp ett slutligt tack-meddelande till användaren för sina insatser under spelets gång.

När spelet avslutats så återvänder man efter ett tag till den allra första introscenen - och här kan nästa användare ta vid.

Modellering: Igenkänning

Inlevelse är en viktig del i hur användaren tar till sig den pre-senterade informationen, och i hur roligt eller underhållande den upplevs. För att uppnå bra inlevelse kan man gå tillväga på olika sätt: exempelvis kan man sikta på att återge en bild likt verkligheten, där tanken är att användaren ska uppleva det som om det vore på riktigt och därmed ta till sig information och intryck. Problemet med detta grepp är att det har en väldigt hög "kostnad": det tar mycket resurser i anspråk att modelera realistiska föremål.

I EnVRment valdes istället att fokusera på igenkänning och få användaren att kunna relatera genom att trycka på sociala och kulturella konventioner, snarare än genom att lura de perceptiva sinnena. Detta bär en mycket mindre resurskostnad, då miljöernas och objektens realistiska utseende blir av mindre värde och vikten av vilka och hur de presenteras blir mer markant.

Arbetsprocess

I början av kursen lärde projektmedlemmarna känna varandra genom att hålla korta, informella möten. Då alla projektmedlemmar har olika bakgrunder, både kunskaps- och erfarenhetsmässigt, fördelades arbetsuppgifter med hänsyn därtill. Också vad projektmedlemmarna var intresserade av att lära sig spelade in. Tidigare hade två projektmedlemmar utvecklat inom VR och hade erfarenhet om vad som skulle krävas. Det bestämdes i tidigt stadie vilka verktyg som skulle användas för utveckling och modellering:

Utveckling - Unity: Båda de projektmedlemmar som hade kunskap om VR-utveckling sedan tidigare, föredrog att utvecklingen skulle ske via grafikverktyget *Unity*. I Unity finns stöd för att utveckla på många plattformar, varav VR är en. Det finns även stöd från olika s.k ”Tool-kits”, och av dessa föreslog en av medlemmarna användning av ”VRTK - Virtual Reality Tool Kit”, som denne hade goda erfarenheter av. Unity har även utförlig dokumentation samt en väl etablerad community, som kan stödja utvecklingsprocessen.

Modellering - Blender: Blender valdes även baserat på de två projektmedlemmarnas tidigare kunskaper. De båda hade använt Autodesk 3DS Max tidigare och tyckt att verktyget varit klumpigt och ointuitivt. Därmed föreslogs Blender som ett bättre alternativ, baserat på gott rykte bland vänner, kollegor samt i diverse webbforum.

Följande arbetsfördelning gjordes:

- **Utveckling**: Två projektmedlemmar som stått för utvecklingen av själva prototypen i den utvalda utvecklingsmiljön. En av medlemmarna med tidigare kunskaper (och ledaransvar) samt en medutvecklare (med intresse för att lära sig mer om VR utveckling och tidigare kunskaper av annan mjukvaruutveckling).
  Arbetsuppgifter: Kodning och allt relaterat till utvecklingsmiljön.

- **Modellering**: Två projektmedlemmar som står för all modellering av 3D objekt, dess design samt material. En av projektmedlemarna (med tidigare kunskaper samt erfarenhet inom design och en känsla för estetik) samt en medarbetare (med intresse för att lära sig mer om 3D-modellering).
  Arbetsuppgifter: Modellering och texturering av 3D objekt.

- **Konceptuell design**: En projektmedlem som har huvudansvar, men med stöd av och gemensam beslutsfattning med resterande delar av projektgruppen.
  Arbetsuppgifter: Utveckla konceptuell design för de uttänkta scenerna och spelen.

- **Projektledning & kommunikation**: En projektmedlem som står tar ansvaret att hålla i möten och föra projektet framåt i dess diskussioner, samt att sköta kontakten med handledarna och därvid framföra det som sagts till gruppen.
  Arbetsuppgifter: Styra diskussioner och sköta kontakt med handledarna.

Arbetsfördelningen ändrades efter halva projektets gång, beroende på hurvida projektmedlemmen fördjupa sig i andra arbetsuppgifter. Denna förändring ledde till en mer strömlinjeformad och effektiv arbetsprocess. Förändringarna blev i stora drag att

- Konceptuell design togs över av projektmedlemmen som suttit som modelleringsansvarig

- Utvecklingsansvarig i Unity fick ta hand om allt kodarbete på egen hand

- Huvudansvaret för modelleringen togs över av den som varit en hjälpande hand i utvecklingen i Unity

- Två personer gick över till att hjälpa till där det behövdes, vilket gav en mer flexibel organisation. Därmed kunde större resurser läggas på modelleringen, vilket tidigare varit en flaskhals, samt att man hade avsatta resurser i till att genomföra formella tester när det krävdes.

Bägge arbetsfördelningarna som gjordes följde samma process som gick igenom tre olika faser:

1. Uppstart-fasen: Här skedde all utveckling på olika håll, i vilka utvecklingsteamet undersökte de möjligheter som fanns för tilltänkta koncept och började prototype:a utan några skapade 3D-objekt från modelleringsteamet. Det senare satte sig in i det utvalda 3D-modelleringsverktyget och påbörjade sin inlärningsprocess för att kunna skapa 3D modeller. Den konceptuella designen försöktes konkretiseras (tecknas upp och skrivas ner) för att enklare kunna refereras till och itereras igenom, dessa dokumenterades i en ”Backlogg” i planeringsverktyget Trello.

2. Aktiva utvecklings-fasen: Under denna fas var de olika del-team:en i sync genom att önskade 3D-objekt levererades till utvecklingsteamet, som sedan kunde implementera och lägga på den önskade funktionaliteten. Den konceptuella designen utvecklades iterativt och nya koncept kom fram allteftersom de behövdes för vidarutveckling.

3. Finjusterings-fasen: I denna fas saktades själva utvecklingen ner för att man skulle gå in mer på testning och finputsning av den nuvarande prototypen. Modelleringen av 3D-objekt fortlöpte lite längre i den mån objekt saknades och behövde kompletteras. Feedback från testning på olika nivåer följdes upp, med ändringar för både prototypens funktionalitet, utseende och konceptuella design.

Utvecklingsprocessen har följt en agil modell där komponenter av upplevelsen successivt har byggts ihop, testats och sedan utvärderats för vidarutveckling.

4 Testning

Testning av produkten är ett essentiellt inslag i utvecklingsprocessen. Även om det kan vara kostsamt vad gäller både tid och pengar, är också en liten datamängd värdefull. I längden räcker det inte alltid att hänvisa till att man som utvecklare vet vad kunden önskar: för att resultatet ska bli som bäst för båda parter gäller det att också ha testat sin produkt på gruppen av användare. Detta minskar också risken att vägledas för mycket av övertygelsen att man själv "vet bäst" om hur den rationella användaren borde agera[2]. För att få vidare bekräftelse på att projektet var på väg i rätt riktning besöktes därför en skolklass för att testa programmet, samt se och höra testpersonernas reaktioner på produkten.

Testning: Tyringe

Under onsdagen 2018-10-17 besöktes en sjundeklass på Tyringeskolan i Tyringe. I den kvalitativa forskningsmeto-den finns flera olika alternativ när det gäller produkttestning. De tva ˚ som valdes ut till exkursionen var 1) fokusgrupper och 2) direkt observation. Det ansågs att dessa kunde kom-plettera varandra: diskussioner i fokusgrupper kan visserligen p˚ averkas av dynamiken deltagarna emellan, men de kan också ge värdefulla insikter om varför det beteende (som visat sig vid observationerna) uppkommit[2].

Målet med testet var att, genom att spela in själva användningssessionerna och plocka ut nyckelord från de olika gruppdiskussionerna, besvara följande frågor:

- ¨Ar det roligt att spela?
- ¨Ar någonting svårt att förstå?
- Lär man sig något, och i så fall vad?
- Vilka är kunskaperna och tankarna kring miljöarbete?

Tre projektmedlemmar var på plats för att utföra testet, valdes att under ett helt lektionspass (på ca 50 minuter) låta eleverna lösa av varandra i ett angränsande rum, där VR-utrustningen monterats upp. På plats var hela tiden en handledare, som hjälpte till med instruktioner samt headset och handkontroller. Under tiden utförde de båda andra - en antecknande, och en i rollen av moderator - tre stycken gruppdiskussioner: först två med 4 respektive 4 elever som *sedan* skulle testa prototypen, och därefter en med 5 elever som *redan testat* den. Efter elevernas lunchrast fick så ytterligare 3 elever testa produkten, utan att sedan intervjuas.

I diskussionerna försökte de som intervjuade att konsekvent ställa öppna frågor med anknytning till de övergripande frågeställningarna ovan. De elever som intervjuades *före* testet fick frågor som:

- Vad har ni tidigare lärt dig om miljöarbete?
- Gör ni själva något för miljön?
- Vad gör era familjer för miljön hemma?
- Hur tror ni att miljötänket kan skilja sig mellan olika familjer?

De som intervjuades *efter* testet fick svara på:

- Vad tyckte ni om spelet?

Figure 9: Livebild från en testsituation

- Var det någonting som var särskilt svårt? (Både spel och utrustning)
- Fanns det något som var speciellt roligt?
- Var hamnade ni efter själva spelscenen?
- Har ni provat Virtual Reality tidigare?

Testning: VR-labbet på IKDC

I VR-labbet på IKDC var målet med testet att de justeringar och åtgärder som genomförts utifrån Tyringe-elevernas respons skulle provas (genomfördes 2018-11-09).

Energispelet var ännu bara i konstruktionsfasen och därför inte möjligt att klara testa, i vidare mening än att man hade möjlighet att stänga dörrar och vrida av kranen.

De personer vi tillfrågat var alla elever vid Lunds Universitet, med nästan genomgående bakgrund inom antingen data- eller kognitionsvetenskap, och det mål vi från början satt upp (på ca 10 individer) uppfylldes med råge.

Istället för muntlig respons fick deltagarna efter avslutat test ge sin feedback i form av betyg på en 5-gradig skala, avseende ett antal olika aspekter, rörande (i ordning) de olika inslagen i sopsorteringsspelet:

- Hur väl man kunde identifiera föremålen.
- Hur det gick att greppa och kasta objekten
- Huruvida man noterat agenterna och förstått skyltarna i konsekvensscenerna.

Testning: Svaneskolan

Testtillfället på Svaneskolan i Lund (2018-11-27) skilde sig fr˚ an det i Tyringe i ett par avseenden. För det första var upplevelsen nu utökad med **energispelet**, vilket gjorde att varje testperson fick upp till 10 minuter på sig att prova sig fram i upplevelsen.

För det andra hade vi därför också valt att inte föra diskussioner i *grupp* efteråt, då eleverna var tvungna att bege sig tillbaka till lektioner och inte heller kunde stanna kvar i upp till (eller återkomma efter) en halvtimme för att sedan diskutera upplevelsen. Istället fick var och en av dem, efter just sin egen session, svara på kortare frågor från en av testledarna, som därunder förde anteckningar.

För det tredje var nu också eleverna tillfrågade på *förhand* - och schemalagda under en hel förmiddag - vilket innebar att de som tackat nej på frågan t.ex. inte fick chansen att ändra sig och göra testet efter att vi själva fått presentera oss och vårt projekt. Urvalet blev därför lite annat, och upplägget lite mer ordnat, än vid det första testtillfället. Av medlemmarna i projektgruppen var nu endast två stycken på plats, vilket med tanke på tidigare testning - ansågs kunna räcka.

Tidsomfånget blev också större: testen utfördes under *tre* timmars tid, i princip utan avbrott, med deltagarna som avlöste varandra efter 10 minuter testning. Likheten med första tillfället var att vi använde oss av samma VR-kit och därför kunde räkna med liknande avvikelser från vad vi upplevt med utrustningen i labbet.

Frågorna som ställdes till eleverna gällde i första hand *energispelet*. Vad vi ville veta var om de förstått vad själva uppgiften var:

- Vart man skulle röra sig, och hur handkontrollerna skulle användas.
- Om man identifierat och kunnat använda sig av poängräkningen (i form av klockan och batteriet).
- Om man förstått budskapet med denna spelscen och om den i allmänhet var svår.
- Dessutom fick man kommentera övergångs- och konsekvensscenerna till energispelet, och produkten i stort - och därmed alltså även sopsorteringsspelet, som var i fokus vid förra tillfället.

Vad som nu istället *inte*, till skillnad från i Tyringe, berördes i samtalen med deltagarna var hur (och om) man annars själv tänker på miljön eller aktivt handlar miljövänligt, vad man fått lära sig i skolan om miljöarbete och vad man tror att man själv, som barn och skolelev, faktiskt själv *skulle kunna* uträtta.

Testning: Demo på Wisdome i Malmö

Efter de mer strukturerade testerna har vi också genomfört en större demonstration för ett antal personer på luft- och sjöfartsmuséet i Malmö som tillhör/representerar intressenten för projektet; Wisdome (där den fullbordade produkten kommer att installeras).

Under en tvåtimmarssession beskrevs bakgrunden till och våra resonemang bakom hur den utformats - och på grund av utdragen teknisk förberedelse gavs introduktionen extra mycket tid.

Felkällor för samtliga test

Städerna Tyringe och Lund är storleksmässigt inte jämförbart med Malmö (platsen för själva Wisdome), vilket kan innebära att användarna på ena hållet har andra hemmiljöer än på det andra. Därtill är det tänkbart att det genomsnittliga tekniska kunnandet påverkas av demografin.

Under testerna som skedde utanför VR-labbet på LTH - användes en mobil enhet, både i form av en laptop och ett mobilt VR-kit. Användandet av laptop gav inga synliga problem, då dess hård- & mjukvara var tillräckliga för uppgiften. Däremot uppfattades det mobila VR-kitet (Samsung HMD Odyssey [7]) ge sämre upplevelse och resultat än det som programmet utvecklats mot (HTC Vive [8]). Detta lär mestadels vara på grund av dess mobila natur jämfört med att ha uppsatta sensorer för tracking och inte på grund av ”*wearables:en*” i sig själva.

Problemen som stöttes på gällde mest handkontrollerna, då dessa var mycket svårare att hantera än de tillhörande Vive:en. Det var mycket svårare att kasta föremål, samtidigt som det var mindre intuitivt med hur ”*triggers:en*” var placerade på kontrollerna, så pass mycket att testledaren fick instruera väldigt noggrant hur dessa skulle hållas.

Samtidigt speglar detta resultat vad som framhävdes i teoridelen: en teknisk begränsning hos utrustningen (t.ex. ett extra avgränsat synfält eller sämre tracking) kan innebära att användaren inte upplever den presence som kan krävas för att förståelsen av situationen ska bli tillräckligt bra.

När det specifikt kommer till testningen på Svaneskolan i Lund - så var de tekniska felkällorna som sagt detsamma som i Tyringe, medan demografin kan tänkas ha varit lite annorlunda. 6 av 19 elever hade uttryckligen testat VR tidigare, vilket inte lär vara fallet hos den slutliga målgruppen. Vad gäller intervjusvaren tror vi också att det kan ha funnits en osäkerhet kring vad testledaren menat med frågan om vilket “budskap” de olika spelscenerna haft.

Svaren på dessa frågor blev ganska tvekande, men vi tror att detta snarare har att göra med skygghet - sannolikt har intervjun, när den gjorts mellan bara fyra ögon, upplevts mer som ett förhör än i gruppdiskussionerna i Tyringe - än att man inte förstått att den tänkta lärdomen är att spara resurser.

5 Resultat

Interaktion 1 (MAMN10)

Här nedan beskrivs det slutliga resultatet av projektet som togs fram under Interaktion 1, med start i utkomsten av testet i Tyringe, följt av de uppdateringar som gjorts på produk-ten utifrån testresultaten. Slutligen återfinns en kortfattad beskrivning av slutprodukten.

Testtillfället i Tyringe

Resultatet av de första två diskussionerna visade ett par tydliga tendenser. I princip alla försökspersoner hade lärt sig om både utsläpp, global uppvärmning och sopsortering sedan tidigare, även om det inte tagits upp så mycket just i skolan. Flera hade också uppfattningen att det viktigaste för miljön var att minska koldioxidutsläppen.

På frågan om vad man själv gjorde för miljön var sopsortering det mest genomgående. En av försökspersonerna formligen rabblade olika sorteringskategorier: brännbart, färgat eller ofärgat glas, pant, och tidningar.

Försökspersonerna förstod snabbt hur VR och produktens övergripande koncept fungerade. Det som däremot visade sig vara ett problem var den generella ”*historian*” som produkten försöker berätta, och därför ansågs det svårt att veta vad som förväntades utföras av användaren.

Sopsorteringsspelet ansågs ha ett klart och tydligt syfte som eleverna förstod, och bortsett från vissa svårigheter med kastandet av skräp (vilket delvis kan tillskrivas det mobila VR-kitet) flöt denna scen på bra. Åtminstone en av delta-garna hade uppfattat tidräkningen ovanför sopkärlen i själva spelscenen. Alla eleverna hamnade sedan i konsekvensscenen ”Skrotupplaget” (vilket tros vara på grund av svårigheten att kasta objekten, och att metallkärlet är placerat längst bort), men i denna scen trodde många att de kommit till ännu en spelscen, och missade helt enkelt vad agenten försökte berätta för dem.

Åtgärder från testtilfället

Direkt efter att testresultaten var sammanställda påbörjades revideringar av produkten i enlighet med sagd feedback. De stora punkterna som behövde ändras eller förbättras var:

1. Förbättrad introduktion till handkontrollerna
2. Tydligare direktioner i köket
3. Få agenterna att framträda mer i konsekvensscenerna
4. Finjustera sopsorteringsspelet

För att behandla punkt 1 från listan med åtgärder infördes, för att bättre introducera användaren till handkontrollerna och VR som helhet, en ny startscen. Istället för att användaren först hamnar i köket (som i den första prototypen) hamnar hen istället i ett enkelt vitt rum med endast två tavlor och en dörr som inredning. Detta för att minska det visuella brus som i testerna bidrog till att försökspersonerna (då fokus låg på att ta in köksmiljön) missade den då aktuella introduktionen till handkontrollerna. Tavlorna är till för att både introducera konceptet som ledtråd då liknande metoder planeras att användas i nästa kurs, samt att då ge den faktiska hinten om hur handkontrollerna fungerar. I Figur 10 kan man se denna scen, i vilken användaren har sett tavlan med instruktioner, tryckt på den högra handkontrollerns ”*trigger*” men ännu inte den vänstra.

Om handkontrollerna inte vidrörs kommer de efter en viss tid att börja vibrera lätt, för att dra till sig användarens uppmärksamhet. I denna bild kan även dörren ses, och när användaren väl har ”aktiverat” båda händerna genereras ett knackljud. Detta upprepas sedan tills användaren öppnar dörren. Om detta lyckas så kommer hen sedan(efter att scenen bleknat bort) att befinna sig i köket och där kunna börja interagera med världen.

De största problemen som upptäcktes under testerna var att användarna önskade tydligare direktiv i köksscenen samt att de verkade sakna sammanhang och fokus i konsekvensscenen. I och med att tavlorna nu introduceras tidigare så är chansen större att användaren kommer att kunna ta till sig denna information. En rimlig förväntan är dock inte att detta kommer vara fullt tillräckligt, så eventuella lösningar är att efter en specificerad tidsperiod få luckorna under vasken att börja blinka, lysa eller uppföra sig dylikt, för att dra användarens

Figure 10: Uppdaterad första scen

Figure 11: Sopsorteringsspelet - Version 1

uppmärksamhet till sig. I nuläget innehar luckorna endast en lysande ram runt sig, vilket alltså visade sig vara otillräckligt.

Att agenterna inte fick den nödvändiga uppmärksamheten kan atgärdas genom att man släcker ner allt annat och lägger till en scenlampa (som på tavlan i Figur 10), vilken lyser på det agenten ska se. Alternativt kan allt annat blekna ut till vitt.

Sopsorteringsspelet fungerade som ovan nämnt bra i sin helhet, även om det kan behöva förändras för att få en bättre spridning på de olika konsekvensscenerna, då de som testade alltid hamnade på ”Skrotupplaget”. En lösning vore att placera alla sopkärlen på samma avstånd från användaren, och att de alla ger samma mängd poäng. Nackdelen med detta vore dock att spelet tappar en del av sitt lärande, dåden tidigare uppställningen bättre reflekterade på vad som var viktigast att sortera av de olika föremålen. På grund av otillräckliga tidsresurser är detta ett beslut som skjuts på framtiden.

Prototypen

Efter revidering som gjorts till följd av testningen så har en prototyp färdigställts för Interaktion 1 (MAMN10). Denna prototyp innehåller de olika delarna:

- Introduktionsrum

- Kök (Lägenhet)
- Sopsorteringsspel
- Konsekvensscener
  - Papp
  - Plast
  - Metall

Scenerna har olika syften. Det första rummet, *Introduktionsrummet*, lär användaren som nämnt ovan i avsnitt 5, att hantera kontrollerna till Virtual Reality och hur interaktion med föremål fungerar. Härefter presenteras *Miljonprojektsköket*, vilket agerar som ett nav i prototypen. Här kan användaren utforska flera möjligheter att interagera med sin nya omvärld, bland annat att öppna luckor och lådor, eller lyssna på radion eller barnen som leker på gatan utanför.

När användaren först kommer hit ligger det skräp och sopor i större delar av köket. Det är även här som kopplingen vidare till *Sopsorteringsspelet* finns, genom att de båda luck-

orna under vasken ska öppnas. Öppnas luckorna expanderas köket, och ena väggen flyttas bort ifrån användaren så att en bana bildas. Framför användaren, i slutet av banan, dyker även ett tråg med diverse sopor upp som användaren ska kasta i olika sopkärl som är utplacerade på banan. Klarar användaren av att sortera tillräckligt mycket av skräpet sååtervänder användaren till köket, men denna gång är köket rent och utan skräp. Klarar användaren *inte* spelet så tas denne istället till en av de tre *Konsekvensscenerna*.

Vilken scen som visas bestäms av vad användaren var sämst på att sortera och användaren kommer att ges en ny chans att försöka sig på spelet. De olika scenerna visar platser där konsekvenserna av att inte återvinna har allvarliga följder - såsom i skogen, havet och grundvattnet. Scenerna bebos även av en agent som försöker förklara en historia om hur sambandet mellan att inte återvinna sopor, och hur deras hemmiljö (i konsekvensscenen) har förändrats till följd av detta.

Interaktion 2

Här beskrivs det slutliga resultatet av hela projektet som togs fram under de bägge delkurserna Interaktion 1 & Interaktion 2, med start i utkomsten av testet på studenter på IKDC, följt av de uppdateringar som gjorts på produkten utifrån dessa testresultat. Slutligen återfinns en kortfattad beskrivning av slutprodukten.

Testtillfället på IKDC

Här var resultatet av testerna för de olika deltagande i de flesta fall positiva: de skimrande och blinkande signifiers vi lagt in (eller skruvat upp) på både skåpsdörrar, agenter och scoreboard, gjorde att de nu upplevdes som klart tydligare.

Vad som fortfarande var oklart (frånsett objekten och effekterna i energispelet, som vi ju var fullt medvetna om) var indikatorn för plastskräp i sopsorteringsspelet - som inte ansågs vara fullt tydlig. I stort var responsen dock för det mesta mycket positiv.

Testtillfället på Svaneskolan

Svaren vi fick på en del testfrågor belyste ett flertal detaljer i **energispelet**. Flera av eleverna tyckte att det hade varit svårt att "komma in" i spelet, dels på grund av osäkerhet kring *hur l˚angt* man fysiskt kunde förflytta sig, dels blinkande ljus från olika håll i rummet, och dels att ljudsignalerna var otydliga. Flera hade noterat klockan och batteriet, men inte följt dem för att veta hur väl man prestererat.

Konsekvensscenerna för energispelet upplevdes inte rakt igenom som otydliga, men när de gjorde det berodde det främst på att bakgrunden (med vagnar och kraftverk) varit dunkel.Övergångsscenen till **energispelet** hade i sin tur flera g˚anger orsakat svårigheter under själva testet, vilket dock *inte* framkom i diskussionerna. Detta antar vi berodde på att denna faktiskt framstod som en "sekundär" uppgift - vilket ju också stämmer med avsikten.

Vi kunde också glatt konstatera att de allra flesta tyckt att produkten och upplevelsen i stort varit *rolig*. Några ytterlighetsfall förekom även; någon (som också testat VR tidigare) klarade alla uppgifter helt felfritt och utan vidare startsträcka, medan någon annan nästan rakt igenom uttryckte att det varit både svårt och läskigt - och vi antar att det inte är till just dessa vi i första hand vänder oss.

Demotillfället på Wisdome

De av besökarna som ville prova på produkten efter att vi givet vårt anförande om projektets bakgrund - gavs tillfälle till detta. Entusiasmen och den positiva responsen var omfattande.

Genom att ännu en gång få prova att sätta upp en test-miljö på en annan plats än på IKDC - så fick vi upp ögonen för flertalet förbättringar när det kommer till att göra produkten dynamisk när det kommer till dimensionerna på testarean.

Vi sattes även efter vår demonstration i kontakt med några av deltagarna på demotillfället som gärna ville veta mer - och undrade ifall de kunde bidra till vårt projekt.

Åtgärder från testtilfällena

Efter varje testtillfälle så satte vi oss ner och hade ett gruppmöte - där vi tog fram konkreta punkter av feedback som vi behövde åtgärda från testningen; vare sig det handlade om en bugg, eller ifall det handlade om att en modalitet/interaktion ej fungerade som vi hoppades på.

Dessa punkter av feedback gjordes sedan om till "Cards" på webbtjänsten Trello (en digital "kanban board") - och delegerades till en gruppmedlem som hade hand om det området.

Lyckligtvis berörde de allra flesta punkterna av feedback mindre delar av våra scener som upplevde otydliga eller att interaktioner inte fungerade som tänkt, inte för att vi hade övergripande problematik med vårt koncept.

Punkter av feedback som togs upp på testillfällena/demotillfället och senare åtgärdades innehöll bl.a. följande:

- Nya ljud för aktiva objekt i energispelet.

- Tillagd, tydligare progress indikation i sopsorteringsgsspelet.
- Förändrad svårighetsgrad på energispelet & sopsorteringsgsspelet.
- Omgjort batteri och klocka i energispelet.
- Förtydligad och omgjord scoreboard på sopsorteringsgsspelet.
- Fler feedback-ljud inlagda vid spelslut/spelbörjan.
- Flertalet omfärgad objekt för tydligare signifiers/indikatorer.
- Förändrad logik för pussel-objekten i klassrumsscenen.
- Utbytta modeller i klassrumsscenen.
- Dörren till skolbyggnaden bytt från ”slide”- till ”pull”-typ.
- Förändrad interaktion med skolväska i hallscenen.
- Förkortad tid i förflyttning under transportscenen.

Då punkter av feedback åtgärdats - så loggfördes dessa - för att sedan tas med i nästkommande ronder av tester (vare sig det gällde ett formellt testtillfälle eller iterativ testning som skedde jämte utvecklingen/implementationen).

Slutprodukt

Nu, efter att den sista delkursen (Interaktion 2, MAMN15) nått sitt slut - så har vi tagit fram en slutprodukt som innehåller följande scener/delar:

- Introscen 1: Lära sig kontrollerna
- Kök (Lägenhet)
- Sopsorteringsspel
- Konsekvensscener för sopsorteringsspelet
  - Papp
  - Plast
  - Metal
- Introscen 2: Lära sig ny typ av interaktion
- Energispel
- Konsekvensscener för energispelet
  - Negativ kolkraftverksscen
  - Positiv kolkraftverk- och skogsscen
- Hall (”Transport”)
- Innergård (”Transport”)
- Konsekvensscener för ”Transport”
  - Skolgård: negativ
  - Skolgård: positiv
- Klassrum
  - Solkraftverk & biome
  - Vindkraftverk & biome
  - Vattenkraftverk & biome

Nytt för den sista delkursen var samtliga delar efter konsekvensscenerna för sopsorteringsspelet. Således skapades 11 unika scener - med ett antal stora delmoment/funktionaliteter inom varje scen; en stor tillökning gentemot första delkursen (Interaktion 1, MAMN10).

Det stora fokuset i denna slutfas av projektet var dock att röra produkten framåt när det kom till berättelse-aspekten; *”A Day in your Life”*. Samtidigt lades det ner stora mängder resurser på att förbättra och iterera på de delar av projektet som mer eller mindre färdigställts i föregående delkurs (Interaktion 1).

Till att börja med så togs det fram en *”positiv” konsekvensscen för sopsorteringen* - där man bl.a. skulle visa för användaren exakt hur mycket av varje typ av objekt som man sorterat - och vad detta resulterade i för poäng. Implementationen av denna del slutfördes snabbt när 3D-modellerna för scenen levererats - då den i stort kunde återanvända animationer och transitions från tidigare scener i projektet.

Därefter påbörjades implementationen av *energispelet* utefter den konceptuella design och specifikationer som tillägnats denne. De flesta objekten som skulle manipuleras i detta spel hade redan modellerats sedan tidigare - då de var en del av köket i sig. Dock behövdes det tas fram nya reglage, knappar och spakar som behövde placeras ut vid objekten - som användaren skulle interagera med.

När energispelet implementerats färdigt - skickades det direkt ut till testning (se ”Testillfället på IKDC). Feedback från detta test (och andra) användes till att iterera på de knappar, reglage, spakar och indikationsljud som tidigare attributerats scenen. Ytterligare tester på de åtgärdade delarna av scenen - resulterade i att vi kunde lämna arbetet med scenen bakom oss och fortsätta.

Hallen i lägenheten hade sedan tidigare varit något omöblerad och odetaljerad. Då denna nu skulle användas som en scen för att lämna lägenheten - så behövdes det göras ytterligare 3D-modellering och implementation av ett antal funktioner; genom att använda sig av VRTKs färdigbyggda API:er gällande ”sticky placement”, kunde vi snabbt implementera ett system där användaren behövde placera ett par skolböcker i skolväskan som stod i hallen - innan användaren tilläts öppna dörren och lämna lägenheten.

Nästa led i ”Transport”-delen/scenerna i projektet utspelade sig på *innergården* nedanför lägenhetens bostadshus. Här genomfördes en detaljerad planritning och planering för hur 3D-modelleringen skulle se ut. Planeringen bar frukt, det krävdes relativt lite 3D-modelleringsarbete för att sätta ihop scenen då många modeller kunde återanvändas, och implementation kunde sättas igång omgående. Innergårdsscenen, likt hallen, krävde specifik interaktion från användaren för att denne tilläts fortsätta till nästa scen. När detta implementerats, polerades scenen med animationer på fordon som körde förbi utanför innergården, samt flertalet feedback-mässiga ljud som lades in i scenen.

När användaren gjort sitt val av vilken typ av fordon denne skulle använda för att ta sig till skolan - presenterades användaren med en *skolgårdsscen*. Denna skolgårdsscen gjordes distinkt genom inkluderingen av flertalet objekt, ”particle effects” och animationer - beroende på vilket fordon som valts. Ifall bilen valdes, så gjordes scenen full av smog, samtidigt som bilen i sig kunde ses köra bort från användaren, komplett med avgaser och ljudeffekter. Ifall användaren valde cykeln som färdmedel, så var scenen istället ljus, livlig -och komplett med animerade fjärilar och fåglar som rör påskolg˚ arden. Efter att användaren befunnit sig på skolgården en stund och haft möjligheten att se sig omkring, ringer skolklockan in för att användaren ska förstå att det är dags att röra sig in i skolan, vartefter dörren in till skolbyggnaden indikerade till användaren att den nu kunde öppnas.

Den sist implementerade scenen/delen i projektet - utspelar sig i ett *klassrum*. I detta klassrum behövdes det byggas en stor mängd logik som skulle hantera de olika bordens miniatyr-landskap/biomer och dess delar. Ett system för uppbyggandet av objekt - med hjälp av mindre del-objekt - implementerades. Detta system gjordes även robust genom implementationen av ett slags ”out-of-bounds-detection”-system - som upptäckte ifall objekt som användaren behövde interagera med för att ”komma vidare” i scenen - nu låg för långt bort från VR-systemets spelarea. För att göra de små biomerna på de olika borden tilltalande och levande - lades även resurser på att animera de olika delarna av biomerna; traktorer körde omkring, kor och kameler som går runt i landskapet eller vriter på huvudet, sol och moln som rör sig ovanför biomen, o.s.v.

Nu vid projektet och de bägge delkursernas slut kan vi konstatera att **samtliga** 3D-objekt och modeller i hela projektet är gjorda för hand av projektgruppen. Samtlig kod använd i projektet är även gjord för hand av projektgruppen (förutom den kod gällande manipulation av objekt i relation till handkontrollerna - vilket tillhör VRTK-ramverket). Hela konceptet för produkten har också tagits fram och finslipats av projektgruppen, med stöd av dess handledare.

6 Diskussion

I projektets syfte nämns tre nyckelbegrepp: hållbarhet, integration och inspiration. Det är de bägge första begreppen som hamnat i tydligt fokus, medan den sistnämnda antagits indirekt ha kunnat tillgodoses genom användning och presentation av en intresseväckande ”ny” teknik (VR). Att vid kommande testtillfällen istället bjuda in personer till universitetet och VR-labbet skulle kanske ge en bättre, direkt koppling till just ingenjörs- och teknikstudier. Men ytterligare scener där mer fokus ligger på inspiration till att studera vidare kan också vara en vidareutveckling som kan göras i framtiden.

Då en av avgränsningarna från projektsyftet var att främja integration så har detta haft en påverkan av hur buskap förmedlas, därför återfinns ingen text och tal i något språk i slutprodukten (förutom ordet ”Skola” - som vi antar att besökande skolklasser i vår målgrupp känner till). Detta just för att inte utesluta någon som inte kan det potentiellt valda spr˚ aket. Dock skulle text och tal hjälpa att förmedla buskap i spelet om det fanns stöd för varje spelares språk (vilket det finns stöd för i vår slutprodukt). När uppgiften och syftet med en scen är oklart kan också mycket implicit vägledning ges av ett enda ord, t.ex. ”Sopsortera!”. En sådan information skulle inte bara ge ett tydligt fokus utan också exempelvis förståelse för varför skåpsdörrarna nedanför vasken är mer syngliggjorda än andra skåpsluckor. Som det nämndes under rubriken ”Designprocess” har vi gjort ett aktivt val att undvika all form av tal- och skriftspråk och istället använda rörelsemönster, bilder, ikoner, ljudnivåer/tonlägen och ljus för att försöka göra VR-upplevelsen förståelig. Dock har val av språk implementerats på begäran av handledarna, då tanken är att VR-upplevelsen kommer att komplementeras med tal när den väl installeras på Wisdome. Beroende på hur de talsekvenser som kommer läggas till i framtiden utformas, kommer även upplevelsen av produkten se annorlunda ut.

När det gäller utvecklingsmiljöerna har inga vidare hinder uppstått. Inlärningen av Blender har gått fort och inte fått någon del i processen att stanna upp. Vissa faktorer, som t.ex. inställningen av gravitation hos objekt som faller, sjunker och kastas har vållat större problem än andra, men till slut ändå kunnat kontrolleras bra genom vidare optimering.

I samband med slutpresentationen av projektet (i Interaktion 1, MAMN10) uppkom frågor om huruvida man kunde ändra poängsättningen i spelscenerna, för att eventuellt kunna uppmuntra och främja inlärningen mer. I nuläget ger som sagt metallsortering bättre utdelning än papperssortering, på grund av att det krävs mer energi att producera metall än papper. Kanske skulle dock t.ex. negativ *enf* vid ett misslyckat kast av en metallburk göra att man som spelare undviker dessa objekt istället för att, genom att försöka kasta dem bättre nästa gång, maximera utdelningen. För att undvika en sådan effekt kan direkt kompletterande *elf* behövas, så att spelaren genast får insikt i vad som behövs för att höja sitt resultat. Den feedback som används behöver inte bara vara av rätt typ utan också i många fall ges *direkt*, för att optimal effekt ska uppnås.

Vid valet av sopsortering i en köksmiljö togs särskild hänsyn till igenkänningsfaktorn; de påtänkta användarna antogs sedan tidigare vara bekanta med just denna. Resultatet av testet visade att de även i hög grad är medvetna om fordonsutsläpp. Korrekt återvinning av material är en viktig faktor och kan indirekt (genom att t.ex. mindre metallvolymer behöver utvinnas) medföra färre transporter, vilket skulle potentiellt kunna ge den största långsiktiga vinsten.

Frågan om något sådant vore möjligt i det aktuella projektet är relevant. Med tanke på resursbegränsningarna skulle svaret, åtminstone i det här skedet, bli nej: då sopsorteringen kan tänkas ske relativt intuitivt (skräp på golvet och ett sopkärl inom synhåll ger en tydlig koppling), kräver ett val av transportmedel mer av explicit information. Därtill är kanske inte en **spel**situation optimal för denna typ av inlärning: i sopsorteringsfallet verkar det rimligt att ge t.ex. *enf* när det kastade objektet slår ner i ett sopkärl, men det är inte lika lätt att säga i vilket skede feedback ska ges på ett val av transportmedel. Uppstår felet när man börjar röra sig mot flygplatsen, eller när man börjar gå igenom gaten? Kanske är kunskap om mer generella attityder ämnad att istället läras ut explicit.

Framtidsvision

Tanken med projektet är att det ska vara så pass välutvecklat att det kan stå på egna ben och visas upp på Wisdome då kurserna (Interaktion 1 & 2, MAMN10/MAMN15) är avslutade. Men då det kommer dröja upp till tre år innan Wisdome står klart på Malmö museér, så kommer EnVRment troligtvis fa ˚ en utställningsplats på Vattenhallen Science Center påLunds Universitet så länge. Där kommer det finnas möjlighet att testa EnVRment under en längre tid, och på så vis kan ytterligare värdefull testdata samlas in och användas till att iterera på produkten. Om planen att inkludera tal i produkten blir verklighet, är Vattenhallen en utmärkt testmiljö för att utvärdera olika designförslag av talresurserna.

En del av konceptet för produkten har varit att jobba modulärt så att spelet är lätt att expandera på och bygga ut - något som varit ett fokusområde för gruppen tidigt under projektet. Några av de idéer som skulle fungera bra som ytterligare moduler att implementera i framtiden vore:

- **Handla klimatsmart**: Produkter som t.ex. framställs på olika platser i världen - men säljs i Sverige - har olika nivåer av klimatpåverkan. Hur handlar man klimatsmart?
- **Utöka skolan med fler klassrum** som har andra teman än förnyelsebar energi. Hur kan man göra ”skolbaserad utlärning” och teknik engagerande och tilltalande?

I dessa scenarier kan användaren fortsätta ställas inför olika miljöval och problematik. Då projektet i detta fallet skulle expandera in på djupare typer av inlärning och liknande, och avgräsningarna därmed ändras, kommer även teorin kring detta område att omvärderas och med största sannolikhet behöva utökas/revideras.

7 Sammanfattning

EnVRment har haft som mål att ge kunskap och upplysningar om miljö och hållbar utveckling samt integrera barn och ungdomar i åldern 10-15 år som har olika kulturella ursprung, och inspirera dessa till att så småningom utbilda sig vidare inom tekniska områden. För att uppnå detta har ett VR-spel tagits fram. Detta VR-spel har konstruerats till att berätta och förmedla en s.k. *”A Day in your Life”* - hos en möjlig användare i vår målgrupp - där denna kan lära sig om detta.

Först, introduceras användaren till en köksmiljö, i vilken den första uppgiften består i att sopsortera på rätt sätt. Om försöket misslyckas förflyttas man till en s.k. konsekvensscen, som speglar miljöeffekterna av att materialet hos just det skräp man sorterat sämst, behöver utvinnas i högre grad.

Nästa uppgift för användaren är att se till att minska sin energiförbrukning i hemmet - genom att stänga av diverse objekt i köket som drar ström. Om användaren misslyckas, så tas man vidare till en konsekvensscen där ett stort kolkraftverk som driver objekten i köket, kan observeras. Ifall användaren lyckas med att förminska sin energiförbrukning - så kommer denne i samma konsekvensscen - få möjligheten att trycka på en stor knapp som stänger av och nedmonterar kolkraftverket - samtidigt som miljön ses åter florera.

Därefter får användaren ta ställning till vilket transportmedel denne använder för att ta sig in till skolan. Valet står mellan bilen eller att cykla. Beroende på vad man väljer så kommer skolgården som man hamnar på - att se annorlunda ut (detta fungerar som en slags konsekvensscen likt ovan).

Det sista användaren får i uppgift - är att lära sig om olika typer av förnybara energikällor. Genom att konstruera olika anläggningar på små miniatyr-landskap/biomer på ett antal bord i ett klassrum på skolan - så skall detta förmedlas.

Vår värld står inför enorma utmaningar när det kommer till miljö och hållbarhet. Många av dessa utmaningar och dess konsekvenser - kan ofta bli alltför stora och komplicerade, och därmed abstrakta - i utlärningssituationer. Vi vill därmed se till att presentera problematik och lösningar som är relevanta för v˚ ar målgrupp (10-15 åringar). Vi vill visa att det finns saker i *deras* vardag - där de kan göra enorm skillnad och påverka miljö- och hållbarhetsarbetet på ett positivt sätt.

References

[1] Betty Tärning (2018): ”Review of Feedback in Digital Applications - Does the Feedback They Provide Support Learning?”

[2] Kim Goodwin (2009): *Designing for the Digital Age: Creating Human-Centered Products and Services*. Indianapolis: Wiley Publishing, inc.

[3] Jesse Fox, Sun Joo (Grace) Ahn, Joris H. Janssen, Leo Yeykelis, Kathryn Y. Segovia & Jeremy N. Bailenson (2015): ”Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence”, *Human–Computer Interaction*, 30:5, 401-432, DOI:10.1080/07370024.2014.921494

[4] James J. Cummings & Jereny N. Bailenson (2015): ”How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence”, *Media Psychology*, 00:1–38, DOI: 10.1080/15213269.2015.1015740

[5] Don Norman (1988): The Design of Everyday Things: The Revised and Expanded Edition, *Design Psychology*.

[6] Sun Joo Ahn (2011) ”Embodied experiences in Immersive Virtual Environments: Effects on pro-environmental attitude and behavior”, hämtad 2018-10-31, http://vhil.stanford.edu/mm/2011/ahn-embodied-experiences.pdf

[7] Samsung, Samsung HMD Odyssey, hämtad 2018-10-19, https://www.samsung.com/us/computing/hmd/windows-mixed-reality/xe800zaa-hc1us-xe800zaa-hc1us/

[8] HTC & Valve, HTC Vive, hämtad 2018-10-19, https://www.vive.com/eu/product/

[9] Världsnaturfonden, hämtad 2018-10-23 http://www.wwf.se/wwfs-arbete/klimat/det-har-kan-du-gora/1609668-det-har-kan-du-gora-klimat

[10] Planlösning, hämtad 2018-10-23, https://sv.wikipedia.org/wiki/Miljonprogrammet

[11] Köksinredning, hämtad 2018-10-24, http://byggconstruct.se/miljonprogrammet/

A Demonstration av prototypen

Här återfinns två stycken videor som fungerar som enkla demo's till prototypen för Interaktion 1 (MAMN10); en kortare och en lite längre.

Youtube - Kort Demo (2min)

Youtube - Lång Demo (5min)

B Testspecifikation

Test 17/10
Vad vi vill få veta:

- Är det roligt att spela?
- Är någonting svårt att förstå?
- Lär man sig något, och i så fall vad?
- Vilka är kunskaperna och tankarna kring miljöarbete?

Den sista frågan är intressant men lite sekundär. Allra främst är det VR-upplevelsen som ska utvärderas. Dock har ju grundantaganden i vår produktutveckling varit att kök är ganska lika disponerade (avs placeringen av sopkärl) och att man vet vad kärlen är och att det finns ett samband mellan de produkter man slänger och att någonting händer i vår omgivning (dvs att det behövs råvaror för att framställa dem).

50 min till förfogande, test med VR kit. Dela upp gruppen i 2, den ena halvan (grupp 1) får testa VR, andra (grupp 2) diskuterar med Magnus och ytterligare någon. Efter halva tiden byter man. Grupp 1 får svara på frågor kring hur dom tänker kring miljö, grupp 2 får prata om VR-upplevelsen. Diskussionerna för alltså i fokusgrupper, rimligtvis om 5-6 personer. (Vi tar höjd för att testen tar ca 5 minuter per person, och därmed kan vi få 50-(5x5)=25 minuter för diskussion. Vi tror dock att detta kanske är väl lång tid för en diskussion.) Vår erfarenhet säger att diskussionerna ska föras i grupper bestående av personer som känner varandra (vilket de lär göra rätt hyggligt, då de går i samma klass). Detta ger bättre förutsättningar för ett fritt samtal. Vi låter vår kontaktperson (läraren) på skolan välja ut några elever som både känner varandra, och "funkar" och kan fokusera i grupp.

Diskussionsfrågor grupp 1 (20 min):

- Vad har ni fått lära er om miljöproblem och klimatförändringar? 2 min
- Brukar ni tänka på miljöproblem? 3 min
- Vad tycker du är den viktigaste miljöfrågan? 4 min
- Gör ni något hemma, för miljön? Vad för något? 6 min
- Sopsorterar din familj? Hur går det till? 5 min
- (Om man behöver utfyllnadsfråga: Tror du att det skiljer sig mellan din familj och dina kompisars familjer?)

Diskussionsfrågor grupp 2 (20 min):

- Var tyckte ni om det?
- Varför tyckte ni det/vad var kul eller tråkigt?
- Var någonting svårt? (Ex. tekniskt att använda VR, spelkoncept eller förstå vad ekorren menade)
- Vilken miljö hamnade ni i efter att skräpsorteringsspelet tog slut?
- Varför tror ni att ni hamnade där?
- Såg ni djuren? Vad ville de säga?
- Vad tar ni med er från spelet?

Under diskussionerna är tanken att moderatorn låter samtalet "gå den väg det tar", såvida vi får med våra frågeställningar inom tidsramen. Frågan är om en observatör ska föra anteckningar under diskussionen, eller om vi ska spela in den? Kanske räcker det att notera vilka olika punkter (t.ex. att det var svårt att gripa tag om skräpobjekten) som dyker upp i samtalen, och om någon punkt är genomgående (dvs om t.ex. detta problem gällde för många eller alla).

Under test med VR-set:

- Generella observationer
- Få screencaps

Förberedelser
Testa med någon som aldrig har testat VR förut, hur lång tid tar det? Kan vi använda för att uppskatta hur många personer vi hinner gå igenom. Gruppdiskussioner - vad ska vi ställa för frågor? Vilka "observationspunkter" ska vi ha under VR-sessionen? Hur ska vi ta ut dom som ska testa? Fixa manus - va ska vi exakt säga? Screen caps (OBS)

- Genomförande - Innan 10:05
  - Fixa uppsättning av VR-utrustning
  - Sätta upp screencap
  - Recap testplanering
  - Intro med Olof
  - Kan vi stanna kvar och testa på rasten?
  - Han får ta fram de 8 testpersonerna - vikt på varierande socioekonomiska/kulturella bakgrunder.
- Intro med klassen 10:05 - 10:10
  - Vad vi heter och vart vi kommer ifrån
  - Vi vill genomföra tester av vårt VR-projekt som riktar sig mot deras åldersgrupp.
  - Olof får ta fram testpersoner, fyra personer går med Magnus & Johannes, tre sitter kvar i klassrummet och en person följer med Felicia och testar VR (när denne är klar går den tillbaka till lektionen).
- Test 10:15-10:40
  - **Pre-VR diskussion**
    Ni kommer senare att få testa en VR-miljö som handlar om miljö och hållbarhet. Vi kommer ställa några frågor till er som vi kommer diskutera kring tillsammans. (Ställer frågorna ovan)
  - **Post-VR diskussion**
    Vi kommer ställa några frågor till er som vi kommer diskutera tillsammans. Men innan vi börjar, vill vi bara säga att vi har inte personligen utvecklat den här miljön, så ni kan vara ärliga med det ni säger. Vi kommer inte ta illa upp om ni ger oss kritik, utan er ärliga åsikt är viktig för oss. (Ställer frågorna ovan)
  - **VR test**
    Jag heter Felicia, och jag kommer vara här under hela testet, så har du några frågor eller funderingar så hojta bara till. Innan vi börjar vill jag bara säga att vi testar det här spelet, och inte hur duktig du är. Så känn ingen press. Är du redo?

C Skyltar i Konsekvensscenerna

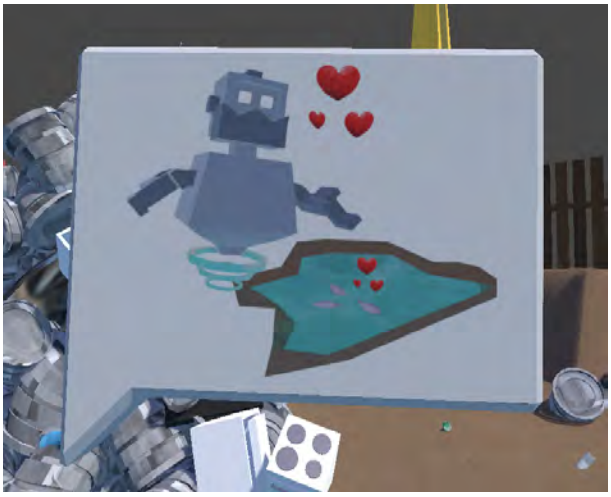


Figure 12: Metall: 1

Figure 13: Metall: 2

Figure 14: Metall: 3

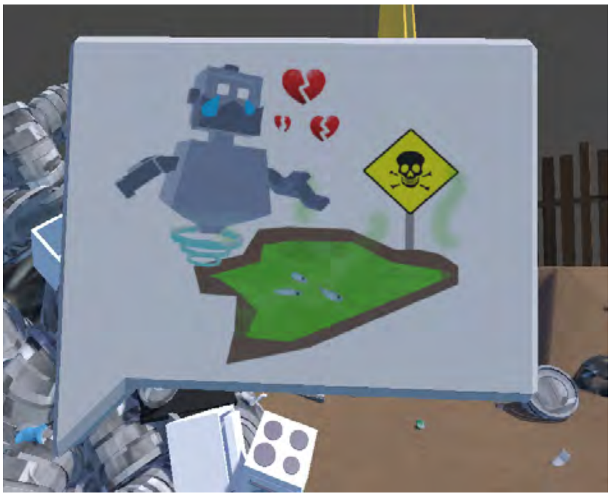


Figure 15: Metall: 4

Figure 16: Papper: 1

Figure 17: Papper: 2

Figure 18: Papper: 3

Figure 19: Papper: 4

Figure 20: Plast: 1

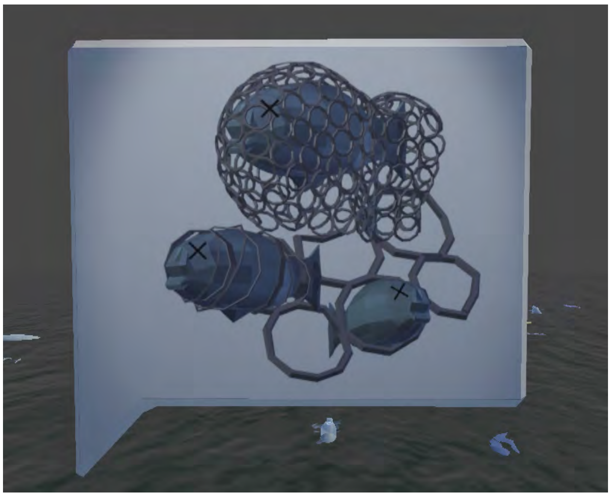


Figure 21: Plast: 2

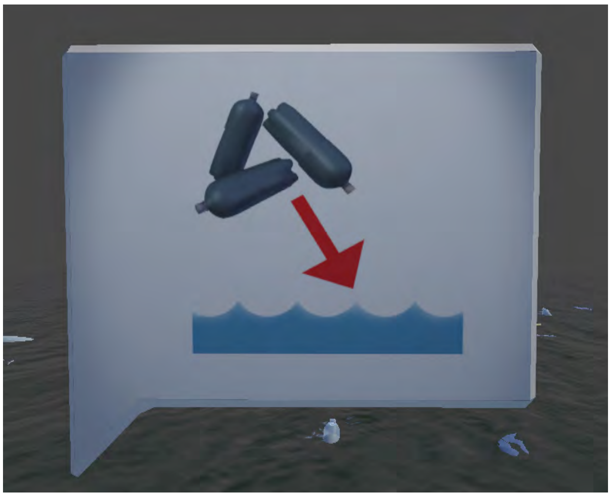


Figure 22: Plast: 3

Figure 23: Plast: 4

Figure 24: Infotavla i köket som pekar till Sopsorteringsspelet

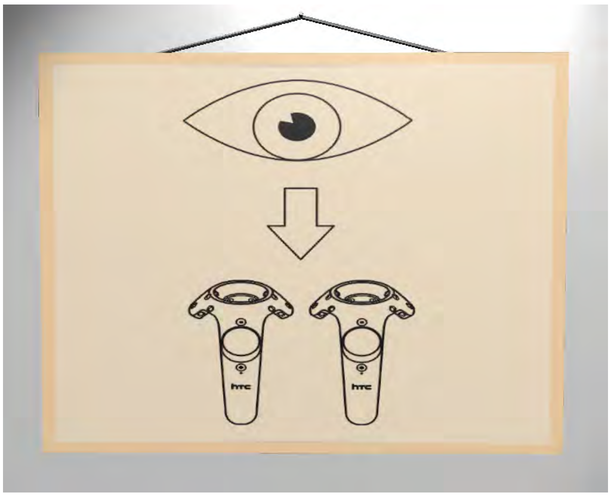


Figure 25: Infotavla som visar handkontrollerna

Figure 26: Infotavla som visar EnVRments logga

Going through the motions of conversation: Social movements in interacting robots

Christian Golcic, Gustav Harrysson, Hampus Johansson, Per-Magnus Ekström, Tora Hellgren

2019-01-20

During social interaction, individuals tend to adjust to each other behaviourally. These adjustments include phenomena such as mimicry and interactional synchrony, meaning that the persons imitate each other or adapt to a mutual rhythm. To investigate the dynamical aspect of social interaction, social movements of interaction partners were mapped onto two robots. Interactions were filmed using Microsoft Kinect, and depth data from the sessions were later used to simulate robot social interaction. The movements of the two robots gave the appearance that they were interacting, but the context was not apparent. A survey showed that a majority of the participants thought that the robots were communicating. As a development of the project, the recorded depth data was used in conjunction with a Bayes classifier, resulting in the ability of a humanoid robot to adapt its head movements to a human interaction partner. The humanoid robot was displayed at an interactive technology convention at Lund University, and visitors were allowed to interact with the robot. The overall response from visitors was positive, though some commented that the robot frequently looked away from them. While the social skills of the robot needs improvement, its current capacity hints at a possibility for socially resonant robots. Robots with a capacity for social resonance may in the future function as a conversation partner for elderly, or as a therapeutic tool for learning social skills in children with autism.

1 Introduction

Imagine two persons engaged in a conversation. As one of the interactants talks, the rhythm of speech is accompanied by body movements in both interactants. Focusing solely on the body movements of the interactants, it is possible to deduce how engaged they are in the topic of discussion or whether they agree or disagree. It is also possible to get a feeling of the overall atmosphere of the ongoing interaction. Simply put, social movements can tell a lot about a conversation.

Now, imagine instead two robots engaged in a conversation with each other. What would you see? Something like interaction sketched above? Or two rigid creatures exchanging monotone speech back and forth, possibly accompanied by one or two awkward gestures? Our guess is the latter alternative. Even so called social robots today lack many of the aspects of human social interaction. But what if it is possible to have robots dynamically adjust their movements to the rhythm of the conversation? Thus making possible an interaction more similar to human conversation.

Aim and social dimensions

The aim of our project has been to investigate the body movements that accompany casual social interaction. This has been accomplished in two ways. Firstly, the coordination of human body movements has been mapped onto two small robots, thus simulating the behaviour patterns that are involved in conversation. Secondly, with the aid of a naive Bayes classifier, a machine learning algorithm, the humanoid robot EPI has learned to respond dynamically to social movements in real-time.

The project group consisted of five members, excluding the project supervisors. Three members had a background in engineering and programming, while two members were cognitive students with a background in the behavioural sciences. The distribution of the different tasks within the project was thus an easy thing to accomplish. The programmers took care of the technical implementation, and the others worked on the theoretical background. The forming process of the group was almost self-regulatory, and the norming as well went on smoothly, with the supposedly storming part resembling just a gust of wind, resulting in an atmosphere with a focus on performing.

Bodily mimicry and interactional synchrony

While the ability to use language separates humans from other animals, non-vocal behaviour has been shown to play an important role in human social interaction. This means that interactants do not just respond verbally to their interaction partner. Instead, bodily mimicry and synchronization of body movements are important aspects of social interaction.

Studies have shown that people unconsciously mimic the behaviour of their interaction partner. For instance, Chartrand and Bargh [1999] found that the tendency to rub one's face or to shake a foot increased as a result of interacting with a person performing one of these behaviours. That is, face rubbing increased in the test subjects when interacting with a person who, repeatedly during the interaction, rubbed his face. When interacting with a person who instead shook his foot, foot shaking increased in the test subjects. Interestingly, none of the test subjects being manipulated in this way reported being aware of these behaviours being performed by their interaction partner. This suggests that this form of bodily mimicry is a nonconscious phenomenon.

Chartrand and Bargh [1999] refer to this type of mimicry as the *chameleon effect*. Besides bodily mimicry, interactants also tend to synchronize the rhythm of speech and gestures to match their partner. This means that interactants adjust not only to what behaviour their partner is performing, but that they also adjust temporal aspects of speech and body movements to match the partner. This coordination of body movements has been called *interactional synchrony* [Bernieri et al., 1994] or *nonverbal synchrony* [Tschacher et al., 2014]. Synchronization of body movements in dyadic interaction appears to be related to the degree of agreement between the interactants. Studies have shown that people tend to value the degree of rapport in a dyad as higher when the interactants show a high degree of synchrony. This effect was found even when the test subjects were watching a video simulation of two walking stick figures [Miles et al., 2009]. Similarly, a high degree of experienced rapport is correlated with interactional synchrony [Bernieri et al., 1994].

In a study investigating the effect of conflict, it was found that interactional synchrony is disrupted when the interactants differ strongly in opinion [Paxton and Dale, 2013]. If interactants instead are engaged in an affiliative, fun task, there is a higher level of interactional synchrony [Paxton and Dale, 2013] [Tschacher et al., 2014]. Finally, there is a link to affect, where positive affect was associated with synchrony [Tschacher et al., 2014]. That is, test persons who reported being in a good mood also exhibited a higher level of synchrony.

The existence of mimicry and interactional synchrony might at first glance seem like a peculiar aspect of social interaction. Lakin et al. [2003] suggest that mimicry has evolved to function as a social glue, making interaction partners better attuned to each other. In accordance with the idea of mimicry as a social glue is the thought of alignment as central to human conversation [Garrod and Pickering, 2004]. Conversation partners are skilled at adjusting to each other, adapting at several levels from behaviour to sentence structure. It is this capacity for alignment that makes conversation easier than holding a speech. Kopp [2010] expresses a similar view, referring to the coordination of interaction partners as social resonance. This resonance acts as a way of telling the interaction partner that *"I know what you mean, I have a similar gestural representation of it, and I know how it feels like"* [Kopp, 2010, p. 592]. In other words, mimicry and synchrony can be thought of as ways to coordinate the cognitive states of the interactants.

Social robots

Knowledge of how humans interact with each other when communicating is highly relevant to the area of social robotics. Since studies show that the degree of mimicry and synchrony has impact on the affective state of the interactants [Tschacher et al., 2014], and also affects perceived [Miles et al., 2009] or experienced rapport [Bernieri et al., 1994], it is likely that interaction with a robot with the ability to synchronize its body movements to the interaction partner will affect how people experience the interaction. If the robot manages to synchronize to its human interaction partner in a successful way, it is likely that the interaction will be perceived as natural and easy-going. It is also likely that humans will experience relating to the robot and that successful synchronization will lead to a feeling of mutual understanding. That is, people interacting with the robot will probably find the interaction more engaging when the interaction has some aspect of synchrony. It is therefore possible that people will prefer interacting with a robot with a capacity for social resonance compared to interacting with a robot lacking this quality. These predictions can be compared to the results of Mörtl et al. [2014], where test subjects described interacting in joint action with a robot that synchronized to its human partner as pleasant. Test subjects also reported that the perceived reactiveness of the robot made it appear lively.

Lindblom [2015] distinguishes two perspectives in social robotics; the aim to develop robots that are able simulate human social interaction and the aim to develop robots that are good at interacting with humans. The first perspective has as its goal to better understand human social interaction, while the second perspective is more focused on designing robots that will function effectively in interaction with humans. Bartneck and Forlizzi [2004, p. 592] give the following definition of a social robot: *"A social robot is 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."*

Interestingly, the focus in this definition is on interaction with humans. The authors further state that a robot that interacts with other robots cannot be defined as a social robot. This would mean that our simulation prototype consisting of two robots engaged in the motions of a human conversation does not, at least according to Bartneck and Forlizzi [2004], count as social robotics. Still, the interacting robots can be seen as a first step towards designing robots that have a better ability to engage in social interaction with humans. Our second prototype, the socially responsive humanoid robot, does on the other hand fit well into the definition of Bartneck and Forlizzi [2004]. Ultimately, a robot with a capacity to coordinate its movements to its interaction partner will make social interactions between human and robot more similar to human-human interaction. A capacity for social resonance is also one way in which a robot can be designed to follow the behavioural norms of people.

2 Theoretical backdrop

Embodied cognition

The field of embodied cognition, which can be summarized as the view that processes that underlie embodied actions are essential for cognition [Lindblom, 2015], is in several ways relevant to the discussion of social interaction and social robots. First of all, the view that behavioural resonance is necessary for social cognition, as expressed by Kopp [2010], is a clear example of how cognition is embodied in that movements of the physical body are a part of the cognitive processes of the interactants. Second, the need for embodied cognition is also expressed in the artificial intelligence field, but how this embodiment is interpreted can have serious effect on how robots are designed and how well social robots interact [Lindblom, 2015]. Lindblom [2015] suggests that embodied cognition in artificial intelligence often has been equated with the need for a cognitive agent to have a physical body, neglecting the impact of how a biological body is an essential part of the cognitive system of humans. While a biological body may not be necessary for a cognitive agent, it may be that the body of a robot in some ways must reflect the capacities for action and interaction inherent in a biological body.

Social embodiment

Barsalou et al. [2003] present four types of social embodiment effects that link cognition to bodily states;

- Perceived social stimuli produce bodily states.
- Perceiving bodily states in others produces bodily mimicry in the self.
- Bodily states in the self produce affective states.

- The compatibility of bodily and cognitive states modulates performance effectiveness.

An example of the first type of social embodiment effect is the fact that subjects primed with an elderly stereotype tend to walk slower compared to control subjects not primed with the stereotype in question [Barsalou et al., 2003]. This means that the subject of the conversation may influence how the interactants move.

The second embodiment effect is essentially identical to Chartrand and Bargh's (1999) chameleon effect; that is, people tend to mimic the behaviours of their interaction partners. Barsalou et al. [2003] distinguish three types of mimicry: bodily mimicry that includes body movements, mimicry of facial expressions and finally communicative mimicry, where conversation partners match the other's speech rate or utterance duration.

The third embodiment effect, that bodily states produce affective states, refers to the fact that a person's mood can be manipulated by manipulating the person's posture or facial expression. For instance, if a person is asked to hold a pen between his teeth while reading a cartoon strip, he will rate it as funnier than if the pen were held between the lips [Barsalou et al., 2003]. This has to do with the fact that facial muscles related to smiling are activated in the first case, while facial muscles related to frowning are activated in the latter case.

The link between bodily and affective states can in the case of social interaction cast a new light on the need for mimicry. To mimic postures or facial expressions of the interaction partner can lead to sharing affective as well as bodily states. The nonconscious mimicry can thus work as a way to align feelings or attitudes, resulting in the interactants being better attuned to each other. Again, an example of social resonance [Kopp, 2010] where the adoption of the partner's physical movements influences affective states in both interactants.

The last embodiment effect described by Barsalou et al. [2003] indicates that performance is more effective when bodily states and cognitive states are matching. For instance, a test subject asked to shake his head repeatedly while studying a list of adjectives, is later better at remembering the negatively valenced adjectives. If test subjects instead are asked to nod their heads as in agreement, they are better at remembering positively valenced adjectives [Barsalou et al., 2003]. This compatibility effect suggests again a clear link between body movements and cognition.

Mirror neurons

There is also neurological data suggesting that acting can be related to understanding. The human brain is thought to include a system of mirror neurons that are essential for action understanding and social cognition [Rizzolatti and Fabbri-Destro, 2008]. Mirror neurons, first identified in macaque monkeys, are motor neurons that are active both during the performance of an action and when seeing someone else perform the same action. Rizzolatti and Fabbri-Destro [2008] claim that the role of mirror neurons in monkeys is restricted to action understanding, while the human mirror system has a wider function including imitation and understanding intentions and feelings of others.

It is worth noticing that the role of mirror neurons has been questioned, especially in relation to action understanding [Hickok, 2008]. Hickok and Hauser [2010] suggest that the role of mirror neurons in monkeys may not be to aid action understanding, but may in fact be essential for low-level mirroring behaviours such as yawn contagion. This idea is interesting in relation to human social movements, since both bodily mimicry and interactional synchrony are unconscious phenomena. It is therefore possible that behavioural resonance in humans is neurologically grounded in a mechanism that is part of our primate heritage.

Gesture and cognition

Another example of how body movements are connected to cognition is the fact that gestures can be seen as a tool for thinking. Cartmill and Goldin-Meadow [2016] use the term *co-speech gesture* to refer to the gestures that accompany speech. These gestures do not carry meaning in themselves, but instead illustrate the spoken word. It has also been shown that persons allowed to gesture when explaining the solution to a math problem, while at the same time remembering a string of letters, performed better than persons doing meaningless hand movements [Cook et al., 2012]. The authors explain this finding by suggesting that free co-speech gesture lightens the load on working memory. These results are interesting in relation to interactional synchrony in that meaningless rhythmic hand movements did not enhance performance. It is therefore not the rhythmic aspect of co-speech gesture that is beneficial to the thought processes of the speaker. This does not preclude that rhythmic movements are important for the interaction as a whole.

Dance metaphor versus information transmission metaphor

Remember the fluid human communication described in the introduction, and the contrasting robot exchange of speech? These two examples can serve to describe two metaphors of social interaction; a dynamic dance metaphor and a traditional information transmission metaphor [King and Shanker, 2003]

The robot exchange of verbal information is a good example of how social interaction has been described as the transmission of information. Communication is thought to occur when a receiver decodes information from a sender [King and Shanker, 2003]. Interaction is thus seen as a linear event, where each partner is either the sender or the receiver of information. The information transmission metaphor is very much in line with how cognition was perceived in the early days of cognitive science, when the mind was likened to a computer. Since this metaphor for a long time was dominant in various fields studying communication, it is perhaps not surprising that we envision robot social interaction to be in accord with the idea of social interaction as information transmission. A possible reason why social robots are not very social may have to do with the fact that communication is still being viewed as first and foremost transmission of information [Lindblom, 2015].

King and Shanker [2003] suggest an alternative to the idea of social interaction as information transmission. The dance metaphor instead focuses on the dynamical aspect of communication. That is, both partners are continuously active as they adjust their behaviour to the partner. In short, communication is a lot like a dance. The dance partners together form a unit, carefully synchronizing their movements. Similarly, communicating partners coordinate their movements. The idea of a link between dance and communication is also expressed by

Michalowski et al. [2007] who suggest that dance can be seen as a formalized version of the rhythmic synchronization typical of normal social interaction. Their dancing robot Keepon, which synchronizes to sound or movement, can therefore be seen as a way to explore interactional synchrony.

Interacting social partners as a dynamic system

King and Shanker's (2003) dance metaphor for social interaction is based on the view that interactants together form a dynamic system. Drawing on dynamic systems theory [Thelen and Smith, 1994] they mean that interactants coordinate their behaviour through a process called co-regulation. Dynamic systems theory has been applied to physical as well as biological systems and states that complex systems tend to self-organize, whereby stable patterns emerge [Thelen and Smith, 1994]. Schmidt et al. [2012, p. 277] suggest that the reason why conversation partners easily coordinate *"may be that a combination of complex rhythmic activity and their interpersonal synchronization provides a basis for two individuals to become a single dynamical system"*. That is, the dyad becomes a single unit through interactional synchrony.

Synchronization in dynamic systems can have two outcomes: in-phase synchrony at 0°or anti-phase synchrony at 180°[Schmidt et al., 2012]. In social interaction this means that interactants can synchronize by acting at the same time, or by taking turns. The most apparent example of turn-taking is the fact that interactants take turns being in speaking or listening mode. The dynamics of anti-phase synchrony has been explored in the case of robot-robot interaction. Prepin and Revel [2007] used oscillatory dynamics to create a situation where each agent would adjust its behaviour to its interaction partner. Perceiving movement in the interaction partner led to an inhibition of the agent's movements, thereby creating a situation where turn-taking emerged. This example of robot-robot social interaction that simulates how humans interact with each other is especially interesting since there is a certain similarity to our project. The difference is that Prepin and Revel [2007] created a situation in which rhythmic aspects of social interaction emerged dynamically, whereas our goal in robot-robot interaction has been limited to mapping the social movements of human interactants onto two robots.

3 Simulating social movements in robot-robot interaction

As described in the introduction, the aim of the project was to map mimicry and synchrony patterns in human interactants engaged in a conversation, with the purpose of using the results to endow a humanoid robot with the ability to adjust head movements to its human interaction partner. A first step in this direction was to use the mapped human social movements to simulate the motions of conversation in two robots. The theory was that by simulating the behaviour patterns of the interactants, the robots would be perceived as having a pleasant conversation. To accomplish this, three main goals were conceived: (1) collecting a data-set with 3D-movies from a Kinect-recording of two persons during a *fika-session*. (2) research for visual segmentation methods in order to find fields of movement in the movies, and (3), writing programs mapping the detected movements onto two robots.

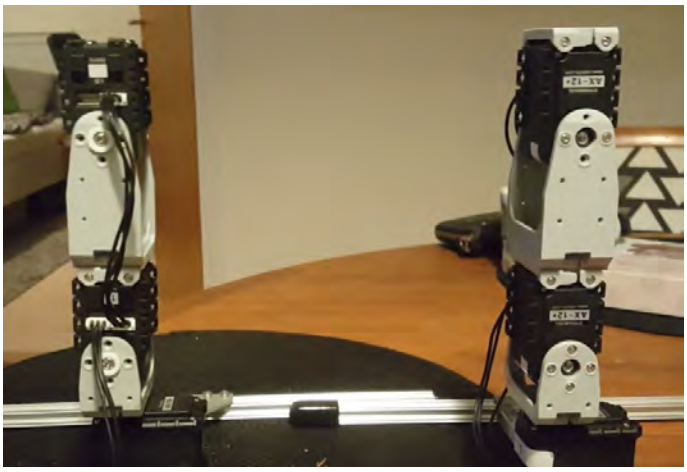


Figure 1: Robot arms

Methods

Recording of social movements was made using Microsoft Kinect, a motion sensing input device developed by Microsoft. It has support for recording regular video as well as depth maps, meaning it senses the distance from the camera to the object it is viewing. For this project Kinect for Xbox 360 was used.

The robots used for simulating social movements were two simple robot arms, were each robot had three servos enabling movement. The two robots are depicted in figure 1. The movement of the robots was controlled using Ikaros, an infrastructure for system-level cognitive modelling developed by LUCS Robotics Group. Ikaros' functionality is based on modules, which may be connected to one another. A large number of modules are built into Ikaros, for example modules for processing images, depth and controlling robotics.

Recording of participants

A video-recording set-up was prepared at the Usability-lab at IKDC. One overhead camera was used during the whole session, together with a Kinect, which was used sporadically. The two participants were placed facing each other, seated at a table, with the Kinect filming them from the side. The test leader was hidden behind a mirror screen. As soon as the instructions to start were given, the test persons were free to start the *fika-session*, which was supposed to last for about fifteen minutes. After five minutes, or when the test leader saw clear signs of synchronization between the participants, recording with the Kinect commenced. Two sessions were recorded, meaning that two different dyadic interactions were included in the data.

The test persons were chosen from fellow students from other programs than ours, minimizing the risk for bias due to knowledge of the purpose of the film. They were already befriended with each other, increasing the likelihood of them having a pleasant conversation and thus increasing the occurrence of synchronization. The participants were given no information about the purpose of the project prior to the recording session. They were just asked to sit down and have a *fika*.

The choice of recording a *fika-session* was based on the assumption that such a social situation fulfills several demands crucial to obtaining an optimal starting point for the robot-mapping. The session was of an open and informal nature, which meant that the participants were free to act as they liked within the natural physical constraints of the situation, thus opening up for a greater amount of spontaneously induced actions and interactions.

Technical implementation

Based on a visual video analysis, one part of the recording done by the Kinect was chosen and served as a basis for the technical implementation. The main reason behind recording with the Kinect was the need for depth data, in order to analyze how the persons acted in the 3D-space, as well as filtering out data not connected to the test subjects.

By using Ikaros built-in Kinect module, the depth maps were saved as an array of 640x480 images in .raw format, with depth values between 0-1. In order to generate the original depth values, a multiplication was done when feeding the images back into Ikaros for further analysis.

In order to filter out data not connected with the test subjects, Ikaros built-in histogram module was used in conjunction with our own segmentation module. The histogram module was used to find "blobs", coherent values in the depth space, and to output the min and max value of the blobs on the depth axis. Since the persons recorded had quite coherent values in depth space, the min and max values returned by the histogram module contained the test subjects, but not for example the background wall.

By feeding the output from the histogram module as well as the original depth map into the segmentation module, we were able to find pixels in the image that had a high probability of belonging to the test subjects. This, since the pixels value was between the min and max values outputted by the histogram module. To separate the test subjects from each other, only one half of the image was processed at a time, since it was previously known that the test subjects were located at different sides of the images. To separate the head from the body, a value on the y-axis was chosen, where all values above the value was mapped as the head and all below to the body.

At this point it was decided that only the torso and head of the participants would be mapped onto the robot, so no further segmentation was performed. This was done due to the technical complexity of finding arms in a video, as well as the limitations of the robot. By having only 3 servos it could not accurately display arm movement without compromising the movement stemming from the head and torso movement. Consequently, the arms were decided not to be included in the mapping.

With a picture representing the head and body for each person, further analysis could take place. The movement of the body was calculated by taking the midpoint of the head and the midpoint of the body and calculating the angle between them. The movement of the head was calculated by taking the rows a quarter length from the bottom and from the top of the head. The rows were traversed towards the front of the head which resulted in two points, one located at the forehead and one on the chin. The movement of the head was then calculated by taking the angle between these points. This roughly gave us the angle of when a person tilts his or her head up and down when engaging in a conversation, enabling us to capture, for instance, the movement of nodding. The value in degrees connected with the movements of the body and head, was outputted from the segmentation module.

By using Ikaros built-in Dynamixel module, the servos on the robot could be controlled. The output degrees from the segmentation module related to torso movement were sent to the second axis on the robot - the axis affecting the robot's "torso" torque. The output degrees from the segmentation module related to the head movement were sent to the upper axis on the robot, the axis affecting the robot's "head" torque. Thus making the robots move in synchronization with the filmed test subjects.

Results

An analysis of the video material was performed and revealed a difference in the interaction patterns of the two dyads. While the interaction in the first session was characterized by a high degree of head movements, the interactants in the second session used a variety of subtle as well as more conspicuous movements that included the use of of head, hands and the whole body.

The interaction of the second dyad was more varied and included examples of both anti-phase and in-phase synchronization. Here both movements of the torso and head movements were subject to in-phase synchrony. There was also a great deal of mirroring of postures and body movements in the second dyad. For instance, the interactants would lean forward at the same time, or would mirror arm positions such as letting the elbow rest on the table.

Based on the analysis it was evident that synchronization of movements between the two interactants took place during both sessions. Though, the clearest cases of synchronization were not filmed by the Kinect. We decided to proceed with the Kinect film that upon visual analysis showed most signs of synchronization. The photographs in .raw format served as input to the segmentation and histogram module. The histogram module provided the following output:

- *Min* - 656
- *Max* - 848

During the segmentation process, the coordinates relevant to the head and body are processed as specified in the method. The outcome of the processing was the degrees relevant for the angle of the torso and head for each of the interactants. The outputted head and body degrees for each of the persons were mapped onto the two robots to simulate them interacting with each other. The result of this is presented in video format in appendix A.

Evaluation of the simulated interaction

A questionnaire was distributed online and received a total of 25 replies. A short video clip was included and the questions focused on how the participants perceived the robot interaction. The first question in the questionnaire required a descriptive answer, and the answers to this question are analysed from a phenomenological perspective in the next section. The rest of the questions, except for an opportunity to add additional comments as a final question, were of a nature that were well suited for a quantitative description. The results showed that 80% of the participants thought that the robots were communicating, but only 64% also replied that the robots were engaged in a conversation. See figure 2 and 3 for a comparison. Due to the low number of participants this 16% difference only amounted to 4 individual answers, which might be a coincidence.

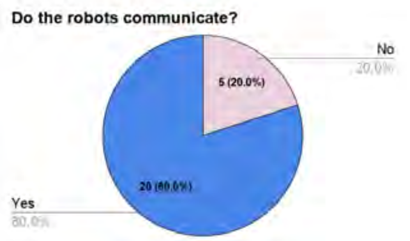


Figure 2: Do the robots communicate?

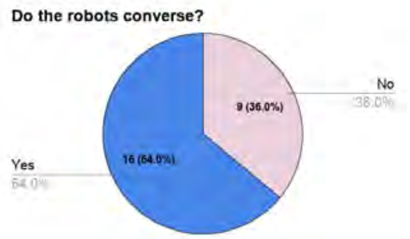


Figure 3: Do the robots converse?

The participants were also asked about how they perceived the robot relationship. 64% responded that the relationship was characterized by friendliness, while 28% thought that a competitive relationship best described the robot interaction. 8% perceived the relationship as hostile. These results, depicted in figure 4, can be compared to the results regarding the nature of the interaction.

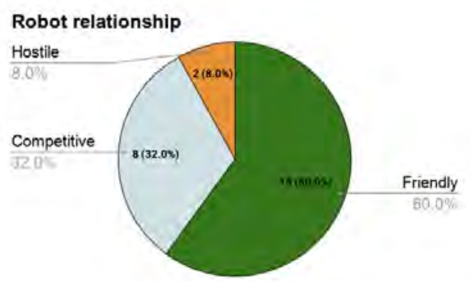


Figure 4: Robot relationship

To get an idea of how the robot interaction was interpreted, the participants were also asked to rate the interaction in terms of friendliness, naturalism, intensity and synchronization. The form was designed to let the participant rate the interaction, using a scale of 1 through 5, in relation to two contradictory descriptions, such as natural versus unnatural or friendly versus unfriendly.

Regarding friendliness, 36% responded that the interaction was quite friendly, while 28% replies were neutral and another 28% thought the interaction was somewhat unfriendly. See 6 for a diagram. These results show that while respondents preferred to describe the robot interaction as friendly when compared to competitive or hostile, they did not rate the interaction as entirely friendly but rather as neutral, or somewhat friendly or unfriendly.

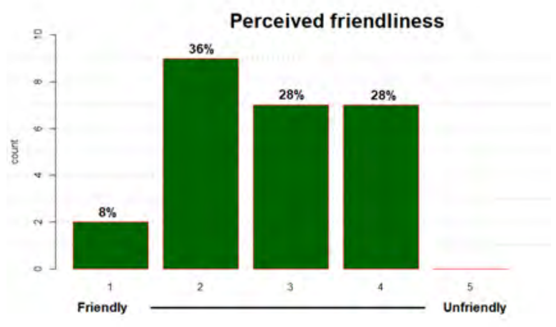


Figure 5: Perceived friendliness

When asked about the intensity of the interaction, the responses were evenly split between quite relaxed (40%) and quite intense (40%). Interestingly only 8% chose to describe the interaction as neutral, meaning that the majority of the respondents saw the interaction as either somewhat relaxed or somewhat intense.

A bit disappointing was the result that only 24% of the participants thought that the interaction was more or less synchronized. Instead 44% responded that the interaction could be described as disharmonious, to a lesser or greater extent. There might be many reasons behind this result. What we believe to be the most feasible explanation is the fact that our analysis and implementation only encompass a few of the involved factors and variables that we know of today.

Similarly, the answers regarding the naturalness of the robot interaction showed a result that was quite evenly distributed around a neutral answer (44%), with 24% and 20% of the answers given suggested a somewhat natural or somewhat unnatural robot interaction. The high number of neutral answers might indicate that the eventual naturalness of the robot interaction was hard to rate based on the short video footage. Additionally, since the nature of the interaction was not easily interpreted, it was perhaps not clear to the participants in what way the interaction would be seen as natural or unnatural.

Phenomenological Analysis

The opening question in the questionnaire was deliberately constructed to ensure that the participants comments were as free from influence from the experimenters as possible. The question was formulated as such: "how would you describe what is happening in the film?" This leaves room for the participant to to be either very economical in the description of the contents of the film, concentrating mostly of using words focusing solely on the observed behaviour of the two robots, or using a more narrative language, implying, implicit or explicit, a deeper level of explanation of the robotical behaviour. The answers to this question, 20 in total, were given a phenomenological analysis in order to examine what kind of impression the film had evoked in the participants.

Phenomenology

The Transcendental Phenomenology was formulated by the philosopher Edmund Husserl in the beginning of the twentieth century. It could very briefly be described as an attempt to focus on descriptions of the world as it is experienced by us humans. Phenomenology is interested in the phenomena as it presents itself to us within certain contexts and at certain times, and is not trying to extract general abstract formulations about reality. From the phenomenological point of view, our experience of an object is inseparable from the object itself. Grounded in philosophy, the phenomenological approach has been in wide use in psychology, mainly because of its focus on consciousness and the personal experience. In practical psychological use, two methodological distinctions are made. The first is phenomenological contemplation and deals with introspective reflections upon personally perceived phenomena. This will not concern us here. The second, and the one put into use in this context, is phenomenological analysis, which deals with interpreting the experience of a research participant. The analyzed phenomena in this context were the participants comments on the robot film.

Method of analysis

The method in use takes its general description from Willig [2008], but was adapted to this special situation.

- All the comments were read and reread.
- Different themes in the text were identified and labeled. These themes were centered around concepts catching something essential in the text.
- The identified themes were structured into different groups, either by their common meaning or reference, or due to their hierarchical nature. They were given appropriate titles, chosen either by the analyzer, or by using "in vivo"-terms, i.e. descriptions found in the comments made by the participants.
- A summary table, table 1, was constructed, consisting of the structured themes together with quotations illustrating them. Normally, a column with the number of the page, and place, where the quotation could be found, would be a part of the summary table as well but due to the relatively limited amount of raw data, this was omitted.

This process makes constant use of the so called "hermeneutical circle", which means that during the process of interpreting the text, there is an ongoing procedure of returning to the original data to ensure that the concepts extracted from the text are valid representatives of the inherent meaning.

Cluster label 1: Definition of activity	**Quote/Keyword**
Theme label 1. Movement	"Two robots are moving"
Theme label 2. Interaction	"Two robots are interacting"
Theme label 3. Communication	"Two robots are communicating"
Cluster label 2: Definition of actor	
Theme label 1. Inanimate subject	"Robots"
Theme label 2. Animate subject	"Meerkats"
Theme label 3. Human being	"Persons"

Table 1: Summary Table

In table 1 the labels of the activity could then be lined up on a scale going from simple to complex, like this: *Movement - Interaction - Communication*.

The labels of the acting part could be lined up in a similar fashion, like this: *Inanimate Subject - Animate Subject - Human Being*.

Conceptual Grid

Using these labels and thematic definitions a conceptual grid was constructed. The numbers in the grid represents how many representations of the conceptual combinations were extracted from the comments. F.ex., the number "1" in the center means that one person thought that the film showed animate subjects engaged in interaction.

	Movement	Interaction	Communication
Inanimate Subject	4	3	4
Animate Subject	0	1	2
Human Being	0	0	6

Comments

Unsurprisingly, no one described the film as consisting of human beings just moving or only interacting. When the robots were described as human beings they were also described as being engaged in communication of some kind, in all cases described as verbal. When the robots were described as animals one case of interaction and two cases of communication was found. When the robots were described as inanimate the description of their behaviour was rather evenly distributed between the three concepts. Most interesting here is perhaps the willingness to attribute the concept of communication to the robots, in three cases even explicitly described as verbal communication, and in two cases described as engaging in sharing emotions! Considering the rather simple and mechanical look of the robots, this is a striking example of the human willingness to anthropomorphize even under sparse conditions.

Discussion robot-robot interaction

If we take a step back to get another, broader and wider view of the project, we can let go for a moment of the strict technicalities of the robots and the mapping process, and reflect for a while about what we are experiencing from the perspective of a human observer, or interactor, in the setting. Hopefully, the robots do what they should, depending on the mapped movements extracted from the film of the two socially interactive persons. But how do human onlookers perceive and interpret the robot-like movements from a, so to speak, more general phenomenological perspective? In what way should, and could, we describe what is happening between the robots?

Some clarifications and terminological explanations might be needed to further enlighten the the perspective. As human beings, we seem to have an inbuilt tendency to interpret several phenomena as being in possession of more than what immediately meets the eye. Humans seem to be hardwired in a way that makes us prone to interpret even rather simple visual phenomena as having a more complex background apparatus than there actually is. Take for example John Conway's game "Life", Gardner [1970]. This is a mathematical game played on a computer, (it could be played with paper and pencil if you have the time and patience). It starts from an initial setting, and obeying a few rules it starts creating patterns in a manner that easily makes you interpret them in terms of a goal-oriented behaviour like "surviving", "dying", "blinking", etc., despite the fact that the whole show is built upon strict rule-following. This is nothing new in itself, any game of chess, for example, could be described in terms of pawns "offering" themselves, or a rook "attacking", even if you know that all the pieces are doing is obeying the rules of the game as they are acted out by the player of the game, which for all we know nowadays even could be a computer. Leaving the human phenomena of anthropomorphizing aside for a while, we can try to zoom in on an appropriate terminology for describing or explaining various behavioural phenomena, be they naturally evolved or artificially constructed.

Stances

The kind of quality or property you attribute to either a fellow human being, an animal somewhere on the phylogenetic ladder, or an artifact, like a robot, could then be dependent upon what kind of stance you take when describing its behaviour. Dennett [1987] describes three different kinds of stances for interpreting behaviour: the physical stance, the design stance and the intentional stance. The physical stance means a focus on the physical constitution of the system, while the design stance means a focus on how the system is designed to function. The intentional stance, finally, suggests that the system should be treated as a rational agent with beliefs and other mental states.

If we take our robot arm as an example, the physical stance would be superfluous, since the construction of the robot is known, and its behaviour is fully determined by the programming and mapping procedures. The design stance, instead would be a better level of description. The robots' behaviours could be described in terms like the following: *"if robot A moves forward, then robot B moves backward"*. Most useful is finally the intentional stance. The robots can be treated as rational agents with goal-oriented behaviour grounded in a world-view. The description could be as follows: *"A leaned cautiously forward and nodded its head slightly, while B turned backwards a bit in surprise"*.

The way the robots are described gives away your opinion, be it explicitly encompassed or not, about what kind of qualities you are attributing to it. This is not the same as believing that the robot actually has has these qualities, but you can choose the level of behavioural description depending on what you want to achieve in terms of behavioural understanding and communication.

Descriptive concepts

Another sort of behaviour-interpretative approach is the *descriptive concept* described by Dennett [1996], that includes four levels of "creature"-concepts, all created with naturally evolved organisms as a matrix for description. The first one is named a Darwinian creature, named after the father of evolution, and consists of an organism with its behavioural repertoire completely hardwired from its birth, without any behavioural flexibilities. The second level is named a Skinnerian creature, named after the behaviorist who made the concept of conditioning his major explanational vehicle, and consists of an organism that responds to environmental stimuli and repeats responses that are reinforced. A third type of creature is the Popperian creature, consisting of an organism equipped with an inner selective environment that is able to pre-evaluate the outcome of an expected event. The last sort is named a Gregorian creature and consists of an organism with the ability of importing "mind tools" from the cultural environment, thus being able to improve both its inner as well as its outer environment.

If we scrutinize our small robot with these concepts in mind, as what kind of creature could it be described? At this point in its life-story, the robot perhaps could be said to place itself between a Darwinian and a Skinnerian creature, with its behaviour of a somewhat rigid kind in a Darwinian manner, with hints of a more Skinnerian kind, as its behaviour sometimes seems akin to reactions to environmental cues of a stimulus-response way. In the next part of its life, perhaps it will be able to interact dynamically in real-time, then giving the impression of being a Popperian creature.

This descriptive apparatus could be applied to the results of the phenomenological analysis of the questionnaire. If you choose to describe the robots behaviour solely in terms of movements, then you could be said to have taken the physical stance towards it. If you choose descriptions involving interaction, you could be said to have taken the design stance, and if you choose a vocabulary involving concepts of communication, you could be said to have taken the intentional stance. If you describe the robot as an inanimate subject, yet capable of involving in responsive interaction, this could be described as a description of a Darwinian or even a Skinnerian creature. If you describe the robot as an animate subject with the ability to interact or even communicate, this could be described as a description of a Popperian creature, and if you describe the robot as a human being involved in verbal communication, then your description is of a Gregorian creature. Within the questionnaires answers to the question of what is happening in the film of the two robots, all the different stances were represented, and all the different creatures could be said to be implicitly represented. Not all combinations were represented though, as the zeros in the conceptual grid shows.

4 Social movements in a humanoid robot: human-robot interaction

The second part of our project meant that we shifted the focus to real time human-robot interaction. Here our aim was that a humanoid robot would be able to use social movements in an interaction with humans, both by mimicking its interaction partner and by responding appropriately to social movements. The robot would therefore be able to engage in human conversations by adapting its head movements to a person standing in front of the robot, thus giving the impression that the robot could silently participate in a conversation between two persons.

Methods

Yet again, as specified in the method for section 3, the Microsoft Kinect for Xbox 360 was used, but this time two units were utilized. Though, the two robot arms were substituted for the humanoid robot EPI. As specified by the Lund University Cognitive Science robotics group: *"Epi is a humanoid robot developed by LUCS Robotics Groups. It was designed to be used in experiments in developmental robotics and has proportions to give a childlike impression while still being decidedly robotic."* LUCS. EPI, in the form we used it, consisted of a head with two eyes, two arms with corresponding hands and a torso, which were all movable. More specifically focusing on the head, EPI could move its head vertically to simulate for example nodding movements as well as rotating its head horizontally and thereby look from side to side.

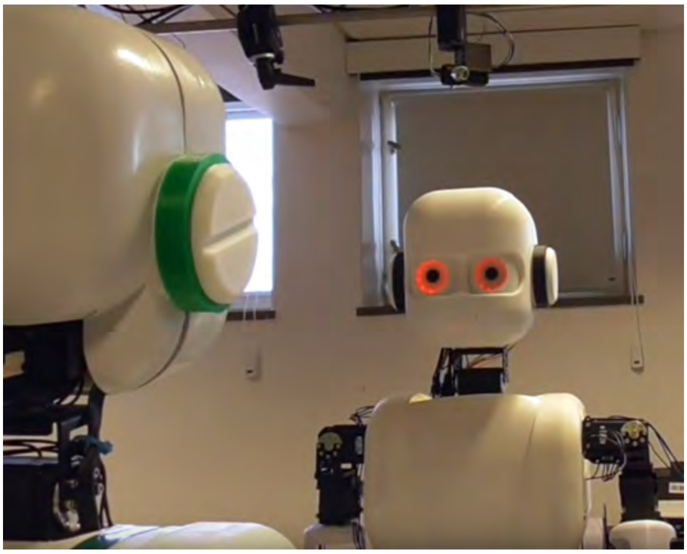


Figure 6: EPI

To control EPI, the Ikaros system developed by the Lucs Robotics Groups, was once again used. As with the robot arms, the movements of EPI were controlled by servos and these servos could be controlled by utilizing modules in Ikaros.

Recording of participants

Due to the fact that the robot should look directly at the person it interacts with, the need to record new videos of people interacting arose. This, since the previous videos only showed the interactants in profile, which not generated enough data regarding the horizontal rotation of the interactants heads. This time, due to time constraints, we recorded ourselves having conversations in a more unregulated manner compared to the previous recording sessions specified in section 3. The recording took place in LUCS Robot Lab and the two interactants were group members involved in the project. In front of each participant, a Kinect was placed to record the interactants from the front whilst having a conversation.

Technical implementation

The same methodology as specified in section 3 up to the head/movement tracking module was used. An adjustment of the module from section 3, which tracked movements and segmented the output data, was developed in order to capture the vertical up and down movements and the horizontal side to side movements of the head. By applying the depth data collected by the Kinect a z-axis could be created, which was needed to appropriately track the head from the front. By splitting the head in two parts, upper and lower, taking the centre point of the two halves based on the x and y values and calculating the angle between them based on the z-values; a representation of the vertical head angle was generated. Similarly, by splitting the head in a left and right part, taking the centre points of these based on the x- and y-values and calculating the angles between them based on the z-axis, resulted in angles describing the horizontal rotation angles of the head. This combined made up the new head tracking module.

In order for EPI to interact with a human, it needed some kind of perception of the human it would interact with. To achieve this, the Kinect was used in conjunction with the head tracking module in order to feed data to EPI that it could base its behaviour on. The choice was made to base the robots head movements on the head movements of the robots interaction partner. So in this case, the data mentioned, is data regarding the head movements of EPIs interaction partner.

In order to generate dynamic and natural head movements that would be based on EPIs interaction partners head movements, machine learning was used. To generate the data necessary to train a machine learning algorithm, the recording of the two group members interacting was processed using the head tracking module in order to generate data corresponding with the head movements of each recorded person. The data consisted of vertical and horizontal head angles for each of the interactants during the conversation. In order for the data to be useful it was split into segments, in our case 20 values long. Each segment could be seen as a short movement.

Furthermore, the mean and variance values per segment for person 1 was saved together with the corresponding segment for person 2 with an offset of 1 segment. The reason for saving the segment with an offset of 1 segment, is due to the fact that the movement of person 2 is a reaction to the previous movement made by person 1. The reason behind saving the mean and variance value was our choice of the machine learning algorithm, the naive Bayes classifier. The naive Bayes classifier was chosen both due to its speed as well as its ease of implementation.

In the final module, which is active when the robot interacts live with a human, two buffers with the same size as the previously recorded movements is filled with values from the Kinect and head tracking module. The mean and variance are calculated on the buffers and feed through the Bayes classifier where it is compared to the previously recorded mean and variance values. This will generate probabilities that the buffered movement is a previously recorded movement. The movement corresponding with the maximum likelihood outputted by the Bayes classifier will be performed by EPI, thereby reacting to the humans head motions. As specified above, the recorded movement stored together with the variance and mean values is the reaction to the movement that generated the mean and variance, therefore the movement performed by EPI will be a reaction to the movement it saw.

We also took into account that the maximum likelihood for a movement, generated by the Bayes classifier, will be low in some cases. In this case the robot performs the buffered movements instead of a previously learned one, thus mimicking its interaction partner with a delay of one segment. Mimicking, as specified in the theory section and section 3, also makes humans perceive the robot as a more interactive agent, which made mimicry a sensible default movement pattern. The threshold for filtering away the previously learned movement was set to 12%.

The movements generated were twitchy though, especially when the observed person was standing still. This was due to minute changes in head movement which made the robot shake. As a solution we implemented a filter which ignored changes which resulted in a very small movements. Outliers in the data were filtered away as well to protect the robot from turning too much too quickly. Lastly, a filter which took the sliding mean of the values of the head movement was applied as well to reduce the aggressive movements. In order to make the robot feel more lifelike, we also implemented a set of pre-programmed arm movements which were randomly cycled through to give the impression it being a real person in a conversation.

Conclusively this made EPI move its head smoothly based on a learned pattern and when no such pattern was found, EPI mimicked its interaction partner.

Results

The final result was a robot that could move its head dynamically based on real time data generated by the Kinect and head tracking module. EPIs most common head movements were small ones, quite similar to movements previously seen in the recorded material. Sometimes though, when interacting with a human, EPI performed large erratic movements; furthermore sometimes the head would get stuck in unnatural positions, for example looking away from the interactant. It also performed pre-programed arm movements to give a more lifelike impression.

Regarding the output from the naive Bayes classifier, the maximum probability usually ranged from below 1% up to 30% likelihood of being a previously seen movement. The most common probabilities was in range between 3% and 15%. Therefore, since the probability threshold was set to 12%, EPI performed lot of mimicry instead of previously learned actions.

Focus group reactions

The final product was demonstrated at an interactive technology convention at Lund University and visitors were given an opportunity to test the robot. They were instructed to stand in a specific place after which we explained to them what the project was. While we conversed with the visitor, an Xbox Kinect was recording their reactions to which the robot reacted. The overall reaction was positive and the visitors seemed to enjoy the experience, but some commented that it turned its head away from them.

Discussion social movements in human-robot interaction

There are several factors regarding the technical implementation that might have affected the outcome. Firstly, the position from which the Kinect recorded the participants and was placed during live in interaction with EPI, played a crucial part. This, since the position of the Kinect affected the input to the head tracking module and therefore also all of the angles outputted by this module. It is by no means certain that the Kinect could film from the same angle both when used to record the conversation and when used in the live robot interaction. Thus introducing errors in the comparison between the stored movements (mean and variance) and live movements made by the Bayes classifier.

Furthermore, the sometimes erratic behaviour of the robot could be explained by the fact that all movements are generated in a black box fashion, by splitting the recording into segments. In the context of the recorded conversation session a certain movement might have made sense, but there is no certainty that the movement made sense in the context with the live interacting robot. A further downside to our approach is that we always select the most probable movement in each case, therefore creating a robot that is highly opinionated in its actions and with less flexibility.

The processing of the real time data, when EPI interacts with a partner live, is also flawed. By only doing calculations for each segment, a significant error source is introduced. This is probably what decreases the likelihood of finding a match for a previously stored movement significantly. If a movement match is not found, the robot will perform a movement that is in the buffer, thus mimicking the human, and once this movement is done and the buffers are filled with entirely new values, the buffers are fed through the Bayes classifier again. An approach that most likely would have generated higher probabilities of finding a matching movement is to do calculations on the buffer each time a new data value is added and only mimicking until a high enough probability is found. Though, this method also has problems since the likelihood that a mimicking movement is interrupted increases, which could lead to unnatural movements.

The naive Bayes classifier is by no means perfect either. Compared to other approaches such as using artificial neural networks, the Bayes classifier, various decision trees etc. the naive Bayes classifier is usually less efficient. The naive Bayes classifier does not take previous movements into consideration either. When splitting the data into short segments in the fashion we have done, the previous movements usually has high correlation with the current movement. This covariance is entirely omitted by the naive Bayes classifier. Another machine learning algorithm might have been more performant, but such an algorithm might have required more data in order to generate a sufficient result.

The intention behind the behaviour of the robot was to have it act as a silent observer of a conversation. Visitors had a hard time acting naturally in front of the robot and would more often that not wave to the robot as their first action. Furthermore, the robot was programmed to mimic their movements in real-time in certain cases and since the "eyes" of the robot was an Xbox Kinect and not the actual eyes of the robot, people tended to look at the robot which made the robot look away, since it was mimicking their head movement in relation to the Xbox Kinect. Perhaps the most glaring problem was that the visitor knew that they were being filmed. The focus of attention may have caused some of them to pretty much stand still with little to no natural movement. Sadly, this caused the robot to sometimes give the appearance of being unengaged.

5 General discussion

Since EPIs response to a human interaction partner at times was far from perfect, this may be a good time to return to the contrasting descriptions of robot versus human conversation presented in the introduction. Is it in fact possible to develop dynamic social movement skills in social robots? The two prototypes developed in our project both hint at this opportunity; the small robots going through the motions of conversation and the humanoid robot that responds dynamically, at least some of the time, to head movements in social interaction all adapt their behaviour to their interaction partner.

In the robot-robot interaction, the robots follow the movement patterns of two friends conversing while having a *fika*. The actions of each robot are thus no more dynamic than the video recordings of the filmed session. There cannot be any new dynamics arising in this situation, since the robots simply follow the recorded material. In contrast, the EPI robot has through machine learning been given a capacity, albeit limited, for acting dynamical in real-time interaction. This includes picking up on head movements performed by the interaction partner and responding by using an appropriate reaction or, alternatively, responding by mimicking the head movement in question.

Even though EPI has only been given a simplistic model of how to respond to a human conversation partner, this dynamic use of social movements could, in our opinion, lead to social robots that act as people expect a conversation partner to behave. An obstacle in achieving this goal is that the dynamics of human social interaction are complex. Each conversation is unique in its unfoldings, meaning that simple rule following of a robot is most likely not enough to effectively simulate the complexity of human social interaction. Still, it is not impossible to imagine that a robot could reach some level of the dynamic complexity that is significant of human conversation.

If we want social robots to interact effortlessly and effectively with their human conversation partners, this means that that the robots must have some capacity for social resonance. Nonverbal behaviours such as joint attention, head movements and gestures are all an integral part of the social resonance in human conversation and are therefore useful to implement in a social robot. Besides integrating nonverbal behaviour into the socially interactive repertoire of social robots, a smooth interaction between robot and human also requires that the robot has the ability to adapt speech to its human conversation partner. It is therefore not enough that a social robot has the ability to understand and produce speech sounds. A robot that also adapts its word choices, sentence structure and prosody to match the conversation partner would most likely make the interaction more engaging and interesting for the human being interacting with the robot. This kind of robot that adapts its speech to its interaction partner may be able to function in a variety of situations, being able to interact with children in a learning situation as well as interacting with adults both in a workplace situation and in a home environment, including both joint action towards a goal and relaxed casual conversations.

A possible area where this kind of advanced social robot could be useful is in health care situations, such as keeping an elderly person company. While a robot never could replace human social interaction, regardless of how advanced it would hypothetically be, it might be a welcome addition in nursing homes, providing interaction opportunities to the residents. Today in Sweden it is not unusual for elderly to live alone, despite being dependent upon support in terms of personal care and other daily tasks. This means that the person might receive short visits by home care personnel, but may spend most of the day alone. In this situation, a social robot could be useful as a personal assistant that can function as a conversation partner as well as being able to remind the person about taking medicines or other routines throughout the day. If the robot is also able to engage the elderly individual in games or fun problem-solving tasks, this could in fact lead to cognitive stimulation that is beneficial in terms of slowing down the development of degenerative brain diseases such as dementia.

There are already several robots specifically designed for an elderly population, such as the robot seal Paro [Wada et al., 2007]. While this particular robot has been shown to have similar benefits as interacting with a pet, a socially resonant humanoid robot could enrich the social environment in a way that may be even more beneficial for a person with limited social contact.

Another target group that could benefit from interacting with a socially resonant robot is children with autism or other social disabilities. Interacting with a social robot that has some of the capacities that the individuals in this group are lacking may be a good way to improve social skills. Given that the robot both in appearance and behaviour is simpler than a human interaction partner, this could be beneficial for children that are easily overwhelmed by too much social information. There are already examples of robots designed for this group, such as the small Keepon robot [Michalowski et al., 2007] that interacts rhythmically in that it adapts its movement to the interaction partner. A humanoid robot robot with a wider repertoire of socially resonant behaviour would most likely be useful for individuals that need to develop their social skills.

Developing the project further

The project in its current state could be developed in several directions, depending on the interests of the persons taking on the project. A main focus, regardless on what road later taken, is to improve the mapping techniques in terms of finding finer details of a variety of body movements to implement in a robot. This can be achieved by identifying relevant points between which the angles of movement can be calculated.

One direction in which the project can be developed is to improve the interaction between the two small robots, making it better capture the behaviour of conversing humans. A possibility here is to make a greater effort to single out social movements in recorded videos, perhaps by extracting non-relevant behaviour. If only behaviour meaningful to the conversation is produced by the robots, it will likely have an effect on how the robot interaction is perceived. Another area that can be improved regarding the small robots is how the movements are performed, which at the current state are not as smooth as would be needed in simulating natural movements.

Yet another another road could lead to the development of a humanoid robot with an increased capacity for social resonance in interaction with a human partner. Besides adding the complexity of a number of different body movements that can be combined in different ways, it is also possible to make a finer analysis investigating how interaction partners respond to the movements of one another. Adding finer details in this area can be achieved by collecting data on many different conversation partners, which can be used to train a network to give more realistic responses to a variety of social movements.

References

LW. Barsalou, P.M. Niedenthal, A.K. Barbey, and J.A. Ruppert. Social embodiment. *Psychology of Learning and Motivation, 43*, pages 43–92, 2003.

C. Bartneck and J. Forlizzi. A design-centred framework for social human-robot interaction. *Proceedings of the 2004 IEEE International Workshops on Robot and Human Interactive Communication*, 2004.

F.J. Bernieri, J.M. Davis, R. Rosenthal, and C.R. Knee. Interactional synchrony and rapport: measuring synchrony in displays devoid of sound and facial affect. *Personality and Social Psychology Bulletin, 20(3)*, pages 303–311, 1994.

E.A. Cartmill and S. Goldin-Meadow. Gesture. in: (ed's) d. matsumoto, h.c. hwang and m.g. frank. *The APA handbook of nonverbal communication*, pages 307–333, 2016.

T.L. Chartrand and J.A Bargh. The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology 76(6)*, pages 893–910, 1999.

W.S. Cook, T.K. Yip, and S. Goldin-Meadow. Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes, 27(4)*, pages 594–610, 2012.

D. C. Dennett. *The Intentional Stance*. Cambridge MA: MIT Press., 1987.

D. C. Dennett. *Kinds of Minds*. Phoenix, London, 1996.

M. Gardner. Mathematical games - the fantastic combinations of john conway's new solitaire game "life". *Scientific American, 223*, pages 120–123, 1970.

S. Garrod and M.J. Pickering. Why is conversation so easy? *Trends in Cognitive Science, 8(1)*, pages 8–11, 2004.

G. Hickok. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience, 21(7)*, pages 229–1243, 2008.

G. Hickok and M. Hauser. (mis)understanding mirror neurons. *Current Biology, 20(14)*, pages 593–594, 2010.

B.J. King and S.G. Shanker. How can we know the dancer from the dance? the dynamic nature of african great ape social communication. *Anthropological Theory 3(1)*, pages 5–26, 2003.

S. Kopp. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication, 52*, pages 587–597, 2010.

J.K. Lakin, V.E. Jefferis, C.M. Cheng, and T.L. Chartrand. The chameleon effect as social glue: evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior, 27(3)*, pages 145–162, 2003.

J. Lindblom. *Embodied social cognition*. Springer International Publishing, Berlin, 2015.

LUCS. Epi. `http://www.lucs.lu.se/epi/`. Accessed: 2019-01-21.

M.P. Michalowski, S. Sabanovic, and H. Kozima. A dancing robot for rhythmic social interaction. *HRI '07 Conference paper*, pages 89–96, 2007.

L.K. Miles, L.K. Nind, and C.N. Macrae. The rhythm of rapport: interpersonal synchrony and social perception. *Journal of Experimental Social Psychology 45*, pages 585–589, 2009.

A. Mörtl, T. Lorenz, and S. Hirche. Rhythm patterns interaction: synchronization behavior for human-robot joint action. *PLOS ONE, 9(4)*, pages 1–17, 2014.

A. Paxton and R. Dale. Rapid communication: Argument disrupts interpersonal synchrony. *The Quarterly Journal of Experimental Psychology, 66(11)*, pages 2092–2102, 2013.

K. Prepin and A. Revel. Human-machine interaction as a model of machine-machine interaction: how to make machines interact as humans do. *Advanced Robotics 12(15)*, pages 1709–1723, 2007.

G. Rizzolatti and M. Fabbri-Destro. The mirror system and its role in social cognition. *Current Opinion in Neurobiology, 18*, pages 179–184, 2008.

R.C. Schmidt, S. Morr, P. Fitzpatrick, and M.J. Richardson. Measuring the dynamics of interactional synchrony. *Journal of Nonverbal Behavior,36*, pages 263–279, 2012.

E. Thelen and L.B. Smith. A dynamic systems approach to the development of cognition and action. *Cambridge: MIT Press.*, 1994.

W. Tschacher, G.M. Rees, and F. Ramseyer. Nonverbal synchrony and affect in dyadic interactions. *Frontiers in Psychology, 5(1323)*, pages 1–13, 2014.

K. Wada, T. Shibata, and S. Hirche. Living with seal robots - its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics, 23(5)*, pages 972–980, 2007.

C Willig. *Introducing Qualitative Research in Psychology*. McGraw-Hill, Maidenhead, England, 2008.

Quasi Galilei

Virtual Reality as a Pedagogical Tool for Teaching Science and History

Felicia Hanserup, Attila Jervelycke, Henrik Persson & Francesca Zermiani

2019-01-20
KOGP06 / MAMN10
Lund University

Quasi Galilei

Virtual Reality as a Pedagogical Tool for Teaching Science and History

Felicia Hanserup, Attila Jervelycke, Henrik Persson & Francesca Zermiani

Virtual Reality (VR) is currently being adopted in the educational domain. While expectations are high, there is currently little empirical support for its efficacy as a teaching tool. In the current project, the ambition has been to maximally utilize the benefits of VR technology by exploiting opportunities for physical manipulation and visualization to learn about the scientific work of Galileo Galilei. In addition, the effects of social presence, in the form of digital agents, were investigated. In collaboration with twin project "Tänka-Testa", the overarching goal was to create a virtual environment specifically designed both for interaction and reflection--for doing and thinking. The scope of the article primarily regards this work in progress, and illuminates aspects to the decision-making process as they relate to theoretical foundations, as well as to previous work. We also describe in some detail the current state of implementation, and our intentions for the future.

1 Introduction

Conceptually, virtual reality is far from new. Immersion in the mental rather than the actual is arguably the defining feature of consciousness, and of states such as dreaming, mind-wandering, and hallucination. Virtual reality is also a defining feature of social life, as interpersonal communication and coordination demands sharing a perspective that to some degree will always differ from that held by each individual agent. Presented like this, it can be concluded that as a species, humans have been in the business of constructing virtual realities for a very long time.

As an explicitly technological innovation, virtual reality (VR) is considerably more recent. Nevertheless, it is not quite as novel as oftentimes conceived; already in the 19th century, rudimentary versions of the technique were being implemented and refined. Currently, there are large investments and rapid improvement in VR technology, and there are high hopes that we are on the verge of a significant breakthrough in terms of how VR is used and experienced.

Among commercial actors, development often appears to be somewhat removed from behavioral and psychological research, with more pronounced focus on technological improvement. Hyperrealism seems to be the stated goal. While realism is desirable to a certain degree, it is nevertheless questionable whether hyperrealism should be considered primary or even equal to more subjective, psychological factors (Blascovich et al., 2002; Fowler, 2015). For example, persuasive storytelling often depends less on the particular features of a given mode (e.g. cinema, literature, dance, etc.) and more on the skillful induction and manipulation of psychological states, such as 'presence' (Slater, 2018). This suggests that psychology and cognitive science, as well as the humanities and arts, should be considered equal to hardware specifications when it comes to developing immersive and engaging virtual environments.

Within the educational domain, the potential of VR technology to complement traditional methods, such as lecturing and laboratory work, is perceived as highly attractive. In particular, the approach has been adopted and well received in science teaching, as it makes possible the visualization of otherwise abstract or microscopic entities, such as atoms and molecules (Bailenson, 2008). It has also been used in medical training, changing how students learn about anatomy and physiology, as well as enabling them to practice surgical operation (Mayer, 2016).

When it comes to content outside of the immediate scope of science, empirical support for the efficacy of VR as a pedagogical tool is currently lacking (Mayer, 2016). To this end, therefore, our aim has been to investigate how virtual reality can be used to enhance learning in a more multidimensional domain: Where history and natural science combine, and where opportunities for manipulation and experimentation are provided alongside more contemplative exercises. In the following sections, we introduce, in much more detail, the conceptual background and theoretical foundations of our project.

2 Conceptual Background

In the current project, the goal has been to build upon and expand previous work done by the Educational Technology Group at Lund University (http://www.lucs.lu.se/etg/). The work in question, "Historiens Väktare" (Eng. Guardian of History), is a desktop game developed for learning about history, with the targeted audience being school children between the ages of 11-12 years. As a pedagogical feature, the game features a 'time elf'—a social, embodied agent (non-playable character)—whom students, as a twist on traditional methods, are then tasked with teaching.

Including such ‘teachable agents’ has a positive effect on learning and comprehension, and teaching content has been proposed to constitute an important stage in the process of structuring recently learned information (Bailenson, et al. 2008). Going further, in Historiens Väktare, the teachable agent is also tasked with taking the actual test later on, to determine how well it has been taught. Among other things, this has been proposed to increase student motivation, as responsibility for the performance of the elf falls with the students. Another benefit appears to be that the sense of ownership for success as well as for failure becomes shared, helping to create a kind of team dynamic between students and agents. As a consequence, anxiety associated with test taking and the risk of failure also seems to be reduced (A. Gulz & M. Haake, personal communication, September, 2018).

Finally, the mere presence of an embodied agent is likely to enhance learning and general engagement by making it a more enjoyable, and intrinsically social, experience. Particularly as students typically play the game on their own, without the rest of their class (Bailenson et al., 2008).

In our version of Historiens Väktare, we chose to focus on the small part featuring Galileo Galilei and his scientific work, and to rebuild his workspace as an immersive virtual environment (IVE). Within the workspace, players may explore freely and interact with various objects, as well as with the social, embodied agents Galileo Galilei and the young girl Anna. As of yet, the functional role of each agent is primarily that of substituting for social presence, rather than being truly interactive. Nevertheless, they are envisioned as being both animated and responsive to some degree, and they feature both in explorative and interactive exercises as well as in the more contemplative exercise.

In terms of content, students can expect to learn about the scientific work of Galileo Galilei, including basic conceptions of gravitational force and air resistance, as well as of optical transformations and astrology. The scope of the project currently includes two exercises, where students are able to test their understanding by relatively freely manipulating objects with the guidance of both verbal and iconic instructions.

In the first exercise, students are tasked with assembling a telescope by placing two lenses relative to each other so as to achieve proper magnification. Done correctly, viewing Jupiter and four of its moons through the telescope will be possible.

In the second exercise, students are encouraged to drop or throw around objects of various shapes and sizes so as to visually perceive their physical properties. As the exercise takes place on the surface of the moon, students may attain an improved conception of how gravity and air under normal circumstances interact, in contrast to how the same objects behave in a vacuum. As an illustrative example, by dropping a feather and a cannonball from the same height, students may to their surprise find out that they accelerate at the same speed.

In this way, it is our hope that the two exercises will allow students to walk away with an improved understanding of the physical world, without overemphasizing the mathematics behind it. By helping to create an interest in science early on, we believe that students will better prepared and more motivated to later learn about mathematics and the actual physics involved.

3 Theoretical Foundations

Can a VR environment be regarded as a learning environment? A central point of today´s pedagogical theories is the view of teaching and learning as social processes, involving social behaviors and activities such as communication, negotiation and interaction (Haake, 2009: 21). A crucial aspect of this process resides in the external environment that triggers and sustains a dynamic learning. In Cognitive Science, the embodied cognition approach claims an active role of the physical environment in human cognitive processes. In other words, according to theories of embodied cognition, it is pointless to isolate all human cognition in the human brain since cognition is indeed a dynamic exchange between mind, physical body and real world. Applying this approach to learning involves considering the physical space as a system that actively supports the student’s internal processes. Nowadays, both real-world and virtual spaces can be considered as learning arenas.

Following the theory of Social Learning, it is thus possible to make digital environments available to education through different strategies. Haake (2009: 22) distinguishes between *intrinsically* and *extrinsically* social learning systems: in the former systems, the interactions and the learning activities involved are supplied within the system thanks to the presence of social actors; whereas in the latter, the social activities are provided by the learners. The potential of *intrinsically* social learning systems can be found in their flexibility: they can be adapted to the specific needs of the single student.

It seems that enhancing such systems with social characters can lead to different benefits in the student, such as increased motivation and comfort, exploration/cooperation/reflection behaviors, increased input of information (e.g. body language and/or facial expressions of the characters) and memory improvement (Haake, 2009: 23-24). The empirical support for these benefits mainly comes from short-term studies. The tendency to set the test group at an average-user level, as well as the lack of long-term studies can lead to an ambiguous interpretation of the results; however, the studies conducted so far sustain the potentiality of such systems as learning supports for certain cases, not as universal solutions (Haake, 2009: 24).

Considering again the Social Learning theory, we might wonder whether the everyday student actually considers learning as a social process while sitting and studying alone before a test. The social component of learning needs to be supported by a more individual component. To complete the learning process, the student also needs to elaborate the received information through attentive reading and reflection (i.e. individual study), according to his/her individual needs. Learning is therefore addressed by the everyday student as a dynamic exchange of social interactions and tasks, as well as periods of individual study (Haake, 2009: 21).

Returning to our initial question, we believe that a VR environment can function as an *intrinsically* social learning system for supporting learning in its different phases. Especially, our attempt is to combine the social and the individual components of learning within VR, with an environment for different interaction activities and another environment dedicated to actual reading and reflection tasks. The concept of individual reflection within VR is completely new in the field and has been implemented by the other group; for this reason, we will not dwell on this topic here.

In terms of overarching pedagogical theory, the implementation of our ideas has been further influenced by the three-stage theory proposed by Fowler (2015). On this constructivist model, learning can be enhanced by emphasizing each of the following three stages:

i) *Conceptualization*, which can be understood as the primary introduction to a novel phenomenon, essentially constituting the foundation of a new concept. In our project, conceptualization is implemented in the form of a brief monologue preceding each task, in which Galileo explains some of the more rudimentary facts needed to understand the experiment.

ii) *Construction*, where students actively participate in generating new knowledge by directly testing how well their initially formed concept maps onto to the actual phenomenon. Construction lends itself particularly well to implementation in VR, and has been the prioritized aspect of our project, essentially constituting each experiment.

iii) *Dialogue*, in which students engage in the co-construction of knowledge by collaborating with one another. Being able to put words on complicated concepts is an important part of a deeper understanding. Dialogue does currently not fall within the scope of our project; however, it is being implemented by our collaborators in the Galileo Galilei group. When the two projects are merged, all three stages will thus be jointly represented.

In the design of pedagogical virtual environments, so called *value-adding features* are defined as features supported by experimental evidence to have a positive effect on learning. Such features include, for example, the use of conversational language, putting words in spoken form, adding prompts to explain, adding advice or explanations, and adding relevant pre-game training (Mayer, 2016; Mayer, 2018). As they relate to our project, *a)* the use of conversational style, *b)* putting words in spoken style, and *c)* pre-game training are considered to be implemented within the stage referred to above as conceptualization; finally, *d)* adding advice and explanations, is implemented during the construction stage in the form of visual instructions, and potentially also through further instructions from Galileo.

4 Methods

The overarching purpose of the project was to combine 'doing and thinking' in order to promote and enhance learning, and figuring out how to best achieve this was one of our main objectives. Importantly, there were two groups working in parallel on initially identical projects, but the vision and conceptual foundations of each group diverged already at the earliest stage. In agreement with our supervisors, we decided that it would be both easier and more efficient if our group focused on the first part of the 'doing and thinking' concept, while the other group (Tänka-Testa) focused on the latter. However, it was always our ambition to eventually combine and merge the two projects into one.

The main focus of our group was thus to design and implement tools for interaction. To provide students with the opportunity to learn about Galileo's scientific discoveries through means of experimentation. To achieve an improved understanding of the concepts involved through active participation.

The project started with an exploratory phase of reviewing the scientific literature. Once the scope of the project had been narrowed, preparations to select and specialize in those areas most relevant for the purposes followed. Attention was focused primarily on models concerning pedagogy and learning, as well as on theories of embodied agents and their impact on social learning systems. In parallel with the strictly theoretical research, the known facts about Galileo Galilei and his most important scientific work were investigated in more detail. By emphasizing theoretical research early on, both the conceptual design and subsequent implementation were better grounded in scientifically valid and historically accurate knowledge.

At the very early stages of the project, it was difficult to set truly long-term goals as we had no clear vision of where we wanted to go. As such, day to day decision-making was mostly concerned with minor details concerning our current appreciation of the situation, rather than with potential revision of the project in its entirety. As better understanding of VR and educational technology was attainted, the project eventually began to make more substantial progress, although it was still a relatively slow-paced process. Within a few weeks, however, it was already easier to talk about the project in terms that everyone could understand, to set new goals, and to test new ideas, all based on a shared vision.

During the subsequent creative phase, various methods were relied on in order to generate ideas that were both useful and entertaining. A lo-fi prototype was created representing the environment of each envisioned experiment, partly to get a better feel for the project, but also to ensure that everybody was on the same page. This prototype helped to more easily visualize the ideas and allowed us to kick-start the processes of conceptual design and implementation.

Early on, the opportunity to test various other VR applications in order to get inspiration—a strategy known as body-storming—significantly benefited the working process as not everyone in the group had previous experience with VR, either testing or development. Hands on experience was thus crucial to familiarize everyone with the equipment, and to provide a more grounded conception of the potential uses of VR. It also helped to create a better appreciation for both pros and cons of VR technology, for example in terms of how applications may be perceived—whether engaging or dull. Following testing, it was determined that the most engaging way to present the discoveries of Galileo would be by maximally exploiting the tools that VR provides, such as visualization and physical manipulation. The possibility of including entertaining 'Easter eggs' that players would be able to discover was also discussed. During the initial phase of the project, frequent brainstorming sessions proved helpful both to generate and to keep track of ideas.

Programming

Once there was a conceptual model that everyone agreed with, the next step was to start implementing the ideas. Working in Unity, Galileo's workspace was the first environment created. In terms of interior design and aesthetic style, the room was modeled to resemble a typical 17th-century Italian workspace, including furniture, maps, and a big rug. Of course, the idea of what such a room should look like was based on a rudimentary understanding of the era, and extensive research was not emphasized. Once a satisfying-enough workspace existed, designing each environment envisioned for each exercise was the next step.

The development and physical arrangement of both the exercise and its corresponding environment were circular processes in which novel ideas regarding either would influence how we envisioned the other. The virtual environment and our conceptual foundations were thus developed iteratively and in tandem during the whole project. Implementation of most relevant aspects was grounded in our currently most up-to-date understanding of the theoretical literature, then tested, and eventually upgraded. Working iteratively is critical for the simple fact that early prototypes are seldom perfect, and in almost all cases, a better product can be created by taking seriously all forms of feedback, and by keeping an open mindset regarding final form. As insight is generated during the entire process, early assumptions underlying and influencing both conceptual foundation and early decisions made, may have to be revised.

Implementation

The implementation was done in several steps and started with the creation and design of a workshop consistent with the overall conceptual design. This involved constructing boundary conditions and adding time-authentic furniture from the asset store and other 3D-modeling websites, as well as painting everything with aptly chosen textures and materials. From the start, the workspace was envisioned as the main environment, and also as the environment in which the *telescope* exercise would take place.

Next step was to create the environment for the *gravity* exercise. The original conceptual design was an exercise where the player could simply interact (e.g. throw or drop) objects of various shapes and sizes. As Galileo's discovery was that two objects of similar mass will fall at the same speed in a hypothetical world without air resistance, we eventually decided that the gravity exercise would benefit from taking place on the moon, where there is actually no air resistance. In addition to being more scientifically accurate, it was also expected to help create a more exciting experience. Indeed, one of the most significant aspects of VR is that that you can create environments and kinds of interactions that are difficult or even impossible to setup in real life.

The moon was created using the terrain feature which added hills and moon tiles as well as a skybox of the earth to make it feel grander and more authentic. A model of the leaning tower of Pisa was added as a nod to the fabled thought-experiment where a lot of people think Galileo proved his theory. Two port keys, in the form of a globe of the earth and the moon respectively, were added, both on the moon and in the workshop. The initial idea was that these would enable the player to teleport from one place to the other. In order to teach the player about why objects in real life do not fall at the same speed—again, due to air resistance — an Earth version of the same moon environment was added to the exercise. This was done by adding identical terrain to the one of the moon and change the ground from moon rock to grass, add trees and flowers as well as an atmosphere using a huge bowl as the sky. The reason for this was to make it able to seamlessly transition between the earth and the moon environment by having the earth ground recede under the moon ground and the skybowl decrease in size until it disappears. And, thus, revealing the moon and its landscape.

The drop experiment was decided to be two cannon balls dropping from the top of the tower of Pisa, with differing levels of gravitational effects depending on the environment. In addition, one of the cannon balls had a parachute, made by bits and pieces of different game objects from Unity, attached to it, and was programmed to have more air drag. Visual and auditory effects were added to when the balls hit the ground which would sound differently depending on the force it hit the ground with and would leave a puff of dust smoke.

Back to the workshop. To add a bit of magical realism, and as a nod to the father of Galileo, a lute playing a self-recorded interpretation of a classical song was added.

The telescope was next, and it was created using a downloaded model from the start but did not look good enough and was subsequently visually enhanced using 3D models created from 3D Studio Max. This was done to create the hull of the telescope and it was had texture change and length adjustment to look good and to fit the overall gold and red color scheme of the game. Lenses were added, first imported from the previous version of this game but later created from scratch using Blender to be able create the mesh to be able to showcase the telescope image as intended. The telescope was adjusted several times and added a function to make the player able to put the lenses in the sockets and adding sound effects to give feedback to the user. When both sockets had been successfully inserted, an image of Jupiter and four of its moons appear on the ocular end of the telescope.

Now the key fundamental functions of the game are in place the next step was to implement the characters.

Character modeling

The second part of the course has been mostly dedicated to the modeling of 3D characters, which has been a long and challenging process involving different moments and methods. In parallel with the programming and design implementation work, the embodied agents of Galileo and Anna have been created with a constant focus on their function (i.e. introducing and guiding the student through the distinct stages of the game).

The first step was the creation of a manuscript—written in Swedish—with all the dialogues between Anna and Galileo, as well as monologues of Anna virtually talking to the student (the reader can find the complete and final version of the manuscript in the *Appendix*). Looking at old cartoons about Galileo Galilei's life and historical time helped to draft the manuscript, according to the goal of writing realistic, concise, but significant dialogues. Before writing the manuscript, two "identity cards" for Anna and Galileo were made so as to give them two distinct and realistic personalities. After this, a long phase of role-playing inside the VR environment has been crucial for redrafting the manuscript and finally agreeing on the final version. This first stage was made in collaboration with the other group so that the content of the dialogues and monologues would be in coherence with both projects.

In parallel with the finishing of the manuscript, the 3D model characters were made in the program Fuse (Adobe Fuse CC beta). The division was made that the other group developed the Anna 3D model (Fig. 1) and that our group made the Galileo 3D model (Fig. 2) since the other group just have Anna inside their room but we also have Galileo in his workspace. At the beginning of the 3D modeling the basics of the software, Fuse was learned. During this time it was discovered that the software did not include typical clothes according to 17th-century fashion. Therefore a 3D model from Arteria3d (arteria3d.myshopify.com) with the help of the supervisor was retrieved. Some changes concerning the appearance of the model were done in order to make the 3D model look more like Galileo. When the 3D model characters were fully developed, both Galileo and Anna were implemented in the game environment. The next step was the animations to the 3D model characters.

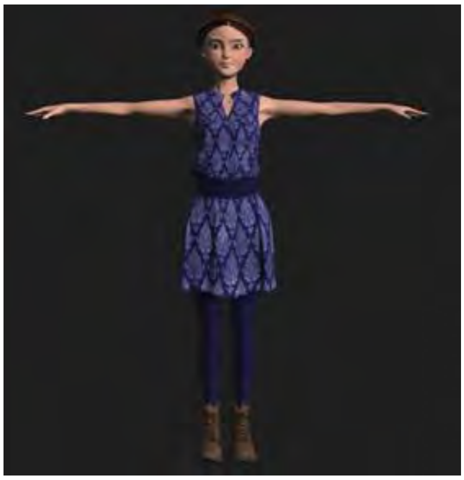


Figure 1. Anna.

Figure 2. Galileo Galilei.

After receiving approval for the manuscript from the supervisors, the next step involved the recording of the virtual characters' movements in the Motion Capture Lab (MoCap Lab). Here, it is possible to capture and transfer the movements of a person to a computer and, later, to a virtual animated character. This is achieved through an optical system of eight cameras and a computer, tracking and recording the motion of special markers (sensor tags) attached to the body of the recorded person (Fig. 3). A misunderstanding about the day of the recording actually became an opportunity to carefully searching for 3D animations available online (Mixamo – mixamo.com). After evaluating several animations, we agreed on downloading some of them and recording instead only five specific movements in the MoCap Lab, since no similar animations were available online. One single member of the group was recorded in the MoCap Lab, while the other members interacted from outside the camera system pretending to be engaged in a dialogue so as to help the recorded person to move as naturally as possible (Fig. 4).

Once the movements were recorded, the data were transferred to the leading software for motion capture systems, Qualisys (qualisys.com). This software provides you with tools for cleaning your data and creating "skeleton"-models by connecting each identified marker (Fig. 5). After learning the basics of the software, all models were created with Qualisys, manipulated with the help of our supervisors (Fuse – fuseanimation.com) and, finally, transferred into Unity as digital characters, with skin and typical clothes according to 17th-century fashion.

Figure 3. Recording movements in the MoCap Lab.

Figure 4. Quasi Galilei after recording moments in MoCap Lab.

Giving a voice to the 3D characters was the last important step for their modeling. All audio recordings were made in the Usability Lab (IKDC – Lund University) and then synchronized with the movements of the characters. Since Anna-agent also appears in the parallel project of the other group, it seemed reasonable to use the same person to give voice to the character. For this reason, a member of the other group was recorded while role-playing the manuscript. Regarding the second character instead, a 75-year-old Italian man volunteered to play Galileo's role and be recorded (Fig. 6).

5 Results

So far we have been able to create a functional prototype which currently includes Galileo's workspace (Fig. 7) and an additional environment on the surface of the moon. Within the workspace, the player may in one place find all parts necessary to assemble the kind of telescope that Galileo himself invented. The actual exercise then consists in learning about elementary optical transformation by manipulating the two lenses to achieve magnification. Once properly assembled, the telescope will provide an opportunity to view Jupiter and the four moons that Galileo also discovered, which thus constitutes *direct* feedback.

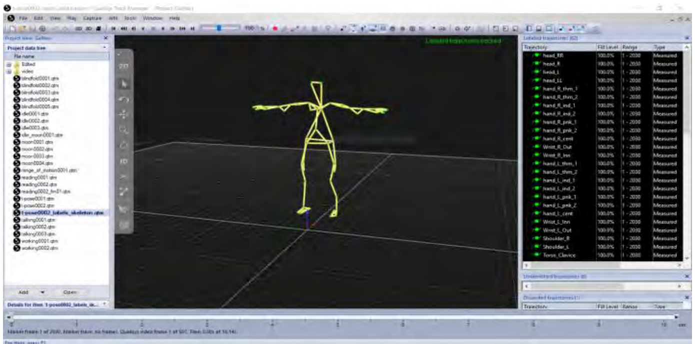


Figure 5. The Qualisys interface.

Figure 6. Quasi Galilei with Giuseppe and Gino after recording Galileo's voice.

Figure 7. Galileo's workspace.

The workspace has been designed to resemble a typical room from the 17th century, although we have allowed ourselves some artistic freedom. It is furnished as such to help create the impression that the players are actually in different place, in a different time. In terms of design, we have also coordinated our efforts with those of our sister projects to create two rooms based on a similar theme and style.

Further, the room contains a small scale model of the moon which constitutes a kind of portal, and which players through interaction can activate. Activation of the portal will transport them to the moon, our second environment. The moon environment includes a table similar to the one in the workshop, on top of which players will find both a cannonball and a feather. These objects are chosen to represent heavy and light objects, respectively, and will be manipulated during our second experiment, in which players learn about gravitational force (Fig. 8). In particular, the two objects have been chosen to illustrate Galileo's idea that, in a vacuum, the acceleration of each two object in free fall is independent of their mass. The moon has also been designed to be exciting and visually pleasing, with an uneven surface featuring large craters, and a starry sky above. To get back to Galileo's workspace, players simply interact with a similar-looking portal-globe hovering above the table, only this time a model of Earth.

Figure 8. The workspace of the gravity experiment on earth.

Regarding 3D digital characters, the final and complete versions of Anna (Fig. 1) and Galileo (Fig. 2) have been inspired by pictures from 17th-century and old cartoons about Galileo Galilei's life. Galileo remains seated at his desk for the whole duration of the game. However, he welcomes the player from his desk at the beginning and gladly encourages him/her to look around and explore his own workspace. On the contrary, Anna is actively engaged: she stands up close to the player and guides him/her throughout the whole exploration till the end, when she introduces the second part of the game (i.e. the secret room).

6 Discussion

With rapid improvements in digital technology, educational practice is likely to change, and arguably the change has already begun. Computers and tablets are becoming ubiquitous in school settings and there is high demand for scientifically valid and tested educational software. Currently, many games and applications are significantly lacking in this regard.

In this article, we have presented a virtual reality learning environment constructed based on scientifically valid models and theories from learning science and cognitive science. In this environment, students are encouraged to explore and learn about the scientific and historical worlds of Galileo Galilei. Specifically, they may learn about physical concepts such as gravitation, optics, and the field of astrology, and also about historical and cultural concepts such as the heliocentric worldview.

The use of social and embodied agents has been motivated both by psychological theories and by empirical investigations of their effects on learning. Indeed, Li, Kizilcec, Bailenson & Ju (2015) discuss the pronounced positive effects of social presence on learning outcome and argue that social, embodied agents in virtual environments have the potential to revolutionize education. To this end, however, it is argued that they must be interactive in a behaviorally realistic manner and, ideally, able to participate in shared reasoning and dialogue.

Implementation of such agents is beyond not only the scope of the current project but also the capabilities of most existing software. Nevertheless, once artificial intelligence has progressed sufficiently, we share their vision and see extraordinary potential in such agents as pedagogical tools.

As a disclaimer, Li and colleagues also note that "artificial agents that have lower agency than a real person (particularly social robots that may appear "neutral" without a particular gender or ethnicity) may provide more standardized and less distracting cues than a human instructor" (2015: 2). This suggests that the pursuit of hyperrealism, in some cases, may in fact be strictly counter-productive.

In a review, Mayer (2016) shows that the use of immersive virtual reality (i.e. VR) as a pedagogical tool is currently not supported by empirical evidence. However, this partly comes down to the fact that studies in this area simply are lacking and that not enough research has yet been conducted. In this sense, all current projects are exploratory investigations; they provide valuable experience, as well as data on which theories and models that apply specifically to virtual reality may be formulated. Once the VR environment has been tested by students it will be possible to evaluate its efficacy generally, and also to evaluate more specific aspects as they relate to theory.

To our knowledge, this project is the first to adopt a multidimensional approach to the design of virtual learning environments in which interactive and contemplative exercises are both included. In addition, our environment combines content from both physics and history and presents it in a culturally relevant setting. As can be seen, our contribution has been an active attempt to move away from traditionally modular approaches to education, favoring instead the integration of material across multiple subjects.

This choice is motivated by the idea that presenting information in such a way will help to make it more meaningful to the student. Ultimately, this approach is rooted in constructivist learning theory, and the idea that learning depends on appropriate scaffolding; i.e., the process of constructing new knowledge is dependent on those cognitive tools already at our disposal. As an analogy, this process of learning is often referred to as the snowball effect, which illustrates the fact that knowledge acquisition often increases non-linearly relative to relevant expertise.

Importantly, evaluation of the learning environment must be a matter of empirical investigation, and user-tests in appropriate settings (i.e. on children aged 10-14, in school) will ultimately determine its efficacy. This is clearly a priority for any future project, as the outcome would provide crucial feedback for any corrections that might turn out to be necessary. It would also provide more general insight into the use of digital learning technology, and into the use of virtual reality learning environments, more specifically.

As previously mentioned, empirical studies on the efficacy of virtual reality as a technological learning tool are severely lacking, while the general interest in adopting the technology is soaring. Inevitably, one consequence of this dynamic will be that schools and institutions invest in technology that has not yet been validated, and thus by definition is unscientific. While there are many reasons to be optimistic about the potential of adopting VR in educational practices, there are plethora ways to go wrong if games and applications are not first scientifically tested to support their use.

8 Conclusion

The working process was characterized by a multitude of different approaches throughout the project. There was some challenging and difficult moments, especially at the early stages, which were partly due to lack of a coherent vision and also due to some technical problems in the VR Lab. However, once those problems were solved, significant improvements could be made. The design process was iterative, and while the core of the original vision are still in place, new ideas were continuously generated and evaluated. Of these, some were subsequently implemented and became part of the iterative process such that potential improvements were always relevant. In general, the first part of the project was characterized more by creativity and generating novel ideas, and the second part was more about implementation, even though both parts to some degree were characterized by both aspects.

The current version of the VR environment has been tested by interested people of all ages, and feedback has largely been positive. Nevertheless, as has previously been stressed, the project would benefit from proper testing and measurements from the intended audience, in ecologically valid settings.

Educational practice is entering a period of change, and if this change is going to be an improvement, resources and interventions must be evidence-based.

References

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C, Lundblad, N. & Jin, M. (2008). The Use of Immersive Virtual Reality in the Learning Sciences: Digital Transformations of Teachers, Students, and Social Context. *The Journal of the Learning Sciences*, *17* (1), 102-141. DOI: 10.1080/10508400701793141

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C. & Bailenson, J.N. (2002). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, *13* (2), 103-124.

Haake, M. (2009). *Embodied Pedagogical Agents*. *From Visual Impact to Pedagogical Implications*. Lund University. Sweden

Li, J., Kizilcec, R., Bailenson, J. & Ju, W. (2015). Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior*, *55* (Part B), 1222-1230. http:// dx.doi.org/10.1016/j.chb.2015.04.005

Mayer, R. E. (2016). What Should Be the Role of Computer Games in Education? *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 20–26. https://doi.org/10.1177/2372732215621311

Mayer, R. E. (2018). Computer Games in Education. *Annual Review of Psychology*, *70* (1). https://doi.org/10.1146/annurev-psych-010418-102744

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, *109*, 431–433. DOI:10.1111/bjop.12305

Appendix

Manuscript

Workspace
1.1: Introduction to the game (*Anna & Galileo*)
1.2: Pre-training—***Gravity*** (*Anna & Galileo*)
3.1: Pre-training—***Telescope*** (*Galileo*)
3.2: Instructions—***Telescope*** (*Anna*)
4.1: ***Culture***/***History***/Reference to ***secret room*** (*Anna*)

Gravity (Earth)
2.1 Instructions (***Anna***)
2.2.1 During, if right
2.2.2 During, if wrong
2.2: Conclusion

Gravity (Moon)
2.3.1 During, if right
2.3.2 During, if wrong
2.3: Conclusion

Workspace
1.1: Galileo is sitting as his desk, working. Anna and Student enters Galileo Galilei's workspace for the first time. Anna greets him cheerily and he looks up.

Anna: *"Hejsan, farbror Galileo!"*
Galileo: "*Hej, kära Anna! Hur står med till med dig idag? Det var inte igår!"*
Anna: *"Jag mår jättebra! Och jag har med mig en ny vän, som jag träffade på gatan utanför!"*
Galileo: *"Jaha, vad spännande! Välkommen hit. Annas vänner är mina vänner. Du får gärna se dig omkring här i min verkstad."*

1.2: Immediately after the first scene, Galileo is still sitting at his desk. Anna asks him about his recent work, and he briefly mentions the telescope and thinking about gravity. Anna then mentions Galileo's proposed thought experiment to prove that objects of different weight fall at the same speed. Galileo replies that it would work better if there were no air resistance.

Anna: *Vad har du haft för dig den senaste tiden?*
Galileo: *Jag har byggt ett teleskop för att studera stjärnhimlen, och funderat på hur gravitationskraften egentligen fungerar.*
Anna: *Ja, just det! Till skillnad från Aristoteles, så tror Galileo att föremål med olika vikt ändå faller med samma hastighet. Och han har föreslagit ett tankeexperiment vid det lutande tornet av Pisa. Om du tar min hand så kan vi testa själva! (Loop)*
Galileo: *Men det hade fungerat ännu bättre i vakuum!*

Earth (Gravity)
2.1: The Student has been instructed to come with Anna, and grabs her hand as she touches the Earth globe. Anna and the Student are teleported to the Tower of Pisa on Earth. Anna appears at the top of the tower, and from there proceeds to give instructions on how the exercise works.

Anna: "*HALLÅ! HÄR UPPE I TORNET! Jag kommer släppa två saker i taget härifrån. Så kan du se vilken av dom som faller snabbast!"*

2.2: The student has completed the exercise. Anna acknowledges that Galileo's prediction did not appear entirely accurate, but then recalls his comment that the experiment would work better with no air resistance. She informs the Student that the next exercise is on the moon.

2.2.1: Anna makes comments during the exercise.
Anna, before dropping (1): "Se upp! Nu släpper jag!"
Anna, before dropping (2): "Här kommer nästa!"
Anna, if right (1): *"Precis som Galieo menar!"*
Anna, if right (2): *"I enlighet med teorin!"*
Anna, if wrong (1): *"Är du säker? Det såg inte ut så här uppifrån."*
Anna, if wrong (2): *"Är du säker? Det såg inte ut så här uppifrån."*

2.2.2: Anna comments after the exercise.
Anna(if right on last): "Ja, det såg verkligen ut så, så..."
Anna(if wrong on last): "Nej här uppifrån såg jag en tydlig skillnad, så..."
Anna: *"... det verkar inte som att de alltid faller med samma hastighet ... Men vänta! Galileo sa ju att hans teori bara gäller i vakuum, utan luftmotstånd. Vi kan testa det också!"*

Moon (Gravity)
2.3: The Student and Anna has teleported to the surface of the moon, and Anna is once again at the top of the Tower of Pisa. She instructs the Student to repeat the same exercise to see if it makes a difference that there is no air resistance.

Anna: *"Visst är det häftigt på månen! Här finns ingen luft alls, så om vi gör om samma sak igen så kan vi se om Galileo hade rätt att kanonkulorna faller lika snabbt här!"*
Anna, before dropping (3): "Nytt försök!"
Anna, before dropping (4): "Det här är de sista!"
Anna, if right (1): *"Precis som Galieo menar!"*
Anna, if right (2): *"I enlighet med teorin!"*
Anna, if wrong (1): *"Är du säker? Det såg inte ut så här uppifrån."*

2:4: Both gravity exercises are completed and it's time to go back to Galileo's workspace.

Anna: *"Så nu vet vi! Saker faller lika snabbt ... Men bara i vakuum! Nu sticker vi tillbaka till verkstaden!"*

Workspace (Telescope)
3.1: Once Anna and Student have completed the gravity exercise, they return to the work space. Anna is excited, and doing the telescope exercise is up next. Standing next to the telescope, Anna asks Galileo if they (Student and her) can try to use it. He agrees, and mentions how there are different kinds of lenses and that their function depends on how they are combined (pre-training)

Anna: ”Hej igen! *Vi har varit vid det lutande tornet i Pisa. Och vid det lutande tornet på MÅNEN! Snälla Farbror Galileo … Kan vi inte få kolla på ditt teleskop också?”*
Galileo: ”*Det är klart ni får! Jag har precis fått nya linser skickade till mig. De är konvexa och konkava, och deras funktion beror på hur man kombinerar dem. Sätter man ihop dem på rätt sätt kan man studera både stjärnor och planeter!”*

3.2: The student is now standing next to the telescope, and has been introduced to the fact that you learn about optics and astronomy by experimenting with the telescope. The student can proceed to manipulate the pieces and try to assemble the telescope.

Anna: "Följ ritningen på väggen. Om du sätter i linserna i teleskoppet på rätt sätt så kanske du får se något häftigt!"

Anna:(If wrong) "Det ser inte rätt ut, testa att byta plats på linserna."
Anna:(If right) *”Om bilden är suddig så kan du försöka att ställa in konstrasten genom att förlänga teleskopet!”*
Anna: *”Teleskopet är riktat mot Jupiter. Kan du se hur många månar som finns runt Jupiter?”*

Workspace (Conclusion)
4.1: Anna mentions that Galileo’s work is controversial because the church doesn’t agree with his conclusions, which are based on his observation of the planets. For this reason, he has to keep some of his work hidden from view. Anna proposes to take Student there.

Anna: *“Folk på gatan viskar om Galileo eftersom kyrkan vägrar acceptera hans idé att jorden snurrar runt solen. De har dömt honom till husarrest. Men han fortsätter att utveckla sina idéer i hemlighet.”*
Anna: ”Jag brukar gå till hans hemliga kontor ibland för att kunna tänka och läsa i lugn och ro. Följ med, så kan jag ta dig dit! Men du måste ha på dig den här ögonbindeln ..."

Real-Time Adjustable Feedback Based on Eye Tracking Algorithms in Educational Games

Amanda Wikholm, Annie Önnered, Christian Mathias Gulmann, Sara Egli and Vincent Wipp Ekman

When processing feedback, students can fall off at multiple instances such as: noticing the feedback, reading the feedback, making sense of it and acting upon it. This project aims at increasing the number of students who successfully complete each of these steps. Our hope is that by completing more steps in the feedback process, a long time improvement of their learning strategies for further progression in educational life is able to take place. This reduction of feedback neglect will be accomplished through smart solutions which take the student's individual learning difficulties into consideration. These solutions are based on theories of scaffolding, using for example subtle gaze direction (SGD) and a teachable agent to adapt to the student's behaviour. The SGD has the possibility of increasing the amount of students reading the feedback, and social factors have the potential of increasing the student's motivation throughout the whole process. All of these solutions will be activated through the utilization of eye tracking technology in a behavior sequencing engine inspired system. This is to insure that the feedback is adapted to the individual users behaviour as a way to help them progress. An essential part of the system therefore relies on remembering the users' earlier actions and adapting the behavior space to better accommodate the individual users' needs in accordance to the situational context of the game environment.

1 Introduction

The project presented in this report has the goal of developing a system which presents and adapts learning game feedback in real time. The system will be able to identify the students who do not attend, process or act upon the feedback and through different intelligent solutions nudge them towards the provided feedback. Receiving feedback is seen as a process of four steps where the student receiving the feedback first needs to notice it, then process it, then understand it and finally act upon it (Tärning et al. 2018). In each of these steps students may neglect the feedback and thus "fall off".

This report presents the final work of developing a real-time adaptive feedback model. This resulted in a theoretical framework of theories relevant for the project, an implemented and tested prototype which illustrates the possibilities of utilizing eye tracking in learning games. This was eventually combined in to a larger conceptual model which describes how one could use these features throughout the game to prevent feedback neglect in students. This report will also present a suggestion for how subtle gaze direction can be utilized in the learning game Guardians of History (in Swedish: Historiens Väktare) to direct students' attention towards the feedback.

The Guardians of History

The task of developing an adaptive feedback system, utilizing eye tracking as a mean of providing real time feedback, is set in the environment of the educational game The Guardians of History. In this game the pupils travel in time to meet the most pronounced physicists and pioneers of modern science to collect information on the history of science in order to help an in game agent with passing tests. It thereby acts as an interdisciplinary game combining the subjects of physics (especially astronomy) with history, by which the discoveries of physics are integrated in a historical context. The game is designed for Swedish middle school children and based on the concept of teachable agents to create a social setting where the user does not only search for information but uses this to educate an in-game peer.

The game is developed in collaboration between the department of Computer Science at Linköping University and Cognitive Science at Lund's University by the research unit Educational Technology Group (ETG), which explores the usage of technologies and educational systems as ways of researching learning processes while still providing pedagogical software of real-world value (Educational Technology Group 2018).

Background

Prior to our project similar research has been done within the same field. One such research project is Tärning et al. (2018) where the goal was to find out when and to what degree students were neglecting feedback while interacting with learning games. The behavior of each student was recorded through eye tracking and data logging. Eye tracking was used to capture information about whether the students noticed and read the feedback. The log data was used to provide information regarding whether the students processed and thus acted upon the feedback. The method used to confirm that the student notices the feedback was based on fixations on certain areas of the screen, a method we will be implementing as well. In Tärning's study as well as our own the reading activity is not measured in detail, but described as a general reading behaviour and not analyzed on a word level. The major distinction between our project and Tärning's study is that in their study all the gathered data was analyzed after the game play had ended, and the goal was to investigate where the students neglected, while we attempt to provide solutions to the observed neglect through real time behavioural analysis. Another important feature was how Tärning's study revealed the significant role the teachable agent play in order for the student to notice the feedback (Tärning et al. 2018). This was done, in one test condition, by replacing the agent with a simple arrow and compare the number of students that notice the feedback. The result was significant; 78% noticed the feedback when the teachable agent was pointing at it, while only 57% noticed the feedback when an arrow was pointed at it. (Tärning et al. 2018). This result clearly speaks for the important influence that the teachable agents has on a student's ability to process feedback.

Aim

The aim for this project is to

1. Create a thorough theoretical framework of previous research relevant to our project (e.g. the use of eye tracking, how focus of attention can be influenced and different characteristics of the teachable agent).
2. Create a conceptual model that describes how to prevent feedback neglect through adaptable agent behavior and gaze directing cues.
3. Implement a system that makes use of some of the solutions presented in the conceptual model and test these intelligent adaptable solutions on users.
4. Bring forward ideas for further possible developments to prevent feedback neglect.

2 Theory

This section present our theoretical framework by exploring different aspects relevant for this project.

Educational Games

Educational games are a special branch of game design attempting to utilize the ideas of technology-enhanced-learning which are situated as a sub-discipline of education (de Freitas 2018). However, the area is positioned with relations to an abundance of different disciplines with contributions ranging from educational science, game science and neuroscience to informational science (de Freitas 2018).

The idea of educational games have not developed in a linear fashion as there have been multiple different approaches trying to influence the subject (de Freitas 2018). The first approach focused heavily on creating definitions (Caillois and Barash 1961; Sutton-Smith and Roberts 1961), while the second one moved more towards developing games of serious substance for very specialized educational purposes (Blumberg et al. 2013). Here the entertaining aspect was set aside for the educational substance, for example military training (de Freitas 2018). The third, and for this report most interesting, approach focused on researching how games could be created for different educational contexts such as school settings, while also looking into the possibilities of educational exploitation of commercial games (de Freitas 2018). This approach illustrated some of the problematic features of implementing games in educational context as games do not fit the traditional learning structure of the schools today with its hourly lessons, single subject, and single teacher organizations (Egenfeldt-Nielsen 2005), making games a disruptive force. Really seizing its merits, requires a change in the current educational organization (de Freitas 2014). To overcome this, The Guardians of history acts directly within a cross disciplinary field of the established subjects of physics and history, while having a temporal organization which accommodate the established school setting.

One important separation to make in the discussion of games is that although games are closely related the term play, which was shown by Jean Piaget (1896-1980) and to be of importance for learning (Piaget 2013), they are not the same. Playing applies to an expressively and free improvisation, while gaming is more rule-structured and goal-orientated (Deterding et al. 2011).

The remainder of the report will explore a distinct part of educational games by exploring ways of utilizing peer teaching of in-game characters as a learning tool for improving students' performance and efficacy in a rule-structured and goal-orientated gaming setting.

TA - Learning by Teaching

Within the area of learning games there have been an increasing focus on the usage and effect of in-game-characters, spanning from the user-controlled avatar, who acts as the user's in game virtual representation (Haake 2009) to the program-controlled agent, who can serve multiple different purposes from conversational to pedagogical (Haake 2009). Although in the scope of this report the term agents will refer to those who serve a pedagogical function for the students interaction with the system. This leads to the real focus of the report: the digital tutee (Tärning 2018) or teachable agent (further referred to as TA), which serves as a protégée by acting in a pupil/teacher relations with the user (Chase et al. 2009; Tärning 2018). It therefore acts as a hybrid between an avatar and an agent as it is both influenced by preprogrammed behaviour and user interaction in form of the user (most commonly a school pupil) teaching the agent to better solve the tasks it is provided (Chase et al. 2009; Tärning 2018). The TA thereby combines the educational benefits of both the agents social presence and the avatars potential for users to adopt properties (Chase et al. 2009). In the context of educational games TAs can serve multiple different purposes for the users, such as: maintaining long term social-emotional relationships (Bickmore 2003); motivating students through polite phrases (McLaren, DeLeeuw, and Mayer 2011) or by challenging the students toward deeper learning and meta-cognitive reasoning (Kirkegaard, Gulz, and Silvervarg 2011), or even by persuading users through the social interaction to engage actively in teaching the TA (Lim et al. 2014) eventually leading to a more positive learning attitude (Lim et al. 2013).

Believability and Agents

James C. Lester's and Brian A. Stone's (1997) research show how animated pedagogical agents play an important role in knowledge-based learning environments, not only through their coupling feedback abilities and strong visual presence, but also through the extent to which the agents exhibit life-like behaviors. This life-like behavior has strong motivational impact on the students and makes them spend more time interacting with the software ((Lester and Stone 1997). They see believability as a key feature of these animated agents, and by increasing the believability in the agent significant rewards can be yielded in students' motivation. However, creating a believable animated pedagogical agent has its challenges. The believable behavior must though never be structured in such a manner that it interferes with student's problem solving, since the agents primary goal is to promote learning. Decreased ability to learn can never be the cost of increased believability. Therefore, believable emotional responses need to be interleaved with the advisory and explanatory behaviors of the agent in an appropriate and realistic manner. If the agent is perceived as acting like a simple automaton, the believability will be partly or altogether eliminated. Instead, the goal of believability is to make the user believe that s/he is interacting with a sentient being that has its own beliefs desires and personality (Lester and Stone 1997). According to Bates, “appropriately timed and clearly expressed emotion is a central requirement for believable characters.” (Bates 1994, p.1). In creating the emotions of Herman the Bug in the extension of the learning game Design-A-Plant, Stelling (2002) used a simplified adaptation of the scheme of 24 emotions reported in Clark Elliot's Affective Reasoner (Elliot 1992). 13 of the emotions were judged to be meaningful in a pedagogical agent as well as possible to convey effectively into different animation and sound combinations (Stelling 2002).

In order to make a life-like system believable, one has to know what is meant by the often loosely defined term "believability". Bogdanovych et. al choose to use the definitions on "believable characters" provided by Mateas (1999) and Loyall (1997)(mentioned in Bogdanovych, Trescak, and Simoff 2016), who consider believability to be about creating a suspension of belief. Stone and Lester define believability as the extent to which the users interacting with the agent come to believe that they are observing a sentient being with its own beliefs, desires and personality (1997, mentioned in Bogdanovych, Trescak, and Simoff 2016).

Bogdanovych et. al's hypothesis is that the believability features suggested by Loyall will increase the perceived agent believability compared to agents who lack the corresponding feature. The believability features provided by Loyall are the agent's personality, emotional responses, self-motivation, adaption to change, creating and upholding social relationships, showing consistency of expression, and creating an overall illusion of life (1997, mentioned in Bogdanovych, Trescak, and Simoff 2016). Bogdanovych et. al created a model which was based on the believability features. The model was then implemented in a technological framework of contemporary AI techniques, letting users rate the believability of an agent displaying these features versus one who did not display any believability-enhancing features. The results show that for an agent to be believable it has to be resource bounded, able to change their behaviour in adaptation to the changes in their environment, exist in the correct social context of the simulation, as well as being overall reactive and responsive. However, the agent's ability to display emotions was found to be less important than other features (Bogdanovych, Trescak, and Simoff 2016).

We expect the effects of believable behavior to be of importance for the users engagement with the feedback as it should lead to a higher degree of social engagement with the TA.

Understandings of Feedback

At this point, through the title alone, one could with confidence guess that feedback is at the centre of all of this. A large problem with feedback is that a lot of students do not get to the point where they actually do what is needed to be done (for statistics on it see (Tärning et al. 2018)). Therefore, we are looking to create a model/system which increases the students' successful use of feedback. To help us with this project we have deployed the model created by Betty Tärning (see fig. 1) where she has divided the interaction with feedback into concrete stages (Tärning et al. 2018). By using this model in our work we can try to employ solutions to specific falloff points. In the following section of this report each separate step of Tärning's feedback model will be briefly presented. The primary reason for presenting Tärning's model is to highlight the amount of students who fall off at each step of the feedback process. The extent of feedback neglect is important to our study because we want to know which steps that are most exposed to feedback neglect, so that we may put extra focus on intelligent solutions on these steps. We are also aware that in order to reach the next step in the feedback process, the previous steps need to have been addressed by the student. Thus, all feedback steps are of importance.

The Steps in the Black Box

Notice is the act of actually noticing, or looking at, the displayed feedback. In our study this means that for a reasonable amount of time the user has had their gaze fixated inside of the feedback box. A large amount of the students presented with CCF (critical constructive feedback) fall of at this stage. In Tärning's study 33% of the students do not even notice the feedback (Tärning et al. 2018). Some of our preliminary ideas of solutions to this fall off are subtle gaze cue, agents pointing, the agent reacting to user actions e.g. opting for attention if the student's gaze focuses on the exit button, as well as more overt gaze directing cues.

Even if the student notices the feedback box Tärning's study shows that 39% of the students did not process the material after noticing it. In the study students who did not read the feedback were coded as not processing the feedback (Tärning et al. 2018). Reading behaviour detection in real-time is not at the moment available for us. Therefore we have chosen to in more general ways determine if someone has read the feedback text, for example on how long their gaze stays within the feedback box. The question we can ask ourselves is: how do we get them to read the feedback text? An example of the type of solution we might use here is socioemotional scaffolding through the teachable agent. The agent might for example appear next to the feedback box saying “What does the text say? Can you help me understand it?”

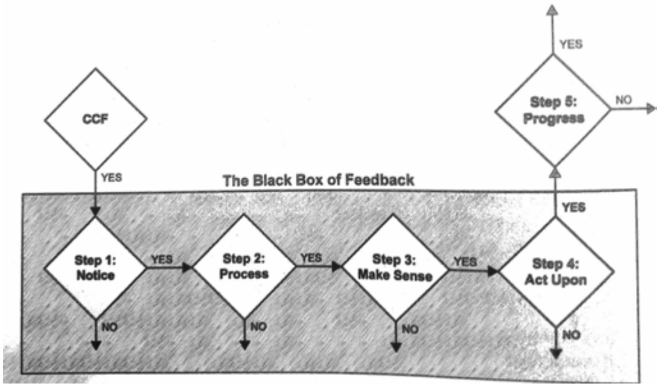


Figure 1: The black box of feedback
(Tärning et al. 2018, p.6)

The next part of the black box is where the student makes sense of the feedback given to them, if s/he comprehends the message. Finding out if a student understands the feedback is not something we can do at the moment as it is difficult in all learning situation to find a way to get this information, something that is not minimized in a in real-time digital system. If the student is motivated, by for example the engaging behavior of the agent, s/he might put in the required effort to understand the feedback and to remember it.

So if the student has made it this far, will they act upon the feedback, will they take the required steps, do they follow through? And if they do, do they use their new knowledge to make progress? Well, sadly the answer seems to be no. Only 23% of the students in this case did act upon the feedback given to them, and only 52% of the 23% who acted upon the feedback actually progressed. (Tärning et al. 2018).

To sum up the feedback neglect in Tärning's study, there is significant fall off at every step throughout the feedback process. One might even despair by the fact that only a very small group of students (4%) go through all the steps and end up using the feedback correctly. On the basis of these results, our project attempts to prevent feedback neglect through the use of intelligent solutions. The goal is that the system can have both theoretical and educational significance on how to improve students' learning behavior by utilizing gaze behavior as an indicator of students progress through the possible steps of feedback neglect.

Elaboration Likelihood Model (ELM)

Another theoretical model we will apply in our study is the Elaboration Likelihood Model (ELM) (Petty and Cacioppo 1986). We believe this model will give us an understanding for why students have acquired a fixed learning strategy which enables feedback neglect to occur to such an extent as it does (Tärning et al. 2018). ELM also helps us in constructing a conceptual model which takes into account how different intelligent solutions that we chose to apply may enable more or less constructive student learning strategies and possible, over time, even (hopefully) increase the student's self-efficacy as a learner (Bandura 1997). The model in short can be described as a model for how attitudes change in people and was in part created to deal with the fact that several results in persuasion theory made sense in one place but not another.

The ELM uses two persuasive pathways, or routes, which both lead to attitude change (see fig. 2). The two routes, the central and the peripheral one, can be understood as "...careful and thoughtful consideration of the true merits of the information presented in support of an advocacy" (Petty and Cacioppo 1986, p. 125) i.e. the central route is when reflection on the subject matter plays a major part in the persuasion. The peripheral route is when other factors are the ones that persuade, for example number of arguments rather than the quality of arguments, insistence from a peer or other factors which are not part of the subject matter. A difference between these routes is among others that the attitude change seems to be more stable in the central route.

A person's likelihood of elaboration is determined by his/her motivation and ability (Petty and Cacioppo 1986). A part of the ELM is how these two factors relates to how we respond to stimuli. The same stimuli can receive different responses depending on the amount of ability/motivation the receiver has. What follows from this is that if we can raise their motivation and increase their ability the students will more often respond to the stimuli by going down the central route, which leads to more stable attitude changes.

A risk to take into consideration is that the same stimuli can give rise to both productive and counterproductive behaviour depending on the motivation or ability of each learner. The conceptual map and the implemented real time feedback system must therefore be able to adapt its intelligent solutions to different levels of these two variables. It is not to be assumed here that ability is some property of the person separate from the structure of the persuasion context, a complex message can be repeated to increase the ability to comprehend, so it would not be too far fetched to assume that the way we formulate, construct and present the material has a contextual impact on the student's feeling of ability (Petty and Cacioppo 1986).

From the ELM perspective our aim should be not only to make the students successfully move through the different steps of the feedback process (notice, process, make sense, act upon) but also to make them do this for the right reasons, such as an willingness to improve oneself and in this case also The TA. We might accomplish this not by simply "nudging" them into performing, but by intelligently and thoughtfully develop solutions that increase motivation as well as ability. As we have mentioned earlier we will use an adaptive TA as a way to increase students' motivation through believable interactional behavioral patterns in the TA and increase students' ability through the explanatory and advisory behaviors from the TA. We will also utilize different responses from the eye tracking system to produce the desirable action e.g. through subtle gaze cues. We will to the largest extent possible adopt solutions which may create long term positive learning behaviours and make students less dependent upon the nudges provided by the system. Initially though, the system may have to make use of more subtle or direct gaze directing cues to get the students to even notice the more intelligent scaffolding which has the potential to lead them down the central route.

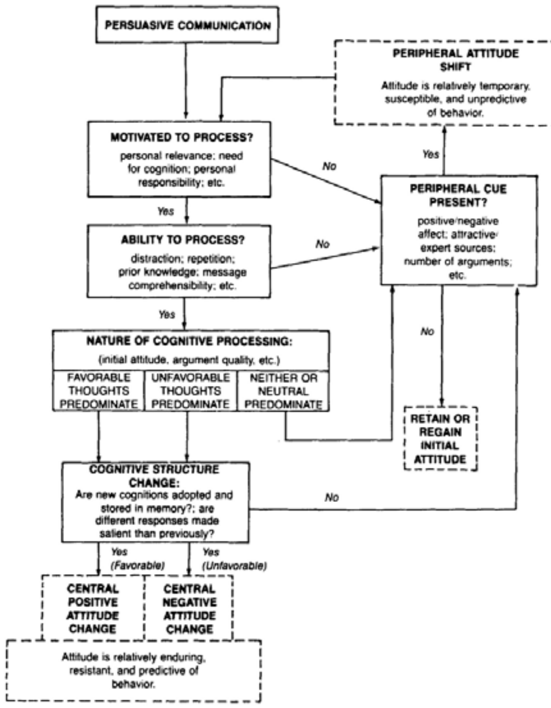


Figure 2: Central and peripheral routes (Petty and Cacioppo 1986, p.126)

So far, this report has dealt primarily with the more pedagogical aspects of our theoretical framework. But in creating an real time adaptive feedback system, we must be able to detect the student's real time learning behavior in order to be able to respond to it. Where previously mentioned studies (Lester and Stone 1997) have used the mouse position or rollover text to determine the student's focus of intention, we intend to use an eye tracking system which is interconnected with the game itself. We believe this method to more accurately measure the student's real time attention than what is possible with a indirect/secondary response such as mouse positioning. In the following sections we will present the more technical aspects of our theoretical framework, that is, systems which will be responsible for detecting student behavior and invoking intelligent digital responses when feedback neglect is detected.

Eye Tracking

The eye tracking technology can be described as a system which measures the movements of the eyes. It allows us to determine which object someone is looking at and which objects that s/he ignores. In order to track this data the eye tracker needs a camera and a light source, which often contains infrared light. The camera is then able to catch the reflected light in order to capture the movements of the eye and calculate the direction of gaze. The new technology also enables the possibility to measure widening of the pupils. New research has shown that this data is relevant in order to measure people's attention, but it will not be taken into account in the solution of this project (Tobiipro 2018a).

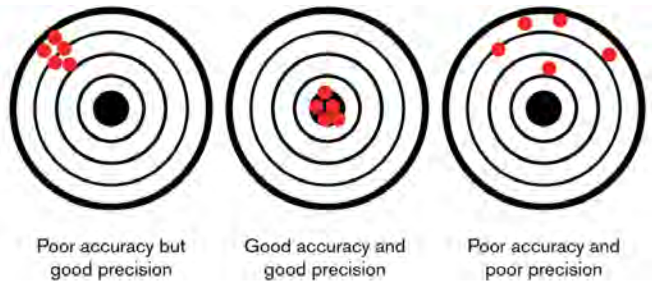


Figure 3: Illustration of accuracy and precision

Precision, Accuracy and Calibration

In order to validate the quality of collected data from the eye tracker, one analyzes the precision and accuracy. If these values are good, the result provides a more reliable collection of data. Accuracy is defined as the average difference between the real gaze position and the measured gaze position calculated by the eye tracker. Precision on the other hand is the eye tracker's ability to replicate the same gaze point measurement. Due to the incorrectness that may occur, the result has to be seen as approximate values of the real gaze direction. A better understanding of these properties is illustrated in figure 3.

Before getting started with an eye tracking session it is necessary to complete a calibration in order to establish a mapping of the coordinates on the screen as well as some other measurements. In order to calibrate, the participant is asked to fixate her gaze at an appearing and disappearing dot that can appear anywhere on the screen, a so called "calibration dot". During this time a camera takes multiple images to analyze the eye movements.

Another relevant factor is the frequency, which determines how many samples the camera is taking of the eye each second. The higher the sample rate, the higher the possibility of calculating the correct eye movements. This needs to be weighted against the cost that these increased number of samples causes has, due to more advanced camera equipment. (Tobiipro 2018b)

Directed Attention

Where we place our gaze direction in not randomly chosen, it depends on both bottom-up and top-down processing. The bottom-up processing occurs when attributes in the picture unconsciously catches our attention, such as contrasts, edges and colors (McLeod 2008). That is the reason why bright colors and details are getting more attention than faded colors and areas that are immutable. Movements in an image is also a significant factor that catches the attention of our gaze direction. All of these factors influence our behaviour without us even noticing it and it all happens very fast. Simultaneously, the top-down processing also effects our attention based on more individual qualities. These are based on our previous knowledge, how we interpret the image and the whole context in which the image is placed. (McLeod 2008)

Area of Interest

Areas of interest (AOI) (Holmqvist et al. 2011) are elements in the game that are of importance for the student in order to solve the task at hand. Example of such elements are the feedback box and parts of a test that contain valuable information about what the student answered right or wrong. Our goal is to make sure the areas of interest are noticed by the students and to implement the use of nudges if they are not.

Subtle Gaze Direction

Subtle gaze direction (Bailey et al. 2009) is a technique utilizing the eye tracking technology to make a person attend to something on a computer screen which they perhaps otherwise would not, by modulating either the luminescence or the shade of the area. Making use of the fact that the eye only sees clearly in a small angle from the gaze direction and that the eye is blind when saccading, the person whose gaze is subtly being directed will be unaware of this manipulation actually taking place.

Within the subtle gaze direction project (Gaze Manipulation Project 2012) several experiments have been conducted using it. One experiment shows that the technique was successful in adapting the test person's gaze patterns when modulating an "uninteresting" part of a static image (Bailey et al. 2009). Likewise, it has been shown that subtle gaze direction had an effect in a searching task even when distractions were present (McNamara, Bailey, and Grimm 2009). That is an important finding in our case, as the game we are using is interactive and animated. In an experiment done by Sridharan et al. (2012) subtle gaze direction was used in mammography training. The subjects were guided to use the patterns of an expert radiologist when reading the mammograms using subtle gaze direction cues (Sridharan et al. 2012). Results show that the students who had been guided by subtle gaze direction through their initiate training performed better than the control group (Sridharan et al. 2012).

Read

Research suggests that eye tracking can also be used to detect different reading behaviours. One paper presents the eye tracking system as being able to distinguish between text scanning behavior and actual reading behavior by categorizing the eye movements depending on whether the gaze pattern indicated "reading" or "not reading" ("pooled evidence")(Campbell and Maglio 2001). This system recognized reading in almost 100% of the cases, but also miscategorized scanning as reading in 40% of the cases (Campbell and Maglio 2001). Another technique which is able to detect different reading behaviors is the hidden Markov model. This model has been used to differentiate between reading and not reading with a precision of 88% and recall of 87% (Kollmorgen and Holmqvist 2007). It has also been presented differentiating the three different tasks of simple word search, finding a sentence that answers a question and choosing the subjectively most interesting title from a list of titles (with an accuracy of 60.2%) (McNamara, Bailey, and Grimm 2009).

Behaviour Sequencing Engine

To combine all these features and theories in our conceptual design we will make use of The Behaviour Sequencing Engine, which was developed by Lester and Stone (Stone and Lester 1996; Lester and Stone 1997). It works as a "competition-based framework for dynamically sequencing animated pedagogical agents' believability-enhancing behaviours." (Lester and Stone 1997, p.1)

The pedagogical behaviours are sequenced by a pedagogical sequencing engine, which is complemented by a believability-enhancing behaviour sequencing engine. The believability-enhancing behaviours are in an ongoing competition with one another, and at each moment, the strongest eligible behaviour is heuristically selected by its exhibition strength. By exhibiting the selected behaviour, the agent will behave in a way that "significantly increases its believability without sacrificing pedagogical effectiveness." (Lester and Stone 1997, p.2). Each behaviour's growth rate of exhibition strength may at each moment be either increased or decreased by the behaviour scheduler, depending on the current context, this is used to create a global behaviour based on multiple local behaviours. By using a behaviour sequencing system such as this, our different solutions (e.g. the agent giving student feedback, advice etc.) can, through a system of classification, be presented in a form that is both adapted to the individual user and appropriate for the situational learning context.

3 Methods of design and implementation

Through our theoretical framework several ideas on how the current game experience could be modified to prevent feedback neglect arose. In this section both the conceptual and implemented modifications and additions to the current game environment is presented. This is meant to illustrate the discrepancy between concept and initial implementation in our methodically approach.

Behavior Sequencing or a Paper Mache Dragon

Our system has been designed through two parallel design processes. One of the processes focuses on the conceptual design of a system-wide model that could be used to transform the interactional structure of the game. This is based on the ideas of the behavior sequencing engine, as a way of structuring believable agent behavior. Our model design have certain modifications to make it compatible with the goal of creating an intelligent and adaptive system. However, the complexity of creating a feedback system that functions in real-time have led us to take a different methodological approach in how we implement our ideas at this point. The implementation is therefore a feature testing system situated in one of the game environments (specifically the work environment for Newton). This environment it used to test and illustrate some of the features of the conceptual model, such as utilizing eye tracking to activate social feedback responses from the teachable agent or behavioral manipulations, such as SGD. The implementation and testing environment is because of this simplification, not relying on the complexity that a sequencing engine requires. The implemented system therefore does not share a structure, but only certain features with the conceptual model in its present form. However, for larger implementations than the scope of this project allows, one should lean more towards the implementations suggested in the conceptual model. The features that the models do share are presented below.

Area Of Interest (AOI)

Where students' focus of attention previously has been determined by the mouse position and the rollover text in the interface (Lester and Stone 1997), this study will make use of the current eye tracking techniques available. By selecting areas of interest (Holmqvist et al. 2011) of different sizes and with different placements in each learning environment, the users' gaze patterns will determine if they are paying enough attention to important features, such as constructive critical feedback (CCF) texts. If they are not paying enough attention, different pedagogical as well as life-like TA behaviors will be sequenced into the learning environment. These TA behaviors are aimed at motivating the student to (once again) pay attention to the feedback provided. The size and placement of the AOIs need to be tested and revised further depending on the different results generated by continuous user tests.

Subtle Gaze Direction

The subtle gaze cue will occur when a student is neglecting to notice the feedback box. A certain area within the box is then accentuated by a barely recognizable shift in either color or light. When the eye tracker notices the student's gaze moving towards the accentuated area, the system turns the area into its original state (Bailey et al. 2009). The student has through her peripheral view unconsciously noticed the accentuated area, and as her gaze is drawn to the area, the disappearance of the accentuated area will lead her to the conclusion that she herself made the choice to look at the feedback box.

We implemented subtle gaze direction according to the conduct used in Bailey et al. (2009). We used the same frequency of the eye tracker and the same calculations for categorizing fixations and saccades. We made one exception though (Bailey et al. 2009);we did not make the colour warmer according to the given formula. Instead we added a dot with a yellow filter which contrasted just enough with the feedback box. (Bailey et al. 2009).

Direct Gaze Cues

If the TA behavior and SGD are not having the desired effect on the efforts made by the student, more direct gaze cue techniques, such as making the entire feedback square occasionally yellow, may be applied.

4 Conceptual Model

With its foundation in the behaviour sequencing engine we have created a sketch of a new system which hopefully will be able to more adaptively handle more complex scenarios. The system that you see in figure 4 is a way to both create lifelike agents and make a system which is based on both user action, context, and previous actions/responses. The system creates a personalized response space that can itself determine when/what/where a direct action or pre-emptive solution is appropriate and execute this response in accordance.

One must keep in mind that this sketch is not a technical one and the details of the implementation and the exact functions for maintaining and updating the response space is not yet established. Instead one should rather see this model as an idea for extending the behaviour sequencing engine to accommodate individualized real time adaptive feedback. Many of the ideas of the system are inspired by our theoretical framework presented previously in this paper. However, some are based on logical assumptions and their real value should be determined by further testing. Next we will look at the parts of the model that are not explained in the behaviour sequencing section.

User profiles

Drawing from the Elaboration Likelihood Model (ELM) we have created a two part user profile, one is a long-term user profile which keeps track of the students overall motivation and ability and the other is a session profile which gauges the current motivation and inclination to engage with and understand the material. The long-term profile is a slower set that uses overall performance from each session to update itself and serves as a basis for determining how certain actions should be interpreted and also track the student's changes in educational attitude. The session profile uses the long-term user profile as a base, adding current actions and results to help decide and update the response space as well as in interpreting current student behaviour.

Behaviour determiner

The behaviour determiner is a subsystem that should serve as a way to make the system's beliefs about the student more accurate by looking at how the behaviour of the student should be understood based on the past actions and the correlation with positive outcomes and the user profiles. One could for example imagine a scenario where the student's gaze is passive. In such a scenario the behaviour determiner first decides that the eye movements are passive using the immediate history of eye movement, next using the user profile, current context, and success storage to interpret the behaviour, as it can be either a result of distraction or contemplation This part of the system therefore both categorizes and interprets student behaviour.

Short term history of responses and success

This part of the system tracks the way the system has responded, how the student has responded, and how successful the solutions have been. It's gathering information over time to create appropriate changes in the response space and also give feedback to the behaviour determiner while updating the user profile.

Updating the system response space and the response space

This is a part of the updated model where we have expanded the response space to include non-agent behaviour. In the response space there are three major groups of responses; relational, situational and comprehensive. The three categories are not mutually exclusive

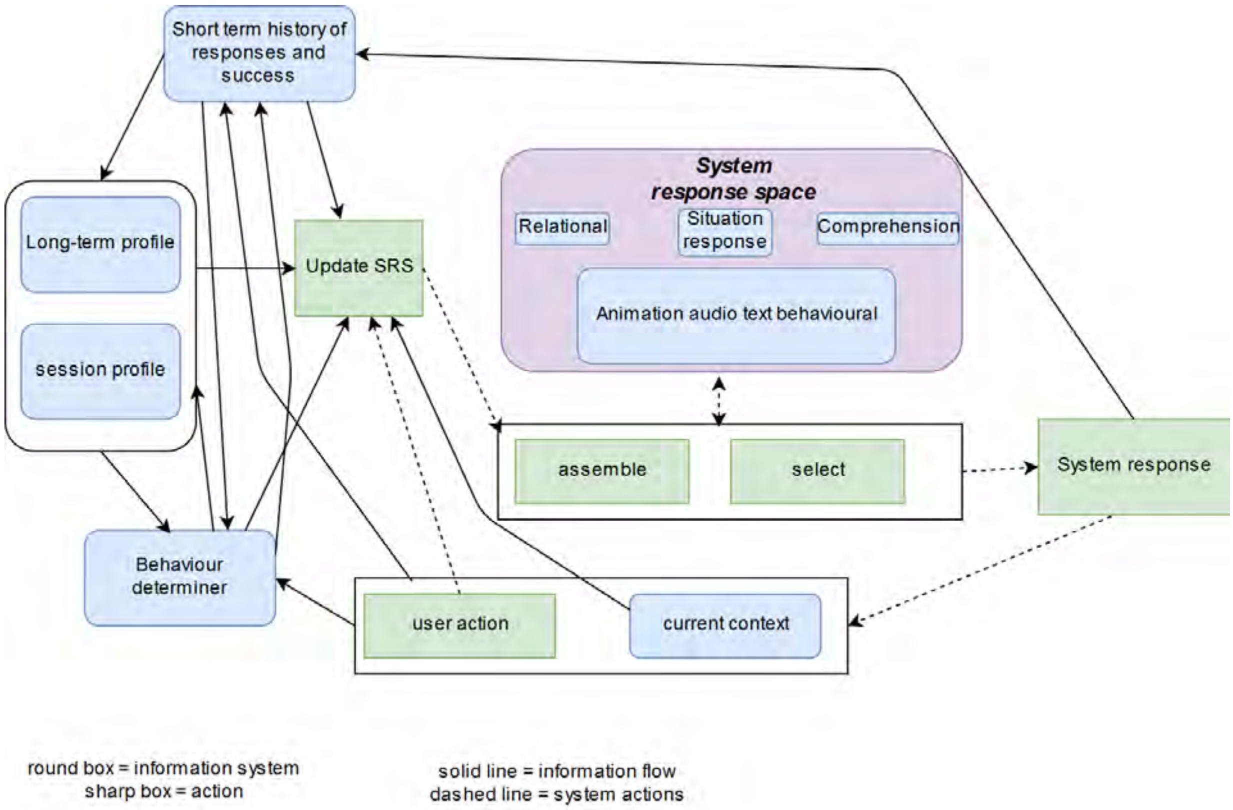


Figure 4: A view of the model

and solutions in one might be part of solutions in the others. Relational responses are designed to increase the value of the relation between the student and the agent. As we have discussed before this is related to better learning behaviour and increased motivation.

Situational responses immediately act in the situation to change current behaviour or disrupt it. When one of these solutions is activated they become prioritized over other solutions given that they are needed here and now. They serve as a way to control student behavior in a more direct way, for example by using subtle gaze direction cues.

Comprehensive responses are designed to help the student engage with the material in a meaningful way. These responses could be the agent reading the feedback to the student, marking important aspects of the information given, using non language solutions such as helpful symbols, using the agent to explain the information etc. (see (de Koning and Jarodzka 2017; Goodwin 2000)).

After the appropriate responses have been selected there are different types of tags differentiating between what type of action it is. Each action can themselves have different subcategories e.g. animations can be happy, sad, bored etc.

So when we are updating the response space we use the information gathered in the current cycle to push responses either away or towards an activation threshold and at the same time changing the threshold values depending on how the student has responded earlier towards system actions e.g. a motivated student with low session motivation might trigger an engaging agent (a bit pushy perhaps) but if this fails to help the student that type of behaviour from the agent might change its threshold value or promote other tags such as relational or it might make the behaviour more likely to occur when the student is highly motivated but distracted in session. The idea is that a system which uses an updating system response space can better match the user and the responses but also use certain responses in novel ways if they seem to have a desired effect.

Discussion on the system

So why is this system helpful? The belief is that by using a two part user profile system, the behaviour determiner and the updating response space creates a system which should be able to be adaptive without to much interference from outside sources. The two part user profile contributes by being able to differentiate between what is occasional mood swings vs long term behaviour patterns and that should lessen the likelihood that unnecessary actions are taken.

Our goal is for the students to be an independent entity in relation to the system, while with the system we wish to be able to accurately determine what is needed through profiling, eye-tracking, and a continuous self-updating, so it responds only when needed. By letting the threshold values rise over time the system can determine the often required responses for each student behaviour and remove unnecessary scaffolding and retrieve it depending on the situation. By using real-time eye tracking we can get real time data on behaviour and make the necessary situational adaptations.

If we look at the black box of feedback we can see that the types of responses proposed should correspond to the drop of points -both through the more passives solutions such as the relational agent and comprehension support but also through more direct actions such as getting the student to actually notice the feedback by using situation responses. We hope the amount of pupils who notice the feedback will be more likely to also read the feedback. They will then have a better chance of finding the answer/solution, sometimes through additional agent support.

5 Implementation Process

Developmental Method and Process

The game The Guardians of History is created with Meteor, which is a JavaScript web framework and React JS which also is based on JavaScript. Since the project already from the start had a huge amount of complex code it took quite a while to comprehend it. To simplify the work we were given a file named eyetracking.js that handled the incoming information from the connected eye tracker. The code had both been read and edited in Visual Studio Code. In order to share the code between multiple collaborators and avoid problems, like double maintenance, we have used git as version management program. This allows us to work on the same project on multiple devices without affecting each others code. This tool allowed us to conveniently make features and create a history of former commits. As the branches allowed us to work with several features in parallel without interfering with each other's work, we were able to simultaneously develop one branch for subtle gaze direction, another branch for a good base environment with a visible dot for eye tracking position, as well as a test branch, which has been used during usability testing. There has only been a few minor issues during merging, but nothing that we weren't able to solve right away.

The development of the code has mainly taken place during October, November and December 2018. A chronological outline was made based on previous knowledge and relevant theoretical background, which provided useful prerequisites for the development phase. This has made it easier to divide the workload between team members and clearly define which functions and actions that are desirable in the game. Changes in the outline have been made along the way, in consultation with the other group members. It has been important that everyone was made aware of changes that might affect their part of the project.

The eye tracker that was used during this project was a Tobii Pro Spectrums, which was placed right under the screen. Along with the eye tracker we used the calibration program Tobii Pro Eye Tracker Manager. A calibration is done in order to adjust the eye tracker to a specific user's eye movements. A computer is also required in order to successfully run the program and to connect the eye tracker to. During the development process we realized that a working memory with more than 4 RAM is necessary. The programs are basically too heavy to run on a computer with low memory.

Designing the System

During the developmental phase different ideas for the system has been discussed but a more prominent one was a tree-like structure, or more like a slime mold structure, where behaviour cues lead towards new decision points where motivation/ability levels and previous reactions to the system determined the choice of strategy. The system could then update the current situation and other context factors to create a new action plans when any significant response was noticed. One problem beyond the complexity involved in such structure was the problem of handling our TA; the structure could call on behaviour from the TA but not "control" it. After a while, we chose to divert from that path and went with the before mentioned system which is conceptually based on the behavior sequencing engine. This route was chosen as it allowed for a more adaptive system where multiple behaviors could be displayed, evaluated and indexed according the the individual users response patterns. The hope is that this will lead to a more effective and individualized feedback system.

At the moment this system seems promising although it is too complex to implement during this project. Still, by adopting and adapting the behavior sequencing engine we will have a system which is able to select and implement different solutions depending on the current context as well as on previous learning history of the student and previous behaviour of the teachable agent. This deep learning system would be far more adaptable than the stricter tree structure we were initially sketching on.

Creating a Prototype

One of the main goals with creating the prototype was to be able to both display the scaffolding we have created and show the technical aspects of the work and the use of eye-tracking in real time. The idea was that we could show a wide variety of scaffolding techniques based on a made-up student profile where we assumed the student to have low ability and low motivation. By trying to create a small system with increasingly strong responses we could make use of an adaptive agent that showed relation building, motivational behaviour, comprehension support in the debriefing and lifelike behaviour. By utilizing real-time eye-tracking we would also be able to let the user's focus (such as looking at Timy) trigger the agent to act. Without the eye-tracker a response would only be able to occur as a result of student action. The scaffolding techniques used were based on our theoretical framework where we have tried to create a believable agent and also use additional non-agent solutions to increase the likelihood of student elaboration and decrease the likelihood of feedback neglect.

In the prototype we use the agent as a way of increasing both motivation and ability by utilizing different types of social behaviour. We also make use of the agent as a way of creating easier access to useful information. Apart from the agent, we have also used techniques described earlier such as SGD when social scaffolding was not sufficient enough. Based on the fact that we are dealing with a low ability/low motivation student, we also use non-agent comprehension support such as making useful information more salient. As one can see these types of system responses are of the same kind as those described in the conceptual model. The responses are just structured differently, due to our simplified implementation environment.

Pre-existing Problems in the Game

The game itself was not an ideally functioning system when we started to develop it further. The code was complicated to understand and to work with. This game is mainly a research tool, which means that it is subject to continuous game development processes by other programmers.

The given feedback was proved to be a bit problematic for the users. The feedback that said that two out of five answers were wrong was confusing in the multiple choice test. This was due to the fact that unchecked answers that should have been marked were wrong as well as marked answers that should not have been marked. This led to inconsistent feedback, since the feedback could say that one should learn more about a subject on one try and a different subject on the next try, although the answer from the first try was still wrong. This meant that the users were frustrated when later realizing that they had not fixed the problem which they first assumed they had.

Another issue that was discovered during the development phase was that the assessment of the quiz answers was not always correct. Once in a while the specific feedback contained inaccurate information, thus expressing the existence of an inaccurate answers in a question that was correctly answered. This led to some confusion regarding what answers were right and what that were wrong. We did not attempt to fix this issue as the code for assessment was not of great importance in our project. Also the assessment involves several different parts of the code and it would be to complex for us to interfere with.

Debriefing

A problematic part of the game was that the quizzes did not include a debriefing part after the test was done. Hence, the user was given no opportunity to actually check the answers to identify the errors or possible corrections after the specific feedback was given. This was something we really wanted to change since we found the time to evaluate the performance before moving on to be very valuable for the students. After the specific feedback was given by Chronos we implemented a break in the game. Here the student was able to review the answers according to the given feedback. We also had to disable the possibility to change the answers since that should only be possible when a new test is starting. Once the student was done looking at the answers, a short dialog with the teachable agent took place, based on the feedback and on the eye tracking data. If the student had read the feedback from Chronos, Timy simply reminded the student about what subject to learn more about and asking the student if s/he was done reviewing and wanted to continue. In case the student had not paid attention to the feedback given by Chronos the dialog from Timy would not contain any information about which subject to focus on, but rather a suggestion from Timy to take a break and come back and learn more later due to him being dizzy.

From an ELM perspective we should support the careful consideration of available information but at the same time we should be try not to encourage using shortcuts. We therefore chose to create separate dialogues for different action patterns where we could support the student by providing information but only if the previous information had already been displayed.

Obstructing Information Termination

Another action that is easily made by mistake is to remove current information in the game by pressing anywhere on the screen. A problematic example is the specific feedback from Chronos that can be removed this way. We considered this feature to be out of our scope, but in order to show our implementations more properly on the exhibition, we implemented a simple solution to prevent feedback being too easily removed. The easiest way to solve this was to enable the student to confirm that the feedback was read by pressing on the dialogue box right below the specific feedback. This makes the student having to press a smaller area and thereby securing the feedback from disappearing by mistake.

Eye Tracking

One of our main goals was to integrate eye tracking into the game. We have received external help with this in order to configure the connection correctly. The game is searching for a device that corresponds to the eye tracker and will be given coordinates for the current gaze position at a specific rate depending on the device's properties. These coordinates are given immediately, which opens up for real-time adjustments in the game. Like many other web applications the game is written in JavaScript. JavaScript has a very useful function called elementsFromPoint(). This function takes to parameters x and y that corresponds to a position on the screen and are provided by the eye tracker. By sending these parameters to the function it returns a list of all the elements that are located at that exact position. We have chosen a few elements of interest, and actions are made if these elements are found to be on the list.

Where the user is looking is not displayed on the screen while playing, but has been marked as a red dot during development. This has been very beneficial for testing functionalities and to confirm that the eye tracker is working properly.

Timy's Reactions

These functionalities of eye tracking, among other, is utilized in regard to the TA's actions. Since Timy is the student's TA he plays a crucial role in the game. After the quiz is completed and the feedback is presented by Chronos we have implemented a few features to ensure that the student will notice the feedback. If the student ignores the feedback but instead looks at Timy there will be three responses to nudge the student to look at the feedback box. The first response is Timy discreetly glancing and pointing towards the feedback. If the student does not turn her attention towards the feedback but continues to look at Timy a second response will occur. This time Timy will whisper to the student "Pssst… Chronos seems to have some advice for you!". The reason why he is whispering is due to the fact that the student is hiding and should not be seen by Chronos. The third action is another dialogue and this time Timy is asking for help, "I don't understand what Chronos means, can you help me? What else do we need to do?"

The reasons behind our design are basically that scaffolding should only be used if needed (making Timy respond only when feedback is neglected) and to showcase three different types of social scaffolding. The initial pointing and looking will hopefully make the student follow Timy's directional hinting. The second response is more direct and refers directly to Chronos through a text box, while staying true to the underlying narrative of the game. The last scaffolding technique is more of a relational nudge where we try to increase student motivation by focusing on his/her relationship with Timy and the benefits of the protegee effect (Chase et al. 2009) .

Timy's Believability

Timy is a time elf who is to be taught history and physics by the student user. S/he therefore has to be perceived as less knowledgeble than the user. S/he also has to be perceived to be about the same age as the student for him/her to be able to identify with Timy as a learner. Apart from these required learning-by-teaching features of the tutee Timy, we also want to make his/her behavior situated in the current context of the game as well as in the interaction with the user.

Due to a heavily restricted time-frame, limited technological knowledge of animation-processes, as well as a need to prioritize among what intelligent solutions to implement, we choose to display only a few believable features in Timy. Hopefully, they will be sufficient enough to motivate the user to make the effort needed to look at, read, understand and act on the provided feedback within the game. We chose to illustrate, animate and implement the not-verbal believability features of body language, facial expression and gaze, as well as the verbal feature of Timy making a joke.

The non-verbal features of Timy were created or enhanced to make him/her act in an appropriate way in the social context of the tutor-tutee interaction. This was accomplished by enhancing the joint attention by modifying Timy's body position and arm gestures, as well as adding directional gaze patterns. Timy's facial expressions were adjusted according to his current actions to make his/her overall behaviour coherent to the user.

According to Gary Dean Stelling's early work on agent-inhibited interfaces, an agent should minimize his/her face time during problem solving, in order to be perceived as well-meaning and on the user's side, instead of behaving as an observer of student behaviour (Stelling 2002). We agree with Stelling that having the agent facing the task instead of the student user, would make it more believable that Timy was watching and learning from the student. We therefore illustrated three picture frames that when animated depicted Timy turning around, facing the assignments that were to be solved. If Timy is clicked on by the user while facing the task, he will simply turn around and face the screen. Timy will then change position again, depending on whether the user decides to continue with the quiz or proceed to the correction part. Hence, Timy will always face the student during conversations and always face the board when the student is teaching Timy about Newton.

After finishing the test and receiving feedback from Chronos, if the student doesn't look at the feedback provided, Timy would then look at the feedback while at the same time pointing with his left arm towards it (Timy is now facing the student again). His/her facial expression has during the gazing/pointing behavior gone from a smile to a more neutral expression, hopefully signaling that it is important for Timy that the student pays attention to Chronos feedback. Otherwise he won't learn enough to be able to pass the tests and join the team.

During this project we initially choose not to focus on the written text boxes provided by the agents of the game. Working on the conceptual model, not knowing how to make very low motivated students stay within the game, a study by Agata Janiszewska was recalled (Janiszewska 2017). The results of her study showed that an entertaining pedagogical agent were able to affect learning, motivation and attention in a favorable way in students with low academic performance. A joke or off-task interaction could also provide a brief cognitive break in the learning process.

If the system notices a student neglecting feedback over and over again implementing an entertaining behavior in Timy would possible keep the student looking at the screen, giving them a cognitive break, while at the same time giving Timy the believable behavior of a peer/friend (sometimes joking or being off-task).

Relation Building Actions

We have also implemented a relation building sequence if the student, despite both levels of subtle gaze direction, does not pay any attention to the feedback. Timy will then make a joke where he wants the student to guess how old Chronos is. The joke is divided into three dialogue boxes provided by Timy and no other scaffolding is active during this period of time. These responses are implemented as a last resort when the system detects a gaze behaviour that indicates that the student has lost interest or focus.

Graphics

A few different versions of Timy have been created so far in order to make the agent more believable (Bogdanovych, Trescak, and Simoff 2016). Further on, several frames of each behavior will be animated to convey a life-like TA behaviour.

Distinguishing Between Noticing and Reading

Our solutions mainly focus on getting the students to notice feedback. In order to find out if a certain element has been noticed by the user, we utilize the eye tracker to see if the element is among those that the student fixates his/her gaze on. However, we have also implemented measures to detect reading. We rely on the assumption that if the user spends more than a certain number of

Figure 5: Different versions of the teachable agent Timy from The Guardians of History

seconds within a certain AOI, it will qualify as reading. Although there are more reliable ways to detect reading behaviour, we have chosen not to implement these as they are not presently accessible to us. However, in our solution the reading does not necessarily have to be continuous, since the algorithm is based on the total time. The amount of seconds is calculated by dividing the number of times that the element has been seen by the user with the frequency of the eye tracker. The frequency corresponds to how many times the eye tracker is sending coordinates each second. This has usually been 20 or 30 times, depending on the equipment used.

Integration Testing

One thing that could have improved the quality of our scaffolding responses would have been to use integration testing, which means testing our implemented functionality in a bigger context. This is due to the fact that our features worked well independently but should have been tested in the game context, with credible actions before and after our additions to the code earlier. We also eventually detected some errors with the game scenario that had to be solved in order to test the game together with the eye tracking and the improved feedback. To test individual implementations in the bigger game context more frequently to discover these kinds of issues is therefore recommendable.

6 User Testing

Method

We held five user tests, each taking between 30 minutes and an hour. All test subjects were university students, around 20 years old.

During the tests one of us functioned as the test leader and one of us as the test observer, both of which sat with the current test subject. The test leader gave instructions and answered questions. The test observer made notes and conducted the eye tracking calibration.

Data on how long the test subject looked at the feedback, as well as when SGD was activated and deactivated, was logged.

The first two test subjects had only one task; to do a quiz about Newton and then continue doing it until they succeeded.

After the first two tests, a second test was added to the initial test. The SGD scaffolding was not triggered in the first two tests. The intention with the initial test was for the test subjects to fail, but not to look at the feedback. We wanted to see if the test subjects noticed the subtle gaze direction and if their gaze was directed towards the feedback despite actively not looking at the feedback.

7 Results

Task 1

The results of the initial test, done by five test subjects are presented in table 1. Furthermore none of our scaffolding responses were triggered in any of these tests. The test subject simply closed the feedback box instead of reading it.

Task 2

Result of the second task could not be confidently gathered as a bug made the logging flood with messages, making it impossible to figure out which one that was correct.

What can be said about the second trial is that all three test subjects triggered and deactivated the SGD dot. Out of these three, two subjects reported on noticing the SGD dot when being asked about noticing anything different with the feedback box.

Discussion of Results

Target Group

Our target group of the implementation was low motivation and low ability middle school children.

Due to time constraints we did not do the user testing on the target group.

We tried to give the test users low ability by not letting them access the necessary information about Newton beforehand. This resulted in none of the test users succeeding on their first try.

Due to not using the intended target group we cannot see if our adaptions led to any general improvement of users' ability to notice, read and act upon the feedback. Differences we have seen between our tests and previous tests of the game done on the correct target group by Tärning et al.(Tärning 2018) is that in our tests all test users read the feedback and acted upon it at least the first time they received feedback. Another reason for this is that in our test the task was to complete the quiz, rather than playing the game, so that the only way the test person could finish the quiz without succeeding was by giving up.

Testing of Subtle Gaze Direction

Our plan on testing the SGD was by logging the data regarding when SGD was triggered and deactivated.

One issue with this was that SGD was not triggered when task 1 was made.

Thus task 2 was added, solely to trigger SGD. Results from this trial would not be as valuable as they would

Table 1: Results of task 1 of the user testing.

	Number of times detailed feedback was shown	**Number of travels to Newton**	**Felt finished when**	**Read (sec per round)**	**Comment**
1	2	1	When succeeded quiz with detailed feedback.	7 / 4	
2	4	3	Did not feel finished when detailed feedback was still presented.	10 / 6 / 2 / 8	Got incorrect feedback from Chronos.
3	8	3, after the first second and fourth round of the quiz	Gave up due to disinterest.	10 / 13 / 10 / 7 / 3 / 8 / 7 / 6 / 6	In the end: tried to pass the quiz by reading Chronos feedback and trying the quiz again based on that.
4	3	2	Was hesitant about being finished as detailed feedback was still presented, decided to be finished after travelling back once after succeeding.	6 / 11 / 6	
5	3	1	Did not feel finished as detailed feedback was still presented.	10 / 10 / 9	Did the quiz once more after passing and then succeeded with 100% correct.

have been from doing the task in the intended game context. Still, we were able to see whether SGD made the person look at the feedback even though it was told not to look at the feedback. We were also able to ask the subject if s/he noticed the SGD dot or not.

Also the execution of task 2 was problematic due to the logging being bugged by the activation of the SGD several times. Due to this we deemed the data unreliable. To prevent this we could have tested the feature more thoroughly beforehand, and also timed the activation and deactivation with a stop watch, which however might have been hard to do without drawing the test subjects' attention to it.

However, it can be noted that a majority reported noticing the SGD dot when being told not to look at the feedback.

Findings from the User Testing

Some findings from the user tests that were valuable and we think applies to users of all ages, is that a reason for not looking at the feedback is closing the feedback box instantly. This is easily done in the game as the feedback box is closed with only one mouse click. When closed, it is not possible to open it again. This also seemed to be the reason for our scaffolding not being triggered, either the user read the feedback, or they closed it immediately. To trigger scaffolding one would have to not look at the feedback while the feedback box was showing.

Some behaviours we noticed were usage of brute force testing, or brute force based on Chronos' feedback and being confused about whether one was finished or not when passing the quiz but still receiving CCF.

How the User Testing Could be Improved

Had we done the project again we would do three iterations of user testing. The first of which would be testing the game without our changes, to learn more about usage behaviour, find usability flaws in the program and be able to use it as a baseline for further testing. Then our solutions would be tested for usability flaws and after solving those, the pedagogical effects should be tested.

Also, it would have been beneficial to focus more on user testing by letting one person on the team be solely focusing on the user testing, giving more time to plan and conduct tests, including making contact with a group of children through for example a school. To do this we would have needed to start with the user testing much earlier, already in the first part of the course.

8 Discussion

Success with Fulfilling our Aim

At this moment of development, we have designed a model for the system that should be the basis of further implementations, leaning on the idea of the behavior sequencing engine. To allow for a more adaptive user experience however, an updating indexing would need to be implemented. We have implemented different scaffolding effects of both social and behavioral nature in the existing system to better understand the advantages and disadvantages of using these. It has for example become clear that the AOIs and behavioral thresholds for activating scaffolding might need to be adapted as there were issues in testing with no scaffolding being activated. One could argue that this might be due to the test subjects being adult students and therefore mastering a better study technique and showing stronger perseverance, which would allow them to avoid the mistakes possibly made by middle school children.

However we are satisfied with how far we have gotten, as we managed to implement a prototype of the feedback model that provided adaptive feedback in real time based on the users' eye movements, while also providing a conceptual model for the entire system. This was our main aim as we from the beginning of the course were aware that we moved in to uncharted waters. A complete implementation of a fully functional system would not have been possible under the provided circumstances.

Social Elements

Although we have not yet been able to test the system thoroughly enough to determine the pedagogical value of our social scaffolding, we, based on our theoretical framework, expect it to be an effective tool in the system as it relies on the TA as a social presence within the game. Hereby utilizing this presence for scaffolding pupils' behavior towards more fruitful learning strategies. One way this could be done is by asking questions to the feedback as a way of directing the pupils' attention towards important features as a basis for reflection and further progress. Yet it is important that the TA does not overstep its intended function of a gullible tutee, as one risk of using this method is that the TA instead ends up acting as a tutor. However, an important advantage of using social elements as a strategy for scaffolding learning process, in opposition to simply influencing behavior with SGD, is that the students are consciously choosing whether to act upon the scaffolding stimulus. This consciousness can be fruitful for the goal of altering student behavior in the long run as they become aware of the productive act of noticing the feedback itself or even by directing their attention towards crucial parts of the feedback. Nevertheless, the method can be criticized for being experienced as intrusive as it confronts the pupil with her/his absent action. Intrusiveness can never be fully avoided in learning/teaching situations as the process of working with a new material does in one way or another influence the student. The art is to create a situation where the student welcomes this intrusion.

Social effects and scaffolding can also be implemented in less direct fashion relying more on bodily communicative cues and the implications of these, such as the TA pointing and gazing towards the AOI. These cues have been shown to have a positive influence especially on students' tendency to notice the feedback, in opposition to using non-social cues such as arrows (Tärning et al. 2018).

In utilizing the TA's presence to create a tool for social scaffolding his/her behavior has to be developed on a micro scale with small social cues, while also implementing more complex social scaffolding to get the pupils to reflect upon their own strategies on a meta level. In the end, the students hopefully will be able to accommodate a more beneficial behavioral strategy.

In creating and implementing Timy making a joke, we also discussed the overall language of Timy. Was his language age-appropriate for a 10-12 year old? Or was it too pedagogically adapted? According to William Yang Wang et al. peers challenge each other in learning-by-teaching situations by using challenges ("see I am helping"), insults ("you are so weird"), condescendations ("nothing you have done has effected me what so ever"), message enforcers ("Earth to Erin"), Pet names ("What's up homie") etc (p.3, Wang et al. 2012). These challenging interaction is not something that occurs in the early stages of a relationship when one is trying a make positive first impression. The Tickle-Degnen and Rosenthal model shows how positivity then decreases over the course of a relationship and instead becomes more challenging (impolite) (1990, mentioned in Wang et al. 2012). If the relationship with Timy during the game is considered by the student not to be extended enough to deepen the relationship, a positive yet age appropriate language will suffice. If the time in the relationship is considered extended enough, will Timy have to start challenging the user in an friendly yet impolite way to stay believable? Could pedagogical considerations stand in the way of the believability of Timy? And would a challenging TA behavior enhance or decrease student motivation and effort?

Subtle Gaze Direction

Our other method of directing attention is SGD, this for sure does not fall into the category of a social cue, as it relies fully an a bottom-up processing of a salient stimuli. However, we see multiple possible gains and risks of using such a method.

Possible Gains of Using Subtle Gaze Direction

The reasons for using subtle gaze direction as we see it are:

- Subtle gaze direction can be used several times, in contrast to for example increasing the saturation of the important elements in general.
- It has the potential to change the student behaviour without them noticing it, possibly feeling less intrusive for the user than alternative solutions.
- It is a technically interesting technique, testing the limits of the speed and precision of the available eye tracking system and integrating it in an interactive online game.
- The mammography experiment (Sridharan et al. 2012) suggests that the results of using gaze manipulation were present even after the training. This suggests that the students may better attend to the feedback boxes later in the game without the subtle gaze direction cue being present.

In a broader perspective, the idea is that by attending to the feedback, a larger part of the students will be able to continue processing the feedback and in the long run both learning more as well as gaining a positive experience of attending to feedback. The results from the experiment of scanning mammograms suggests that the students' gaze patterns could change even when subtle gaze direction is no longer present (Sridharan et al. 2012).

Possible Risks of Using Subtle Gaze Direction

It is not given that subtle gaze direction will give the desired results. There is a risk of the students not noticing the cue, which was the case with one of the test subjects. However, we need the logged data to determine this with certainty as the cue could have been noticed and attended to unconsciously. The reasons for the cue not being noticed, could be due to technical difficulties such as the sampling frequency being too low to turn on and switch off the modulation when needed. However, results from a previous experiment suggest that the gaze of the subject could still be successfully nudged into the desired location (McNamara, Bailey, and Grimm 2009).

There is also the obvious risk that the students not noticing the feedback have already moved on to another setting before the subtle gaze manipulation can direct them to the area of interest. In order for subtle gaze manipulation to be performed successfully the game might for example have to be edited so that the user cannot leave the page until she has attended to the feedback.

Also, even a perfect implementation of subtle gaze direction may not lead to the desired results in a broader perspective. Although it is possible that more students would notice the feedback in this game, even when subtle gaze direction is not present, the mere change of their gaze pattern is not something that is linked to better noticing feedback in a broader sense. The same study also showed that there were no long term effects based on the learners' results.

An Assumption about Learning Attitudes

A problem that needs to be mentioned is the possible consequences of the assumption that having positive learning experiences results in lasting positive change. If we look at some of the solutions we have proposed they can almost all be understood as peripheral persuasion. The problem we are looking at is the fact that social scaffolding and direct action both use persuasion techniques that nudges people toward engaging with the material because of non-subject-matter factors. In the short term it might look good if students start to act upon and make progress when engaging with feedback but if their relationship with feedback is solely based on these manipulations their attitudes toward feedback will not change in the long run and instead there is a risk that they will be dependent on the system for further progression in school work instead of acquiring actual strategies which allows for them to succeed on their own. However, as briefly mentioned earlier, changing a students self-efficacy (Bandura 1997) regarding how she sees herself as a learner may be interpreted as her taking a central elaboration route. By seeing her teachable agents knowledge as incremental (Mangels et al. 2006); changeable and improvable due to the quality of her own teaching efforts, the student's view of her own potential to develop and grow as a learner may be positively effected.

So to justify our model we need to assume that our techniques lead to meaningful contemplation on the material and the processes or that positive learning experiences have some effect that increase motivation/ability and possibly even self-efficacy. If this is not the case, one would be able to make the claim that our system could be a "curling method" that will merely get local results and not lead to meaningful lasting learning strategies. The knowledge they acquire may have just a short temporal impact if their learning experience is based on performance rather than elaboration. This is an important issue to explore in further research.

About the Game

The fact that we have already been given a complete game has both its advantages and disadvantages. Since the game is based on previous research we can assume a thorough theoretical background, qualitative research and well motivated design choices have been made. The game The Guardians of History combines traditional classroom settings and tests with time travels and motion sick elves. Our project therefore had to proceed from these prerequisites. A different game with a different benchmark could have lead to other conditions and issues to work with and may also have implied a different type of user interaction.

9 Reflection of Team and Team Work

Our Team

Our team for this project consists of five students from two different programs: three master students of cognitive science and two students of information and communication technology (ICT). The students of cognitive science have different focus and studying methods based on their bachelor studies. The ICT students have a more homogeneous background as they both focus their studies on usability and design. We have worked in two sub-teams where the cognitive science students have focused on exploring and utilizing relevant theories, and designing the conceptual model, while the ICT students have focused on getting into the code of the game The Guardians of History and implemented different scaffolding methods within this framework (both theoretically and practically). We have also had help from an exchange student who has taken part in the practical work of the programming sub-team.

Our Way of Working with this Project

In the first part of the course we explored the field theoretically to get an understanding of earlier work in the field and begun to get an understanding of the technical aspects of the game. While the second part of the course where dedicated to integrating this in to a conceptual model and a prototype for testing that would allow us to demonstrate some of our ideas. This process have required us to overcome both conceptual and technical obstacles. Here our work on the initial course allowed us to loan ideas from our theoretical framework and gave us time to come to terms with the existing code of the game and eye tracking as a technique.

When looking back at the work we have done so far within the project, we have come to terms that different team members have very different approaches on how to work iteratively - some preferred to have the theoretical framework done before making a draft, while others were used to making drafts even though they might have to be completely remade later on. we also struggled with how to make a prototype when we already have such a clear structure from the game. However, these issue was overcome in the second course since we changed our method of working to more individualized responsibilities, where each member had the responsibility of one or more areas and the deadlines related to those. This let to a smoother and more goal orientated approach.

Even though we had more individual responsibility it did not lead to less cooperation as we worked with weekly meetings of both the entire team and subteams. However, the smoother progress of the second course would not have been possible without our many discussions during the first course as these were helpful in opening the subject for us and giving us a broader understanding of the field that we were moving towards.

In this project we have had a lot of issues with the technology, for example it took a lot of time for us to get access to the code we needed, and then the computers we had access to were not good enough to handle the development we wanted to do. However, once this was sorted out, with the help of one of our tutors, we made some quick progress. This progress was also due to a change in work method, as we divided the work more between the participators with different responsibilities.

What we have realized in this project is the value of working interdisciplinary - we would not have been able to achieve the same results if we had worked individually or mono-disciplinary as our different educational programs let to a broader fan of methods, theories and skills which allowed us to combine many different approaches and thereby creating a more thorough and complete system.

10 Further Research and Development

Further research and development within this project should focus on a thorough implementation of the behaviour sequencing engine. Here the main focus should be on implementing it through the entire game, while further developing a more complex range of possible actions for the system to take, and exploring and developing a true reading detector, so the system could act appropriate based on a larger an more precise amount of data. This would allow for a more believable interaction with the TA, something that would hopefully lead to more engaged students with a lesser degree of feedback neglect. However, this leads us to our final and most important advice for further research with in this system, as one must conduct a number of studies and tests on the target group in order to explore the actual pedagogical value of the presented solutions. The optimal format for this research would be to study both the immediate implications of the solutions, do the students neglect the feedback less? But also the long term implications of the system, do it influence the student's study technique so they also reduce feedback neglect in situations where the scaffolding is not present.

There are though time to conduct this research and development as the system is before it's time, as the eye tracking technology is not yet a common feature in commercial equipment that would normally be seen in a school setting.

References

Bailey, Reynold, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm (2009). 'Subtle Gaze Direction'. In: *ACM Transactions on Graphics* 28.4.

Bandura, A. (1997). 'Self-efficacy'. In: *In V. S. Ramachaudran (Ed.), Encyclopedia of human behavior* Vol. 4, pp. 71–81.

Bates, Joseph (1994). 'The Role of Emotion in Believable Agents'. In: *Communications of the ACM* 37.7, pp. 1–2. DOI: 10.1145/176789.176803.

Bickmore, T. W. (2003). *Relational Agents: Effecting Change through Human-Computer Relationships.* Massachusetts Institute of Technology.

Blumberg, F C, E A Altschuler, D E Almonte, and M I Mileaf (2013). 'The Impact of recreational video game play on children's and adolescents' cognition'. In: *New directions for child and adolescent development* 139, pp. 41–50.

Bogdanovych, A., T. Trescak, and S. Simoff (2016). 'What makes virtual agents believable'. In: *Connection Science* 28:1), pp. 83–108.

Caillois, R and M Barash (1961). *Man, play, and games.* University of Illinois Press.

Campbell, Christopher S. and Paul P. Maglio (2001). 'A Robust Algorithm for Reading Detection'. In: *ACM Workshop on Perceptive User Interfaces*, pp. 1–7.

Chase, C C, D B Chin, M A Oppezzo, and D L Schwartz (2009). 'Teachable agents and the protégé effect: Increasing the effort towards learning'. In: *Journal of Science Education and technology* 18.4, pp. 334–352.

de Freitas, S (2014). *Education in computer generated environments. Research in education book series.* New York, NY: Routledge.

de Freitas, S (2018). 'Are Games Effective Learning Tools? A Review of Educational Games'. In: *Educational Technology & Society* 21.2, pp. 74–84. DOI: http://ludwig.lub.lu.se/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=128981057&site=eds-live&scope=site.

de Koning, B B and H Jarodzka (2017). 'Attention Guidance Strategies for Supporting Learning from Dynamic Visualizations'. In: *Learning from Dynamic Visualization.* DOI: https://doi.org/10.1016/S0378-2166(99)00096-X..

Deterding, S, D Dixon, R Khaled, and L Nacke (2011). 'From Game Design Elements to Gamefulness: Defining "Gamification'. In: *MindTrek'11, September 28-30, 2011, Tampere, Finland.* ?, pp. 9–14.

Educational Technology Group (2018). *Educational Technology Group - Homepage.* http://www.lucs.lu.se/etg/.

Egenfeldt-Nielsen, S (2005). *Beyond edutainment: Exploring the educational potential of computer games.* Game-Research.com.

Elliot, C D (1992). *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System.* PhD Dissertation, Northwestern University, Evanston, Illinois.

Gaze Manipulation Project (2012). *Project Overview.* https://sites.google.com/a/g.rit.edu/gaze-manipulation/.

Goodwin, Charles (2000). 'Action and embodiment within situated human interaction'. In: *Journal of Pragmatics* 32.10, pp. 1489–1522. DOI: https://doi.org/10.1016/S0378-2166(99)00096-X..

Haake, M (2009). *Embodied Pedagogical Agents – From Visual Impact to Pedagogical Implications.* Sweden: Dept. of Design Sciences – Lund's University.

Holmqvist, K, M Nyström, R Andersson, H Jarodzka R Dewhurst, and J van der Veijer (2011). *Eye tracking: A comprehensive guide to methods and measures.* Oxford University Press: New York.

Janiszewska, A. (2017). 'Is funny better? The effect of an entertaining pedagogical agent on motivation, attention and performance in learning in a digital teaching aid'. In: *Master Thesis. Lund University.*

Kirkegaard, C, Gulz, and A. Silvervarg (2011). 'Introducing a Challenging Teachable Agent'. In: *Computers & Education* 56, pp. 574–584.

Kollmorgen, S and K Holmqvist (2007). 'Automatically Detecting Reading in Eye Tracking Data'. In:

Lester, James C. and Brian A. Stone (1997). 'Increasing believability in animated pedagogical agents'. In: pp. 16–21.

Lim, S F, Ailiya, C Miao, and Z Shen (2013). 'The Design of Persuasive Teachable Agent'. In: *IEEE 13th International Conference on Advanced Learning Technologies.*

Lim, S F, Ailiya, C Miao, and Z Shen (2014). 'Persuasive Teachable Agent with Goal Net'. In: *IEEE 14th International Conference on Advanced Learning Technologies.*

Mangels, J, B Butterfield, J Lamb, C Good, and C Dweck (2006). 'Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model'. In: *Social, Cognitive and Affective Neuroscience* Sep; 1(2), pp. 75–86.

McLaren, B M, K E DeLeeuw, and R E Mayer (2011). 'Polite web-based intelligent tutors: Can they improve learning in classrooms?' In: *Computers & Education 56 (2011) 574–584* 56, pp. 574–584.

McLeod, Saul (2008). *Visual Perception Theory.* https://www.simplypsychology.org/perception-theories.html.

McNamara, Ann, Reynold Bailey, and Cindy Grimm (2009). 'Search Task Performance Using Subtle Gaze Direction with the Presence of Distractors'. In: *ACM Transactions on Applied Perception* 6.3.

Petty, Richard E. and John T. Cacioppo (1986). 'The Elaboration Likelihood Model of Persuasion'. In: *Advances in Experimental Social Psychology* 19, pp. 123–205.

Piaget, J (2013). *Play, dreams and imitation in childhood (Vol. 25).* New York, NY: Routledge.

Sridharan, Srinivas, Reynold Bailey, Ann McNamara, and Cindy Grimm (2012). 'Subtle Gaze Manipulation for Improved Mammography Training'. In: *Proceedings of the 7th ACM Symposium on Eye Tracking Research and Applications.*

Stelling, G D (2002). *Affective behavior control for lifelike pedagogical agents.* Master thesis, North Carolina State University.

Stone, Brian A. and James C. Lester (1996). 'Dynamically Sequencing an Animated Pedagogical Agent'. In: pp. 424–431.

Sutton-Smith, B and J M Roberts (1961). 'The Cross-cultural and psychological study of games'. In: *International Review for the Sociology of Sport* 6.1, pp. 79–87.

Tärning, B (2018). *Design of teachable agents and feedback in educational software – Focusing on low-performing students and students with low self-efficacy.* Sweden: Dept. of Cognitive Science– Lund's University.

Tärning, B, Y J Lee, R Andersson, K Månsson, A Gulz, and M Haake (2018). 'Looking into the black box of students (not) handling feedback on mistakes.' In: *Design of teachable agents and feedback in educational software : Focusing on low-performing students and students with low self-efficacy.*

Tobiipro (2018a). *Eye tracker accuracy and precision.* `https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/`.

Tobiipro (2018b). *Eye tracker accuracy and precision.* `https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/what-affects-the-accuracy-and-precision-of-an-eye-tracker/`.

Wang, W. Y., S. Finkelstein, A. Ogan, A. W. Black, and J. Cassell (2012). 'Love ya, jerkface": using Sparse Log-Linear Models to Build Positive (and Impolite) Relationships with Teens'. In:

Spatial Organisation

Richard Magnusson, Kristoffer Mårtensson,
Anton Wrisberg

dic14rma@student.lu.se, fpr14kma@student.lu.se,
an0823wr@student.lu.se

This paper presents the work of a course at Lund Univerity revolving around academic solutions to real-world problems. The project presented in this paper seeks to solve the problem of the ever-increasing number of applications and services, their internal indexation, and discoverability of new services. The proposed solution uses a spatial understanding as foundation for a new take on indexation challenging the conventional way of thinking about categories based on common one-word associations.

1 Introduction

This section presents the theme of this paper, the problem description central to our investigations, the solution we have been developing and investigating, and relevant research.

The Work of Research and Development

This paper presents the results of the work within a course on research and development at Lund University in the fall semester of 2018. Participating in the course are master students of cognitive science and master students of information and communication technology engineering at Lund University, Sweden. The course is the second of two inter-connected and related courses commonly known as "the research and development courses" or "R&D-courses for short". The courses revolve around problems from the industry which the students are to ponder academically. The problems are to be interpreted, researched upon, and investigated for a solution based on cognitive science and human computer interaction design. The final design solutions are presented at a fair at the end of the course. This paper deals with one of these problems: that of how the growing number of smart services is convoluting their indexation and discoverability.

Growing Number of Applications Increasing Complexity

There is no denying that we have yet to reach the maximum amount of digital services that humanity will experience. Despite this, we can already see that the amount of services are starting to become overwhelming for the average user. Just for smart phones, more than 5000 applications are added to the applications download hubs Google Play™ Store and iOS® App Store® for Android- and IOS-based smart phones respectively (42Matters, 2019). While this huge amount of new and varyingly relevant services is in no doubt creating new opportunities for the user, they also cause trouble. Not only are consumers having trouble maneuvering the services they already possess and use, they are also having troubles discovering new services. This observation has led to discussions concerning how we index and manage our services in order to not drown in an ocean of services. This paper presents an approach based on spatial cognitive abilities: using maps for indexation instead of conventional categories.

Maps as a Means of Organisation

This paper presents and investigates a new concept for indexing, navigating and managing digital services. The concept revolves around using maps instead of conventional common one-word, theme-based associational categories for organising services in our lives. While relevant for all types of services and originally inspired by the growing number of Internet of Things (IoT) devices, this paper uses smart phone applications as a medium for testing the approach.

Spatial Codes for Human Thinking

While this is an entirely new approach to cognitive indexing, we have succeeded in finding a theoretical framework somewhat related to our conceptual idea. It has long been an intriguing question of cognitive scientists how intelligent organisms find their way around. Research into rodent brains has added a central puzzle piece by revealing what has come to be known as place and grid cells. Place cells are neurons that fire at a specific geographical location (Humphreys et al., 1908). Grid cells are neurons that fire in a hexagonal pattern in every location giving the organism a sense of distance travelled (Hafting, Fyhn, Molden, Moser, & Moser, 2005). Research suggests that these findings from rodent brains may be transferable to the human hippocampus (Humphreys et al., 1908). A recent paper by (Bellmund, Gärdenfors, Moser, & Doeller, 2018) investigates the possibility of linking hippocampal-entorhinal processes linked to spatial navigation with (Gärdenfors, 2004) notion of cognitive spaces. The main idea of Bellmund et al. (2018) is that as the hippocampal-entorhinal cells are able to re-code themselves to represent different spatial spaces they may as well possess the ability to represent an unlimited array of conceptual spaces.

While the claims of Belmund et al. (2018) does not include use cases such as the one presented in this paper, they open up for new ways of contemplating human cognitive abilities. Their linking of conceptual organising and the hippocampal-enthorinal area offers justification to question the classical semantics-based organising of concepts. If there is indeed a link between how humans navigate spatial and conceptual spaces, research efforts should go into developing methods of conceptual organising taking advantage of this new knowledge and existing knowledge suddenly relevant for the field. This paper exemplifies this new area of research by testing a new conceptual organisation of digital mobile applications based on locations.

2 Method

This section presents a theoretical overview of the main activities our team has performed as means of moving from the initial problem space of growing application complexity to the notion of maps as a tool for organising applications. The activities are primarily of co-creational character: workshops and prototyping.

Workshops with Brainstorming

In the early stages of problem solving, it may be challenging to know which approach to solve the problem is most suitable. The project and its problem space may not be well defined or clearly understood. To efficiently develop a better and deeper understanding of the problem, to form ideas, and to formulate solutions, a brainstorming session may be a proper approach. In a brainstorming session, a group of people meet to bring out new perspectives and ideas to gain further understanding of the problem. By taking notes of all ideas given by the group of people, while keeping the brainstorming environment free from criticism and judgmental thoughts, more ideas and solutions may be generated.

Prototyping

Prototyping is roughly creating an early approximation of the finished product. Prototyping can be used to share a vision for the product or to share an idea for a certain design, functionality, or user-case. Developers can use the feedback from prototype tests to evaluate the usability and functionality of a product. Most often, the early prototypes are to be shared with the design team or main stakeholders. Late prototypes with extended functionality are very similar to the released product. The prototype is tested and evaluated to an extent until the prototype is acceptable and the requirements for usability is achieved.

A prototype can be either low- or high-fidelity (lo-fi and hi-fi, respectively). In the former case, the prototype is made using a simple to alter material such as drawings on paper. Low-fidelity prototypes are often used to create easy-to-evaluate visualisations and may result in new alternative design solutions. Making changes in a low-fidelity prototype is less costly and more efficient than in later stages, when the prototype moves from low-fidelity to high-fidelity status. A high-fidelity prototype is often made in the same medium as the final product to allow for real users to test and evaluate it in a true-to-reality user setting.

3 Activities

This section presents the content of the different activities carried out during the research and development course: workshops, lo-fi prototype development and usage, data processing, and hi-fi prototyping development and usage.

Brainstorming in Workshops

Most of the work done during the brainstorming phase was done during workshops. The goal of these workshops was to agree upon the actual problem and to understand it, in order to decide upon how we could solve it. The brainstorming phase took long time and required much discussion and compromise for the session members in order for the project to progress.

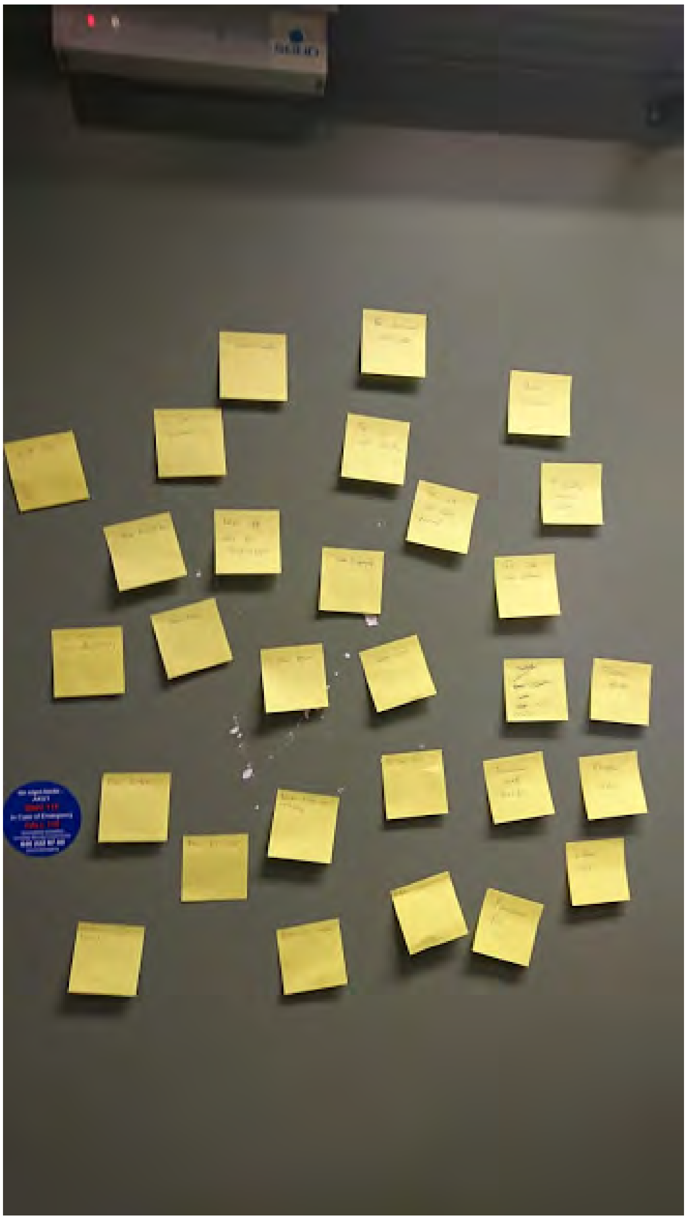


Figure 1: This picture shows the result of one of the workshops that was held, in which we brainstormed over different kinds of services and how they could potentially be categorised through spatial locations.

The format of the workshops varied much. Some of the workshops were simple discussions over a cup of coffee where we tried to conceptualise our view of the concept to an extent allowing us to discuss our suggestions to approach it. Other workshops were more formal involving classic brainstorming tools such as post-it notes, white boards, pens, and paper. During a certain session, we filled out post-it notes with different kinds of services and tried to make sense of categorising services using locations rather than conventional categories, see figure 1.

Lo-Fi Prototype Development and Usage

To get insights into the applicability of our initial thought processes, we developed a paper-based lo-fi prototype. The lo-fi prototype tool we developed comprised of three parts used in two stages. The first part was a set of 20 post-it notes with pre-written names of services. The services chosen were existing smart phone applications. The second part was a set of blank

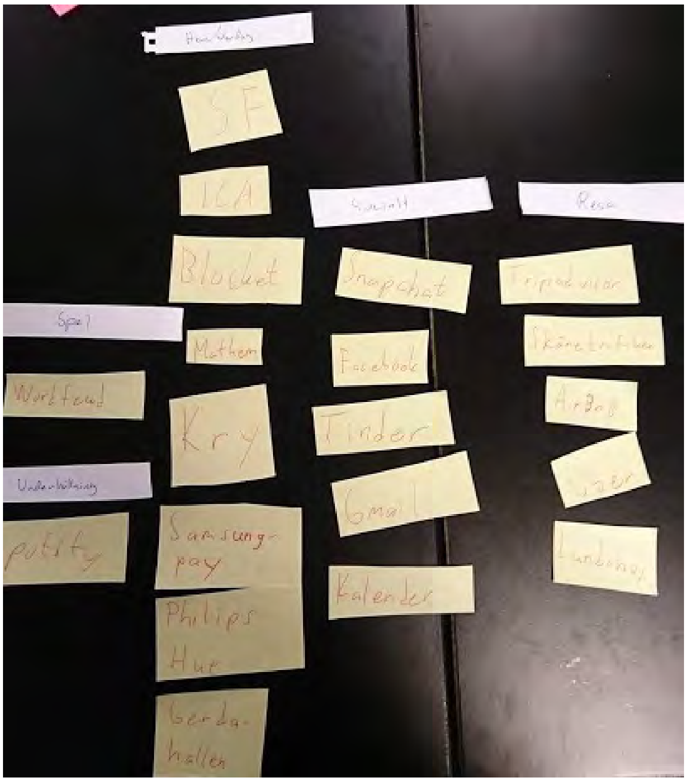


Figure 2: This picture shows the services placed by a user under the users own written categories.

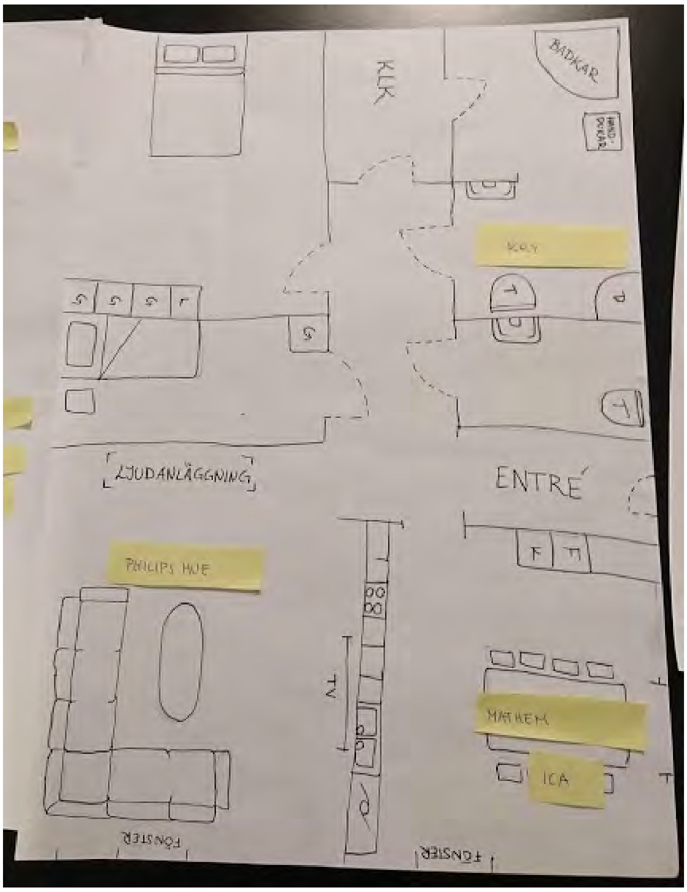


Figure 3: This picture shows the services placed by a lo-fi test subject on the map representing the home vicinity.

notes on which the user was asked to write category names suitable for the services which had been presented to them. The last part was a set of maps that were hand-drawn by us. The different maps pictured different sized areas. These areas were: the Öresund region, Skåne, Lund, a construction plan of a regular apartment, and lastly a picture of a human body.

In the first stage, the user was asked to categorise the services and write down category names for their organisation. We told the user not to feel limited to assumed expectations when choosing category names or how many categories they wanted. We also told them that they were allowed to create as many categories as they felt necessary, and even allowed them to create a category such as "other" where they could place services they failed to fit into any other category. The results were then documented for later data processing. Figure 2 shows one of these user-produced application organisations.

In the second stage, the test subject was asked to categorise the services using the maps provided. The instructions were as follows: "Try to associate each service with a location and try to find that location on one of these maps. Place the service in that location on the map and try to motivate why you chose to place it there". The results of this part was also documented for later data processing. Figure 3 shows a part of one of these map-based application organisations.

Processing of Lo-Fi Prototype Usage Data

The lo-fi tool was tested on fellow students of cognitive science and of information and communication technology engineering. In total, eight students took part in the test. The collected data for each test subject includes a category and a map location for each service. The time spent was not registered. The test subjects were asked to "think aloud" and their thoughts noted along the duration of the test. Their categorisation and spatial placement of the services was documented through pictures, see e.g. figure 2 and figure 3, and put into an excel sheet.

Two main factors were interesting when examining the lo-fi prototype test data. The first one was, for each service, how much did the category association differ between users, or more precisely, under how many different conventional categories did the users place the service? The second one was, for the most popular category for each service, how many users chose this particular category? These two questions were combined into a single index score: $I = (1 + \frac{C}{T}) * (1 + \frac{1}{N})$, where I is the index score, C the number of categories for each service, T, the number of users participating in the lo-fi testing and N the number of users choosing the most popular category. In cases where the most used conventional category was "other", the second most frequent category was used for the calculations instead.

The possible maximum index scores are equal for both maps and categories, but as the number of categories is potentially as high as the number of services presented (20), while the subjects where limited to choosing between only four maps, it is expected that the index scores are generally higher for the maps, as the minimum value varies slightly (1.13 and 1.4 respectively). However, the index score formula is designed to best counteract this fact. The results of the index scores were later taken into consideration when creating the hi-fi prototype tool.

Hi-Fi Prototype Development and Usage

Based on the results from the lo-fi prototype tests, we developed the hi-fi prototype test tool to allow us to make precise

Figure 4: This figure displays the test interface screen of the hi-fi prototype tool where a test subject is to choose under which category to place the service "Snapchat".

Figure 5: This figure displays the test interface screen of the hi-fi prototype tool where a test subject is to choose at which map to place the service "Uber".

Figure 6: This figure displays the personal results and overall statistics screen from the hi-fi prototpye tool allowing the test subjects and researchers to get easily accesible insights into the test results throughout the data collection.

measures of test subjects' abilities to localise digital mobile applications within the two paradigms of organising through conventional categorisation and organising through spatial locations. The hi-fi prototype was implemented as a simple structure web application exploiting the fast scalability and modularity of web technologies such as HTML, CSS, javascript, and PHP. The test tool presents three main screens to the test subject: 1) the introduction screen, 2) the test interface screen, and 3) the individual results and statistics screen. Each of these screens will be described in details below.

Screen 1: introduction screen

When opening the test tool, the test subject is presented with a simple screen informing him or her about the aim of the test as well as instructions of how to complete it. A big button allows the test subject to commence the test. Initially, the test was designed so that the test subject could choose between two versions of the test via two blue start buttons: the full test with 20 services and a shortened version with 5 random services. Later, during data collection, we decided to limit the test subjects' options to ensure a higher completion rate with a slightly higher number of tests per subject. Thus, the remaining option was a version of the test with seven randomly chosen applications for each organisation paradigm.

Screen 2: test interface screen

When initiating the test, the test subject is presented with a screen that randomly loads the name, the logo, and a short description of one of the 20 services part of the test. On the right, the test subject is asked to chose which of the categories or spatial locations the subject thinks that the service belongs to. See illustration 4 and 5 for examples of the categories and maps respectively.

When clicking on an answer option, a colour change to either green or red indicates whether the answer is right or wrong respectively. "Right" and "wrong" refers to whether the option is the option chosen by the most test subjects prior to the current test subject. The tool continuously evolves and corrects its database so that the "right" answer is always the option having been chosen most frequently as the first answer for that service. Initially, the right options is based on data from the lo-fi prototype.

If the test subject chooses a wrong answer, the options turns red and the subject is to continue choosing until he or she reaches a point in his or her logic, where he or she agrees with the majority of previous test subjects. When the right answer is chosen, the screen automatically loads a new service until all services (or the max number in the shorter versions) are through. All test subjects go through both organisational paradigms (maps or categories), but is randomly assigned to start with the one or the other.

The six categories available are "Handla" (shopping), "Hem/Vardag" (home/everyday), "Transport/Resor" (transportation/travel), "Socialt" (social), "Träning & Hälsa" (sports and health), and "Underhållning" (entertainment). The six spatial locations available are that of the personal space, the home, the neighborhood, the town, the region, and the world.

Screen 3: individual results and statistics screen

When completing the test, the test subject is presented with a screen describing the subject's personal results as well as overall statistics allowing the test subject to investigate its own performance with that of the general public. The statistics include data on how much time users test subjects spend on finding the "right" answer in each organisational paradigm, how many attempts test subjects sue to find the "right" answer, and how many percent choose the "right" answer as their first option. See illustration 6. We will elaborate on these statistics later in the paper.

Application / Service	Right choice in first attempt		Attempts before finding right		Time spent to find right	
	Categories	Maps	Categories	Maps	Categories	Maps
Actic	100%	64%	1.0	1.7	4.9	15.2
AirBnB	100%	82%	1.0	1.5	9.2	6.2
Blocket	88%	40%	1.1	2.0	8.9	18.6
Facebook	90%	42%	1.1	1.9	5.6	12.6
Filmstaden	67%	71%	1.7	1.4	10.1	10.0
Gmail	69%	27%	1.4	3.1	12.8	16.7
ICA	93%	40%	1.1	2.8	5.2	19.6
Kalendar	64%	37%	1.7	3.2	15.4	36.0
KRY	85%	77%	1.2	1.4	10.1	14.7
Lundahoj	79%	92%	1.4	1.1	7.2	8.3
MatHem	75%	36%	1.3	2.4	7.9	26.9
Philips Hue	90%	80%	1.1	1.5	7.0	8.1
Samsung Pay	86%	33#	1.1	3.5	7.9	20.0
Skånetrafiken	100%	72%	1.0	1.5	4.3	8.1
Snapchat	86%	58%	1.1	2.5	11.8	12.9
Spotify	88%	43%	1.3	2.8	7.6	22.0
Tinder	100%	36%	1.0	2.9	6.7	27.1
TripAdvisor	100%	100%	1.0	1.0	5.5	11.6
Wordfeud	69%	15%	1.5	5.1	12.0	17.7
Uber	100%	31%	1.0	1.9	6.2	11.8
Average	87%	54%	1.2	2.3	8.0	16.0

Table 1: Results from the hi-fi prototype tool tests

4 Results

This section presents the results from the lo-fi and hi-fi prototype tests with subjects presented in the sections above.

Lo-Fi Test Results

Table 2 presents the results from the data processeing for the results form the lo-fi prototype test: services ranked by index scores for each organisational paradigm. The service at the top received the highest score, meaning this service had the most favourable combination of few different indexations and a high number of users choosing the most popular indexation.

Categories		Maps	
Facebook	2.33	Lundahoj	4.00
Snapchat	2.33	Gerdahallen	4.00
Lundahoj	2.17	Regiontrafiken	2.81
Regiontrafiken	2.17	Mathem	2.81
Tinder	2.17	Philips Hue	2.81
Wordfeud	2.03	Kry	2.63
AirBnB	2.00	Spotify	2.44
Tripadvisor	2.00	ICA	2.44
Über	2.00	AirBnB	2.33
Gmail	1.88	SF	2.33
Kalender	1.72	Facebook	2.25
Blocket	1.65	Wordfeud	2.25
Spotify	1.65	Über	2.25
ICA	1.60	Snapchat	2.17
Kry	1.50	Gmail	2.17
Mathem	1.50	Samsung Pay	2.17
Gerdahallen	1.46	Tripadvisor	2.03
SF	1.46	Tinder	2.00
Philips Hue	1.35	Kalender	2.00
Samsung Pay	1.29	Blocket	1.83
Average	1.81	Average	2.49

Table 2: Results from the lo-fi prototype tool tests

Hi-Fi Test Results

Table 1 shows the clean results from the hi-fi tool test. On average, the test subjects chose the right category 87% of the time when using conventional categories and 54% of the time when using maps. On average, the test subjects used 1.2 attempts to find the right category when using conventional categories and 2.3 attempts when using maps. On average, the test subjects spent 8 seconds finding the right category when using conventional categories and 16 seconds when using maps. Times for the individual applications/services can be found in table 1.

5 Discussion

In this section, we will dissect and discuss our process, activities, results, and the concept in general.

The Initial Problem

In the workshop section of this article, it is mentioned that much of the time was spent work shopping, brainstorming, and developing the problem and our concept. There were more than one reason for this. It was not very difficult to grasp the fact that today we have many digital services and that it is very likely that the number will only increase. It was far more troublesome to understand what specifically was the problem with conventional categorising and following this: what we wanted to achieve by redefining it?

One of the things that was brought up was how differently you can organise things when using conventional one word categories. Take *Spotify* playlists for example. It is possible to categorise all your songs in *Spotify* using the decades they were released, but you could just as well categorise them after genres, moods, the occasion at which you first heard the song, beats per minute, etc. Thus, one of the things we wanted to achieve was to find a new way of organising services that was more mutual for all users.

Another issue that was even more noticeable during the lo-fi testing, but also discussed during our workshops, was the issue of coming up with categories for niche services. Placing *Candy crush*, *Wordfeud*, and *Angry birds* in your *"games"* folder is rather straightforward, but where do you place your calendar service or your GPS service? What this often leads to is one big *"general"* folder where all of these "others" are placed. This was something that we wanted to solve as well.

Furthermore, discoverability was something that was discussed during the workshops. Having services placed in normal categories does not open up for many possibilities of discovering new services. Thus, another challenge that was presented for our new concept was exactly that: improving discoverability. Is there a way of categorising services that can allow users to discover new services in a smarter way?

Organisation Based on Location

During one of the very first workshops we had, a thesis was presented by one of the people who initially predicted our problem. The thesis was this: People have a more mutual understanding towards pairing their services with locations, than they do when pairing them with categories. This turned out to be one of the primary research topics of our study; trying to figure out if this was true or not.

During the workshops, we carried out exercises where we all tried to think of locations and which services that we relate to these locations. At the time we were not specifically looking at digital smart phone application, but just any service that is available to us in our surroundings. A frequent example in our discussions at this points was buying a ticket for a train on the train station. What we figured out from this was that yes, we do have similar understandings of the services available to us at certain locations as long as we are somewhat familiar with the location. This was, however, pretty far from understanding how this would work when it came down to organising digital services.

We agreed that we needed to conduct tests on other users in order to start investigating the thesis. During our workshops, we often worked with abstract services that was not necessarily applications or solutions that were accessible today. This was something we felt had to change if we wanted to test it on users. So in order to make it all more understandable, we started to look at services such as mobile applications that exist today and allows for a far better understanding among test subjects. With that in mind we began our work with the lo-fi tool.

The Universality of Spatial Cognition

When we were to develop a new way of conceptualising categorisation, we stumbled upon the problem that as categorisation is a rather common activity, its underlying behavioural patterns are so ingrained in our minds that it was a rather difficult task. However, the choice to go for a spatially based organisational principle, is based on the idea that as most humans are able to wander about and that we share this ability with our ancestors, spatial cognition seems to be a widely shared ability. If we are to find a paradigm for categorising items such as applications that may prove superior to that of conventional categorisation, it appears as a fruitful approach to begin in areas where general abilities are tremendously profound.

Developing the Lo-Fi Prototype Tool

The main purpose of the lo-fi prototype was to obtain insights in free service-category associations and service-map associations of users with little knowledge of the project. The main research question for the test was: *How strong are the mutual association patterns for services and maps, and services and conventional categories? Do they differ for different services?*. While the workshop sessions were shaped around organising general everyday tasks such as *doing laundry* or *renting a car*, the lo-fi tool included specific branded digital smart phone applications. Such specialisation decreases the level of abstractness for the test subjects while increasing the resemblance to a mapping system applied in real life where branded services for task assistance is the reality. The range of services was deliberately widely chosen in order to maximise the output of user associations during the tests. Allowing the test users to phrase the categories of their conventional categories themselves allowed for freer associations.

The main limitation of the lo-fi tool was the manual workload required for compiling the test results, which made the lo-fi tool suitable for only a limited number of test subjects. Another limitation of the data analysis from the lo-fi tool test results was the imprecise registration of service placement on the four maps. Rather than recording where on the map the service was placed, the analysis was formed based on on which map is was located. Such generalisations were made with the motivation that the precise placement might not affect the discoverabilty in a future system with well-designed affordance while non-continuous registration of map placements during the tests allows for a much more intuitive comparison between conventional categories and maps.

Results From the Lo-Fi Tool

An immediate and interesting finding from the lo-fi test is that the applications with high index scores differ depending on whether based on the organisation through maps or through conventional categories. A general rule of thumb is that the applications receiving a high index score in the spatial paradigm are applications with an overt "belongingness" to a physical place. These include (from the top of the list of table 2: *Lundahoj*, *Gerdahallen*, *Regiontrafiken*, *Mathem*, and *Philips Hue*. The applications receiving a high index score under the conventional categories are application that have limited physical presence and that represent more abstract phenomena. These applications include (again, from the top of the list of table 2 *Facebook*, *Snapchat*, *Lundahoj*, *Regiontrafiken*, and *Tinder*.

Discovering this was the primary driver for developing a new theory of the organisation and discoverability of services: it may not be that one organisational paradigm is suitable for all types of services. Different organisational principles may prove to be better at govern one type of services, while other principles may govern other types of services. Developing and investigating this thesis further is, unfortunately, outside the scope of this paper, bu we hope that the insights from this paper may guide such research in the future.

Developing the Hi-Fi Tool

The hi-fi tool was developed to obtain further user data for answering the research question. While the challenge was still to figure out whether one of the organisational paradigms had better applicability, this test would let the test users find the services rather than *place them*. The correct answer would always be the most popular answer among previous users of the tool, making the tool flexible. The tool was initially calibrated based on the lo-fi test results. The categories of the test was carefully crafted and selected based on the results from the lo-fi tests, to best fit the majority of our data sets. To minimise probabilistic advantage for maps or conventional categories, both of the test parts included six alternatives, thus "residential area" as well as "the world" was introduced as maps. As for the lo-fi, a discrete map paradigm was chosen in order to compare the conventional category and spatial location results. As a consequence, no data about *where* on the maps, but only data about which of the maps was collected.

The hi-fi tool's main advantage over the lo-fi tool would include automatic data registration and processing allowing for a much higher number of test subjects, and time registration. The simple GUI of the hi-fi prototype test tool allowed for few distractions and unintended impact on the choices of the test subjects.

Alternative Hi-Fi Design Options

Our implementation of the spatial organisation paradigm showcase six maps ranging from *the user* over *the home*, *the neighbourhood*, *the town*, and *the region*, to *the world*. While this map based implementation make easy-distinguishable spatial locations, what they truly represent is a continuous "distance from self". The map-based interface translates this continuous variable into discrete spatial locations easily comparable to the discrete conventional word-based categories that they will be compared to. Another way to implement the "distance from self", would be to make an input scale, where the one end represents *the self* and the other end represents *the world*. A scale-based implementation may suggest other types of and more correct statistical inferences, however, when needed for comparison with discrete conventional categories, we deemed that discrete partitions of the continuous scale would prove more fruitful. Tests investigating how a continuous understanding of the spatial organisation may affect the indexing of services is outside the scope of this paper, but a highly interesting future research agenda.

Results From the Hi-Fi Tool

The results presented in figure 1 above show that generally, users are both quicker, more accurate, and agree more when finding the common organisation of digital smart phone applications when using conventional categories than when using spatial map-based locations. Like the results from the lo-fi prototype tool tests, the services ranging high differs based on what organisational paradgim was used. However, five out of the nine services ranging above average on correctness on first attempt in the map paradigm is also present above the average threshold of the conventional category paradigm. It is also interesting to see that while correctness on first attempt in the conventional category paradigm varies just between 64% and 100%, for the map paradigm the interval is much bigger: 15% to 100%. This somehow indicates that taken generally, it appears as if the conventional category paradigm is superior to the map paradigm.

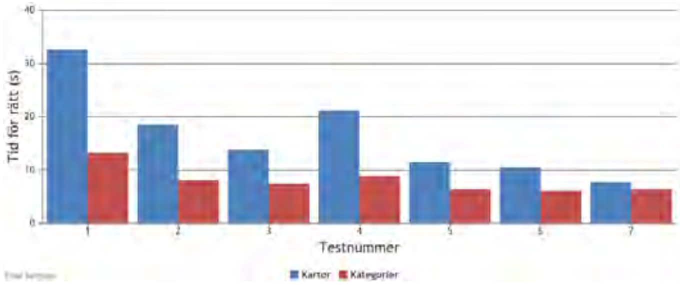


Figure 7: This figure shows the learning effect of multiple trials on maps and categories respectively.

The Apparent Superiority of Conventional Categorisation

From the data above, one could easily get the idea that categorisation based on maps is far inferior to classical conventional categorisation. However, this may be due to the fact that the users tested in the tests reported upon in this paper are more commonly used to categorising applications (and other things, such as digital files) via conventional one-word based associations. During our test of our hi-fi prototype, we recorded how much time each test subject spent before reaching the generally accepted category (whether map or category). An interesting finding is that as users neared the end of the test, their ability to deduce the most logical spatial location of an application equals that of their ability to deduce its correct category. Figure 7 shows the average learning effect of the tested subjects.

The Prototype Design

The lo-fi prototype and later also the hi-fi prototype were designed to answer the very specific question *How strong are the mutual association patterns for services and maps, and services and conventional categories? Do they differ for different services?* where the key word is *association* rather than *finding*. This question is very much relevant as the initial step of exploring spatial service association patterns is crucial to designing an interface for discoverability. However, it does not cover some practical aspects of such interfaces. A very distinct difference between the test tool and an interface for discovering services is that you have to actually find the service within the category in an interface, most often by skimming a list. This does not make a considerable difference for twenty services but it most surely does for two-hundred services. The tests presented in this paper would therefore favourably be complemented with tests in which the users would have to click the actual service in an interface where the amount of services is much larger than the twenty presented here. The major focus of such test would be discoverability measured as time taken for discovery. Such test results would, however, be very sensitive for the specific design solution introducing a major set of parameters making the result very troublesome to assess without introducing several tests in which the effect of certain design elements could be determined.

6 Conclusion

Both the lo-fi and the hi-fi tests revealed that different organisational paradigms was found to be favourable for different services. The test results did however prove for higher average accuracy using conventional categories. How heavy the influence of learned categorising patterns biased the score in conventional categories favour would require further investigation. Further testing would also be required to obtain information of how a continuous spatial variable would effect the test results and how such tests or products would be designed and constructed. Further tests would also require a much higher set of services in order to assess how such adjustment would affect the discoverability test results in tests which do not focus on *association* but *finding*.

References

42Matters (Ed.). (2019). *Store stats.* Retrieved 2019-01-20, from `https://42matters.com/stats`

Bellmund, J. L. S., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*, 801–806.

Humphreys, G. W., Duncan, J., Treisman, A., O'Keefe, J., Burgess, N., Donnett, J. G., … Maguire, E. A. (1908). Place cells, navigational accuracy, and the human hippocampus. *Transactions of the Royal Society of London. Series B: Biological Sciences*.

The Augmented Flight Training

Daniel Bogdan Andersson, Elin Hanstorp, Karl Johan Klang,
Gustav Lilja & Amar Vrbanjac

2019-01-20

This report aims to describe the process, within the both project courses MAMN010 and MAMN015 2018, in which Augmented Reality tech could come of use during pilot training and education. Previous studies, starting with the "Super Cockpit" in 1986, have investigated the use of Augmented Reality glasses for commercial and pri-vate pilots, but using AR for cockpit orientation and pro-cedure training is open to further investigation. Usability restrictions such as shakiness, selection of objects, camera position and the field of view are typical areas of improvement in AR. Checklists are powerful tools, and are shaped using logical steps based on experience, but working with so called flows form the base for a more spatially intu-itive and human centered approach. Pilot training em-phasize the demand for a more extensive cockpit procedure training to meet the demands of modern airline industry. Technical and timely limitations which constrain the extensiveness of these project courses are lifted. Method: The design process of the project follows six steps where the problem is defined, information collected, solutions developed, feedback and improvement applied and back to a new definition of the problem. As a result a solution of a basic digital "paper tiger" has been implemented in Unity for Microsoft HoloLens, where a cockpit procedure, the general scan flow, can be performed at a flexible phys-ical location. The solution has been tested on 5 pilot stu-dents and one instructor at Lund University School of Aviation. The test was performed using qualitative inter-views as well as System Usability Scale and NASA-Task Load Index surveys. The results show a system usability of above average but not very good. Qualitative feedback is clear in the appreciation and opportunity of future pos-sibilities based on Augmented Reality such as flow guid-ance as well as detailed practice of procedures, but more interactive features are needed, such as illumination of panels. The spatial pattern of the demonstrated flow was efficiently remembered by the participants after the test. Finally strengths, weaknesses with AR and future possible developments of the result are discussed.

1 Introduction

Today in 2019, augmented reality (AR) is a huge buzzword and compared to virtual reality (VR), AR tech can already be found in most smartphones. Apart from being available in most people's pockets, AR headsets much like VR headsets are starting to emerge from several big tech companies. In this project the AR headset Microsoft Hololens have been used to try and help the pilot students at the Lund University School of Aviation TFHS (Trafikflyghögskolan) in Ljungbyhed, Sweden.

In the early stages of the pilot student's training at the academy, they must learn how to operate an aircraft from the cockpit using nothing but a simulator made out of printed paper panels and the operations manual. This is meant to be a period for the students when they learn about procedures done in the cockpit before, during and after a flight, the teamwork between pilots and also learn different flow patterns for each phase of the flight. By using these flow patterns the pilot students can then use a checklist to cross check that nothing has been missed. This is a more reliable way than using a checklist throughout which is both slower and offers a bigger risk the pilot might miss something.

According to the school there is a big learning gap between pilot students using the paper simulator and using a full flight simulator. This project aims to help pilot students learn particular flows in a digitally enhanced paper simulator, which is called a "paper tiger" at TFHS, using AR through the Microsoft HoloLens. The goal of the project is to ultimately reduce the gap between the paper tiger and the full flight simulator. Making the learning experience more intuitively, more fun and finally to reduce the abstraction level for students just starting out. Instead of sitting by a paper tiger, AR can offer the student real-time feedback, and according to f.e. (Black & Wiliam, 1998) feedback has been deemed to be very effective when it comes to students learning.

2 Related Work

Previous studies in AR environment

In 1986 Tom Furness described a virtual crew station concept called the "Super Cockpit" (Furness, 1986). It discussed how flight information could be portrayed virtually and interactively. This resulted in new areas for research in engineering and the discipline called Human Factors.

Previous research in AR tools for cockpits have often focused on guidance of the actual flight phase by providing information from the Primary Flight Display (PFD) regarding speed, altitude, attitude and heading (Goteman, 2006, Haiduk, 2017). Heads-Up-Displays (HUD) are already used in many fighter planes such as the JAS39 Gripen, and in modern commercial aircrafts such as the Boeing 737 Next Generation (Goteman, 2006). "Smart glasses" have also been developed for private pilots, for example Recon Jet, Vuzix M3000, Google Glasses and Movierio BT-200 to help guiding them in a visual traffic pattern. Microsoft HoloLens is not the most portable of the mentioned and it can be heavy to wear. Still it is one with the widest field of view so far with 30 degrees, even if that is also a limited field of view (Haiduk, 2017). There is one experiment where the landing procedure was performed by 28 pilots in Virtual Reality (VR) and in a research flight simulator at Graz airport. The conclusion was that even though it worked out well using the VR simulator, pilots were frustrated and the system did not react as quickly as the regular simulator. It was therefore preferable for the tested pilots to still use the conventional simulator for actual flight training (Oberhauser, 2018). In august 2018, within the course MAMN40 at Lund University, a detailed 3D model of the Full Flight Cockpit Simulator was scanned in, and the group successfully implemented the Electrical Power up Flow with the relevant buttons to interact with. The implementation was guided by a text letting the user know the next button or item to adjust or verify, for example "Next, Gear Down" (VaIR, 2016). The use of AR to help orientate within a cockpit and in a learning phase seemed to be an area with less research and implementation conducted.

In an article from 2009 (Iordache, 2009) a project called ARiSE (Augmented Reality for School Environment) is described to be an attempt of evaluation the pedagogical value of AR in a school environment. This is related to this project since it also aims to evaluate if the use of a HoloLens could possibly help the flight students to understand flows better. The article mentions a method which is called formative usability evaluation which is based on both a heuristic evaluation as well as user testings. The goal is to compare and integrate both quantitative and qualitative measurements in order to increase the results credibility as well as enhance the descriptive parts of the evaluation in order to make it easier for designers to use it to improve the product. The conclusions from this article was that many of the problems which existed with AR in 2009 have also been experienced in this project, even though different systems have been tested, f.e. shakiness, the users start position, blurriness and the lack of intuitive interactive features.

3 Theory and Background

Semi-structured interview

Instead of an interview consisting of a pre-determined set of questions, the test leader conducts the interview on a basis certain focus points. This is perfect for tests where the test scenario is combined with an observation because the scene is ever evolving. The test leader is asking questions as certain events happen instead of following a strict question protocol. Using semi-structured interview allows the user to tell the test leader how they are interacting with a system when they are doing it instead of trying to remembering how they did it (Arvola, 2014).

Think-aloud

As a test subject is given a set of tasks to do in order to test the system, a common practice is to allow the test subjects to tell a silent test leader everything. The test subject can f.e. tell the test leader what they are feeling, what they are doing and why they are doing a certain thing while performing the given tasks. This can be recorded and later transcribed or just written down directly. Based on the recorded protocol one can learn how the test subject reason when they are performing a specific task. However this can sometimes be hard to do since some tasks require the test person to focus on that thing, f.e. a phone call. A workaround is to allow more than one test subject to perform a task, then they can reason with each other and explain to each other what they are thinking (Arvola, 2014).

The Design of a Checklist

Checklists are powerful tools. They can increase precision of a specified behaviour and reduce errors. An example of a checklist used in the cockpit can be seen in figure 1. Still, some things on the checklist are dependent on systems and people outside the cockpit. For example cabin crew, refueling- and maintenance personnel, gate and check-in personnel. Certain items therefore cannot be performed until a specific phase of the flight, for example first when the cargo- or cabin door has been closed at the gate. This *operational logic* is good to be mindful about when designing a checklist and when its chronological and logical parts are to be organized. The article '*What's in a checklist. Human Factors of Aviation* (Degani, 1991), lists important advice for future work within Human Factors, design and learning in a cockpit environment. One important advice is to '*Sequence the most important first*'.

Figure 1: Checklist on yoke

While pilots work with a checklist the risk for distraction increases the longer the list is. They may loose concentration and direct their focus towards something else which needs their attention. One effect of this is to keep the actual checklists as short as possible. Instead aircraft manufacturers such as Boeing have set up a working method based on so called flows (Jal, 2013). During a flow you prepare the cockpit for different configurations such as start up, takeoff, descend and landing.

Previous research highlight the development and a need for a more human-centred interaction design in pilot training (Degani, 1991, Norman, 2013, Bitenieks, 2018). One example is during a type rating course, a course in the end of the pilots training when he or she learns how to fly a large complex commercial aircraft. Before the course starts it is good to have as much experience as possible. Due to substantial economical costs for training, approximately 16000 - 34 000 EUR for a type rating (Mjåtveit, 2018) and therefore high time pressure in the learning phase, the fresh student can be overloaded with information during a type rating course. An interactive learning environment may serve as a threshold and a more intuitive approach and frame to the content, than starting with reading pure text and exhaustive lists in the initial learning phase of a type rating. The discussion about cockpit work based on flows will be taken up again later in the report.

Cockpit Resource Management

Most commercial passenger aircraft are flown with two pilots. The pilots share the workload in the cockpit to cope with all the tasks, information, communication and to provide redundancy and minimize the risk of errors. This is done by defining areas of responsibilities for the Pilot Flying (PF) respectively Pilot Not Flying (PNF).

Cockpit/Crew Resource Management (CRM) is a multi crew working method where social processes, coordination and combined performance shapes the basic principle to manage an aircraft in a "multi-crew- system". This method was given increased support when it became clear that "Human Error" was a big reason for many accidents, as aircraft technology and systems became better and more reliable over time and no longer the responsible cause in many accidents (Degani, 1991). One example, that can be seen in figure 2, is the "point and look" -method.

A cockpit is full of information and it is of great importance that it is well designed to be able to orientate and acquire relevant and necessary information as a base for good decision making for the pilots. One way to enhance the cockpit-work and to stay organized is the cooperation between the two pilots in the cockpit. The basic process could begin by the Pilot Flying (PF) requesting the Pilot Not Flying (PNF) to execute a specific task. For example: Air Traffic Control (ATC) gives an instruction to the pilots of the Scandinavian flight 1501 to climb to 4000 feet over the radio by the following phraseology

ATC: "Scandinavian 1501, set altitude 4000 feet". Thereby the PNF who is responsible for the radio replies

PNF: "Altitude 4000 feet, Scandinavian 1501". PF who also hears the radio communication then requests the PNF to put 4000 feet in the speed window of the Mode Control Panel by saying

PF: " Set altitude 4000 feet".

PNF now sets 4000 in the speed window and notifies the PF when this is done by saying

PNF: " Altitude 4000 feet set". At the same time PNF holds the finger under the "speed window" and does not remove it until PF has visually cross checked and looks at the speed window and the PNF finger and verifies with the word.

PF: "Check". By verbal as well as spatial verification and confirmation the cognition is constantly lifted and shared in the cockpit system. Thereby the risk of human error is minimized.

To present the "cognitive scientific" aspect the article "How a Cockpit Remembers its Speeds" by Hutchins

(1995) introduces the term "thinking system" where the interaction between human-machine and verification human-human can be perceived as one thinking system.

Figure 2: Point and look

Designing for Error

One common way to design for errors is by adding constraints, signifiers and affordances and the clever use of shape and size to block errors. For example: when refilling different fluids in a car, gasoline, oil, cooling fluid, separation is made through having different shape, size and location for the different fluids. Hence the risk for selecting the wrong input is minimized (Norman, 2013).

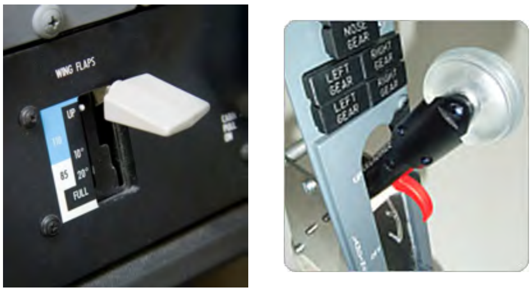


Figure 3: Left: Flap lever, Right: Gear lever

In the Aircraft: An early example of HCD and intuitive understanding in the cockpit was the design of the gear lever as a wheel and the flap as a wing. It was implemented after many accidents had occurred when pilots accidentally had retracted the landing gear instead of the flaps after landing. These different levers can be seen in figure 3. It can sound as a ridiculous mistake, but when sitting in the cockpit of a real aircraft in a moving and dynamic outside environment there is a lot of information where you need to be proactive in every step. This means that your mind is already quite occupied by substantial cognitive workload. Especially in the phases of takeoff and landing. Seen from this context almost every mistake can be made. Sometimes an extending safety procedure is applied by verbally confirming "flaps up" and holding the hand on the flaps, whereby the instructor or copilot verifies by looking and confirming.

Motor Flows, Look, Touch

Intuitive logic can be seen in the way how the instruments and displays in a cockpit are organized in panels and placed in the cockpit according a specific "geographical" arrangement. This arrangement is designed with regard to how frequently the instruments and panels are being scanned and monitored and in proportion to their critical value for the flight. To enhance a logical flow during the configuration of the aircraft it is necessary to perform the actions in a specific sequence of motor-eye coordination. These co-ordinations are called *flows*.

Further enhancement of procedures are obtained by the patterns shaped by the sequences of the flows. Flows follow a logical order and the items which are to be set up are organized in *chunks* within every panel. For example the flight control panel. The part-panels on the overhead section build up the "overhead panel" and are scanned in "rows" from top to bottom. Using a top-bottom order for checking the panels and items provides a structure and thereby reduces the work to learn the items, at the same time as it provides bio-mechanical enhancements as they are less exhaustive to move your arm from top to bottom than in the other direction. The same spatial technique can be used when verifying items of a checklist. Sensory organs and movements have the potential to produce an optimal experience, and confronting tasks which one has a chance to complete are key factors for obtaining enjoyment and satisfaction (Hoang, 2016).

Flows are challenging and a bit mechanic to learn in the beginning. However they can be quite rewarding once learned. A comparison can be made to learning a dance. First you struggle with getting all the steps right, and you easily forget what comes next. But after a while you get into a flow and one move seem to lead into another. The same is applied for a flow. When the pattern is recognized one intuitively connects one item to the next. At the same time you start experiencing a nice and positive feeling of flow.

Testing Usability

There are several ways of testing the usability of a product of system. Two surveys that can be used to test the usability in a quick and easy way is the System Usability Scale (SUS) and the NASA Task Load Index (NASA-TLX). The SUS-survey is based on ten specific questions where you choose on five options on a scale between strongly agree to strongly disagree. All the responses from the survey is converted to numbers and then a total number from all responses is generated. If the total score is above 68 the system is above average and the higher the better. Above 80.3 is very good. Highest possible is 100. (Bangor, 2009) The NASA-TLX-survey consist of six questions with the subjects Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration where you answer on almost all from very low to very high. Just like the SUS-survey the responses are converted to number but here a total score for each question area is generated. (NASA, 1986)

4 Pilot Training

Conventional Pilot Training

Today, just as in the early days of flying, much emphasis is put on the pilot students practical flight training and navigation. To obtain a commercial pilot's licence (CPL), which is still by far the most common training given at flight schools, pilot students need a minimum of 150-200 flying hours in total (Easa, 2016). However, practice in the simulator of a multi-crew aircraft is significantly less. It consists of approximately 20 hours during the Multi Crew Cooperation course (MCC) during the final stage of the training. This is aided by approximately 30-40 hours procedure preparation in a paper tiger. A paper tiger is a model made of pictures representing a cockpit and all its buttons. The name "paper tiger" derives from the cold war and was used to describe the power and threats of something that appears to be strong and fierce, but in the reality turns out to be all show and nothing to be feared (Lary, 2015). A picture from the paper tiger at TFHS can be seen in figure 4.

Figure 4: A Paper Tiger at TFHS

Multi-Pilot Licence Concept

To become competitive and attractive candidates on an international pilot market, CPL students often need extended training, for example by a costly additional Jet Orientation Course (JOC). This extended training emphasizes the need for more cockpit procedure training in conventional pilot training.

At TFHS the students learn based on an education plan called Multi Pilot Licence (MPL) instead of the conventional CPL education plan. During MPL training the effort is put into the multi pilot training from a very early phase (approximately 150 hours of simulator practice) and less practice is done in a single pilot aircraft (approximately 70 hours). TFHS uses the airplane Cirrus SR20 for the single pilot training. A pilot with a MPL is not allowed to fly commercially in a single pilot aircraft. The main difference from the conventional training (CPL) is a more market targeted approach within MPL. The aim is to prepare the students in an early stage of the learning phase towards working in a multi-pilot crew in the airline industry, and receive practice of this in a similar context. Still there is a big step in the learning process from practice in a paper tiger to flying a Boeing 737 full flight simulator.

5 Purpose

The purpose of this project was to investigate the possibility to improve the learning process for pilot students at TFHS with AR.

6 Limitations

In this section the limitations which were set for this project given the framework of the course, the technical limitations of the HoloLens, the minimal experience of Unity in the group and the needs from TFHS will be described.

The project has several limitations due to the framework of this course and due to the wishes and suggestions from the people at TFHS. There is a specific part of the education which the pilot students receive, where they are supposed to learn different flows through the cockpit. This project focuses on that specific part of the education in order to bridge the gap between training in a paper tiger and a full flight simulator. The HoloLens offers several technical limitations on it's own. Because of the narrow field of view the user can't use its peripheral vision, which in turn must be considered when developing an educational model with it, other guidance's must be added to help the user. Another technical limitation of the HoloLens is the few number of had gestures available to the user. The pinch movement which is used for clicking might not come natural to a pilot student when clicking on buttons in a cockpit.

Also we have been limited in such a way that no one in the project group had previously worked in Unity. Unity is a cross-platform game engine used for developing applications to HoloLens. Therefore time had to be allocated to actually learn the program instead of using that time to work on the actual project.

7 Method

In this section both the general process of the project and the design process are described.

Design process

In this projects design process the following steps have been followed following the iterative design process shown in figure 5:

- Define the problem
- Collect Information
- Brainstorm & Analyze
- Develop Solutions / Build a model
- Feedback & Testing

Figure 5: Design Process (Chicago Architecture C, 2018)

Define the problem

To define the problem, discussions were held both before and after the visit at TFHS. Before the visit the discussion was based mainly on the project description, *The Augmented Checklist* and what the actual problem was. The project is run in cooperation with TFHS who is also a potential client of the developed product. In order to more accurately define the problem and understand their interest in the project and how AR could be used at their school, the group visited TFHS. For example: One thing which were discussed on the first meeting at TFHS was the logic behind a checklist and why it is there.

After the visit more discussions were held in order to define the direction and goal of the project. The conclusion was that the goal should be:

- Improve the learning process for pilot students
- Focusing on flows rather than checklists and bridge the gap between the paper tiger and the full flight simulator

Collect Information

To aid our theoretical knowledge more than browsing through earlier studies and projects within this field of study, the main focus of the visit at TFHS was to collect information. One member in the group also had previous flying experience (CPL), and simulator experience from the Boeing 737, relevant for theory and primary testings.

Some practical information gathering was done by following a tutorial on how to implement an AR-program in Unity. The program implemented was called Origami and found through Microsoft's tutorial site for Hololens development (Microsoft, 2018). This contributed with information on how to develop in an AR environment for the HoloLens and gave the group an idea of how difficult different parts of the development could be. From the information given at TFHS it was clear that the paper tiger could be the main focus in the project. Because that was the part of the learning process with the highest level of abstraction, since the students did not have access to a real simulator at this stage in their education.

Brainstorm & Analyze

When all group members felt fully informed and synchronized the brainstorming started. Up until this point discussions regarding a solution and ideas to solve the problem were held to a very low quantity. This was done to prevent the group members from affecting each other.

The type of solution was discussed through a brainstorming session where a mind map was created. What part of the learning process could be improved and in what way? First everyone got to think quietly individually and then all possible solutions that came to mind were written down on a second mind map. Strengths and weaknesses were discussed and each idea was analyzed before anyone voiced their opinion about the best solution. Almost all ideas were sorted to either a solution mapped towards the physical paper tiger, or a solution separated from the physical paper tiger. One discussed idea related to the checklist which could be implemented with AR technology was to put out things in the cockpit for example an empty coffee cup on the throttles, to act as a reminder that there still was a point at the checklist which the pilots need to come back to. Another topic discussed was to help pilots get back into a flow by guidance of the HoloLens.

To select solution, everyone first selected the solution they personally believed in most. This resulted in five different solution options. These were then further discussed. Since the technical competence of Unity was limited in the group a discussion regarding the possible technical depth for the project was held, which gave very valuable input to the different ideas. Finally a conclusion was made. It was decided to start developing a *solution separated from the real paper tiger demonstrating flows in the cockpit digitally.* It was a good start and was also the underlying foundation to some of the other ideas.

This idea was decided because testing and visualization of the other ideas could be based on the digital paper tiger, which could be further modified to the other ideas as well.

To visualize the flow in the cockpit another brainstorming session was held. Different design ideas were suggested with Norman's seven design principles in mind: *Visibility, Feedback, Constraints, Mapping, Consistency & Affordance* (Norman, 2013). Apart from the seven design principles, the risk of disturbing the pilot rather than helping was considered. Since a pilot needs to put a lot of focus into the practice and what happens in the surrounding environment, there is a big risk of putting too much information or events being triggered in the AR environment. This might disturb the practiced way of working in the cockpit. The different designs were sketched on a paper with gray and colored pencils. The final result was a design were a specific area's outlines was marked with a green line and the start of the area was marked with a simple arrow.

Design iterations

Our goal is to bridge the gap between the paper tiger and an actual realistic cockpit. The gap consists mainly regarding the interface on which the students practice. During our first study visit, the instructors explained that when a student is done with his or hers 40 hours in the paper tiger. The goal of the paper tiger is to familiarize the student with the cockpit instruments, to create a smoother learning curve for the transition into the real cockpit simulator. While it achieves that goal of familiarization, the students still experience a period where they have to adjust to the new cockpit. We attempt to reduce that time by creating a visual representation of the flow, to enhance the mapping from the paper tiger to the simulator cockpit. This way, when the student sits down in the new cockpit, he or she will be able to visualize the flow procedure more easily.

For the second part of the course one of our main focus areas was to improve functionality and usability of the virtual cockpit.

One major iteration included all the remaining borders for the panels that are part of the flow, which consists of 15 steps. This also includes the corresponding arrows that point to the starting point of a flow panel and the button for each step. After some internal testing we decided to include an extra arrow that pointed to a specific step of the whole flow procedure. The main reason for that addition was due to the position of the panel in relation to the panel in the previous step. The sequence shifts from being in the upper part of the cockpit, to a lower part that is not immediately visible after the button press. This additional arrow is different from the other arrows as it is the only that shows the direction of the next step.

We implemented a "Demo" function that demonstrated the flow sequence. When the "Play Demo" button is pressed, the flow is visualized by sequentially showing a border around each step in the flow sequence together with an arrow corresponding to that step. This function is accessible by interacting with a floating button, in near vicinity of the floating cockpit. We chose a location that is not in the line of sight while interacting with the virtual cockpit.

We also implemented a "Reset" function, that allows the user to reset the state of the application to the initial state. This was developed in case the user got stuck during some part of the flow procedure, either due to being unable to finish the current step or due to an unforeseen technical malfunction. It was also developed with testing purposes in mind, as it allowed us to easily restart the application from within, instead of having to exit it and executing it again.

During our testing process we often used the Microsoft Mixed Reality headset, as it was a much faster procedure for testing the software. Testing in the HoloLens requires building the project in Unity, then uploading it to the hardware, then navigating to the app and waiting for it to load. The whole process took around 5 minutes. Using the Mixed Reality headset the process overhead is greatly reduced, enabled by the testing functionality of the Unity software. However, this testing method is only suitable for testing out small iterations, such as script changes. Anything involving placement and changes in the rendered models required testing in the HoloLens. This was mainly due to hardware capabilities and differences between the two platforms. During one iteration we added a glow onto the markings which highlighted the current panel the user was at. This was done to increase visibility and clarity of the marking. When we tested the implementation in the Mixed Reality headset, it worked as intended. However, when we tested the same build in the HoloLens we experienced heavy frame rate issues, which rendered the program unusable.

All of the logic is contained in one C++ script. That includes the start-up initialization and the functionality of all the buttons. The start-up initialization consists of hiding all the borders and arrows that are visible by default, as well as all the corresponding buttons, leaving only the "Play Demo", "Start Flow" and "Reset" buttons visible. The coding was usually done applying pair programming, with all present members having the possibility to test.

Develop Solution / Build a model

The analyzing and brainstorming resulted in a clear idea of what to develop in the first iteration in the design process. To make it even clearer a Mid-Fi model was built. This was used to verify that our idea looked like the intended model when it was digital.

A Mid-Fi prototype was created from the simple sketch to visualize the planed marking of the flight panel. The online interactive prototype (see link in references "Adobe link to online prototype", 2019) is shown here below in Figure 6-8 as pictures from the different steps.

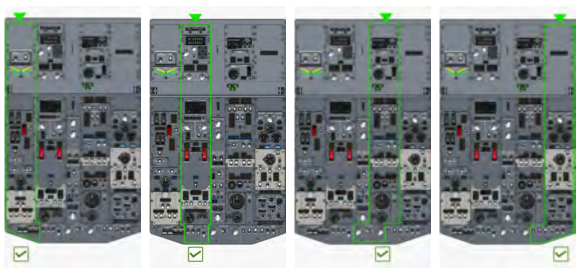


Figure 6: Flow in top panel marks. Step 1-4

Figure 7: Middle and bottom flow. Step 5-6

Developing the solution in unity was mainly done with pair programming. First a 2D picture (PM flight, 2014) of the paper tiger was put in the 3D space creating a virtual paper tiger. The mapping between the different parts of the paper tiger were adjusted several times by trying the model in the HoloLens and then changing it until it was in a place corresponding to the real paper tiger.

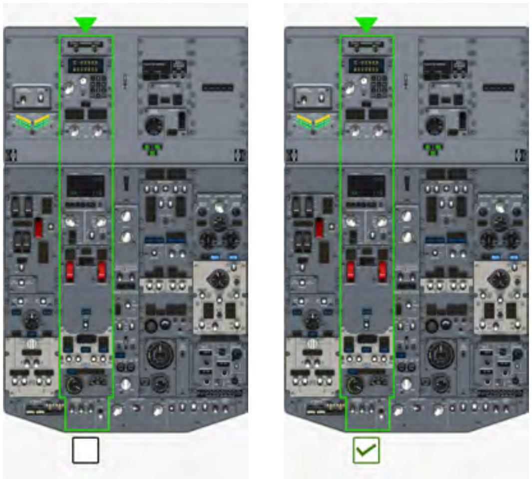


Figure 8: Prototype step 2. Before and after checking box

This made up the base of the solution. Then a check-box was added to the model and was made clickable. This makes the solution interactable. As mentioned above the development of the solution continued and is therefore not yet complete. A further description follows in the result section.

Test at TFHS

When finished with a working prototype the application was tested at TFHS. It was tested on six student that were used to flows and the paper tiger and one instructor. All test persons were male, all student went to the same class and no one had tried a HoloLens before. The test was divided into four significant parts:

- Introduction to test and HoloLens
- Testing the application
  - Looking at the Demo
  - Interacting with the flow
  - Quick debriefing
- Answering surveys
  - SUS
  - TLX
- Deep and subjective interview

All testers went through the same procedure and got the same information, assignments and questions at each part of the test. All members of the project group participated in different part of the test as test operators. On several occasions during the test it was clearly stated that it was the application that was tested and not them.

Introduction to test and HoloLens:

First the students were introduced to the test and how the whole procedure was planned to work. Then each student got an ID to use throughout the test to make it anonymous and with that a short survey regarding background information was done. Then the student got an introduction on how to use the HoloLens. The different hand gestures and features were explained and tried out. A different program than the application that were tested was used for this purpose. This was done especially to take a way the so called "wow-factor" with the AR-technology and the HoloLens. The program used for introduction was better suited to familiarize with the functions of the HoloLens than our application. After the introduction the student was led in to the test room where test operators took over to minimize the effect of how the first interaction with the HoloLens went, on the actual test.

Testing the application

Before the actual test was performed the tester got to sit down by the real paper tiger to explain and show what he or she knew from before. This was done to have a reference on their level of knowledge and to have a point of comparison for after the test. Then the test person got seated in a chair a couple of meters away from the physical paper tiger, facing an empty wall. Here they opened up the application and got somewhat used to the interface. The test operators made sure that they noticed all different panels and buttons before continuing with the test by instructing the test participants to look up to see the overhead panel, look left to see the thrust and communication panels and look right to see the oxygen mask and asking for a vocal confirmation.

In the first step of the test the Demo was played. The test person stood up and walked a couple of meters behind the chair to get all panels in his field of view at the same time. Then the test person played the Demo and observed. When the Demo was done the test person got seated again and started to do the flow. When pilot students normally practice towards the physical paper tiger they pretend they are interacting with buttons and metrics according to a list they have in their hand. This was done in the application as well. The test person pretended to adjust all buttons within the outlined green marked area. When the test person was done in one area the "next-button" was clicked and then the area where the next coming buttons on the list to adjust was showed. When the whole flow was finished a cheering sound was played to give the test person a feeling of doing well and to make it clear that they were done with the task. After the testing procedure in the HoloLens the test person was brought back to the real paper tiger. Here the test person got to show the flow just practiced in the HoloLens and some spontaneous first thoughts were talked about and noted by the test operators

Answering surveys

Right after the test two surveys were filled out (see Appendix for the whole questionnaire). Both with focus on the usability of the application. The first one was SUS and the second were NASA-TLX (the surveys can be seen in Appendix). These surveys were done before the the deeper questionnaire to get a result based on first instinct rather than a deep analyze.

Deep and subjective interview

The last part of the test was a deeper questionnaire with open questions with focus on the general experience with the application and what they personally thought about the experience. One test operator held the interview by asking the questions (see Appendix) and sometimes encouraged them to expand their answer. The other test operator wrote down all answers.

8 Result

When the simulator is started the user has a 3D replica of the paper tiger to his or her left, as the user starts from the right seat point of view in the training. It is assumed that the user is in a seated position. There is also ambience sound of an ambient airplane sound playing on a loop. As a tracking point, a standard "Gaze" pointer is used and this resembles the cursor on a PC. Instead of following hand movements, it will follow head movements. The simulator needs this so that it knows where the user wants to perform an action, e.g. mark one of the flows as done. Underneath the top-panel there is a button that will be used as a green "Done"-button. When the user clicks on this, the next step in the flow will show.

Result from first debriefing

The test involved 6 participants - 5 students and one instructor. All of the students were on their third semester. Everyone was familiar with the paper tiger and the flow procedure and was practicing in a real cockpit simulator. Everyone completed the flow successfully, with 13 to 15 items correct of 15 possible items, and an average completion time of 6 minutes, ranging between 2 minutes 58 seconds to 8 minutes 20 seconds. Issues related to shakiness and image resolution were brought up during the testing. After a student completed the demo we conducted a quick debriefing session, asking for their initial thoughts and first impressions. The most common comments were wishes for more interactivity, for example being able to change the state of switches and see light indicators change on the panel.

Other common comments were regarding minor inconsistencies of the panel images we used in our application compared to the ones used at TFHS. Many found the inconsistencies to be distracting or confusing, since they potentially represented undesired states of the panels. One participant mentioned the request for some form of receipt, indicating if the correct input had been made, and explanatory information near the panels. One participant also reported mental strain and indications of nausea after working with the flow in the HoloLens.

Main Interview

Similar feedback and answers were given by the subject during the main interview as in section *Result from first debriefing*. The subject thought that both the concept and idea was interesting and would be very desirable to close the gap between simulator and paper tiger. They also believed that the use of an fully developed application in the HoloLens would enhance the learning process and make it more enjoyable. However the prototype was not developed enough for it to be desirable as is, this opinion was reoccurring among the test subjects. The main concerns that the subjects had were the lack interactability, flows, image resolution and shakiness.

Result from SUS

The SUS-survey were done by all six test subjects. The final score differed between 60 to 85 and gave a total average score on 75,42. This result is above average but not very good.

Result from NASA-TLX

The first question in the NASA TLX-survey revealed that the application was, on average, not very mentally demanding, i.e. the task was not too complex and quite easy. The test subjects were also in agreement that the application was not very physically or temporally demanding, i.e. the test subjects did not feel much time pressure and the task was not to laborious. Almost all of the test subjects were satisfied with their performance in accomplishing the task in the application. The effort to accomplish this level of performance which the test subjects reported were not that high on average and almost none of the test subjects felt much frustration while performing the task. In figure 9 the result from the NASA-TLX survey can be seen.

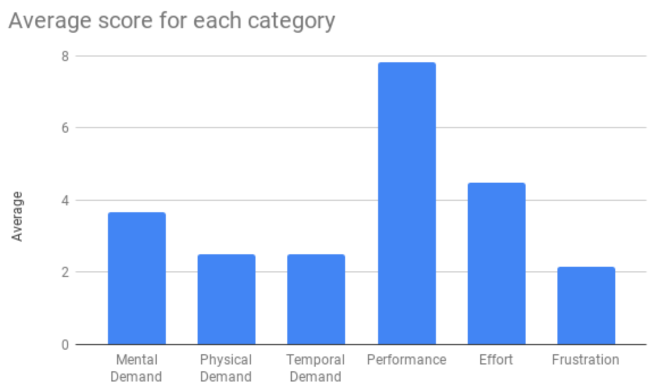


Figure 9: Results from NASA-TLX survey

Deep interview conclusion

Overall the testers found the technology to be very fun and interesting. The subjects saw a lot of potential in further development, however there were mixed thoughts about the current implementation and the use of it in the current curriculum. Another issue that several of the subjects pointed out was the shakiness in combination with a low resolution on the poster that was used. Some had difficulties reading what the panels said and what switch they were "flipping". Something that all of the subjects desired was more interaction, such as being able to flick a switch and a lamp for the switch glowing, or not glowing. A common denominator was that for the prototype to be a good substitute to the classical paper tiger was that further development was needed, especially with more flows and more interactive flows.

Results from design iterations

The main goal of this application was to reduce the gap, which students are experiencing, between the paper tiger and full flight simulator. The coding was done in C++

and all the logic was contained within one script, the visual implementations were done in Unity. All the coding were done mostly using pair programming. The application only contains one flow at this moment, the general scan flow, which allows the student to test the application in the whole cockpit. The students will start the application and be guided through the flow by green markings on specific areas in the cockpit, the markings will change place when the student presses one button as can be seen in the top image in Figure 10. The general scan flow consists of 15 steps in this application, therefore borders which matched the specific panels had to be made as well as corresponding arrows to tell the student where to start. At one stage in the flow there is a big gap between two panels and because the HoloLens has a narrow field of view the students would have a hard time figuring out where to continuing working. Therefore the big arrow pointing towards the goal was added, it can be seen in the bottom image in Figure 10. In order to help the student more a demo was added which demonstrated how the marking would move through the cockpit and navigating arrows were also shown here. In addition to providing the students with a demo button a restart button was also implemented in order to make it easier for the students to try a flow again fast if they should fail, it also made it easier to perform tests without having the test subject restarting the application every time.

Some technical issues arose during the development such as getting the sequence script of the demo working and hiding the buttons not relevant for each phase of the flow, as well as shakiness of the paper tiger. These difficulties were worked with until they were eventually solved.

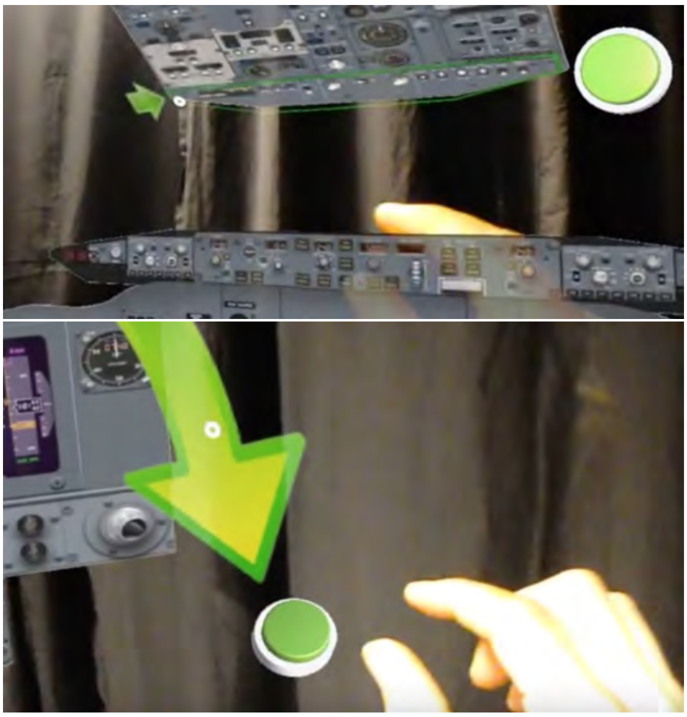


Figure 10: Flows with pointing arrows

9 Discussion

The goal of this project is to a provide a tool that enhances the learning experience of using the paper tiger at the early stages of pilot training program.

One of our main goals was to develop a proof-of-concept model. Developing for AR was a new area for all members of the team- and we opted in for importing high resolution images of the cockpit panels and rendering them in front of the user once the application is started. See the Result chapter for further information. Another aim was to bridge the gap between the training in the paper tiger and the training in a cockpit simulator.

Benefits in Augmented Reality

One major benefit with AR is the mobility of the device. Currently in pilot training, a pilot student has to study a specific flow in an operations manual before going to the physical paper tiger and test his or her knowledge. Such a physical paper tiger is limited to be used at a fixed location for example in the practice room at TFHS which can be seen in Figure 4. As described earlier in this report the final result of the developed application is a wireless and portable 3D model which can be used to sit and practice the locations of panels, buttons and layout of the cockpit with the augmented paper tiger in a similar way as with a physical paper tiger. As the application used in HoloLens can be placed anywhere in a room, the pilot students could hypothetically also borrow the HoloLens and practice at home, for example the day before a training session in the real simulator. The extended mobility and hence more available time to practice is one way of usage which can contribute and help bridging the gap between the paper tiger and the simulator.

Another advantage in the HoloLens is the fact that you can see yourself and especially your hand in the application in AR. This also enables you to see and communicate with a copilot as well as seeing and working with a physical operations manual or checklist in your own or your copilot's hand. The biggest contribution with the developed application of an augmented paper tiger is to offer choices of different scenarios. While using the application you can see a demo of the flow, or interactively receive flow guidance, with the purpose to learn the flow pattern in the cockpit. In our case we began with a general scan flow to demonstrate the spatial structure of a scan flow, which you interactively can click your way through. Our solution allows for a potential merge between the source of knowledge - the manual, the practice interface and the paper tiger. This starting point forms a very good structure for future developments toward even more interactive possibilities for learning which are further discussed in the section Future Work.

Still unsustainable for multi crew practice

The digital nature of the application allows for extensions to further enhancing and augmenting the learning experience. While the present advantages in the HoloLens can be summed up as seeing your surrounding and receiving guidance while using the application at a flexible location, which is also the whole purpose with Augmented Reality, some major important features of the application in AR still remain unsolved.

One discussion that kept popping up throughout this project was how suitable this application actually is for the HoloLens. Firstly, in the HoloLens the field of view is still quite small and the user has to turn his or her head in some unnatural ways. In Virtual Reality it would be possible to see the cockpit model in a more natural way.

Secondly, which part of reality is actually augmented? Especially for the interactive practise between pilot and copilot. Synchronization can be made possible using two HoloLenses so both the pilot and copilot can view the same digital "Paper tiger". However, a point to take into consideration is the price of the Microsoft HoloLens. As of this time of writing it is priced at USD 3000 for the development edition which has been used during this project. For an educational institution this should be a reasonable price if they find great value in the application, but getting two of them may be more than the economic budget for this purpose allows. Without access to an extra HoloLens there is no paper tiger visible for anyone else than the user. It thus seems like the person is just pretending setting buttons and values in the panels, which makes it rather hard for an interactive training session. One of our original ideas, which could solve this problem, was to use the capabilities of the Unity Engine to detect when a user is looking at the actual paper tiger and augment it directly with a digital layer. This would allow for a better experience by mapping the flow guidance to the real paper tiger and providing guiding arrows and other symbols to help the user navigate across the paper tiger panels. With this capability, two pilot students would be able to practice together with the only difference that one of them would have extra guidance and help. The lack if this possibility is still an important deficiency with the application in AR. Even though it was one of our original ideas, this solution was never developed partly due to the limited access to the paper tiger, which was located at TFHS in Ljungbyhed. To sum up: With the adjustment of the application so it would map perfectly to the real paper tiger it could be a great AR application for pilot students enabling the possibility to augment their individual and multi pilot practice in the physical paper tiger, but up until today that solution as is not yet sustainable.

Test of the application - First Debriefing

The group of students that were available for the testing at TFHS was homogeneous. One major point that had impact on the results was the degree of completed education and current experience with the physical paper tiger to provide greater insight about the usefulness and relevance. It would be interesting to have a more diverse group that consists of students at different stages of their education.

We received a chance to conduct quick informal tests on three first semester students that were on a break from their usual lectures. At their stage of the education they had not yet had any experience with the physical paper tiger. From the quick debriefing that was conducted the feedback was positive. The students were able to correctly replicate the flow pattern that they experienced in the HoloLens on the physical paper tiger. In one case a student completed only the first 4 steps of the pattern, and the next student that continued from that point was also able to correctly show the steps he completed. Since these tests were unplanned for and conducted in a hastily manner, the results are unfit to be used for any kind of analysis. However, they did show that a more thorough study involving groups of students at different stages of the education should be of great interest for future work.

Our goal was to enhance the learning of the general scan flow on a high level. Providing interactivity with different switches and buttons would shift the focus away from a high level learning model to a lower one. A low level interactivity practice could potentially be offered as an extension of the main functionality of the program, for example as a separate setting or mode that can be selected. More on this in the future work section.

Main Interview

In the main interview the feedback on the concept was good, where all of the test subjects believed that a similar but fully developed application would be of use in their early education. Since our test subjects, often more than once, in one way or another said that the current application would not be desirable because did not cover bigger parts of the curriculum. We think that the participants might have missed understood what they were testing, it seems like they expected a finished product instead of a small part of what could be a finished product. This leads us to think that we should have been even more precise during our introduction. The test subjects were very coherent during the interview many of the issues that they raised was more or less the same thing. So what could be the reason for such an coherent result during the interview? We believe that it could be that they all had the same experience thus leading to the similar expectations. This, in combination with that they also seemed to expect an finished product could point us towards why they all thought that there should be more than one flow. The image resolution on the poster is definitely something that could be explored more, especially by finding an even higher resolution poster and use that. But a smarter solution, that would also increase intractability, would be to use a 3D model of the cock-pit. We also believe that the image resolution could be what is causing the shakiness when inside the HoloLens. A problem we tried to solve by changing settings before building the application. The shakiness improved but was still noticeable.

SUS & TLX

As can be read from the results from the NASA TLX survey the application was not very complex which resulted in users not finding the task neither physically or mentally demanding. Since the application only contained one flow and the students could not interact with any special buttons or controls on the dashboard only one button at the end of every marked part of the dashboard the task was very easy which resulted in the students not feeling any time pressure did not find the task to laborious. If the buttons and controls were intractable and part of a challenge which the students would have to complete within a time frame the results from the NASA TLX would probably have looked very different. The fact that the test subjects were older students which had already passed the part of the education which our application was designed for must also be seen as a contributing factor. Because the students were already familiar with the procedure they knew beforehand where to look and they recognized many of the controls and buttons, since they could not interact with the buttons the effort was not particularly high to achieve the level of performance which was reported from the test subjects. The frustration was not high either, if the buttons and controls had been intractable there might have been greater frustration and more effort from the test subjects to achieve the same level of performance. The SUS score of 75,42 is considered to be above average, however it was only a small test group which could result in an imprecise estimate of the usability of the product. The test subjects, as mentioned before, also knew the contents of the application beforehand which gave them an advantage so the score might have differed if it had been tested on students which had no prior knowledge of flows.

10 Future work

In this section different expanding possibilities or ideas for other projects in the area are described and discussed.

The simulator could be further expanded in many ways to enhance the experience when using the simulator. Panel screens could be animated, this also applies to the surroundings i.e. moving clouds and changing light conditions. Panel lights could be turned on or off, based on what flow the user is training. This is also suggested in the testing section of the report.

Different target groups

As it will be mentioned in the subsection Gamification Implementations, several levels could be added. Depending on the users knowledge different levels of interaction could be added. I.e. if the user is a total novice, it could be enough with the flow being demoed and the user can then be requested to do the flow in a simplified manner, much like the prototype that has been the result of this project. However if the user is advanced, as shown during testing of the prototype, there is a need for more interactive parts on the paper tiger. This could be lights going on and off on the panels and thus allowing the student to follow the flow in a detailed manner.

Seen from a broad perspective, the application could potentially be valuable also for conventional (CPL) flight training where practice in a real simulator is even less than in the Multi Pilot Licence (MPL) concept, and the time in the paper tiger is shorter. A further developed application could even bridge to the Jet Oriention Course (JOC) or even a type rating.

Further 3D modelling

To have 3D models of the panels would make it possible to use the instruments (knobs) on the panels, thus giving the ability to change values on panel screens in a realistic manner. By being able to turn on knobs the simulator could check values against recommended values and thus passing or failing a user based on the set values. However, this would not be the same as flow training, this would be more detailed.

For a full version of the simulator there would be a lot more flows available to train on, approximately 18, with a menu where the user would be able to choose desired flow.

Mixed reality

In the future when the VR development has gone trough the stages where a users hands properly can be tracked or even seen inside the VR headset, very immersive experiences can be created. For pilots this is especially important since they need to see their hands while doing a flow. Then users can simply be loaded into a full-sized cock-pit, with all the proper animations and also different scenarios could be experienced (i.e. engine failure).

Vuforia Engine

In comparison to current implementation, the use of *Vuforia Engine* ("Developing Vuforia Apps for HoloLens", 2018) would instead use the poster that is used in the physical paper tiger and augment it by using *image targets* ("Optimizing Target Detection and Tracking Stability", 2018). With image targets, 3D objects could be placed (and locked in place) on top of the paper tiger, thus truly augmenting the perceived reality. Buttons could be lit up, levers could be added and made interactable and animations (the guided flow) could be added on top of the paper tiger. Example how on how this can be done can be seen in figure 11.

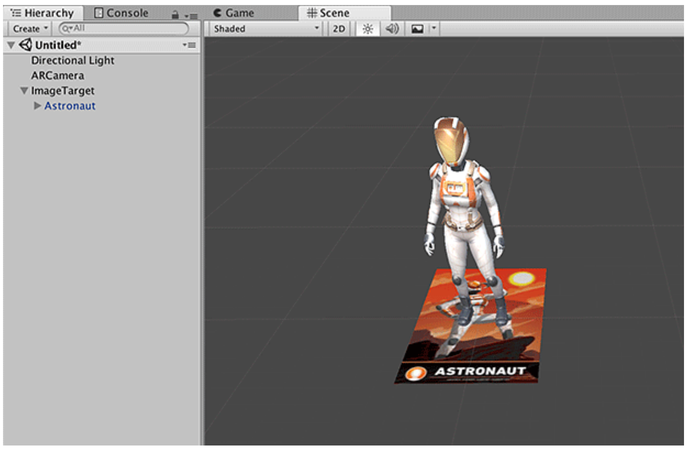


Figure 11: Vuforia Example using image targets ("Getting started with Vuforia in Unity", 2018)

This approach would be restricted to the user needing to have access to a physical version of the paper tiger and the developer team having access to the paper tiger. Another restriction that has not been explored in this report is how good image tracking characteristics the poster has, if they are bad, this solution is of low value. The software would not be able to track the poster reliably enough for implementations to be made.

Walk-around flow

Similar to the flows in the cockpit, are the flows that a pilot does before stepping into the airplane. One potential area of application of AR for flight training is to emulate a plane for a "walk-around", which is an outside check of an aircraft, see figure 12. The pilot checks that everything i.e. control surfaces and tires, are as they should. By using a 3D model of a airplane this could be practised in AR. One advantage of using AR for this application is that it is possible to re-scale and rotate a 3D-object to fit the constraints of the environment the user finds him or herself in. The pilot can then walk around the airplane and practise the flow. This can be further developed by adding random variables (such as ice or an bird nest in the jet engine) that the pilot should notice and handle. Because of hardware constraints in VR this application is much more feasible for AR. The digital nature of the application could provide a useful and also economical addition to the curriculum of any pilot school. Primarily because the price of a Boeing 737-700 is USD 85.5 million (Boeing, 2018). Because of that, the walk-around is not a part of the pilot training that is offered at i.e. TFHS.

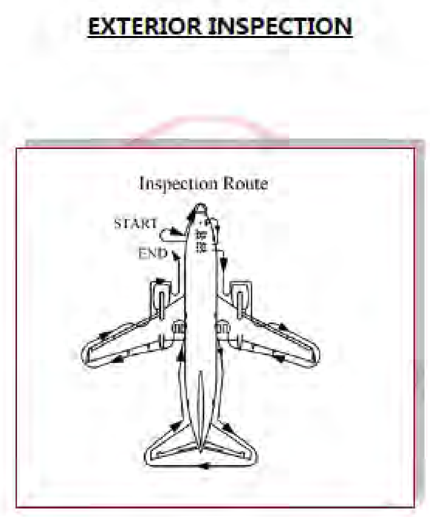


Figure 12: Preflight Preparation Flow

Operational logic

Operational logic could be implemented from a student perspective, to learn strategies to cope with distractions and the management of a flight, which contains of much more than controlling the movements of the aircraft in itself. For example the sound of a cabin crew member knocking on cockpit door with a request. In the simulator training these distractions are practiced, but it could potentially also be taught in an earlier phase of training.

Spatial Sounds

Spatial mapping of guiding sounds to the flows, and dynamic movement of the flows like droplets moving down aided by a intuitively matching sound could further aid in directing the users gaze toward the area in the cockpit of the flow to be worked with. One example of a sound could be similar to the song Descent Waterdrops, A Sence by TimeWarp (Kraak, 2018)

Gamification Implementations

To further promote learning based on the gamification principles it has been found that retrieval practice and spaced separation, where knowledge is rehearsed after a certain time interval, are powerful learning tools. When studying for a driving licence or a pilot's licence the theoretical studies are often closely connected to multiple choice questions where relevant knowledge must be brought forward to distinguish between slight differences in the options, and to be able to select the correct one (Easa, 2016). This method uses both retrieval practice and spaced separation .

In the context of our project future scenarios and areas where to implement gamification could be:

1. *Initial*: The mere fact to have the privilege of using an exclusive and highly technical device such as the Microsoft HoloLens could inspire an initial interest of the student.
2. *Retrieval Practice*: After the the first guidance of a flow, i.e. "general scan flow", an optional test to "show the flow" with the hand without guidance could be added. If the student is successful he or she can be rewarded with additional points.
3. *Unlocking levels*: After having showed the flow successfully a new flow is unlocked, following the same principal. If the student performs the flow correctly after the guidance the next flow in the sequence will be unlocked. Positive audio feedback such as "Well done!" And "Keep up the good work!" could be added after every successful flow.
4. *Spaced separation*: After having unlocked the flows which are performed on the ground (at the gate and before take-off). The option of a "mini-game" can be unlocked. Now the student can play the game and go through all the flows within one level, in the correct order. The student will get points for every correct flow and sequence and extra points for performing every flow within a well estimated time episode.
5. *Visual reward*: When the student has passed the examination of the ground flows he or she will be rewarded not only by unlocking the next levels and the first flow of the air-born procedures. The student's cockpit will also be upgraded with a nice vision over a beautiful city, mountain area or clouds for example. The cockpit will now also more resemble the real cockpit with the thrust section becoming 3D, cockpit window frames, circuit brakes behind the chairs, cockpit door etc. The paper tiger could transformed into something which closely resembles the full flight cockpit flight simulator. Even operational logic features such as receiving a cup of virtual coffee from a cabin attendant after reaching cruise level, and having performed the climb and cruise procedure, could be included. Now this could certainly add up to a magic, rather odd and hopefully inspiring and positive experience.

Public Feedback

Some additional feedback was gathered during a day of public testings at an open day at Lund University th 17 january 2019. Generally a positive enthusiasm to the possibility of interactive cockpit procedures with AR was noticed, and many of the people we talked to who tested the application could see future opportunities in the concept, students as well as companies working with Augmented Reality. One Pilot currently flying the Airbus 320 agreed with the function of showing a pattern of a flow, he also mentioned an additional flow for the emergency procedure rapid depressurization. He could also see potential value in the mobility of the device to be able to practice more, as today there is a lot to learn in a short period of time during a type rating, One person showed a video of how AR could be used in a similar way for the automobile industry to name components and show how parts were mounted together. The public testings were quicker and by far not as formal as those performed at TFHS but the majority of those who tried the application managed to click their way through the whole flow. They understood the concept even if it sometimes was difficult to start the application and to find the buttons "start flow" and "restart" and to click on the green buttons. Many also commented on the limited field of view. The public testings also generated some interesting discussions about ergonomic aspects of the eye and how we perceive distance in real life, how does it develops naturally and what is important to take into regard when developing an application in Unity. Finally further ideas developed by other students during these two project courses during the autumn 2018 also contributed with inspiration and further ideas of development of the application. For example attention guidance developed by the group "Real time feedback- your eye movements influence how you receive feedback", digital agents coaching a user through a learning scenario developed by the Galileo groups, and a demonstration of image recognition using the Vuforia engine developed by the Augmented Security group.

11 Conclusion

At the present time the developed product could be used mainly for early familiarization and introduction to a normal paper tiger as well as an introduction to the concept of flows. More development and interactive features are needed for any further applications. The development procedure has followed six steps where the problem is defined, information collected, solutions developed, feedback and improvement applied and back to a new definition of the problem. This has served as a good procedure for combining development and testing, including Lofi and Midfi prototyping. Results from the testings showed a system usability of 75,42 which is above average but not very good. Most test subjects were also in agreement that the application was not very mentally, physically or temporally demanding, although one participant reported mental strain and effect of nausea, and many mentioned difficulties in reading the panels. Almost all of the test subjects were satisfied with their performance in accomplishing the task and almost none of the test subjects felt much frustration while performing the task. Additionally all participants remembered the demonstrated flow with 86 to 100 percent accuracy. In the future many promising applications are possible and encouraged according to the conducted testings, primarily interactive and illuminated panels corresponding to the aircraft cockpit. In the time beeing the developed application consists of the general scan flow which can be easily followed, but reading panels necessary for more interaction can be difficult. In general terms the applications of AR learning tools could likely be assumed to rise along with increased quality of AR hardware.

To sum up, our project showed that the learning process for students can be improved with AR. Our final result, with a prof of concept model, is not enough but it showed the interest and possibilities with AR.

References

[1] Adobe link to online prototype: https://xd.adobe.com/view/156b6736-eb30-42aa-59a3-295850f555e9-635a/

[2] Bitenieks, R. Buckely, P. (2018) An Introduction to the Concept of Gamification and its Use in Commercial Pilot Training. Lund University School of Aviation.

[3] Black,P., & William,D. (1998). Assessment and Classroom Learning, Assessment in Education: Principles, Policy & Practice, 5:1, 7 — 74. https://doi.org/10.1080/0969595980050102

[4] Boeing (2018). About Boeing Commercial Planes. Retrieved (2018, 29 October) from https://www.boeing.com/company/about-bca/

[5] Chicago Architecture Center.(2018). The Design Process Consists of 6 Steps [picture]. Retrieved 2018-10-15 from https://www.discoverdesign.org/sites/default/files/inline-images/CAC-Design-Process-Chart_900x867px-01.jpg

[6] Degani, A (1991). Human Factors of Flight-Deck Checklists: The Normal Checklist. NASA Contractor Report. San Jose State University Foundation.

[7] Furness, T. A. (1986) The Super Cockpit and its Human Factors Challenges. Proceedings of the Human Factors Society-30th Annual Meeting-I986. Armstrong Aerospace Medical Research Laboratory Wright-Patterson Air Force Base, Ohio 45433-6573

[8] Developing Vuforia Apps for HoloLens. Retrieved (2018, October 22) from https://library.vuforia.com/articles/Training/Developing-Vuforia-Apps-for-HoloLen

[9] Easa Part FCL (2016). Retrieved (2018, October 29) from https://www.easa.europa.eu/sites/default/files/dfu/Part-FCL.pdf

[10] Getting Started with Vuforia in Unity. Retrieved (2018, October 22) from https://library.vuforia.com/articles/Training/getting-started-with-vuforia-in-unity.html#add-targets

[11] Goteman, O. (2006). Airborne cognitive systems in search of an appropriate context: Studies of the introduction of new technology in aviation using a cognitive systems engineering approach. Department of Design Sciences, Faculty of Engineering, Lund University.

[12] Haiduk, P. (2017). Display Formats for Smart Glasses to Support Pilots in General Aviation. Technische Universität, Darmstadt. Retreived (2018, 29 October) from http://tuprints.ulb.tu-darmstadt.de/id/eprint/6650

[13] Diem T.N. Hoang. (2016) BOOK REVIEW: Csikszentmihalyi, M. (2008). Flow: The Psychology of Optimal Experience (1st ed.). New York, NY: HarperCollins. 336 pp. ISBN 978-0-06-133920-2.

[14] Hutchins, E. How a Cockpit remembers Its Speeds. University of California, San Diego. Cognitive Science 19, 265-268 (1995).

[15] Japan Airlines Virtual Boeing 737- 800 Standard Operating Procedure (SOP). Edition 1- March 2013.

[16] Kraak Records(2018).Timewarp, Descent Waterdrops. Catalouge no. KRK101. Release Date February 16, 2018. Digital Media CDR.Retrieved 2018-10-26 from www.kraakrecords.com/krk101.html

[17] Lary, Diana (2015). China's Civil War: A Social History, 1945-1949. Cambridge: Cambridge University Press. p. 80. ISBN 978-1-107-05467-7.

[18] Microsoft Dev Center. (2018). MR Basics 101: Complete project with device. Retrieved 2018-10-15 from https://docs.microsoft.com/en-us/windows/mixed-reality/holograms-101

[19] Mjåtveit,S. (2018) Vad är en Type Rating? Published 2018-06-06. Retrieved (2018, 29 October) from https://www.osmaviationacademy.com/sv/blogg/vad-ar-en-type-rating

[20] Norman, D.(2013). The Design of Everyday Things. Basic books. New York.

[21] Bangor, A. Kortum, P. Miller, J (2009) Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale.Journal of Usability Studies Vol. 4, Issue 3, May 2009. Austin, Houston. Retrieved (2018, 13 December) from http://uxpajournal.org/wp-content/uploads/pdf/JUS_Bangor_May2009.pdf

[22] Human Performance Research Group. (1986). Nasa Task Load Index v.1. NASA Ames Research Center Moffett Field. California Retrieved (2018, 13 December) from https://humansystems.arc.nasa.gov/groups/TLX/downloads/T

[23] Oberhauser, M.,Dreyer, D., Braunstingl, R., Koglbauer, I., What's Real About Virtual Reality? Comparing the Fidelity of Virtual Reality with a Conventional Flight Simulation Environment. Aviation Psychology and Applied Human Factors (2018), 8(1), 22-34 https://doi.org/10.1027/2192-0923/a000134

[24] Optimizing Target Detection and Tracking Stability. . Retrieved (2018, October 22) from https://library.vuforia.com/content/vuforia-library/en/articles/Solution/Optimizing-Target-Detection-and-Tracking-Stability.html

[25] PM Flight. (2014). Free poster reference for 737 simulator sessions. Retrieved (2018, September 25) from http://www.pmflight.co.uk/737-simulator-cockpit-diagrams/

[26] Wikander, R., Dahlstrom, N. (2014).The Multi Crew Pilot Licence - Revolution, Evolution or not even a Solution? http://www.tfhs.lu.se/fileadmin/tfhs/doc/The_Multi-Crew_Pilot_License.pdf/

[27] Iordache, D., Pribeanu, C. (2009). A Comparison of Quantative and Qualitative Data from a Formative Usability Evaluation of an Augmented Reality Learning Scenario. Informatica Economica, vol.13, pp 67-75.

[28] Arvola, M. (2014). *Interaktionsdesign och UX: om att skapa en god användarupplevelse*. Studentlitteratur AB, Lund, Sweden.

A Questionnaire

Questions asked in interview:

- Beskriv hur du upplevde att det var att använda applikationen?
- Beskriv Demot och vad som visades, (den första delen av testet).
- Vad tror du syftet med demot var?
- Uppfyllde applikationen det syftet?
- Beskriv Flödet och dess funktion, andra delen av testet.
- Vad tror du syftet med flödet var?
- Levde applikationen upp till det syftet?
- Fastnade du på någon del i flödet? (Varför/inte? utveckla")
- Vad upplevde du för skillnader mellan demot och flödesträningen ?
- Hjälper områdsmarkeringen för att hitta knappar tillräckligt mycket eller behövts snarare knappspecifikt-markering?
- Vad kan du se för för och nackdelar med båda dessa lösningar?
- På vilket sätt skulle applikationen kunna användas vid studierna?
- För att få fram JA/NEJ) Tror du att applikationen hade varit hjälpsam?
- Hade du personligen använt den här applikationen under din studietid om den fanns tillgänglig? (i så fall vad tror du att det hade gett dig?)
- Hade du personligen använt en utökad applikation, till exempel med 20 flöden att öva.
- Hur ser du på fortsatt utveckling av applikationen? Med fler flöden och fler detaljer i instrumentpanelerna. (samt: För ett specifikt flöde; relevanta knappar och områden upplysta?)

B SUS

The following ten SUS-questions were answered on a five graded scale where 1 was "Strongly disagree" and 5 was "Strongly agree"

- I think that I would like to use this application frequently
- I found the application unnecessarily complex
- I thought the application was easy to use
- I think that I would need the support of a technical person to be able to use this application
- I found the various functions in this application were well integrated

- I thought there was too much inconsistency in this application
- I would imagine that most people would learn to use this application very quickly
- I found the application very cumbersome to use
- I felt very confident using the application
- I needed to learn a lot of things before I could get going with this application

C NASA-TLX

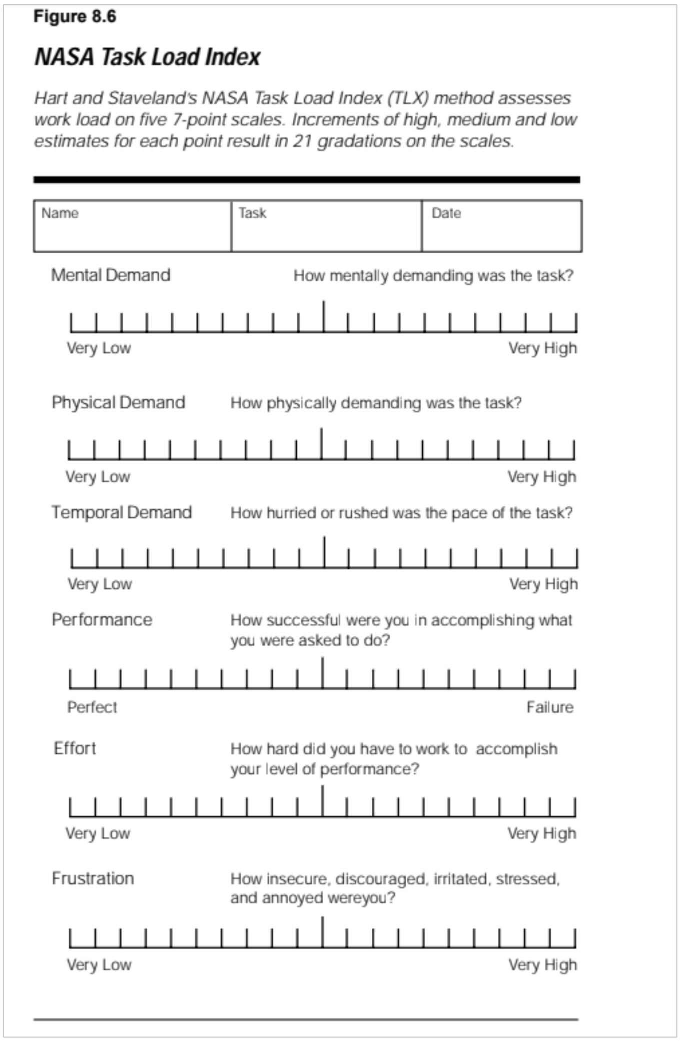


Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 13: Nasa-TLX form

Tidig programmering
Utveckling och utvärdering av ett lärspel för barn i åldrarna 4-6 år

Ebba Busch
dic14ebu

Dennis Jin
dat14dji

Lia Karida
dic13lka

Joel Ottosson
dic14jot

Abstract

Rapporten beskriver arbetsgång, resultat och slutsats efter ett projekt genomfört i kurserna Interaktion 1 & 2 vid Lunds universitet. Syftet med projektet var att utveckla ett programmeringsspel för barn i åldrarna 4-6 år. Programmering införs under 2019 i läroplanen för förskolan och därmed växer behovet av digitala hjälpmedel för tidig programmering. I dagsläget finns en mängd digitala lärspel men många är bristfälliga då de inte ger barnen någon användbar, informativ och lärande återkoppling eller feedback. Under projektet har tre LoFi-prototyper arbetats fram: Loopspelet, Saftspelet och Tårtspelet. Alla tre behandlar grundläggande programmeringskoncept på ett lekfullt och enkelt sätt. Prototyperna har sedan utvärderats tillsammans med pedagoger i förskola och förskoleklass. Efter testningen valdes Loopspelet ut för vidare utveckling. Baserat på utvärderingarna gjordes en rad förbättringar av Loopspelet som sedan resulterade i en HiFi-prototyp som kunde testas på målgruppen. Efter testerna togs en rad förbättringsförslag fram vilka ligger till grund för rapportens diskussion. Slutsatsen efter genomfört projekt blev att spelet har kommit en bra bit på vägen för att bli ett bra programmeringsspel, men att mycket arbete kvarstår innan spelet är redo att lanseras som ett läromedel. Förhoppningsvis kan resultaten från projektet ligga till grund för vidare utveckling av liknande spel.

1. Inledning och syfte

I takt med digitaliseringen av skolmaterial och utbildning har många dörrar öppnats upp för forskning om utveckling av datorbaserad utbildningsmetodik inom *tidig matematik* och *tidig programmering*. En del tidigare forskning har lagt fokus på att utforska skillnader mellan fysiska och digitala system [1].

Detta projekt är en del av kurserna Interaktion 1 & 2 (MAMN10 resp. MAMN15) vid Lunds universitet. Projektets syfte är att utveckla ett programmeringsspel för barn i åldrarna 4-6 år. I projektet ska ett eller flera delspel till *Magiska trädgården*, ett lek- & lärspel för barn, utvecklas.

2. Bakgrund och teori

Den här sektionen behandlar bakgrunden till varför utvecklingen av ett programmeringsspel för små barn behövs samt de teorier och den forskning som ligger till grund för de beslut som tagits under projektets gång.

2.1. Digital kompetens i läroplanen

I samband med den tekniska utvecklingen och digitaliseringen som samhället genomgår krävs det även att skolväsendet följer med i förändringen och ger alla elever en god *digital kompetens*. I mars 2017 beslöt regeringen att det behövs bättre förutsättningar för elevernas utveckling inom digital kompetens. Detta ska ske genom förtydligande och förstärkningar i skolans styrdokument. I bestämmelserna, som tillämpats 1 juli 2018, ingick bland annat att ”programmering införs som ett tydligt inslag i flera olika ämnen i grundskolan, framför allt i teknik och matematik” [2].

2.1.1 Digital kompetens i förskolan

Vidare bestämde regeringen att den tjugo år gamla läroplanen för förskolan också behövde förnyas. Digitaliseringen bidrar till att även förskolan behöver bli bättre anpassad till dagens moderna samhälle och framtidens utmaningar. Kvaliteten i undervisningen behöver därför ökas för att kunna uppnå en ökad måluppfyllelse [3].

I och med den reviderade läroplanen, som träder i kraft 1 juli 2019, behöver förskolan ta ansvar för att lägga grunden för att barn ska ha möjligheten att anamma de kunskaper som krävs för att kunna leva och verka i dagens samhälle. Det medför att även förskolan behöver använda digitala verktyg i det pedagogiska arbetet för att kunna utveckla barnens lärmiljöer. För att kunna genomföra denna förändring i barnens pedagogiska utveckling krävs det att förskolepedagogers digitala kompetens fördjupas och förbättras. Denna förändring är fortfarande relativ ny och kompetensutveckling inom digitalisering, i form av en webbkurs, erbjöds av Skolverket först i november förra året (2018) [4].

2.2. Programmering i förskolan

Som tidigare nämnt ställs det nu annorlunda krav på förskolan. Att väcka intresset för teknik hos barn och att ge en första uppfattning av att många saker i barnens omvärld är programmerade, är några av förskolans uppdrag. Barn ska tidigt få en förståelse för att världen de lever i kan påverkas; alla kan interagera med datorer oavsett kön och bakgrund [5].

När det talas om programmering i förskolan innebär det inte att alla barn ska bli programmerare. Det är själva logiken bakom att programmera som är viktig att komma åt. Förståelse för hur ett mål formuleras, hur ett problem uttrycks och hur instruktioner ges behöver utvecklas. Det är viktigt att tidigt förstå att den ordningen som moment utförs i spelar roll för slutresultatet [5]. Vilket vidare mynnar ut i två begrepp som är relevanta att känna till då programmering i förskolan diskuteras: *datalogiskt tänkande* och *analog programmering*.

2.2.1 Datalogiskt tänkande

Datalogiskt tänkande (eng. Computational Thinking) är en problemlösningsprocess. Det är en process som med datorers hjälp används för att beskriva, analysera och lösa problem [6]. Med andra ord handlar det om olika sätt att lösa uppgifter och bryta ner ett problem i mindre delar (*söndra och härska-tekniken*, eng. Divide & Conquer). Datalogiskt tänkande kan vara att sortera information, hitta mönster eller att förstå att långa instruktioner kan ersättas med kortare och lättare kommandon.

Begreppet datalogiskt tänkande har börjat användas för att beteckna en generell färdighet, snarare än en ämnesspecifik färdighet. Programmering blir därför mer ett pedagogiskt verktyg, för att träna upp det datalogiska tänkandet, som kan appliceras i flera skolämnen [7].

Många moment i förskolan tangerar redan datalogiskt tänkande i form av att instruktioner ges, mönstermatching sker i lekar och ramsor, rutiner och villkor ställs på barnen så att de gör saker i rätt ordning för att få det önskade resultatet. Utmaningen för förskolepedagogerna blir att själva kunna förstå sambandet och hur det kan vara en grund för digital kompetens [5].

2.2.2 Analog programmering

Analog programmering syftar till att ta programmeringens begrepp ut i verkligheten och låta barn programmera utan dator. Exempel på analog programmering kan vara att barn styr varandra med hjälp av talade instruktioner eller pilar. Att sätta ihop olika rörelser till en danssekvens kan också vara en typ av analog programmering [5]. Analog programmering uppmuntrar till att använda kunskaper om programmering i det vardagliga livet och kan vara ett viktigt komplement till allt digitalt material som används idag.

2.3. Digitala lärspel

För att underlätta och uppmuntra till digitalisering och datalogiskt tänkande i både förskola och skola finns en uppsjö digitala lärspel och applikationer. Det visar sig dock att en stor majoritet av de digitala lärspel som används i svenska skolan idag inte går i linje med det som forskning påvisat krävas av ett spel för att det ska vara ett spel för lärande och inte ett ”spel för testande”. Det vill säga ett spel som enbart testar en elevs redan existerande kunskaper [8]. En annan benämning på testande spel är ”träningsspel”, som syftar på att spelet tränar det en elev ska lära sig. Oavsett om ett spel kallas för tränings- eller testande spel är det viktigt att komma ihåg att spelet inte nödvändigtvis är lärande. En viktig faktor som är avgörande för att ett lärspel ska vara meningsfullt är dess hantering av feedback.

2.3.1 Feedback i digitala lärspel

Feedback handlar om att förmedla resultat och framsteg efter en handling. Att använda feedback i digitala system är väsentligt för att den som använder systemet eller spelet, användaren, ska förstå vad som händer och varför. För att feedback ska vara meningsfull måste den dock utformas på rätt sätt. Feedbacken måste vara informativ, planerad och prioriterad så att den viktigaste informationen uppfattas först. Vidare får feedback inte vara för omfattande, utan måste innehålla lagom mycket information. Annars kommer användaren att sluta ta den till sig. Feedback måste också presenteras inom rimlig tid efter den genomförda uppgiften, annars kan en användare inte avgöra vad den får feedback på [9].

Vid konstruktion av feedback i digitala lärspel finns det dessutom ett antal andra faktorer som det måste tas hänsyn till. Feedback som tillåter en elev att testa sig fram till dess att korrekt svar uppnås är ett exempel på feedback som i stor utsträckning inte bidrar till lärande. Feedback som ger konstruktiv kritik om den utförda uppgiften och hur den kan utföras bättre är däremot ett exempel på feedback som bidrar till lärande. Bra feedback bör innehålla information om huruvida uppgiften löstes korrekt eller inte samt varför. Några olika typer av feedback som lämpar sig olika bra att använda i digitala lärspel är *verifierande feedback*, *korrigerande feedback*, *utvecklande feedback* (eng. Elaborated Feedback), *uppmuntrande feedback* och *resultatfeedback* [8].

Verifierande feedback berättar bara för användaren huruvida svaret var rätt eller fel. Den här typen av feedback uppmuntrar till trial & error-beteende, det vill säga att användaren testar sig fram till dess att rätt svar uppnås. Digitala lärspel som endast tillämpar verifierande feedback med möjlighet för användaren att använda en trial & error-strategi bidrar enligt flera studier inte till lärande. Korrigerande feedback ger användaren mer information om vad som gick fel, tillsammans med det rätta svaret. Korrigerande feedback har visat sig vara bättre än verifierande feedback i kontexten av lärspel. Utvecklande feedback är sådan feedback som ger användaren information om vad som gick fel, men istället för att ge det rätta svaret ges ledtrådar eller andra implikationer på hur det rätta svaret uppnås. Uppmuntrande feedback handlar om att berätta för användaren hur väl den presterat, sådan feedback ska dock användas med försiktighet eftersom den inte ger någon feedback relaterad till uppgiften och därmed inte bidrar till lärande. Resultatfeedback presenterar bara poäng eller ett betyg för användaren och bidrar, på samma sätt som uppmuntrande feedback, inte till lärande eftersom feedbacken inte säger något om den utförda uppgiften. Dessutom kan resultatfeedback ha negativ inverkan på motivation om den används för att jämföra resultat mellan användare. [8]

2.4. Existerande programmeringsspel

I detta avsnitt kommer några av de mest populära programmeringsspelen för barn presenteras. Det finns många olika variationer på spelen, både digitala och fysiska. De flesta är dock digitala vilket tillhandahåller ett enkelt och skalbart sätt att lära ut programmering på, både i klassrum och över Internet.

2.4.1 Bee-Bot & Blue-Bot

Både Bee-Bot och Blue-Bot är fysiska, programmerbara robotar. På robotarna finns det fyra fysiska knappar: en för rörelse bakåt, en för rörelse framåt samt två för rotation (höger/vänster). Knapparna används för att mata in kommandon. Kommandon som matas in lagras i minnet så att användaren kan bygga en sekvens av kommandon. Det finns även en *Go*-knapp som används för att köra kommandona i minnet, en *Clear*-knapp för att rensa minnet och en *Pause*-knapp för att pausa. [10]

Till Blue-Bot finns en mobilapplikation som kan användas för att kontrollera roboten via Bluetooth. I applikationen finns en digital representation av Blue-Bot som rör sig synkront med den fysiska roboten. Det finns även olika typer av rutbaserade spelfält. Exempelvis finns ett med alfabetets bokstäver, där idéen är att lära barn att stava. Dessa spelfält har sedan olika svårighetsgrader som lägger till hinder och begränsar mängden kommandon.

Både Bee-Bot och Blue-Bot fokuserar huvudsakligen på sekvensiellt tänkande och introducerar barn till tidig programmering genom de olika spelfälten och den relativt primitiva mängden kommandon. [10]

2.4.2 Lightbot

Lightbot är ett mobilspel som går ut på att sätta ihop en sekvens av kommandon för att förflytta en robot från en punkt till en annan i ett rutbaserat spelfält. Lightbot är lite mer avancerad än Blue-Bot eftersom roboten utöver rörelse framåt och bakåt samt rotation också kan hoppa för att förflytta sig på spelplanen. Till skillnad från Blue-Bot så finns det mer tydliga programmeringskoncept som loopar och subrutiner (och därmed rekursion); det finns alltså möjlighet att skriva ett huvudprogram med subsekvenser som går att loopa igenom. [10]

2.4.3 Scratch

Scratch är ett tvådimensionellt programmeringsspel som MIT Media Lab tagit fram. Scratch är mycket mer avancerat än de tidigare nämnda spelen eftersom det efterliknar nodbaserad programmering. I spelet byggs ett exekverbart program upp med hjälp av både primitiva och avancerade block. De enklaste blocken låter spelaren kontrollera rörelse av objekt i spelet samt vilka texturer objekten ska ha. Det ger spelaren möjlighet att lägga till egna karaktärer som denne själv ritat. Tillvägagångssättet vid animation av karaktärers rörelse är också mer komplicerad än i Lightbot eftersom spelaren måste definiera diskreta pixelvärden på en spelkanvas. Utöver dessa grundblock finns även olika typer av sensoriska block som låter objekt reagera på kollision och färger i dess omgivning. [10]

2.5. Magiska trädgården

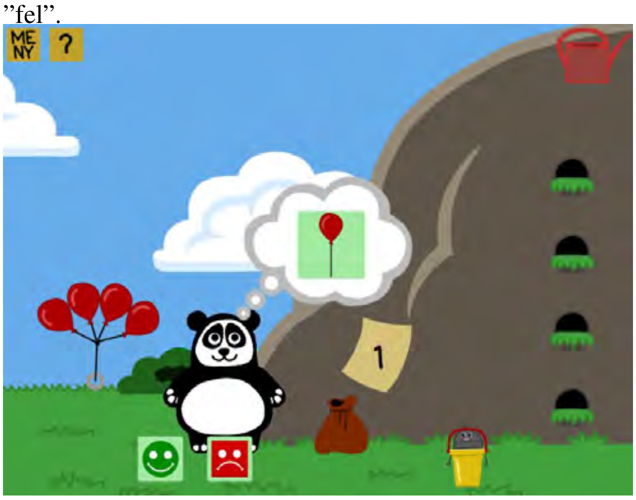
Magiska trädgården är ett lek- & lärspel utvecklat av Educational Technology Group på Lunds universitet. Spelet fokuserar på tidig matematik för barn i åldrarna 3 till 6 år. Målet med spelet är, ur spelarens perspektiv, att samla vatten för att odla en egen trädgård. Detta görs genom att utföra uppgifter i spelet, vilka alla tränar tidig matematik. I Magiska trädgården anpassas svårigheten dynamiskt efter hur spelaren presterar, så att alla spelare kan träna matematik utifrån sin individuella förmåga. Detta sker dock utan att spelarna själva får veta vilken nivå de spelar på, så att olika spelare inte ska kunna jämföra sig med varandra [11].

Som redan nämnt består Magiska trädgården främst av en uppsättning olika banor som successivt ökar i svårighetsgrad. Utöver detta finns en uppsättning så kallade *kalasspel*. Dessa spel är separata från de olika banorna och syftet är att ordna ett överaskningskalas åt en digitala karaktär. Även här fokuserar alla spel på tidig matematik.

Utgångspunkten för de spelidéer som utvecklats i det här projektet har varit att de ska passa in bland kalasspelen, men de ska fokusera på tidig programmering istället för tidig matematik. Idéerna som presenteras handlar om förberedelserna inför spelkaraktären *Panders* kalas.

2.5.1 Teachable agent

I Magiska trädgården används en så kallad *teachable agent*. En teachable agent är en digital elev i spelet som spelaren själv får lära upp. Efter det att en spelare själv löst uppgiften ska denne instruera sin digitala elev att också lösa uppgiften, i syfte att förstärka lärande [11]. Efter det att spelaren har testat en uppgift i spelet några gånger dyker den digitala eleven upp i spelet för att titta på och lära sig. När den digitala eleven sedan observerat spelaren ett tag börjar den ge förslag på lösningar. Då blir spelarens uppgift att svara på om den digitala eleven tänker rätt eller inte, snarare än att spela själv. I Magiska trädgården sker detta genom att den digitala eleven tänker på ett alternativ som dyker upp i en tankebubbla, detta illustreras i figur 1. Spelaren får sedan ta ställning till om den digitala eleven tänker rätt eller fel genom att trycka på den gröna smiley-knappen för ”rätt” respektive röda smiley-knappen för ”fel”.



Figur 1. Den digitala karaktären Panders tänker på en lösning.

3. Arbetsgång

Under uppstartsfasen av projektet diskuterades gemensamma hållpunkter och gruppregler inför arbetet. Tidigt bestämdes det att gruppmöten skulle hållas veckovis, dels med projektgruppen och dels med projektgruppen tillsammans med handledarna. En målsättning för projektet sattes upp och en tidsplan med deadlines och veckovisa uppgifter gjordes. Det beslutades att målet för delkurs 1 var att ha gjort förstudier, LoFi-prototyper till två eller tre spelidéer samt användartest av dessa för att sedan välja en spelidé att gå vidare med i delkurs 2. Arbetssättet som valdes var en iterativ designprocess där idéer, delresultat och tankar kontinuerligt diskuterades.

3.1. Förstudie

Alla i gruppen tog gemensamt ansvar för att ta reda på vilka programmeringsspel som existerar, i allmänhet och för målgruppen. Vidare gjordes en studie om målgruppens kompetensnivå och intressen i syfte att bestämma svårighetsgrad och tema för spelet. På vilket sätt programmering förekommer i den svenska läroplanen för både skola och förskola undersöktes. Med utgångspunkt i den egna programmeringskompetensen diskuterades sedan vilka koncept och begrepp inom programmering som är väsentliga och lämpliga att inkorporera i ett programmeringsspel för barn.

3.2. LoFi-prototyper

Efter att det beslutats vilka programmeringskoncept som skulle finnas representerade i spelen började olika spelidéer diskuteras. För att spelen skulle passa in i Magiska trädgården valdes *kalas* som tema för spelen, eftersom Magiska trädgården bland annat består av en uppsättning kalasspel. Efter diskussion om kalasrelaterade aktiviteter, programmeringskoppling och idéernas betydelse i utbildningssyfte valdes tre spelidéer ut. De utvalda spelidéerna delades upp mellan gruppens medlemmar som gjorde varsin LoFi-prototyp av respektive idé. Prototyperna gjordes på papper med viss interaktivitet så att det skulle vara möjligt att testa dem. Till varje prototyp skrevs ett manus med ljudinstruktioner för användning vid test.

3.3. Test av LoFi-prototyper

När de tre LoFi-prototyperna var klara och delvis reviderade efter diskussion inom gruppen samt med handledare förbereddes två testsessioner. Handledarna tillhandahöll kontaktuppgifter till en förskola som tidigare varit involverad i utvecklingen av Magiska trädgården. Vidare kontaktades en skola där två pedagoger i förskoleklassen var intresserade av att medverka i testet. Både förskolan och skolan var belägen i södra Sverige. Testerna hölls vid två olika tillfällen, ett test på skolan där två pedagoger medverkade och ett test på förskolan, även där medverkade två pedagoger. Testerna av prototyperna genomfördes individuellt så att pedagogerna inte skulle bli påverkade av varandra under sessionen. Efter att alla tre prototyper testats fick de fylla i tre enkäter, en för varje prototyp, också detta individuellt. Enkätmallen finns i bilaga A och innehåller sju påståenden som graderas på en femgradig Likert-skala från ”Håller inte alls med” till ”Håller helt med”. Syftet med enkäterna var att samla in data om kvalitet, meningsfullhet i utbildningssyfte och enkelhet hos de olika prototyperna. Sessionen avslutades med en gemensam diskussion där prototyperna och pedagogernas tankar och åsikter kring tidig programmering diskuterades. Underlaget för diskussionen finns i bilaga B.

3.4. HiFi-prototyp

Efter genomförda tester sammanställdes resultaten och det beslutades att ett av de tre spelprototyperna skulle väljas ut för vidare utveckling. Med testresultaten och diskussionerna med pedagogerna i åtanke valdes Loopspelet ut för detta ändamål. En rad förbättringar och förändringar diskuterades. Efter detta började arbetet med HiFi-prototypen, den utvecklades i *React* med spelmotorn *Phaser* och kördes på en iPad. Utvecklingen av HiFi-prototypen involverade programmering, grafiskt arbete samt inspelning av ljudinstruktioner. Målet med HiFi-prototypen var att den skulle innehålla så mycket fungerande spellogik som möjligt så att den skulle gå att testa på målgruppen. För att utvecklingen ändå skulle rymmas inom ramen för projektet valdes att endast satsa på en svårighetsnivå.

3.5. Test av HiFi-prototyp

Testerna av HiFi-prototypen utfördes på samma skola och förskola som testerna av LoFi-prototyperna, båda belägna i södra Sverige.

Den här gången skedde dock testet med eleverna och inte med pedagogerna. Precis som vid testet av LoFi-prototyperna genomfördes två tester vid två olika tillfällen. Totalt medverkade 16 elever i åldrarna 4 till 6 år. Några genomförde testet enskilt och några genomförde det parvis. Testpersonerna blev ombedda att spela spelet på en iPad och fick sedan frågan om det var roligt. Skärmen spelades in under hela testets gång. Under testet fick de guidning av en testledare. En protokollförare närvarade också under hela tiden för att anteckna händelseförloppet. En mer detaljerad specifikation över testet finns i testplanen i bilaga D. När testerna genomförts sammanställdes resultaten, vilka ligger till grund för diskussionen senare i rapporten.

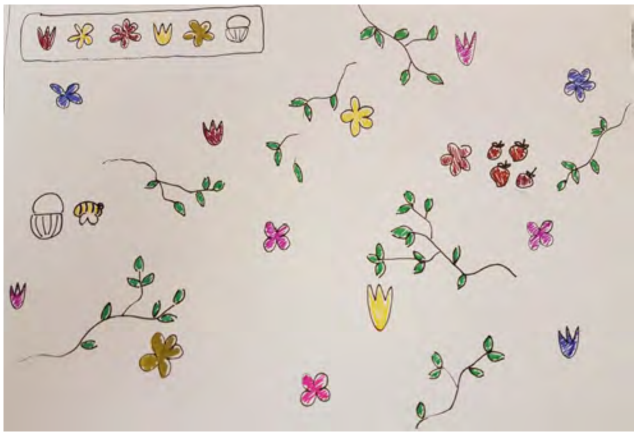
4. Utveckling av LoFi-prototyper

Den här sektionen beskriver dels de tre LoFi-prototyper som är resultatet av första iterationen i projektet och dels resultatet efter genomförda tester med efterföljande diskussion.

Under projektets förstudie beslutades vilka begrepp inom programmering som spelen skulle utgå ifrån. Dessa var:

- Instruktioner
- Villkor
- Mönstermatchning
- Loopar och upprepningar
- Konsekvenstänk och tänka i förväg

Begreppen ansågs vara lämpliga att bygga ett spel kring eftersom de är fundamentala delar av programmering och därmed viktiga att ha förståelse för. Detta mynnade ut i tre spelidéer: *Loopspelet*, *Saftspelet* och *Tårtspelet*, som alla passar in bland kalasspelen i Magiska trädgården. Några begrepp som tillkom senare under processen och som också ansågs relevanta för tidig programmering är *söndra och härska*-tekniken samt *exekvering*.



Figur 2. Loopspelet

4.1. Loopspelet

Spelplanen består av ett antal blommor i olika färger och former. Vid en av blommorna finns ett antal jordgubbar. Spelplanen återfinns i figur 2. Spelaren ska hjälpa ett bi att flyga till jordgubbarna och tillbaka till start, där det finns en korg. På spelplanen finns en ruta med instruktioner i form av en uppsättning blommor i en speciell ordning. Spelaren ska klicka på de blommor på spelplanen som motsvarar instruktionen, alltså rätt blommor i rätt ordning. Biet berättar detta för spelaren på följande sätt:

“Jag är på väg till Panders kalas och har lovat att ta med mig jordgubbar till tårtan. Kan du hjälpa mig att plocka jordgubbarna? Följ instruktionerna i rutan och visa mig vilka blommor jag ska flyga mellan. Vilken blomma ska jag flyga till först? Tryck på den!”

När spelaren klickar på en blomma dras ett streck till den valda blomman från den senaste blomman. Biet kommer för varje blomma fråga *“Vilken blomma ska jag flyga till nu? Tryck på den!”*. När spelaren klickat sig tillbaka till start börjar biet flyga och säger *“Nu flyger vi!”*, se figur 3. När biet flyger förbi en korrekt blomma utropar den *“Jippi!”*. Om biet passerar en blomma som är felaktig kommer den att flyga tillbaka till start och spelet börjar om. Då informerar biet *“Oj, nu blev det fel... Vi får flyga tillbaka och försöka igen”*. Om biet passerar en blomma som är felaktig efter att biet plockat upp en jordgubbe kommer jordgubben att lämnas tillbaka innan biet flyger tillbaka till start. I detta fall ändras biets föregående kommentar till *“Oj, nu blev det fel... Nu får vi lämna tillbaka jordgubben och försöka igen”*. När biet passerar blomman där jordgubbarna finns tar den en jordgubbe och utropar *“Åh, här är jordgubbarna! Synd att jag bara kan ta en i taget”*. Biet kan bara bära en jordgubbe i taget.

När biet kommer tillbaka till start och en jordgubbe har plockats upp blir spelaren frågad om hur många gånger till biet måste flyga för att hämta alla jordgubbarna. *“Men*

Figur 3. Lösning på Loopspelet

det finns ju fortfarande jordgubbar kvar, hur många gånger måste jag flyga för att plocka upp alla?". När spelaren har svarat kommer biet att flyga så många varv som spelaren angett. Det finns ett alternativ som är rätt och ett alternativ som anger för få varv. Om spelaren väljer fel alternativ och biet flyger för få varv får spelaren svara på samma fråga igen, om hur många varv som är kvar.

Svårighetsgraden i Loopspelet kan varieras genom att variera längden på instruktionen. Det vill säga antalet blommor som biet ska flyga mellan. Utöver detta kan också antalet blommor på spelplanen som inte ingår i instruktionen varieras i olika grad för att göra det svårare för spelaren.

4.1.1 Feedback

I Loopspelet tillämpas utvecklande feedback eftersom en spelare efter ett feldrag får information om att det blev fel samt får försöka igen. Efter ett antal feldrag kan det tänkas att spelaren får extra hjälp att se nästa instruktion (blomma) genom att denna t.ex. lyses upp i instruktionsrutan. Om spelaren svarar fel när den ska välja hur många varv till biet ska flyga, kommer spelaren själv att visuellt se att det finns jordgubbar kvar och spelaren får sedan en chans att rätta till felet genom att svara igen. Spelaren får feedback på huruvida den har valt rätt blommor först efter det att loopen har slutits och biet börjar flyga, det fångar dels programmeringskonceptet "skriv först exekvera sen" men det gör också att spelaren inte kan tillämpa en trial & error-strategi. Om spelaren istället hade fått veta direkt om den valda blomman var korrekt eller inte hade den kunnat testa sig fram och trycka på olika blommor tills dess att loopen är korrekt sluten.

4.1.2 Programmeringskoncept

För att klara Loopspelet måste spelaren förstå en instruktion. Det är viktigt att spelaren följer denna instruktion så att biet flyger till rätt blomma i rätt ordning. Instruktionen kan därför ses som en algoritm, något som är mycket vanligt i programmering. Den inbördes ordningen mellan blommorna i instruktionen, liksom ordningen mellan olika steg i en programmeringsalgoritm, är avgörande för att problemet löses korrekt. Här får spelaren möjlighet att träna konsekvenstänk och får en visuell återkoppling på vad som händer när instruktionen inte följs. Att följa instruktionen en gång i Loopspelet är dock inte tillräckligt för att samla alla jordgubbar och därmed klara spelet. Mönstret som biet flyger måste upprepas identiskt flera gånger. I ett programmeringssammanhang skulle mönstret motsvara kod som återkommer flera gånger. Ett vanligt sätt att korta ner och göra programmeringskod mer kompakt är att använda loopar. För den erfarne programmeraren kan koden också i vissa fall bli lättare att förstå. En loop är ett skrivsätt som göra att ett kodblock upprepas ett antal gånger. Vilket är det scenario som utspelar sig i Loopspelet, biet ska flyga på ett speciellt sätt ett upprepat antal gånger. Sättet som biet flyger på ger också en visuell illustration av hur en loop ser ut. När spelaren har följt instruktionen och biet börjar flyga motsvarar detta i programmeringssammanhang att koden är färdigskriven och exekveras. Illustrationen kan ses i figur 3.

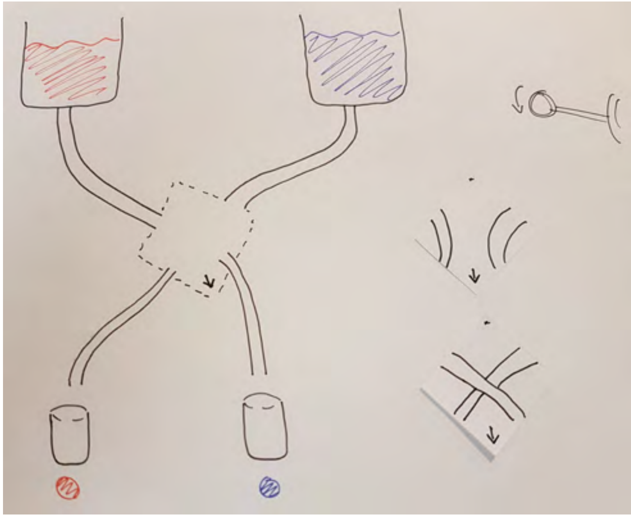
4.2. Saftspelet

Spelplanen består av ett antal behållare med saft i olika färger. Från dessa behållare finns det rör som leder saften till olika glas. På rören finns det avbrott som gör att saften inte kan rinna hela vägen ner till glasen. Glasen är i sin tur färgkodade med färgerna som återfinns på saften i behållarna. Spelarens uppdrag är således att leda rätt saft till rätt glas genom att koppla ihop rörsystemet på rätt sätt. För att sluta rörsystemet har spelaren ett antal rördelar till sitt förfogande som skall dras till rätt avbrott i rörsystemet. När spelaren har gjort detta skall denne dra i spaken, som finns på spelplanen, för att saften ska börja rinna genom rören från saftbehållarna ner till glasen. Rördelarna kan flyttas och tas bort från rörsystemet så länge spelaren inte har dragit i spaken. Spelplanen illustreras i figur 4.

Spelet inleds med en ljudinstruktion som berättar ovanstående spelmekanik för spelaren. Denna instruktion ser ut som följer:

"Jag har lovat att hjälpa till med att hälla upp saften på Panders kalas! Alla gästerna vill ha olika smak på sin saft. Kan du hjälpa mig att bygga ihop rören så att rätt smak hamnar i rätt glas. Dra rördelarna till rätt plats. Dra i spaken när du är klar och vill hälla upp saften".

Under spelets gång får spelaren uppmuntrande kommentarer i form av *"Bara en bit kvar!"* och *"Åh, vad gott det ska bli med saft!"*. Gör spelaren ett fel som leder till att ett glas inte fylls med saft, ges kommentaren *"Oj, nu blev det fel... Igelkotten fick ingen saft..."*. Om rördelarna placeras så att det kommer för mycket saft i ett glas



Figur 4. Saftspelet

får spelaren ljudfeedback:en *"Oj, nu svämmar det över...".* När rördelarna har placerats korrekt och spelaren har klarat spelet berättas detta med följande fras: *"Ja, super! Nu har alla fått sin saft!".*

Spelet har många möjligheter till anpassning av svårighetsgrad. Spelet innehåller ett antal parametrar som kan förändras för att uppnå detta. Antalet behållare med saft kan ökas och minskas. Samma justering kan göras för antalet glas som ska fyllas. Utöver detta kan också antalet avbrott i rörsystemet förändras men även antalet rördelar som spelaren kan välja mellan för att placera i rörsystemet.

4.2.1 Feedback

Feedbacken i Saftspelet ges efter det att spelaren har dragit i spaken och saften börjat rinna. Om det har blivit fel kommer saften att hamna i fel glas eller på golvet. Kanske hamnar två sorters saft i samma glas, så att saften får en helt ny färg och glaset svämmar över. För att undvika verifierande feedback kommer spelaren inte att presenteras för rätt svar om den gör fel, utan kommer istället att få testa en gång till. Eftersom feedbacken kommer efter det att spelaren drar i spaken kan spelaren inte tillämpa en trial & error-strategi och testa olika bitar tills dess att rätt bit läggs. För att undvika att en trial & error-strategi tillämpas mellan spelomgångarna kan banan och bitarna varieras så att en spelare som gjorde fel inte får samma bana och samma bitar vid nästa försök.

4.2.2 Programmeringskoncept

Saftspelet tvingar spelaren att bryta ned problemet och titta på varje rörkoppling för sig. I programmering så kallas denna nedbrytningsteknik för söndra och härska. Den går ut på att angripa problem genom dela upp dem i mindre delproblem. Spelaren måste för varje rörkoppling fundera på vad som kommer hända beroende på vilken rörbit som placeras där. Det vill säga, olika rördelar kommer ge upphov till olika konsekvenser och saften kommer att rinna på olika sätt baserat på detta. Rörkopplingarna och rördelarna kan därför ses som villkorssatser där rördelarna anger olika villkor för hur saften ska rinna. Villkorssatser är något av en grundbult i programmering och är mycket vanligt förekommande.

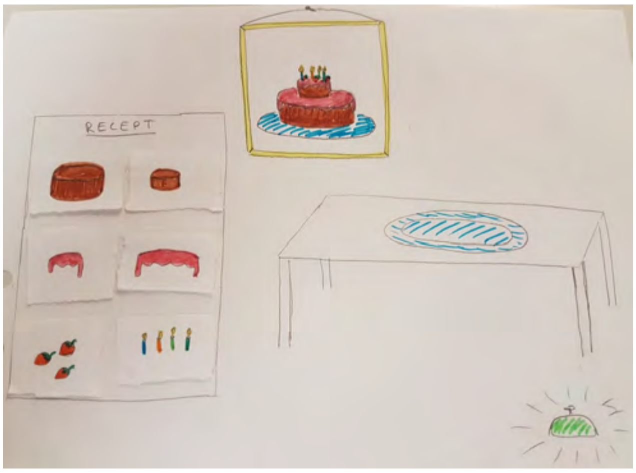
I Saftspelet ger spelaren en visuell representation av hur en villkorssats fungerar i form av saft som rinner på olika sätt. Saften i spelet börjar rinna först när spelaren drar i spaken vilket kan liknas vid att kod exekveras när den är färdigskriven. Kod exekveras vanligen efter att programmeraren har tänkt igenom ett problem och kommit fram till en lösning. Spelet fungerar på samma sätt, spelaren tänker igenom hur saften ska rinna och placerar rördelarna innan den drar i spaken. Beroende på svårighetsgraden så finns det olika antal rörkopplingar och spelaren tvingas då tänka i flera steg. I detta fall måste det tas hänsyn till hur rörkopplingar på olika ställen i systemet påverkar varandra. En förändring på en rörkoppling högre upp i systemet kan göra att en rörkoppling längre ner leder saften på ett oönskat sätt och också måste förändras. Detta blir också en återspegling av hur programmering fungerar där olika kodblock kan vara beroende och påverka varandra.

4.3. Tårtspelet

Målet med detta spel är att låta spelaren följa ett recept och bygga upp en tårta. Spelaren får tillgång till ett antal olika ingredienser som tårtan ska byggas upp av. På spelplanen finns receptet på tårtan i form av en bild som illustrerar hur de olika ingredienserna skall sättas ihop. För att bygga upp tårtan skall spelaren dra respektive ingrediens till tårtfatet på spelplanen. Detta meddelas initialt till spelaren genom ljudkommentaren *"Dra en ingrediens till tallriken".* Under spelets gång ges ytterligare kommentarer som *"Vad ska vi lägga på härnäst?".* Ingredienserna måste placeras på tårtfatet i rätt ordning ovanpå varandra. Om en del placeras fel, kan denna dras bort från fatet och en ny ingrediens kan placeras på dess plats. Spelaren uppmuntras också med följande kommentar i spelet: *"Det här ser gott ut!".* När spelaren känner sig färdig med tårtan ska denne plinga på klockan som finns på spelplanen. Detta förmedlas med en instruktion som säger *"Plinga på klockan när du är färdig".* Spelplanen kan ses i figur 5.

Om spelaren har följt instruktionen på bilden och lyckats bygga upp en korrekt tårta förmedlas det med orden *"Åh, vad fint det blev! Precis som Panders vill ha det!".* Om den uppbyggda tårtan däremot inte blev som instruktionen förmedlas det såhär: *"Oj, den ser inte riktigt ut som på bilden... Försök igen!".*

Utöver ljudinstruktionerna beskrivna ovan ges också en



Figur 5. Tårtspelet

inledande spelbeskrivning vid spelets början. Den ser ut på följande sätt:

"Vill du hjälpa oss att baka en överraskningstårta till Panders? Här är en bild på Panders favorittårta. Vi ska använda alla ingredienserna. Placera ingredienserna på tallriken så att tårtan ser ut som den på bilden. Plinga på klockan när du är färdig."

Svårighetsgraden i tårtspelet kan huvudsakligen anpassas genom att antalet ingredienser som tårtan kommer att bestå av ökas. Detta ger en ökad komplexitet då det finns fler ingredienser att välja mellan. Utöver detta så kan det också införas fler ingredienser än vad som ska ingå i tårtan. Antalet ingredienser att välja mellan blir ännu större och spelaren måste aktivt hålla koll på hur många ingredienser som finns i receptet. Spelaren måste sedan välja ut dessa bland ingredienserna och dra dem till tårtfatet.

4.3.1 Feedback

Även i Tårtspelet får spelaren feedback efter det att den lagt ut alla bitar, för att undvika att en trial & error-strategi används. Om tårtans bitar ligger i fel ordning när spelaren plingar på klockan kommer tårtan att bakas ihop fel. Om t.ex. den stora tårtbottnen ligger ovanför den lilla kommer tårtan att tippa och ”dega ihop”. Spelaren får därmed visuell feedback på vad som gick fel och kan få förståelse för varför den stora biten måste ligga under den lilla. Spelaren får se konsekvensen av att utföra instruktionen i fel ordning och får chansen att testa igen. Spelaren får dock inte det rätta svaret rakt av och på så sätt undviks verifierande feedback.

4.3.2 Programmeringskoncept

En viktig del av tårtspelet är förmågan att följa en instruktion, det vill säga receptet på tårtan. Det är viktigt att en spelare enbart gör de saker som står och att de görs i rätt ordning steg för steg. I programmering är detta förfarande mycket vanligt, men brukar där kallas för att följa en algoritm. I programmeringsalgoritmer är det ofta viktigt att saker sker i en specifik ordning för att önskat beteende ska uppnås. Om något skulle göras för tidigt så kan ett program bete sig konstigt och likaså är fallet i spelet. Om spelaren till exempel försöker att lägga dit tårtbottnen sist så blir resultatet inte alls som på receptbilden. Om instruktionen inte följs får detta alltså konsekvenser för hur resultatet blir, det vill säga tårtans utseende. Tårtspelet ger således en introduktion till algoritmer och ett tankesätt som är viktigt i programmering. När spelaren anser sig färdig med sin tårta plingar denne på klockan, vilket motsvarar att färdig kod exekveras.

4.4. Resultat från LoFi-testerna

I detta stycke kommer resultatet från de två testtillfällena presenteras. Detta inkluderar en sammanställning av pedagogernas enkätssvar samt en kort redogörelse från diskussionen som tar upp väsentliga tankar och idéer som uppkom med avseende på prototyperna.

Totalt var det fyra personer som testade prototyperna, två från Lund och två från Veberöd. Pedagogerna från Lund arbetar i en förskoleklass på en grundskola (F-klass till åk 9) med *Montessoripedagogik*, medan de två pedagogerna från Veberöd arbetar på en förskola (år 1 till år 6) med inspiration från *Reggio Emilia-filosofin*. Tabell 1 listar en testare med en viss arbetstitel och plats med ett nummer. Detta för att underlätta hanteringen av enkätsvaren samt kunna identifiera personerna i testdokumenationen.

Tabell 1. Sammanställning av testpersonernas arbetsroll och plats.

Person	*Plats*	*Arbetsroll*
1	Lund	Montessoriförskollärare (F-klass)
2	Lund	Montessoripedagog (F-klass)
3	Veberöd	Förskolepedagog, IKT-ansvarig
4	Veberöd	Pedagogista

När Loopspelet (figur 2) testades missade tre av fyra test-personer instruktionsrutan med de olika blommorna uppe i vänstra hörnet av spelplanen. Alla tre personerna placerade istället ut den snabbaste vägen till jordgubbarna och sedan antog de att biet skulle ta samma väg tillbaka till korgen. Person 1 och person 3 drog med fingret mellan blommorna när de skulle placera ut rutten för biet, medan de andra två tryckte på de olika blommorna. Majoriteten av testperson-erna hade svårigheter med att förstå var de skulle börja.

Det som utmärkte sig under testerna av Saftspelet (figur 4) var att samtliga testpersoner tog sig an uppgiften på olika sätt. En person behövde fundera först innan denne provade sig fram, en annan tog direkt den översta biten och testade medan övriga två kunde se direkt vilken bit som passade in, men en av dem valde också att vrida på den för att se om den passade in på flera sätt.

Samtliga testpersoner klarade av Tårtspelet (figur 5) direkt. Det enda som skiljde sig mellan testpersonerna var att personerna från Veberöd placerade ingredienserna i rätt ordning men i en hög medan pedagogerna från Lund placerade ut ingredienserna liggandes vertikalt i nivåer.

Tabell 2. Totalpoängen för respektive prototyp.

Fråga	*Loopspelet*	*Saftspelet*	*Tårtspelet*
1	10	5	5
2	10	6	5
Totalt	**20**	**11**	**10**
3	16	12	19
4	14	17	19
5	15	16	17
6	17	17	18
7	18	18	19
Totalt	**80**	**80**	**92**

Tabell 2 sammanställer alla testpersonernas enkätsvar. Varje testperson svarade på tre enkäter med samma frågor, men med avseende på de olika spelen. Enkätmallen kan ses i bilaga A. I tabellen identifieras frågorna med ett nummer mellan 1 och 7 vilket korresponderar med enkätfrågornas nummer. Personerna behövde ta ställning på en femgradig Likert-skala där 1 motsvarade ”Håller inte alls med” och 5 motsvarade ”Håller helt med”. Fråga 1 och 2 uttrycktes med negation medan de andra frågorna (3-7) formulerades med den inledande texten ”Jag upplevde...”. Siffrorna i kolumnerna under de tre olika prototyperna (Loopspelet, Saftspelet, Tårtspelet) är totalpoängen från alla fyra testpersonernas svar på respektive enkätfråga. Med andra ord har varje persons svar med värde mellan 1 och 5 adderats ihop.

I och med att de två första frågorna är formulerade annorlunda än de övriga, har tabellen två totalsummor. För att få det bästa resultatet för den översta totalsumman ska siffran bli så låg som möjligt, medan för den undre totalsumman ska siffran bli så hög som möjligt.

4.4.1 Resultat från diskussionerna

I bilaga B presenteras de olika diskussionsfrågorna som ställdes till testpersonerna efter att de genomfört testerna och svarat på enkäterna. Sammanställningen nedan kommer redovisas med data från båda testtillfällena gemensamt då denna undersöknings syfte var att få återkoppling på de tre spelidéerna.

Alla testpersonerna poängterade vikten av att framhäva instruktionsrutan med mönstret bättre i Loopspelet. De föreslog att en visuell effekt, som exempelvis att rutan lyser upp, skulle underlätta för användaren. Vidare var majoriteten av personerna osäkra på var på spelplanen de skulle börja. Testpersonerna från Lund föreslog att det i den upplästa spelbeskrivningen skulle kunna nämnas att det är biet som behöver hjälp. Samtliga pedagoger ansåg också att det var ologiskt att biet skulle behöva flyga till så många blommor för att komma till jordgubbarna. Testpersonerna från Veberöd kom med lösningsidén om att biet skulle kunna få näring vid blommorna eller att biet samtidigt ska plocka en bukett med Panders favoritblommor. En annan idé som de hade var att introducera något sorts hinder framför den snabbaste vägen och på så sätt förklara varför biet behöver flyga en omväg (runt hindret). Vidare föreslog de att spelaren i förlängningen själv skulle få programmera biet.

Saftspelets största utmaning, som diskuterades under båda testtillfällena, var att djupseendet i spelet kan vara svårt att förstå. Med andra ord 3D-effekten då två rör korsar varandra och det ena röret går bakom det andra. En annan utmaning som pedagogerna misstänkte, var att barnen kommer vilja prova sig fram tills de lyckas. Ifall de skulle misslyckas hade de en idé om att saften skulle spillas ut på golvet för att ge den visuella återkopplingen att det har blivit fel.

Tårtspelet uppfattades som roligast av pedagogerna. Det de ansåg svårt med konceptet var att det kan bli utmanande för barnen att förstå att de gör rätt eftersom tårtan inte kommer att se ut som på bilden förrän de plingar på klockan och tårtan ”bakas ihop”.

Under diskussionen med pedagogerna i Veberöd framkom också att de saknade ett spel som introducerar barnen till programmeringsbegrepp. De arbetade mycket med analog programmering, det vill säga programmeringstänk i vardagen. Ett exempel på detta kan t.ex. vara en instruktion för i vilken ordning barnen ska klä på sig ytterkläderna då de ska gå ut. Pedagogerna menade att de själva försöker introducera termer som t.ex. *bugg*, då någonting går fel i den analoga programmeringen. De menade att ett spel som på ett liknande sätt introducerar bregrepp skulle vara till nytta för dem. De ansåg också att koncept som *mönster* och *loop* inte var speciellt förekommande i de existerande spel som de använder.

4.5. Slutsatser efter LoFi-testerna

Som det går att avläsa från tabell 2 efter en summering från båda totalsummorna, fick Tårtspelet bäst resultat av pedagogerna. Den uppfattades både som lättast att förstå och att genomföra. Saftspelet hamnade i mitten men fick sämst betyg på fråga 3 som handlade om egenbedömning på hur väl uppgiften genomfördes. Loopspelet fick dock sämst och uppfattades som minst intuitiv. Däremot diskuterades det mycket om Loopspelet under diskussionerna.

Efter testerna beslutades det att Loopspelet skulle bli det spel som tas vidare för utveckling av HiFi-prototyp. Följande stycken beskriver och motiverar de tankar som ligger bakom beslutet.

4.5.1 Jämförelse med andra programmeringsspel

Testningarna gav mycket värdefull information om vad som saknas hos existerande programmeringsspel. Förskolan i Veberöd arbetade mycket med analog programmering och införandet av programmeringsbegrepp i vardagen. De ansåg att programmeringsbegrepp saknas i existerande spel för målgruppen.

För att matcha det behov som pedagogerna såg valdes införandet av programmeringsbegrepp som en uppgift för Loopspelet. Spelet ska stegvis introducera programmeringsbegrepp i takt med att spelaren utvecklas och spelets svårighetsgrad ökar. Aktuella begrepp kan t.ex. vara *bugg* (när någonting går fel), *loop* (i samband med att biet flyger) och *sekvens* (som synonym till instruktion). I samband med att flera begrepp införs kan också komplexiteten hos instruktionen (blommönstret) öka. Just mönster och loopar var också sådant som pedagogerna saknade i existerande spel vilket var en bidragande orsak till att Loopspelet valdes för vidare arbete.

De existerande programmeringsspel som beskrivits tidigare handlar alla om att på mer eller mindre avancerade sätt styra en karaktär rätt genom ett rutnät med hjälp av enkla instruktioner som ”höger”, ”vänster” och ”framåt”. Scratch behandlar programmering på en avancerad nivå och lämpar sig inte för målgruppen 4-6-åringar. Spelen tycks vara designade så att en spelare ska uppleva att den håller på med just programmering. I Blue-bot, Bee-bot och Lightbot är de uppenbara programmeringskoncepten instruktioner, konsekvenstänk och exekvering. Det handlar om att instruera en digital eller fysisk robot om hur den ska röra sig för att nå ett visst mål, sedan får roboten utföra detta och spelaren ser resultatet. Koncept som saknas i de här spelen är loopar och mönster, vilket också är ett argument för att välja Loopspelet för vidare utveckling då det behandlar dessa koncept. Att just loopar och mönster saknas påpekades även av pedagogerna i Veberöd.

En av tankarna bakom de framtagna prototyperna har varit att de inte ska upplevas som programmeringsspel. Existerande programmeringsspel uppfattas som starkt kopplade lite programmering och kan ha en avskräckande effekt för de som är obekanta eller obekväma med ämnet teknik. Några bidragande faktorer till detta är att många spel handlar om robotar som förknippas med avancerad framtidsteknik. Dessutom har många av programmeringsspelen ett ”stelt” upplägg med ett rutnät som spelplan och några instruktionsknappar.

De framtagna spelidéerna i detta projekt utspelar sig i en hemtrevlig och bekant miljö, som en del av ett större spelkoncept; att planera ett kalas. Syftet med idéerna handlar inte enbart om att lyckas förflytta sig från punkt A till punkt B, utan programmering och logiskt tänkande är inbakade i spelen. Konceptet är att allt som spelaren lyckas åstadkomma i slutändan leder till ett färdigt kalas, vilket innebär att det finns en större mening för spelaren jämfört med att enbart styra en robot i ett rutnät.

Att spelaren inte ska uppfatta spelet som ett programmeringsspel samtidigt som spelet introducerar en rad programmeringsbegrepp kan låta motsägelsefullt. Tanken är dock att programmeringsbegreppen i spelet ska användas i ett vardagligt syfte. Ett fel kan t.ex. presenteras som en bugg så att spelaren med spelets hjälp får förståelse för vad ordet bugg innebär i en vardaglig kontext, så att denne i ett senare skede kan förstå begreppet i dess egentliga kontext. När spelaren sedan lär sig ”riktig” programmering kommer denne lättare förstå koncept som bugg och att en sekvens är en rad instruktioner.

Under testningen av Loopspelet föreslog några pedagogerna att spelaren själv skulle kunna programmera biet. Det för visserligen tankarna till Blue-Bot och Bee-Bot, men det skulle vara intressant att undersöka om, och i sådana fall på vilket sätt, det skulle kunna implementeras med hjälp av teachable agent-metodiken.

4.5.2 Tankar inför utveckling av HiFi-prototyp

Det kan konstateras att Loopspelet ansågs vara det spel med mest utvecklingspotential. Loopspelet var det spel som diskuterades mest under testningarna och som pedagogerna dessutom såg en meningsfullhet i. Det beslutades därför att Loopspelet skulle tas vidare för utveckling av HiFi-prototyp, samt att ett av målen för spelet skulle var att stegvis introducera programmeringsbegrepp eftersom pedagogerna såg ett behov för detta då det inte förekommer i de spel som används i nuläget. Dessutom diskuterades möjligheten att bygga ut instruktionerna i Loopspelet så att de stegvis blir svårare. Till en början kan instruktionen bestå av en kort sekvens som allt eftersom spelaren blir bättre kan göras längre, för att ytterligare lite senare bli en uppsättning sekvenser: en *algoritm*. Detta öppnar dels upp för en stegvis ökning av svårighetsgraden men också för att successivt introducera spelaren för fler och fler programmeringsbegrepp. Vidare konstaterades det att Loopspelets spelplan behövde designas om till att ha ett konkret hinder som stoppar spelaren från att ta snabbaste vägen till jordgubbarna. Detta har tidigare diskuterats och infördes i form av kvistar som är tänkta att skymma biets sikt, men det var inte tillräckligt konkret och tydligt. Instruktionsrutan måste också vara tydligare, kanske ska den lysa upp när den nämns i ljudistruktionen så att spelaren förstår att den inte är tryckbar (en del av spelplanen). Något sätt att kunna rensa spelplanen om en spelare gjort fel behövs också och efter diskussion med pedagogerna föreslogs en ikon med en soptunna för detta ändamål, eftersom det är en symbol som barnen förstår. Sist måste det undersökas vidare vad som ska hända om en spelare inte förstår instruktionerna och inte sluter loopen, hur ska detta förmedlas och vilken feedback är passande?

5. Utveckling av HiFi-prototyp

Den här sektionen beskriver dels HiFi-prototypen för Loopspelet som är resultatet av den andra stora iterationen i projektet och dels resultatet efter genomförda tester på målgruppen 4- till 6-åringar.

Inför utvecklingen av HiFi-prototypen diskuterades resultatet från LoFi-testerna och följande punkter identifierades som förbättringsområden:

- Tydligare ruta med instruktioner
- Tydligare att det är biet som spelaren ska hjälpa
- Hur ska det hanteras om spelaren klickar sig snabbaste vägen till jordgubbarna och inte förstår att denne måste sluta loopen?
- Inför visuellt hinder så att det är logiskt varför det inte går att ta snabbaste vägen till jordgubbarna
- På vilket sätt ska programmeringsbegrepp införas?
- Undersök möjligheterna att utöka instruktionerna till att bestå av flera sekvenser: en algoritm
- Undersök möjligheterna att ha en teachable agent

Utöver detta ansågs följande punkter vara viktiga för utvecklingen av HiFi-prototypen:

- Röster och ljudinstruktioner
- Spelets grafik

5.1. Revidering av Loopspelet

Att rutan med instruktioner upplevdes som otydlig under LoFi-testerna kan vara en följd av att prototypen var skissad på ett otydligt sätt, där rutan t.ex. inte skiljde sig i färg och djup från övriga spelplanen. Därför beslutades det att instruktionsrutan i HiFi-prototypen skulle vara mer utstickande och placeras på en träskylt som dessutom lyses upp när den omnämndes i ljudinstruktionerna.

För att förtydliga att det är biet som spelaren ska hjälpa, samt introducera spelaren till vem biet är, beslutades det att biet ska ge spelets instruktioner. När spelet börjar ska biet vara stort på skärmen och prata med spelaren, så att det kan presentera sig själv samt förklara spelet, för att sedan minska och placeras i utgångsläget för spelet.

Någonting som diskuterades mycket inför utvecklingen av HiFi-prototypen var hur spelet ska hantera fel, mer specifikt om spelaren inte förstår att denne måste sluta loopen och ta kortaste vägen till jordgubbarna. Tidigare var det tänkt att biet skulle börja flyga av sig själv när loopen var sluten, men det skulle innebära en del problem om spelaren inte förstått att denne måste sluta loopen. Då skulle spelet behöva upptäcka att spelaren inte förstått, t.ex. genom att låta spelaren ha en viss tid på sig att lägga ut loopen innan ledtrådar ges. Det skulle dock vara problematiskt eftersom en spelare kanske lägger ifrån sig spelet en stund eller bara tar längre tid på sig än vad spelet förväntar sig. Det var svårt att komma på en teknisk lösning som skulle vara passande för att avgöra om en spelare missförstått instruktionerna, varför idén om att biet skulle börja flyga av sig själv övergavs. Istället beslutades det att spelaren ska trycka på en knapp när denne känner sig färdig, då kan det direkt upptäckas om loopen är sluten eller inte och om så inte är fallet får spelaren information om detta. Knappens placering valdes till att vara intill instruktionsrutan så att den hamnar efter instruktionen. På så sätt uppmuntras ”instruktion först, knapptryck sen” vilket är en logisk händelseföljd och dessutom ger knappen en tydlig mappning. Knappen bestämdes att vara grön med en play-symbol vilket ger en tydlig affordance. Dessutom ser det ut som att knappen ”sjunker” vid nedtryckning i syfte att ge feedback.

För att göra spelet mer logiskt behövdes ett visuellt hinder, som ska uppmana spelaren att inte ta kortaste vägen till jordgubbarna, införas. Det diskuterades mycket kring vad det skulle kunna vara för hinder. Till en början var planen att det skulle vara Panders som sitter och läser, så att spelaren måste flyga runt honom för att inte bli upptäckt. Eftersom spelet dock skulle testas fristående och inte som en del av Magiska trädgården skulle testpersonerna inte känna till Panders och därför övergavs den idén. Istället valdes en sten som hinder. Planen var att stenen skulle vara så pass stor att det skulle upplevas som att jordgubbarna döljs bakom den. Det är lite riskfyllt eftersom spelaren ser planen snett uppifrån och därför själv måste tänka sig att stenen är i vägen, men stenen valdes ändå som hinder. För att göra spelet ytterligare lite mer logiskt och tydligt beslutades det att ljudinstruktionerna också skulle uppdateras. De ska beskriva att jordgubbarna finns på ett hemligt jordgubbsställe och att biet behöver med hjälp av instruktionerna hitta dit, likt en skattjaktskarta.

Eftersom ett av syftena med spelet beslutades vara att införa programmeringsbegrepp hölls också en diskussion om vilka begrepp som var relevanta att introducera. Med utgångspunkt i de egna programmeringskunskaperna samt i resultatet från LoFi-testerna och diskussionerna med pedagogerna om vilka begrepp och koncept de saknade i existerande spel, togs beslut om att följande begrepp var relevanta:

- Loop
- Instruktion
- Bugg
- Mönster
- Sekvens
- Algoritm

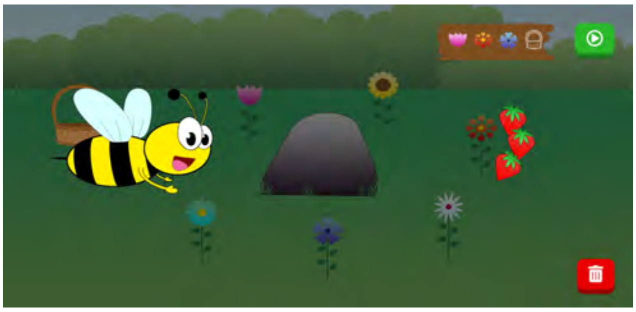
Där begreppen *sekvens* och *algoritm* bör introduceras senare i spelet då svårighetsgraden ökar. Begreppen bör introduceras på ett förklarande vis i ljudinstruktionerna så att en förståelse kan byggas. De bör också underbyggas med visuell återkoppling, som t.ex. att biet flyger i en cirkel (loop) samtidigt som begreppet introduceras och förklaras.

Utöver detta diskuterades också hur frågan om hur många fler varv i loopen biet behöver flyga för att hämta upp alla jordgubbarna skulle ställas. Detta testades inte under LoFi-utvärderingen och lösningen som används är därför snarlik lösningen i LoFi-prototypen, eftersom inget underlag för att designa om lösningen fanns. Det beslutades dock att de olika alternativen skulle presenteras som knappar med siffror på, där de siffror som finns representerade är den siffra som är korrekt (rätt antal varv) samt alla siffror mellan 1 och denna. Det vill säga, om korrekt antal varv skulle vara 4 finns siffrorna 1, 2, 3 och 4 att välja på. Det diskuterades mycket om målgruppen skulle förstå siffrorna eftersom det inte är självklart att barn på förskolan har förståelse för detta men det beslutades att detta i sådana falla skulle få visa sig under testningen.

Vidare beslutades det att utbyggnad av svårighetsgrad samt möjlighet att ha en teachable agent i spelet låg utanför projektets omfattning, varför detta inte implementerades i HiFi-prototypen. Områdena diskuteras dock i sektion 6 Diskussion.

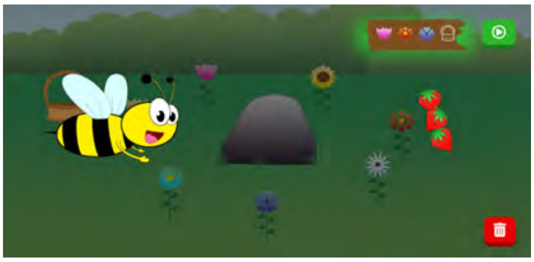
5.2. Loopspelet 2.0

Allt det som nämnts i föregående avsnitt resulterade i en ny version av Loopspelet med reviderad spelplan och delvis reviderade ljudinstruktioner. Loopspelet som beskrivs i detta avsnitt är den version som testades under användartesterna (se avsnitt 5.3).



Figur 6. Biet berättar om programmeringsuppgiften.

Spelet börjar med att biet är förstorat på skärmen och berättar för spelaren om spelets uppgift, se figur 6. Un-der tiden biet pratar är spelplanen i bakgrunden ”utgråad” och det går inte att trycka på några knappar. Biet förklarar uppgiften såhär: *“Jag ska på kalas och har lovat att ta med mig jordgubbar. Jag har ett hemligt jordgubbsställe, men för att hitta dit behöver jag din hjälp! I rutan finns en instruktion för hur jag ska flyga för att hitta till jordgubbarna och tillbaka. Följ mönstret i rutan och visa mig vilka blommor jag ska flyga mellan. Tryck på den gröna knappen när du är klar, så att jag vet när jag ska börja flyga.”*



Figur 7. Instruktionsbrädan blinkar.

Figur 8. Den gröna knappen blinkar.

När meningen som beskriver rutan med instruktionen läses upp blinkar det grönt runt instruktionsrutan, detta går att se i figur 7. När sedan meningen som beskriver den gröna knappen läses upp blinkar det istället grönt runt denna knapp, se figur 8. När ljudinstruktionen är slut minskar biet i storlek och placeras i utgångsläget för spelet.

Spelplanen består av en bakgrund som ska föreställa en grässlätt med buskar vid horisonten och en ljusblå klar himmel. På gräset finns det sju olika blommor som är placerade i en cirkel, alla med olika färger och former. I mitten av cirkeln finns det en stor grå sten. Vid en av blommorna, den orange-röda blomman, finns det tre jordgubbar. På andra sidan om stenen, mitt emot där jordgubbarna finns, är en stor korg placerad som också är biets utgångsläge. Spelplanen kan ses i figur 9.

Figur 9. Biet berättar om programmeringsuppgiften.

Uppe i högra hörnet av spelplanen finns en bräda med en uppsättning av tre blomformer och en korg. Dessa utgör instruktionsrutan med blommönstret som spelaren ska följa. Till höger om rutan finns en grön knapp med en ”play-symbol” som spelaren ska trycka på när denne känner sig klar med uppgiften och vill att biet ska börja flyga.

Nere i högra hörnet av spelplanen finns en röd knapp med en soptunna-symbol, som är till för att rensa spelplanen ifall spelaren känner sig missnöjd och vill börja om.

När biet har förklarat speluppgiften och placerat sig vid utgångspunkten intill korgen, säger biet: *"Vilken blomma ska jag flyga till först? Tryck på den!"*.

Figur 10. En blomma är markerad.

Tanken är att spelaren, genom att titta på instruktionsrutan, ska följa blommönstret genom att i tur och ordning klicka på de motsvarande blommorna som finns ute på spelplanen. När spelaren trycker på en blomma dras det ett vitt streck till denne och den valda blomman får grönaktigt "glitter", se figur 10. Första gången spelaren trycker på en blomma kommer strecket dras från korgen till den valda blomman. Därefter kommer det vita strecket dras från den föregående valda blomman till den nyvalda blomman. Efter att en blomma är vald kommer följande ljudinstruk-tion: *"Vilken blomma ska jag flyga till nu? Tryck på den!"*

Figur 11. Alla blommor är markerade i rätt följd.

Spelet fortsätter genom att spelaren klickar på nästa blomma, ett nytt vitt streck dras till den nyvalda blomman med samma uppmaning som tidigare från biet. Detta fortsätter tills spelaren har slutit loopen genom att avslutningsvis trycka på korgen, se figur 11. Därefter trycker spelaren på den gröna knappen bredvid instruktionsrutan. Då kommer biet utbrista *"Nu flyger vi!"* och börja flyga till den första valda blomman. Är denna korrekt, i jämförelse med den angivna instruktionen på brädan, säger biet *"Jippi!"* och flyger vidare till nästa valda blomma, se figur 12.

Figur 12. Alla blommor är markerade i korrekt följd.

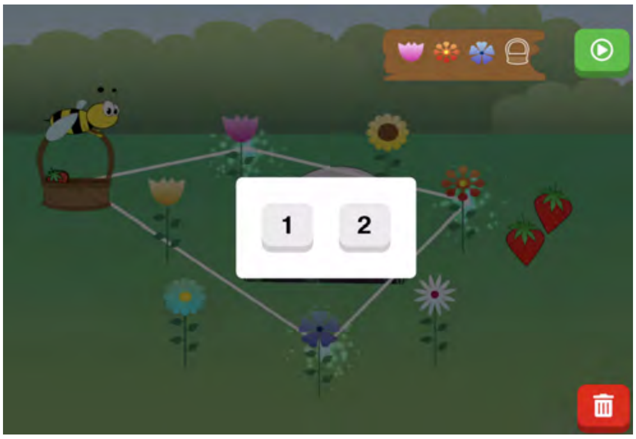
Skulle det vara så att spelaren valt att klicka på en blomma som inte är med på brädan (i instruktionsrutan) kommer biet, efter att spelaren tryckt på den gröna knappen, flyga i den ordning som spelaren markerat blommorna. När biet anländer till den första felaktiga blomma kommer biet säga *"Oj, här var en bugg, så det blev det fel... Vi får flyga tillbaka och försöka igen."*. Spelplanen kommer då automatiskt rensas på både blommarkeringar och dragna streck och spelaren får börja om från början.

Lyckas spelaren följa mönstret i instruktionsrutan korrekt och välja blommorna i rätt ordning fram till blomman med jordgubbarna, kommer biet plocka upp en jordgubbe vid den orange-röda blomman. När biet anländer till den blomma kommer biet säga *"Åh, här är jordgubbarna. Synd att jag bara orkar bära en i taget."* innan en jordgubbe plockas upp. Därefter kommer biet flyga vidare till nästa valda blomma. Figur 13 visar en bild på när biet håller i en jordgubbe och flyger tillbaka till korgen efter att ha besökt blommorna i rätt följd.

Figur 13. Biet plockat upp en jordgubbe och är på väg till korgen.

I det fall när spelaren har lyckats plocka upp en jordgubb men nästvalda blomma är felaktig, kommer biet säga *"Oj, här var en bugg, så det blev det fel... Nu får vi lämna tillbaka jordgubben och försöka igen"*. Biet kommer då inta utgångsläget vid korgen, jordgubben lämnas tillbaka och kommer synas vid den orange-röda blomman tillsammans med de andra två jordgubbarna.

När spelaren lyckats välja rätt blommor i korrekt följd samt plockat upp en jordgubbe och återvänt till korgen, lämnas jordgubben i korgen. Därefter får spelaren frågan *"Men*



Figur 14. Dialogrutan med två svarsalternativ.

det finns ju fortfarande jordgubbar kvar, hur många fler varv i loopen måste jag flyga för att plocka upp alla?". Som det framgår i figur 14 dyker det då upp en kostumiserad dialogruta med två knappar. På den ena knappen står det ”1” och på den andra ”2”. Spelaren ska då ta ställning till om biet behöver flyga ett eller två varv till för att plocka upp de resterande jordgubbarna. Med andra ord behöver spelaren ta ställning till om biet behöver göra en loop eller två loopar till.

Väljer spelaren att trycka på den högra knappen (”2”), kommer biet automatiskt flyga två varv till. Biet kommer flyga exakt samma runda som tidigare (blommönstret i instruktionsrutan). Varje gång biet passerar blomman med jordgubbarna kommer en jordgubbe plockas upp och lämnas i korgen. Efter det sista varvet kommer korgen ha tre jordgubbar och därefter är spelet avklarat.

Skulle spelaren istället trycka på den vänstra knappen (”1”) är tanken att biet flyger automatiskt ett varv. När biet sedan anländer till korgen och lämnar den andra jordgubben i korgen kommer biet konstatera att det fortfarande finns en jordgubbe kvar. Samma fråga kommer upprepas och spelaren kommer få två nya alternativ där det ena är korrekt (i detta fall ”1”) och det andra är felaktigt.

Tyvärr hanns det sistnämnda i stycket ovan inte implementeras inför användartesterna. Hur spelet egentligen hanterade de fall då fel alternativ valdes (ett varv), var att spelplanen nollställdes och spelaren var tvungen att börja om från början med att utföra speluppgiften.

5.2.1 Teknisk implementation

Spelet implementerades i React.js och Phaser.js som ett webbspel. React.js (React) är ett ramverk för webbapplikationer och Phaser.js (Phaser) är en spelmotor som kan köras i en webbläsare. Phaser är det som driver spelgrafiken, ljudet och animering av biet genom en typisk *game loop*-arkitektur som återfinns i de flesta spelmotorer.

De statiska elementen i användargränssnittet, till exempel de olika knapparna och instruktionsrutan, är byggda i React och kan kalla på Phaser-metoder för att ändra på spelets tillstånd.

För att underlätta distribution av spelet så användes Netlifly, en tjänst som automatiskt kompilerar kod från Github och laddar upp den senaste versionen av spelet till en server (https://mamn15.netlify.com/). Spelet kan alltså direkt spelas i vilken webbläsare som helst med en Internetuppkoppling.

5.2.2 Grafik

Verktyget som användes för att rita grafiken var Adobe Photoshop CS6. Alla delar i spelet, förutom den gröna och röda knappen, är designade och ritade förhand. En digital *moodboard* gjordes innan grafikritandet påbörjades, se bilaga C. Moodboarden bestod av olika bilder i syfte att visualisera idéer samt för att få en känsla för stämningen och konceptet som eftersöktes i spelet. Bilderna på moodboarden representerade olika färgkoncept, mönster och texturer som skulle vara utgångspunkten för designen av grafiken. Konceptet det strävades efter var att grafiken skulle efterlikna verkligheten med valet av färger och objekt, men vara färgstarkare, samt att färgskalan skulle vara relativ ljus. Gradienter av färger användes flitigt för att ge bakgrunden ett naturligt djup för att öka förståelsen av spelplanen för spelaren. Blomobjekten valdes att vara färgglada och inbjudande och jordgubbarna beslutades vara lite större och mer markanta.

5.3. Resultat från HiFi-testerna

I detta avsnitt kommer resultatet från de två användartesterna på spelets målgrupp presenteras gemensamt.

Testningen utfördes på totalt 16 elever i åldersspannet 4 till 6 år. 8 av dessa 16 elever utförde testet enskilt medan resterande gjorde testet i par. Testet utgick från testplanen som återfinns i bilaga D. Under testet fick eleverna spela igenom prototypen på en iPad. iPadens skärm spelades in under testsessionerna och ett testprotokoll fördes med observationer och kommentarer från eleverna. Detta för att lättare kunna utvärdera och analysera testningen. Maxtiden för varje enskilt test var satt till 15 minuter. I testprotokollet eftersöktes ett antal delmål. Dessa delmål var moment i spelet som ansågs påvisa hur väl eleven hade förstått uppgiften i spelet. Mallen för testprotokollet återfinns i bilaga E.

Innan testet påbörjades blev eleverna frågade ifall de hade använt en iPad tidigare. Det visade sig att majoriteten av deltagarna hade tidigare erfarenhet av en iPad, antingen genom skolan eller hemifrån.

Det första delmålet som eftersöktes i testprotokollet var huruvida testpersonen uppmärksammade spelets instruktionsruta. Instruktionsrutan är nödvändig för att klara spelet och det är därför av stor vikt att denna är tydlig och uppfattas av spelaren. Majoriteten av testdeltagarna la märke till instruktionsrutan. Flera av dessa försökte dock att trycka på blommorna i instruktionsrutan istället för på blommorna på spelplanen. Efter att de hade fått försöka ett tag fick de en ledande fråga av testledaren. Frågan som ställdes var om blommorna som de försökte trycka på i rutan fanns någon annanstans också. De allra flesta hittade då blommorna på spelplanen och klarade av att följa instruktionen. Några av dem som först uppmärksammade instruktionsrutan och klickade på den gick över till att trycka på godtyckliga blommor på spelplanen. De kopplade helt enkelt inte ihop instruktionen med spelplanen och fick därför lite hjälp efter en stunds testande. Några av testdeltagarna missade helt instruktionsrutan och behövde få denna förklarad för sig för att lyckas ta sig igenom spelet. Det fanns även några få testdeltagare som förstod direkt att de skulle hitta motsvarande blommor i instruktionsrutan på spelplanen.

Den sista instruktionen i rutan var korgen, vilket också var startpositionen där biet var placerad. Tanken med detta var att göra ett varv och sedan avsluta med korgen, med andra ord göra en sluten loop. Några av testdeltagarna fastnade på denna instruktion och förstod inte att de skulle tillbaka till korgen efter att de hade tryckt på den sista blomman.

Delmål 2 i testprotokollet uppnåddes i de fall eleverna förstod att de kunde trycka på blommorna. Ungefär en tredjedel av testdeltagarna förstod direkt att blommorna på spelplanen var klickbara. Resterande behövde en ledtråd från testledaren för att inse detta. Någon av eleverna försökte att med fingret dra ett streck från biet till en av blommorna istället för att klicka på blomman.

Flera av eleverna klickade på alla blommor i cirkeln i tur och ordning för att på detta sätt sluta loopen. Andra slöt loopen med godtyckliga blommor utan att passera jordgubbarna och missade således huvuduppdraget med spelet: att hjälpa biet samla in jordgubbarna. En annan variant som observerades under testet var att en deltagare enbart klickade på blommor fram till jordgubbarna och sedan ansåg sig färdig. Alla dessa varianter av att klicka på godtyckliga blommor var en följd av att instruktionen hade missats. Eleverna provade sig helt enkelt fram istället.

Delmål 3 i testprotokollet uppnåddes ifall eleverna förstod att de skulle trycka på blommorna enligt instruktionen. Tidigare nämndes det att en majoritet av eleverna uppmärksammade instruktionsrutan. Andelen av dessa som utan hjälp också förstod att blommorna på spelplanen var klickbara var väldigt få.

Om testpersonen förstod att denne skulle trycka på den gröna knappen för att få biet att börja flyga, uppnåddes delmål 4. Många av eleverna kom ihåg att trycka på den gröna knappen när de hade slutit loopen. Andra tryckte på knappen efter en ledande fråga från testledaren. Frågan som testledaren ställde var ”vad skulle man göra när man var färdig?”. Det förekom också att någon av testdeltagarna försökte trycka på den gröna knappen redan under den inledande ljudinstruktionen. En annan tryckte på knappen utan att ha slutit loopen.

Det sista delmålet avser om testdeltagaren förstod hur denne skulle svara på frågan om hur många loopar biet behövde flyga för att plocka upp de resterande jordgubbarna. Detta moment var det tveklöst flest som uppnåde utan hjälp. Någon behövde en ledtråd om att antal varv i loopen som skulle flygas motsvarade antalet jordgubbar som var kvar på spelplanen.

Alla testdeltagarna utom en fullföljde hela testet. Av de som fullföljde testet klarade alla av att ta sig igenom hela spelet med varierande mängd hjälp från testledaren. Några få förstod alla moment i spelet utan nästan någon hjälp eller ledtrådar över huvudtaget.

Efter att testdeltagarna spelat spelet fick de frågan om de tyckte att det var kul, samtliga testdeltagare svarade ja.

5.3.1 Pedagogens tankar om HiFi-prototypen

Under testsessionen på Maria Montesoriskolan gavs det tillfälle att prata med en av pedagogerna som deltog i testningen av LoFi-prototypen. Pedagogen fick prova HiFi-prototypen och delade med sig av följande tankar och kommentarer.

Pedagogen ansåg att spelet fick eleverna att träna på att se mönster som också är ett moment i läroplanen för matematik, vilket är bra. Vidare berättade pedagogen att spelet därför skulle kunna användas som ett bra kompletterande material ”som vilket annat Montessorimaterial som helst på hyllan”. Pedagogen ansåg också att spelet gav en bra möjlighet till att träna på att följa instruktioner.

En annan kommentar som gavs var att följande del ur den inledande instruktionen var något vilseledande: *Tryck på den gröna knappen när du är klar.*. Pedagogen menade att det var oklart vad ”klar” syftade på och att det inte riktigt framgick om det var loopen som skulle vara klar eller om spelaren skulle trycka på den gröna knappen när denne hade lyssnat klart på ljudinstruktionen. Däremot ansåg pedagogen att det vara bra att instruktionen inte berättade hur spelaren inte skulle göra och att ordet ”inte” ej förekom vid felfeedback.

Avslutningsvis tyckte pedagogen att spelet behövde vara längre och innehålla fler nivåer än i sin nuvarande utformning för att vara lärorikt, meningsfullt och inkluderas i skolundervisning.

6. Diskussion

Instruktionsrutan i Loopspelet är nyckeln till att kunna klara av spelet. Därför är det viktigt att denna omhändertas, designas och implementeras med försiktighet. Det finns förbättringspotential hos instruktionsrutans konstruktion samt hos presentationen för användaren. Hela instruktionsrutan hade kunnat vara större med större korg- och blomikoner. Blommorna hade kunnat förtydligas en aning med exempelvis en ljusare färg på brädan eller en svag skugga runt blomformerna. Detta för att de ska sticka ut från brädan lite mer än vad de gör i nuläget, men samtidigt inte missleda spelaren genom att framställa blomikonerna som tryckbara. Korgen i instruktionsrutan kunde också förtydligas då den kamoufleras med brädan utan en tydlig kant (outline). Den vita kanten användes för att få korgen att sticka ut från bakgrunden, men gjorde framställningen av instruktionen på brädan inkonsekvent. En lösning hade förslagsvis varit att sätta en smal vit kant även runt blomformerna.

Det är också intressant att diskutera ordvalet ”ruta” i ljudinstruktionen. Frågan är om det inte hade varit tydligare om instruktionsrutan istället kallats för *instruktionsbräda* eller *brädan med instruktioner*. Begreppet ”instruktionsruta” uppkom under LoFi-prototypen då den ritades på papper som en ruta. Vidare valdes under revideringen av LoFi-prototypen (se avsnitt 5.1) att instruktionsrutan skulle blinka när den omnämndes i ljudinstruktionen. En idé för att underlätta förståelsen med instruktionsrutan och instruktionen skulle vara att rutan blinkar när den omnämns, men i den efterföljande meningen (*Följ mönstret i rutan och visa mig vilka blommor jag ska flyga mellan*) blinkar blommorna inuti rutan istället, för att förtydliga blommornas (och korgens) funktion. Den blinkande färgen skulle också kunna ändras till en annan färg, exempelvis gul eller orange, för att dra till sig spelarens uppmärksamhet. Den gröna färgen som används i nuläget smälter in med spelplanens färgval. Dessutom utmärkte sig inte den gröna nyansen när den blinkar runt den gröna play-knappen. Det kan dock konstateras att även om den blinkande funktionen kan förbättras, så var det mycket bra att den fanns. Det var tydligt under testerna att när rutan och knappen blinkade tittade testpersonerna på rutan respektive knappen och testade att trycka på dem.

Som konstaterat under redogörelsen av testresultaten var det många som inte förstod hur de skulle hantera instruktionsrutan utan hjälp. Testpersonerna verkade ha svårt att koppla att instruktionerna var en bild eller karta av spelplanen och att blommorna i instruktionen också fanns ute på spelplanen. Flera testpersoner försökte trycka direkt på blommorna i instruktionen. När de fick lite hjälp var det dock en majoritet som förstod konceptet, vilket får ses som ett mycket bra resultat då spelet kommer att vara ett läroverktyg. Eftersom många verkade förstå vad de skulle göra efter ett tag, kan det konstateras att när testpersonerna väl förstått att instruktionen innehöll representationer av blommor snarare än de blommor denne skulle trycka på, var de kapabla att lösa uppgiften.

En del frågetecken dök också upp rörande den gröna knappen under användartesterna och kommenterades även av pedagogen. Symbolen på knappen kan missuppfattas som att spelet sätts igång när denne trycker på den. Tolkningen är i viss mån rätt, tanken är att biet ska börja flyga vid tryck på den gröna knappen. Ett förslag skulle vara att ersätta symbolen med en ikon på biet eller med ett par flaxande vingar. För att ytterligare förtydliga att knappen inte startar spelet utan får biet att börja flyga bör även ljudinstruktionen förändras. I nuläget får spelaren höra ”Tryck på den gröna knappen när du är klar, så att jag vet när jag ska börja flyga”, bara att undvika ordet ”klar” och istället säga exempelvis ”tryck på knappen när jag ska börja flyga” hade kunnat förtydliga dess funktion. Ett annat alternativ skulle vara att ta bort den gröna knappen helt och i ljudinstruktionen istället be spelaren trycka på biet (på spelplanen) när denne vill att biet ska börja flyga.

Vissa testdeltagare valde att trycka på andra blommor efter att de följt instruktionen, istället för att trycka på korgen. Det kan bero på otydligheten med korg-ikonen som tidigare beskrivits. Ett förslag som gavs från en testperson var att byta ut korgen mot ett bi, för att förstärka förståelsen för att spelaren ska återvända till utgångspunkten. Det kan också bero på ljudinstruktionen *"Vilken blomma ska jag flyga till nu? Tryck på den!"*. Efter att den sista blomman i instruktionen valts av spelaren blir det missvisande när ljudfeedbacken uppmanar spelaren att flyga till en annan blomma, när spelaren egentligen bör trycka på korgen för att sluta loopen. I detta fallet bör det övervägas om ljudinstruktionen helt ska skippas eller om den ska bytas ut mot en annan fras.

Visst trial & error-beteende kunde observeras under testningen. När en spelare inte hade förstått att denne skulle följa instruktionen var det vanligt att denne testade sig fram genom att trycka slumpmässigt på olika blommor. Det kan dock konstateras att ingen testperson lyckades klara spelet på detta sätt, vilket är en logiskt följd av att det finns väldigt många olika kombinationer. Det är inte lönsamt att tillämpa en trial & error-startegi i det här fallet, till följd av de många kombinationerna, utan det får antas att de testpersoner som provade sig fram gjorde detta eftersom de inte alls hade förstått vad de skulle göra. Det var alltså förmodligen inte en utarbetat trial & error-strategi.

En intressant iakttagelse som gjordes var att några av testpersonerna valde att trycka på alla blommor i tur och ordning runt den stora stenen för att sedan avsluta med korgen. På ett sätt visar det att det visuella hindret, den stora stenen, uppfyllde sin funktion med att spelaren inte ska välja ”snabbaste vägen” till jordgubbarna. Å andra sidan, som nämnt under resultatdelen från användartesterna (avsnitt 5.3), visar det att spelaren har missat instruktionsrutan och inte förstått speluppgiften till fullo. Att spelare testade att trycka på alla blommor i tur och ordning kan vara en följd av att blommorna var placerade i en cirkel, tanken var egentligen att blommorna skulle vara utplacerade över hela spelplanen och inte endast runt stenen. Det hade varit intressant att testa spelet med blommor över hela spelplanen, eftersom en spelare då förmodligen inte hade upplevt riktigt samma uppmaning att trycka på alla blommor i tur och ordning. Det kan ha uppfattats som att blommorna i cirkeln och instruktionsrutan var lite motsägelsefulla: å ena sidan finns instruktioner men å andra sidan är blommorna tilltalade att trycka på i en cirkel, bl.a. eftersom människor gillar att se slutna mönster.

Eftersom många testpersoner klarade spelet med lite hjälp borde möjligheten att implementera viss hjälp i spelet undersökas. Ljudinstruktionen i början kan göras om så att den innehåller mer av den informationen som testledaren under testerna behövde hjälpa tespersonerna med. Ett helt fungerande spel skulle också kunna ge ledtrådar liknande de som testledaren gav. Om en spelare t.ex. tycks trycka helt slumpmässigt på spelplanen kan instruktionsrutan börja blinka igen och spelet kan påpeka något i stil med ”kan du hitta blommorna i instruktionen på spelplanen?”. Kanske var den inledande ljudinstruktionen även lite för lång. Många testpersoner kunde komma ihåg att de skulle trycka på den gröna knappen när de var klara, vilket var det sista som sades i instruktionen. Det kan tyda på att de mindes det som sagts senast men inte riktigt det som sagts innan. Det var inte heller någon testperson som använde soptunna-knappen. Det kan bero på att den aldrig omnämndes i ljudinstruktionerna. Anledningen till att den inte gjorde det var eftersom den inledande instruktionen redan var lång. Vid en vidareutveckling av spelet hade det varit bra att introducera den röda soptunna-knappen. Ett förslag hade varit att instruktionerna eventuellt hade kunna komma stegvis, så att spelaren inte får all information på en och samma gång.

Eftersom nästan inga testpersoner hade några problem med att svara på hur många fler varv biet behövde flyga för att plocka upp resterande jordgubbar antas den implementerade lösningen för detta vara bra. Det kan dock finnas en poäng med att ha i åtanke att det i prototypen endast fanns två alternativ vilket därför gav spelaren 50% sannolikhet att gissa rätt. Testpersonerna valde dock ofta rätt alternativ utan att tveka, vilket ändå tyder på att de förstod vad de gjorde.

En annan liten förändring som skulle kunna göras är att byta ut biet mot något annat flygande djur, till exempel en fjäril. Detta för att spelet i mindre utsträckning ska jämföras med andra programmeringsspel som använder sig av bin, som exempelvis Bee-Bot.

Som en kommentar på testerna är det under användartester bra att ha med användare som inte har mycket erfarenhet inom området för att representera användare med låg kompetens som potentiellt kan använda systemet, i detta fall programmeringsspelet. På engelska kallas dessa testpersoner för *Least Competent Users* (LCU). I det här fallet är målgruppen barn i åldrarna 4-6 som ska lära sig samt få en introduktion till programmering med hjälp av spelet som körs på en iPad. Därav utgörs spelets LCU av de testpersoner som aldrig använt en iPad förut.

Avslutningsvis är det viktigt att ha i åtanke att spelet som utvecklats är ett pedagogiskt verktyg, ett stöd för pedagoger att inkludera i sin undervisning i syfte att utveckla barnens förståelse för programmering. I och med att det är ett nytt koncept och det var första gången som barnen kom i kontakt med spelet, gör det inget att testledaren fick hjälpa testdeltagarna under testerna. Barn behöver olika mycket hjälp och har olika inlärningskurvor. Loopspelet är inte till för att examinera barnen utan för att de ska lära sig och få ökad förståelse för programmering på ett lekfullt men lärorikt sätt. Därför kan det bli lite svårt att dra några exakta slutsatser utifrån testtillfällena då spelet var för kort för att barn ska hinna lära sig samt för de nya begreppen att sjunka in.

6.1. Utökad svårighetsgrad

För att bygga ut spelet till att innehålla flera svårighetsnivåer och introducera fler programmeringsbegrepp diskuterades några lösningar som t.ex. att spelaren ska följa två olika instruktioner och på så sätt skapa två olika loopar. Utöver att plocka jordgubbar kanske spelaren också behöver plocka blåbär och gör då två olika loopar som ska köras olika antal gånger. Här kan också begreppen *sekvens* och *algoritm* introduceras: spelaren blir presenterad för de två olika sekvenserna som resulterar i en algoritm. Att ta med detta i HiFi-prototypen ansågs dock ligga utanför projektets omfattning.

I Magiska trädgården finns en liten trollkarl som kommer och ”ställer till det” för spelaren ibland, genom att till exempel göra spelet lite svårare. Om Loopspelet skulle göras till en del av Magiska trädgården skulle trollkarlen kunna användas även här. Denne skulle till exempel kunna kasta om ordningen hos blommorna i instruktionen så att instruktionen förändras eller förändra antalet jordgubbar.

6.2. Kompletterande material

En intressant tanke som diskuterades under projektets gång var om programmeringsspel bör användas i förskolan tillsammans med något kompletterande fysiskt material till pedagoger. Materialets syfte skulle vara att ta digital programmering och göra det analogt. Det kan handla om lekar med programmeringstänk som utgår från exempelvis mönster och instruktioner eller förslag på i vilka sammanhang programmeringsbegrepp kan användas i barnens vardagsliv. Materialet ska uppmuntra till analog programmering som komplement till de digitala spelen, för att undvika att kunskaperna blir kontextberoende.

Att utveckla sådant material är utanför det här projektets omfattning, men det är värt att nämna inför framtiden. Utöver att det stärker barnens förståelse för programmering kan det också bidra till att pedagogerna får ökad förståelse för vad programmering innebär. Under diskussionen med pedagogerna nämnde samtliga att de själva kan ha svårt att förstå programmering och att det därmed blir en större utmaning att lära ut till barnen. Material som berör programmering i förskolan måste vara skrivet för pedagoger på ett sätt som inte avskräcker dem ifrån det. Det ska snarare uppmuntra till ett intresse som går att associera till det vardagliga livet och ses som en kompetensutveckling.

Som kompletterande material till Loopspelet skulle en analog spelplan kunna utvecklas, det skulle exempelvis kunna ta formen av en matta med olika blommor på. Till detta skulle en uppsättning lösa blommor medfölja, så att eleverna kan placera de lösa blommorna i olika mönster och låta sina kompisar följa mönstret genom att röra sig på mattan eller förslagvis dra en grantråd mellan blommorna. Med kompletterande instruktioner till pedagoger som introducerar enklare programmeringsbegrepp, samma som de som används i spelet, blir det en analog variant av Loopspelet. Detta kan i sin tur dels hjälpa barnen att få ökad förståelse för programmeringskonceptet och dels uppmuntra till samarbete och gemensam problemlösning.

6.3. Feedback

Arbetet med feedback har varit viktig under utvecklingen av både LoFi- och HiFi-prototyper. Den här sektionen beskriver och diskuterar tankarna bakom den tilltänkta och implementerade feedbacken under projektets båda iterationer.

6.3.1 Feedback i LoFi-prototyperna

I samtliga spelidéer har feedbacken utformats med målet att vara utvecklande. Eftersom spelen är tänkta att vara lärande har verifierande feedback undvikits. Feedbacken är utvecklande eftersom spelaren i alla de tre spelen får se vad som blir konsekvensen av deras felhandling: alla jordgubbar blir inte upphämtade, saften rinner fel, tårtan ramlar ihop. Den utvecklande feedbacken ligger i att spelaren får en ledtråd om vad som gick fel, genom att denne visuellt får se resultatet av sin handling. Om spelaren upprepade gånger misslyckas kan ytterligare ledtrådar ges, t.ex. genom att väsentliga delar av spelaplanen lyses upp.

Under designen av feedbacken har det varit viktigt att fånga programmeringskonceptet "gör först exekvera sen". Samtliga spel implementerar därför detta. Det är också ett sätt att undvika verifierande feedback och trial & error-strategier. Feedbacken är heller inte för uppmuntrande. Visserligen får spelaren en uppmuntrande kommentar när den gör rätt, t.ex. när biet passerar en blomma som är korrekt enligt instruktionen, men detta är för att påvisa att spelaren gjort rätt och är därmed relevant för spelet. Uppmuntrande feedback som inte har någon koppling till spelen finns inte. Spelaren får heller inga poäng på sin prestation, på så vis undviks resultatfeedback.

Tidigare under projektets gång var planen att feedbacken i Tårtspelet och Loopspelet skulle vara mer direkt. I Tårtspelet skulle bitarna falla på plats direkt då de lades ut på fatet och inte, som nu, falla på plats först efter det att alla bitar lagts ut och spelaren plingar på klockan. I Loopspelet skulle spelaren få feedback direkt vid klick på en blomma huruvida den var korrekt eller inte. Efter att ha diskuterat hur den typen av feedback skulle påverka dels möjligheten för spelaren att använda en trial & error-strategi, dels på vilket sätt detta demonstrerar "gör först exekvera sen" togs beslutet att förändra feedbacken. I båda fallen hade detta uppmuntrat till trial & error-beteende och programmeringsmetaforen hade gått delvis förlorad.

6.3.2 Feedback i HiFi-prototyperna

Samma tänk kring feedback som genomsyrade LoFi-arbetet har följt med in i HiFi-prototypen av Looopspelet. Under arbetet med HiFi-prototypen har feedbacken för lärande varit viktig, men också den feedback som gör att spelaren förstår spelet har diskuterats. Exempel på sådan feedback är det blinkande, gröna ljuset runt olika vitala delar av spelaplanen under spelets introduktion samt "glittret" runt markerade blommor. Detta är feedback som hjälper spelaren orientera sig i spelet. Glittret runt blommorna är ett designbeslut som tillkommit under HiFi-arbetet, eftersom det ansågs att endast streck mellan blommorna (som var orginalidéen) inte var tillräcklig feedback. Även ljudfeedback har varit viktigt för HiFi-prototypen, men som tidigare diskuterat bör det övervägas igen hur denna feedback ska utformas för att inte förvirra spelaren. Fallet då biet säger *"Vilken blomma ska jag flyga till nu? Tryck på den!"*, när det i själva verket ska flyga tillbaka till korgen, är ljudfeedback som borde designas om baserat på testresultatet.

Redan under LoFi-arbetet fokuserades det mycket på hur feedbacken skulle bli utvecklande. Det visade sig under testerna att visst trial & error-beteende fanns hos testpersonerna, men ingen lyckades klara spelet med denna strategi varför det kan konstateras att det är kostsamt att lösa spelet på detta sätt. Dock bör feedbacken då en spelare sluter loopen på fel sätt utvecklas, eftersom många inte förstod direkt varför de gjorde fel utan testade igen och gjorde fel även denna gång. Till exempel skulle blommorna i instruktionsrutan som spelaren lyckats välja korrekt kunna lysa upp när biet passerar dessa, för att visa för spelaren att det finns en koppling mellan spelplanen och instruktionsrutan.

Vidare finns utvecklande feedback då spelaren ska välja antal varv som biet ska flyga i loopen för att hämta upp alla jordgubbarna. Om spelaren väljer för få antal varv kommer biet att flyga det valda antalet varv och hämta upp en jordgubbe per varv. När biet sedan är färdigt kommer spelaren se att det fortfarande finns jordgubbar kvar och kan då ange igen hur många varv till biet måste flyga för att hämta upp de resterande jordgubbarna. Spelaren kommer alltså att se konsekvensen av sitt fel och förhoppningsvis förstå varför det blev fel. Tyvärr hann inte detta att implementeras inför testerna och därför finns det inte något testresultat som understödjer hur väl feedbacken fungerar i sammanhanget.

Figur 15. Exempel på en teachable agent i Loopspelet.

6.4. Teachable agent

Att implementera en teachable agent i HiFi-prototypen ansågs ligga utanför projektets omfattning, men på vilket sätt en teachable agent skulle kunna finnas med i spelet diskuterades. Det konstaterades att detta skulle kunna göras på liknande sätt som i Magiska trädgården: att den digitala eleven efter ett tag kommer in i spelplanen och tittar på för att sedan försöka själv. Den digitala eleven kan föreslå vilken blomma som borde vara nästa och spelaren får då ta ställning till om detta är korrekt eller ej. Figur 15 illustrerar hur det hade kunnat se ut med en teachable agent i Loopspelet, där karaktären Panders från Magiska trädgården är den digitala eleven.

6.5. Felkällor

Under båda iterationerna finns det intressanta iakttagelser som det behöver tas hänsyn till i denna rapport. Dessa iakttagelser kan ses som felkällor som både beror på misstag under projektet och observationer från de olika testerna.

6.5.1 Felkällor från första iterationen

Eftersom det totalt endast var fyra pedagoger, från två olika skolor, som medverkade i testsessionerna blir de insamlade enkätsvarens värden inte statistiskt signifikanta och därmed är variansanalys inte relevant i detta sammanhang. Det var dock känt från början och fanns i åtanke vid insamlingen och utvärderingen av resultaten. Både enkäterna och intervjufrågorna användes för att bygga en kvalitativ analys kring de aspekter där feedback efterfrågades.

Det är svårt att testa pappersprototyper på barn och därför har detta inte ingått som ett moment i projektet. Inga slutsatser kan därför dras om vad målgruppen tycker om spelen och därmed är det också svårt att avgöra hur spelens svårighetsgrad bör vara. Gruppen valde att analysera spelkoncepten utifrån pedagogernas inputs och erfarenheter.

Återkopplingen från pedagogerna som har samlats in och analyserats introducerar två bakomliggande faktorer som kan leda till missvisande fakta: pedagogernas gissning på vad de tror att barn tycker om, som de nödvändigtvis inte behöver göra, samt att det finns en risk att pedagogerna eventuellt inte svarade helt ärligt på enkätfrågorna eftersom de kanske ville undvika att kritisera projektet.

I enkäten som pedagogerna fick fylla i var de två första frågorna ställda som negationer medan resten av frågorna inte var det. Det kan diskuteras huruvida detta är bra eller dåligt, men då enkäten konstruerades togs inga beslut om hur resultatet skulle sammanräknas, vilket gjorde att presentationen av resultatet delades upp mellan de olika grupperna av frågor (tabell 2). Resultatet hade istället kunnat presenteras som en totalsumma om värdet av negationerna räknats som sitt motsatsvärde, vilket hade underlättat presentationen.

6.5.2 Felkällor från andra iterationen

Under andra iterationen kan felkällorna delas upp i två delar: faktorer relaterade till implementationen av HiFi-prototypen samt faktorer relaterade till själva användartesterna.

En betydande felkälla var att all funktionalitet inte var implementerad inför testet. Om spelaren trycker på den gröna knappen utan att ha valt någon blomma, borde spelaren blivit uppmanad att titta på instruktionsbrädan samt trycka på en blomma. Feedback borde också ges i de fall då spelaren inte lyckats sluta loopen men ändå tryckt på den gröna knappen, detta exempelvis i form av att biet flyger fram till blomman och säger något i stil med att det har blivit fel.

I de fall då spelaren följt blommönstret korrekt fram till och med blomman vid jordgubbarna men sedan valt fel blomma efteråt, kommer biet inte plocka upp en jordgubbe. Istället flyger biet förbi den orange-röda blomman, som om det vore en ”korrekt” blomma, och vid den felaktiga blomman kommer spelet uppföra sig som det är tänkt att göra vid felvald blomma.

Om en spelare väljer fel antal varv då denne ska ange hur många fler varv biet behöver flyga i loopen, rensas spelplanen istället för att spelaren får försöka igen efter det att biet flugit det angivna (för få) antalet varv. Samtliga faktorer är relaterade till implementationen och kan ha påverkan på resultatet. Om spelarna hade fått korrekt feedback i samtliga fall hade utfallet av testerna kunnat vara annorlunda.

Det finns även en del felkällor att ta hänsyn till från användartesterna. Användartesterna valdes att presenteras gemensamt då målgruppen var barn i åldrarna 4-6 år. Det visade sig dock under testsessionerna att det fanns en skillnad i kunskap, erfarenhet och beteende mellan barnen på förskolan och barnen i förskoleklassen, eftersom det skiljer upp emot två år mellan de yngsta och de äldsta deltagarna.

Det var första gången som någon i projektgruppen utförde användartester på en så ung målgrupp. Det som främst observerades var att ju yngre barnen var, desto mer tystlåtna, blyga och försiktiga var de. Detta försvårade en aning för testledaren och antecknaren då det blev svårare att notera testdeltagarens resonemang och åsikter. Barnen i förskoleklassen var dock mer kommunikativa före, under och efter testerna. Dessutom samlades mycket lite data in om testdeltagarnas bakgrund och erfarenhet. Den enda bakgrundsdata som samlades in var ålder samt om de tidigare hade använt en iPad. Med andra ord togs det inte hänsyn till om en testperson t.ex. led av färgblindhet. Inte heller data om testerpersonernas eventuella tidigare kunskap och erfarenhet om programmering samlades in. Detta ledde till att det blev svårare att avgöra om spelet var på en rimlig nivå eller ej.

Istället för att enbart fråga testdeltagarna om de tyckte spelet var roligt efter genomfört test, kunde också frågor ställts om spelet för att se om de upplevde att de lärde sig något nytt. Till exempel kunde frågor berörande begreppen i spelet ställts, för att se om de observerat programmeringsbegreppen som sades under spelets gång och om de förstått vad de innebär. Å andra sidan var många testpersoner som redan nämnt ganska blyga och försiktiga, vilket gjorde det svårt att ställa komplexa frågor. Om fler liknande tester skulle genomföras hade det till exempel varit bra att besöka skolan eller förskolan flertalet gånger och bekanta sig med eleverna, så att de känner sig mer trygga vid själva testet. Då hade det kanske öppnats upp för mer utvärdering.

Att testpersonerna inte kände sig helt trygga och bekväma påverkade troligtvis hur de svarade på frågan om de tyckte att spelet var kul. Alla som testade spelet svarade att de tyckte att det var kul, men detta resultat bör inte få väga allt för mycket då många deltagare svarade detta utan eftertanke och på ett tveksamt sätt.

Som tidigare nämnt var vissa av testpersoner LCU. I rapporten presenteras dock inte hur dessa personer lyckades navigera i systemet och förstå speluppfiften i jämförelse med de som inte var LCU. Detta eftersom ingen markant skillnad noterades, vilket kan vara en följd av att testpersonerna inte var så många.

6.6. Framtida arbete

För att Loopspelet ska bli komplett behöver dels de diskuterade förändringarna göras och dels behöver de ännu inte implementerade funktionerna fixas. Vidare behöver spelet byggas ut så att de består av flera nivåer, så att det kan introducera fler begrepp och upplevas som meningsfullt.

Det hade också varit intressant att göra fler, omfattande tester. Ett förslag är att till exempel låta barn i en förskoleklass spela Loopspelet samt några andra lärspel och sedan låta dem på egen hand välja vilket av de testade spelen de vill spela. På så sätt skulle det gå att på ett mer objektivt sätt mäta om de tycker att Loopspelet är roligt och ifall de väljer det framför andra spel.

Sedan hade också en utvärdering som konstaterar om eleverna lär sig något på spelet eller inte varit intressant. Hur en sådan skulle utformas är inte aktuellt att spekulera i nu, men det hade varit betydelsefullt för spelets framtid.

6.7. Grupprocessen

Arbetet med grupprocessen inleddes med en gemensam lunch vars syfte var att skapa sociala band inom gruppen. På det första mötet diskuterades lite hållpunkter och gemensamma regler inför arbetet, kring vilka alla var överens. Efter en fjärdedel av tiden hölls en liten session då det diskuterades hur alla medlemmar i gruppen upplevt arbetet dittills och om någonting fungerat bättre eller sämre. Likadant efter halva tiden, då diskuterades tankar inför det fortsatta arbetet med HiFi-prototypen (andra iterationen).

Under andra hälften av projektet hade gruppens medlemmar svårare att hitta gemensamma träffar och valde därför att dela upp arbetet mellan sig i större utsträckning än under första iterationen. Detta resulterade även i att de gånger gruppen sågs var fokus att arbeta effektiv så det gavs inte mycket utrymme åt det sociala. Det fungerade väldigt bra att dela upp arbete mellan gruppmedlemmarna eftersom arbetet under första iterationen i stor utsträckning skett i grupp, så alla hade mycket bra insikt i vad som skulle göras och vad som skulle uppnås. Några gemensamma kaffepauser och en gemensam lunch hölls dock för att jobba med det sociala inom gruppen under andra iterationen.

7. Slutsats

Avslutningsvis kan det konstateras att Loopspelet är långt ifrån komplett. Det behöver förfinas ytterligare basert på testresultaten: feedbacken behöver utvecklas och i vissa fall designas om, svårighetsgraden behöver anpassas bättre och spelplanen behöver till viss del förändras. Vidare behöver, som en av pedagogerna också konstaterade, spelet ha mycket större omfattning för att bli meningsfullt. Det behöver byggas ut till att bestå av fler nivåer och svårighetsgrader, så att det ges mycket tid och utrymme för spelaren att faktiskt förstå de introducerade begreppen. Även kompletterande material bör utvecklas för att både elever och pedagoger ska få ut så mycket som möjligt av spelet.

Testerna av både LoFi- och HiFi-prototyper fick mycket bra respons och gav bra resultat, men det bör också has i åtanke att testerna skett i mycket liten skala och att deras tillförlitlighet inte är jättehög. De har varit en mycket bra utgångspunkt för det här projektet men de skulle behöva utföras i större omfattning vid en vidareutveckling av Loopspelet.

Dock har de resultat som kommit ur projektet varit både intressanta och get insikt i hur ett programmeringsspel för barn skulle kunna se ut. Slutsatsen efter genomfört projekt blir därför att det absolut är möjligt att utveckla lärande spel om programmering för barn i åldrarna 4 till 6 år, men det behöver ta sin tid och måste grundas på forskning och omfattande tester.

Projektet har varit mycket lyckat och kan förhoppningsvis ligga till grund för framtida utveckling av programmeringsspel för barn. Som redan konstaterat är Loopspelet långt ifrån ett färdigt spel, men det har kommit en bra bit på vägen för att undersöka möjligheter inför framtiden. I takt med att programmering blir en allt större del av samhället ökar efterfrågan på den här typen av läromedel, varför projektet kan spela en viktig roll inför framtiden.

Referenser

[1] A. Munches, C. O'Malley, S. Benford, ”The role of physical representations in solving number problems: A comparison of young children's use of physical and virtual materials”, 2010. Computers & Education, DOI: 10.1016/j.compedu.2009.09.023.

[2] H. Hessling, H. Paues & C-M. Vikingsson, ”Stärkt digital kompetens i läroplaner och kursplaner”, *Regeringen*, mars 2017. [Online] https://www.regeringen.se/pressmeddelanden/2017/03/ starkt-digital-kompetens-i-laroplaner-och-kursplaner/. [Hämtad: 29 oktober, 2018]

[3] H. Hessling, ”En förtydligad läroplan ska höja kvaliteten i förskolan”, *Regeringen*, augusti 2018. [Online] https://www.regeringen.se/pressmeddelanden/2018/08/ en-fortydligad-laroplan-ska-hoja-kvaliteten-i-forskolan/. [Hämtad: 29 oktober, 2018]

[4] ”Digital kompetens i förskolan”, *Skolverket*, [Online] https://www.skolverket.se/temasidor/digitalisering/ forskolan. [Hämtad: 29 oktober, 2018]

[5] K. Jäverbring, "Förståelsen börjar i det lilla, ”Förskoletidningen”, vol. # 6, 2017. [Online] http://www.forskoletidningen.se/6-2017/forstaelsen-borjar-i-det-lilla. [Hämtad: 29 oktober, 2018]

[6] F. Heintz & L. Mannila, ”Datalogiskt tänkande för svenska grundskolan – vad, hur och varför?”, *Linköpings universitet*, mars 2015. [Online] https://www.ida.liu.se/ frehe08/FredrikHeintzLinda Mannila-skolverket-20150302.pdf. [Hämtad: 29 oktober, 2018]

[7] F. Heintz, ”Datalogiskt Tänkande i skolan”, *Linköpings universitet*, september 2015. [Online] https://www.ida.liu.se/ frehe08/DatalogisktTankande-Skolverket-20150914.pdf. [Hämtad: 29 oktober, 2018]

[8] B. Tärning, ”Review of feedback in digital applications - does the feedback they provide support learning?”. *Design of teachable agents and feedback in educational software*. Diss., Lunds universitet. Lund: Media-Tryck, 2018, s. 5-18, 43, 47-48.

[9] D. Norman, *The design of everyday things - Revised and expanded edition*. Basic Books: New York, 2013, s. 23-25.

[10] K. Sönnerås, *Programmering i förskolan*. Stockholm: Dimograf, Polen, 2017, s. 24-29.

[11] Gulz, A., Kjällander, S., Frankenberg, S. & Haake, M., ”Early Math in a Preschool Context: Spontaneous extension of the digital into the physical”, (opublicerad artikel utdelad av handledare), 2018.

Bilagor

Bilaga A

Enkät

1. **Det var svårt att förstå instruktionerna**
   Markera endast en oval.

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

2. **Det var svårt att utföra uppgiften**
   Markera endast en oval.

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

3. **Jag upplevde att jag genomförde uppgiften väl**
   Markera endast en oval.

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

4. **Jag upplevde spelet som intuitivt**
   Markera endast en oval.

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

5. **Jag upplevde spelet som meningsfullt**
   Markera endast en oval.

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

6. **Jag upplevde spelet som relevant i utbildningssyfte**
   Markera endast en oval.

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

7. **Jag upplevde spelet som roligt**
   Markera endast en oval.

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

Bilaga B

Intervjufrågor

Arbetar ni med programmering? Hur? Vad är elevernas inställning till detta?

Vad är era tankar om spelidéerna vi presenterat?

Vad tror ni om spelens svårighetsgrad?

Vad var enkelt/svårt att förstå? På vilket sätt?

Upplever ni att spelet har någon programmeringskoppling?

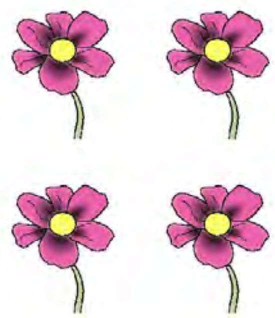
Fanns det något i spelet som du upplevde inte passade in?

Upplevde du spelets tema som passande i utbildningssyfte?

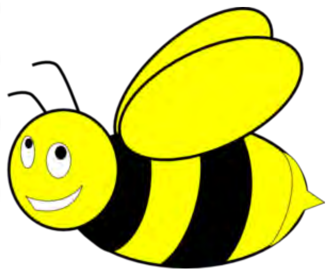
Tror ni att era elever skulle uppskatta den här typen av spel?

Bilaga C


LANDSCAPE
BANNERS
LANDSCAPE BANNERS
designed by freepik.com




shutterstock



Bilaga D

Testplan

Syfte

Syftet med testet är att undersöka om spelprototypen är intuituv, om målgruppen förstår den och om spelets svårighet ligger på rimlig nivå. Resultatet kommer ligga till grund för utvärdering av prototypen och eventuell vidareutveckling av spelet.

Urval av testpersoner

Testet kommer att genomföras vid två tillfällen. Vid varje tillfälle kommer 4 test att utföras varav 2 på enskilda individer och 2 parvis, d.v.s. med 2 testpersoner närvarande. Totalt 6 personer kommer alltså att delta per testtillfälle. Ett testtillfälle kommer att hållas i en förskoleklass och ett testtillfälle kommer att hållas på en förskola. Testpersonerna ska vara i åldrarna 4 till 6 år.

Rollfördelning

På varje test medverkar:

- 1 testledare (konverserar med testpersonen, ställer frågor)
- 1 protokollförare (skriver ner observationer och svar på frågor)

Material

- iPad
- Testprotokoll
- Penna

Testuppgifter

Uppgift	*Deluppgifter*	*Korrekt slutförd*	*Maxtid*
Spela spelet	1. Se instruktionerna 2. Trycka på blommorna 3. Trycka på blommorna enligt instruktionerna 4. Få biet att börja flyga 5. Svara på hur många varv biet ska loopa	Testpersonen klarar spelet	15 min

Genomförande

1. Starta screen recorder
2. Startat spelet och lämna plattan till testpersonen
3. Observera och för protokoll under testet
4. Ställ frågor till testpersonen
    a. Var det ett roligt spel?
    b. Vad var roligast?

Bilaga E

Testprotokoll

TestID: ______
Antal testpersoner: ______
Ålder på testperson(er): ______

Huvudmål:

- ❏ Testpersonen klarar spelet

Delmål:

- ❏ Testpersonen ser instruktionerna
- ❏ Testpersonen förstår att denna ska trycka på blommorna
- ❏ Testpersonen förstår att denna ska trycka på blommorna enligt instruktionerna
- ❏ Testpersonen förstår hur denna gör för att biet ska börja flyga (trycka på grön knapp)
- ❏ Testpersonen förstår hur denna ska svara på fråga om hur många loopar biet ska flyga

Kommentarer och observationer:

Var det ett roligt spel? Vad var roligast?

Balkenius, C., Gulz, A., Haake, M., & Wallergård, M. (Eds.) (2019). Intelligent, socially oriented technology IV: Projects by teams of master level students in cognitive science and engineering. *Lund University Cognitive Studies*, *175*.