

Intelligent, socially oriented technology II:

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

Editors

Christian Balkenius

Agneta Gulz

Magnus Haake

Mattias Wallergård

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

Intelligent, socially oriented technology II: Projects by teams of master level students in cognitive science and engineering. Lund University Cognitive Studies, 162.

ISSN 1101-8453

ISRN LUHFDA/HFKO—5070--SE

Copyright © 2015 The Authors

Table of Contents

Magical Islands: Mer lek – mer lärande?	5
<i>Celie Gunnarsson, Johan Mattsson, Aron Söderling</i>	
Virtuell miljö vs. Verkligheten: Choice Blindness-experiment utformat för en virtuell miljö med en virtuell agent.	27
<i>Emil Boman, Martin Lingonblad, Ludvig Londos, Arvid Nilsson</i>	
Cognition, Feedback and Educational Software Mathematics	41
<i>Ryo Yamazaki, Andreas Stephens, Hannes Johansson, Johanna Petersson</i>	
Oculus Rift: Virtual green environments for elderly	65
<i>Alexander Badju, Elise Ellerstedt, John Larsson, Joel Olofsson</i>	
Digitala läromedel: Feedback för 2000-talet	81
<i>Axel Duvebäck, Line Flank, Fredrik Lindholm, Christoffer Toft</i>	
1st Grade Students Learn Number Sense with Math in Space	93
<i>Per Grahn, Charlotte Leifland, Kristoffer Leo</i>	
The effect of gesture-speech lag on comprehension	113
<i>Sarah Alomari, Andrey Anikin, Joakim Bonnevier</i>	
Quieasy- ett digitalt testverktyg för grundskolan	123
<i>Bernt Christensen, Kalle Palm, Niklas Sjöberg, Olga Fredriksson</i>	
Project MonZ	133
<i>Rebecka Alves-Martins, Frida Bredberg, Niklas Grahl, Felix Hedlund, Carl Larsson Danell</i>	
Behavior Emulating Non-humanoid Decision Enacting Robot – an implemented cognitive model of social attention	149
<i>Thobias Bergström, Erik Jansson, Tim Nilsson</i>	
Animating natural robotic movements	159
<i>Asas Innab, Bjarni Runar Birgisson, Trond Arild Tjøstheim</i>	
Hur tittar människor?	175
<i>Robert Bagge, Olof Knape, Henrik Siljebråt</i>	
Människa-Robotinteraktion I Virtuella Miljöer	191
<i>David Lundberg, Lolita Olesen, Caroline Schouenborg, Eric Wall</i>	
Do you Trust me NAO? Measuring the effect of anthropomorphic behavior on seniors' trust in a social robot	203
<i>Amanda Bjernstedt, Mikael Gråborg, Josefin Karlsson, David Sivén</i>	

Magical Islands

Mer lek – mer lärande?

Celie Gunnarsson, Johan Mattsson, Aron Söderling

Magical Garden är ett digitalt lärospel i matematik med tre- till femåringar som målgrupp. Spelet är utvecklat av Educational Technology Group vid Lunds och Linköpings universitet. Magical Gardens styrka är att det implementerar många starka pedagogiska tekniker: lära genom att lära ut med hjälp av så kallade teachable agents, adaptiv svårighetsgrad genom ett smart system och ett signaleringssystem som hjälper läraren upptäcka barn som halkar efter etc. Uppgiften i detta projekt har varit att bygga vidare på Magical Gardens grund med inriktning på en ny målgrupp, barn med mental utvecklingsålder på sex till åtta år. Uppgiften medför många problem och frågor: Vilka matematiska uppgifter bör finnas med? Behöver spelet ett nytt utseende? Finns det något vi kan göra bättre än Magical Garden? Projektet har gått ut på att lösa dessa frågor genom en iterativ designprocess. Processen har innehållit mycket brainstorming och idégenerering samt testning för att verifiera idéer och designbeslut. Projektet har resulterat i speliden Magical Islands.

Magical Islands utspelar sig i ett tropiskt ölandskap där karaktärerna från Magical Garden blivit strandsatta efter att deras flygplan tvingats nödlanda under semesterresan. Spelarens uppgift blir sedan att samla ihop flygplansdelar från de olika öarna genom att hjälpa den lokala befolkningen. Utöver att skapa en uppföljare till det föregående Magical Garden har det även tillkommit en ny frågeställning. Vad händer om man inför inslag av lek i lärandet? Kan lekmoment användas som motiverande belöning efter avklarade matematikuppgifter eller blir dessa snarare en onödig distraktion?

1 Inledning

Syftet med Magical Islands är att vidareutveckla det befintliga Magical Garden för att anpassa det till en äldre målgrupp, 6-8 år. Med ålder menas i sammanhanget mental utvecklingsnivå, alltså den nivå som förväntas av en elev i angiven kalenderålder. Detta kan exempelvis innebära en femåring som är på en sjuårings nivå eller en nioåring som är på en sexårings nivå. Spelet är tänkt att ge en djupare matematikförståelse än föregångaren Magical Garden och fungera som en uppföljare. De gamla fokuspunkterna i formativ feedback, *teachable agents* och *scaffolding* ska bevaras så att utläringen sker med samma metoder som tidigare. Eftersom spelet är designat för äldre barn krävs det ny och svårare problemlösning där nya matematiska begrepp introduceras. En central del i spelet är att bygga vidare på det *number sense* som

barnen har börjat få ifrån det första spelet och få det att övergå i mer matematisk förståelse. När de har lärt sig hur siffror ser ut, deras betydelse och inbördes relation så är nästa steg att bli bekväm i att manipulera dessa siffror för att få fram en lösning till en uppgift.

Arbetet med lärospelet leds av funderingarna om hur stora lekmoment som kan läggas in med en bibehållen nivå av lärande för barnen lär sig, och i vilken grad lekmoment blir distraherande istället för motiverande.

2 Bakgrund

Det finns en mängd forskning inom området med lärospel. En del av dessa har granskats och analyserats för att skapa en bra grund för projektet.

Spel som läromedel

Användandet av spel i utbildningssyfte är inget nytt som digitaliserandet av samhället fört med sig; spel användes redan på 1700-talet för att träna militär taktik (Linderoth, 2012). Idag är digitala spel i utbildningssyfte både ett växande forskningsområde och en växande industri, men det finns också en stor övertro på dessa spel (Linderoth, 2014). Spel har bevisats ha en motiverande kraft, men det är väldigt svårt att fånga detta i spel med pedagogiskt syfte. Det är även svårt att veta hur ett lärospel ska se ut för att utnyttja den motiverande kraften i spel (Garris et al., 2002). Ett digitalt spel som används som läromedel ger inget mervärde bara för att det är digitalt. För att spelet ska ge något pedagogiskt värde utöver en vanlig lärobok måste det digitala mediet utnyttjas till att implementera kraftfulla pedagogiska tekniker som är svåra eller omöjliga att använda utan digital hjälp; *bok-på-burk* är alltså inte något att eftersträva (Gulz & Haake, 2014).

Regler

Regler är något som varje spel måste ha, varje situation i ett spel kan bara ha ett begränsat antal möjliga interaktions-möjligheter. Detta kan utnyttjas i pedagogiskt syfte. Ett spel kan implementera regler som begränsar elevens möjliga interaktion till just det som är viktigt. Exempelvis kan ett matematiklärospel tvinga eleven lösa uppgifter innan det går att komma vidare. På detta sätt kan elevens uppmärksamhet styras till det som eleven ska fokusera på (Linderoth, 2014). Det räcker dock inte att reglerna tvingar fram situationer där eleven måste lösa uppgifter, de måste även motverka att

eleven kan lösa uppgifterna på sätt som inte ger något pedagogiskt värde. Ett tydligt exempel på detta är om eleven själv kan anpassa svårighetsgrad eller om det är enklare att lösa uppgifter med *trial-and-error* (att bara prova sig fram tills rätt svar hittas) än att faktiskt räkna ut rätt svar (Linderoth, 2014). Detta problem kan lösas med bra feedback och adaptiv svårighetsgradsanpassning.

Feedback

För att motverka *trial-and-error* samt ge eleven större förståelse för hur och varför något fungerar är det viktigt att ett spel ger bra feedback. När eleven har gett sitt svar på en uppgift räcker det inte att bara säga "rätt" eller "fel". Denna typ av feedback kallas summativ och ger endast eleven bekräftelse på om det tänkta svaret är just rätt eller fel utan någon ytterligare vägledning eller konsekvens. Summativ feedback används ofta i digitala lärospel idag (Pilner Blair, 2013). Detta är en stor svaghet då summativ feedback uppmuntrar eleven att testa sig fram tills rätt svar hittas. Vid användning av denna typ av feedback lär sig eleven, i många fall, inget mer än att memorera rätt svar. För att eleven ska ta till sig kunskapen och verkligen lära sig förstå problemet måste feedback ges på ett sätt som förklarar varför något är rätt eller fel. Detta går att göra på ett antal sätt, eleven kan ges feedback som säger om det blev för mycket eller för lite, alternativt feedback som visar en faktisk konsekvens i spelvärlden som uppkommer på grund av det felaktiga svaret. Detta kan kommuniceras med exempelvis vatten som rinner över ur en bägare, en boll som rullar för långt på grund av för mycket kraft eller liknande.

Den feedback som ges efter ett felaktigt svar bör vara informativ och syfta till att eleven ska utvidga sina förmågor och därmed kunna lösa mer avancerade uppgifter i framtiden. Detta kallas formativ feedback eller förbättringsorienterad feedback.

Adaptiv svårighetsanpassning och scaffolding

Formativ feedback är ett bra första steg för att stödja elevens lärande, men att bara ge bra feedback räcker inte. Skulle spelets svårighetsgrad vara för hög skulle bra feedback inte räcka för en svagare elev; här krävs att svårighetsgraden anpassas sig efter elevens förmåga. Detta kallas adaptiv svårighetsgradsanpassning. Väl anpassad svårighetsgrad har visat sig ha positiv effekt på elevens motivation. Är det för lätt tappar eleven intresse och är det för svårt tröttnar eleven. Faktum är att adaptiv förändring av andra element såsom handling, design och uppgifts-utformning i ett spel också visat sig ge positiv effekt på elevens motivation (Sampayo-Vargas et al. 2013).

Med adaptiv svårighetsgrad kan spelets uppgifter hela tiden ligga på gränsen att eleven klarar av dem, med lite hjälp. Denna gräns brukar kallas *Zone of proximal development* (Vygotsky, 1978). Här är det egentligen lite för svårt för eleven själv, men med lite hjälp kan eleven klara av det. Det är just vid denna gräns som eleven lär sig som mest, men detta fungerar såklart endast om eleven faktiskt ges den hjälp som krävs för att klara uppgifterna. Denna hjälp brukar kallas scaffolding (svenska: byggnadsställning). Scaffolding kan till exempel ges med visuella ledtrådar, muntlig guidning eller

direkt hjälp från en lärare. I ett digitalt spel kan scaffolding till exempel implementeras genom färgkodning, ett intro som visar hur man gör, en guidande karaktär etc. I takt med att eleven klarar uppgiften kan hjälpen plockas bort och slutligen klarar eleven av uppgiften helt utan scaffolding och kan gå vidare till nya problem.

Det har även visat sig att ett icke-adaptivt system där eleven själv får välja svårighetsgrad inte är någon bra idé, då eleven kan välja den enklaste nivån hela tiden för att lättare att klara av att ta sig framåt i spelet (Linderoth, 2014). Detta leder till att elevens inläring begränsas då uppgifterna inte är tillräckligt utmanande.

Learning by Teaching

En annan styrka digitala spel erbjuder är dess karaktärer. Den karaktär som representerar spelaren i spelvärlden brukar kallas avatar. En avatar kan skilja sig drastiskt från eleven själv. Allt från genus, utseende, ålder till påhittade fantasiegenskaper kan vara attribut som skiljer hos avataren. Studier av Bailenson et al. visar att en avatars utseende kan påverka en spelares beteende (Bailenson et al. 2009). Blir spelaren exempelvis tilldelad en lång avatar kan denne uppträda på ett mer självsäkert sätt (Gulz & Haake, 2014). Studien visar också att beteende som lärts under spelets gång finns kvar när spelaren återvänder till det riktiga livet. Effekten att en spelare åtar sig beteende från en avatar kallar Bailenson et al. Proteus-effekten efter den skepnadsbytande grekiska guden Proteus (Bailenson et al. 2009).

I pedagogiskt syfte finns en roll som kan utnyttjas i digitala spel, lärarrollen. Eleven får lära genom att lära ut, engelska *Learning by teaching*. Detta är ett mycket kraftfullt pedagogiskt verktyg (Gulz & Haake, 2014). När eleven lär upp någon annan blir det mycket viktigare att det blir rätt, eleven får ansvar för någon annans lärande. Detta medför högre motivation och engagemang hos eleven (Gulz & Haake, 2014). Agenten kan även tillföra ytterligare funderingar hos eleven genom att ifrågasätta och på så sätt tvingas eleven som lär ut granska sin egen kunskap. En tredje styrka hos *Learning by teaching* är att eleven uppmuntras att fundera över hur en annan person tänker. Eleven blir alltså tränad i metakognition, tänkande om tänkande, vilket fungerar bättre när det är någon annans tänkande som funderas över, eftersom det inte kräver parallella processer på samma sätt som när man själv ska både tänka och samtidigt fundera över sitt tänkande (Gulz & Haake, 2014). Eleven får även en större tro på sin egen förmåga om denne får lära ut. Speciellt elever som fått en bild av sig själva som svaga kan gynnas av detta. Eleven får lära ut om ämnet, och att lära ut gör man ju om man kan. I digitala spel kan *Learning by teaching* implementeras genom så kallade *Teachable agents*.

Teachable Agents

Agenter i spel är digitala karaktärer som styrs av datorn. Ofta har de någon form av artificiell intelligens och kan interagera med spelaren. Digitala agenter är något som Scholl of Education, Stanford, *AAALab* utnyttjat för att utveckla konceptet *Teachable Agents* (Gulz & Haake, 2014). Denna typ av agent är specialiserad för att utnyttja *Learning by teaching* genom

att eleven tilldelas en lärling i form av en digital agent kallad *Teachable Agent*, TA. Denna TA lär eleven sedan upp i spelet (Gulz & Haake 2014). För att detta ska fungera krävs det att eleven kan ta till sig konceptet att den digitala eleven faktiskt lär sig, här spelar *The Media equation* in. *The Media equation* är en teori som säger att människor tenderar att behandla datorer som om de har mänskliga egenskaper om datorns beteende liknar mänskligt beteende (Reeves & Nass, 1996). Alltså kan *The Media equation* utnyttjas genom att låta TA:n uppträda som att den lär sig, eleven kommer då behandla TA:n på samma sätt som den skulle ha gjort om lärlingen var en faktisk människa. *Educational Technology Group* vid Lunds och Linköpings universitet har genomfört ett antal studier som visar att lärospel som innehåller TA:s ger bättre lärande än spel utan TA:s.

Om Magical Garden

Magical Garden är ett matematiskt lärospel med målgruppen barn på dagis och förskola, med en mental utvecklingsålder på ca tre till fem år. Spelet ska på ett lekfullt sätt träna upp de underliggande grunderna för talsystemet och en känsla för tal allmänt. Tillvägagångssättet för att åstadkomma detta är att använda teachable agents, formativ feedback och scaffolding. Magical Garden är ett delprojekt av "The Magical Garden" som är ett samarbetsprojekt mellan School of Education, Stanford University och Educational Technology Group, Lunds Universitet och Linköpings Universitet.

Spelidén går ut på att eleven ska samla ihop vattendroppar i sin vattenkanna som sedan används för att vattna en virtuell trädgård. När trädgården blir vattnad blommar det upp olika, mer eller mindre magiska, blommor vilket ger spelaren en progressionskänsla. Vilka blommor som hamnar på vilken plats i trädgården är slumpartat eftersom eleverna inte ska kunna jämföra sig med varandra då tävlingsmoment inte är något som uppmuntras i detta lärospel. Svårighetsgraden i spelet anpassas efter hur bra spelaren presterar. Detta leder till att alla spelare får vattendroppar i samma takt oavsett vilken nivå de ligger på, allas trädgårdar blir alltså lika blomstrande.



Figur 1. Delspel i Magical Garden. Teachable Agent Mille Mus ser på medan eleven anger vilken våning fågelungen ska åka till för att komma hem.

Spelet är uppbyggt utav delspel som är avskilda från den huvudsakliga trädgården. I dessa delspel löser eleven matematikuppgifter, se fig. 1. Spelaren och dess lärling (*teachable agent*) får börja med trädgården som utgångspunkt för att sedan spela ett delspel

Efter delspelet återvänder de till trädgården igen där spelaren antingen kan välja att spela ännu ett delspel och få fler vattendroppar i sin kanna eller att använda sina vattendroppar för att vattna sin trädgård. Delspelen fokuserar på att lära ut olika saker genom att vara designade så att olika matematiska uppgifter ligger som grund för olika delspel, det kan till exempel handla om tallinjer eller andra moment eleverna behöver lära sig. Upplägget i de olika delspelen är att eleven först får lösa uppgiften själv ett antal gånger, då repetition är en viktig del av matematiskt lärande. Efter att eleven har gjort det själv kommer lärlingen in i bilden och frågar om den får titta på när eleven löser uppgifterna. Efter att ha tittat på några gånger vill nu lärlingen försöka själv. Denne föreslår då ett lösningsförslag och eleven får agera lärare genom att säga om svaret är rätt eller fel. Om det skulle vara fel får eleven visa sin lärling det rätta svaret. Ju bättre eleven lär sin lärling desto oftare kommer den svara rätt i framtiden.

Användartestning med barn

Vid utveckling av programvara är det viktigt att noggrant utföra tester på produkten för att lokalisera fel och brister samt se att produkten passar den tänkta målgruppen. Vid den händelse att målgruppen är barn finns det specifika aspekter som är lite extra viktiga att ta i beaktande för att undvika ett missvisande resultat från testerna.

Barn idag är inte noviser när det gäller användandet av datorer och touchplattor. Faktum är att den tid som barn spenderar framför datorn har tredubblats under det senaste årtiondet, enligt forskning av Kaiser Family Foundation (Nielsen, 2010). Den utökade vanan och kunskapen har bidragit till en förändrad syn på datoranvändande och fört den unga användarens beteende närmre de vuxnas inom vissa aspekter. Detta kan exempelvis ses i Nielsen Norman Groups studie där barn numera var negativt inställda till läsning av instruktioner innan påbörjande av spel, till skillnad från när studien utfördes nio år tidigare (Nielsen, 2010). Trots detta finns det fortfarande stora skillnader mellan åldersgrupper. Barn är mycket känsliga för när material är anpassat för en målgrupp som är äldre eller yngre än de själva. Detta kan visa sig i form av negativ inställning och, på något håll, överdriven beskrivning av uppgiftens svårighetsgrad (Nielsen, 2010).

För att klargöra att en produkt är lämplig är det viktigt att designa användartestningen på ett sätt som är anpassat efter åldersgruppen. Det finns många aspekter att täcka in vid utformning av test för barn. Först och främst är det önskvärt att skapa en relation mellan försöksledare och testpersoner. Detta kan ske under en förberedande presentation eller ett inledande samtal vid testtillfället. Det är bra att tidigt få barnen att öppna upp sig och släppa eventuell blyghet genom att exempelvis diskutera favoritämnen, sport eller födelsedagar. Det kan också vara lämpligt att yngre barn får utföra testen i par eller tillsammans med en lärare eller förälder för att undvika att de känner sig osäkra. (Hanna et al. 1997)

Under testen är det mycket viktigt att tänka på att speciellt yngre barn kan ha svårt att uttrycka sina åsikter i ord. Om frågor ställs blir ofta svaret snarare det som barnet tror att försöksledaren önskar, än ett svar baserat på den egentliga upplevelsen. För att få ett önskvärt resultat är det därför viktigt att än mer än vanligt titta efter tecken på positiv eller negativ inställning till uppgiften, såsom leenden eller suckar. Det kan även vara svårt för yngre barn att hålla koncentrationen längre stunder. Det kan följaktligen vara bra att ändra ordningen på uppgifterna om testet är långt. På detta vis undviker man att vissa uppgifter endast testas då barnen är trötta eller har ledsnat på aktiviteten. (Hanna et al. 1997)

När testningen är avslutad är det ofta uppskattat att utdela någon slags belöning som tack. Detta kan exempelvis vara ett certifikat som gör det möjligt för barnen att dela upplevelsen med sina föräldrar (Girard, 2011)

3 Metod

Utvecklingsprocessen har delats in i olika faser där en designprocess följts av en utvecklingsprocess som sedan avslutas med testning. Under designprocessens gång har ett iterativt tillvägagångssätt använts. Utvärderingar har gjorts i samtliga steg och nya idéer har uppkommit allt eftersom problem har stötts på.

Designprocess

Projektet påbörjades med en brainstormingfas. Syftet med denna fas var att få fram en stor mängd tankar och förslag som sedan kunde bantas ner till konkreta spelidéer. Processen innebar bollande och utvärdering av idéer, hopslagning av tankar och diskussioner om hur de bästa alternativen skulle kunna tas fram. Nyckelord som hölls i bakhuvudet under brainstormingen var *scaffolding* och *feedback*. Processen gav väldigt många olika idéer om hur en uppföljare till Magical Garden kunde se ut. Först och främst söktes ett genomgående tema som skulle vara attraktivt för en äldre spelarbas. Därefter kvarstod att komma fram till nya delspel. Dessa skulle inte bara vara nya till utseendet, utan vissa delar av logiken var också i behov av uppdatering. Något som diskuterades flitigt under brainstormingen och kom att utgöra en stor del av projektet i helhet var att tillföra ett antal mindre lekmoment utöver matematikuppgifterna. Detta för att förhoppningsvis öka elevernas motivation och för att tillföra lite action till lärospellet.

Efter brainstormingen gick processen över i en designfas, där de idéer som ansågs bäst ur utvecklingssynpunkt skulle fastställas och bearbetas. De idéer som hade arbetats fram under föregående fas var förslag på delspel, olika matematiska begrepp samt övergripande teman. Målet med den här fasen var att skapa ett sammanhängande spel där del spelen kunde relateras till det övergripande temat. Detta krav medförde att många idéer som ansågs bra fick slopas på grund av att de inte passade tillsammans med något tema.

Designen av delspelen utgick främst från de olika matematiska begrepp som enligt läroverkets kursplan skulle läras ut till barn i målgruppsåldern. Efter att matematiken var fastställd anpassades idén efter det tema som ansågs vara mest lämpat. Vissa delspel var designade fristående från ett

speciellt tema vilket gjorde dem anpassningsbara, medan andra var mer knutna till ett speciellt koncept. När temat väl var fastställt granskades samtliga delspelsidéer ännu en gång. De som var knutna till någon av de slopade temaidéerna bearbetades för att se om det fanns något av användning innan det resterande av den idén skrotades.

Utveckling

Vid designprocessens slutskede inleddes en agil utvecklingsprocess där en spelbar prototyp implementerades. Processen inleddes med en analys av spelets olika delar vilket resulterade i en prioriteringslista för de olika delarna, i sammanhanget ofta kallade *features*. Projektgruppen delades även upp i två delar för att effektivisera arbetet. Det föll sig att den ena gruppen fokuserade på den grafiska presentationen och den andra på själva spellogiken. Implementationen var från början tänkt att bygga på tidigare Magical Gardens grund i *HTML5* och *Javascript*, men gruppen tog ganska tidigt i processen beslutet att istället börja om från början och bygga allt i spelmotorn *Unity* istället.

Spel i *Unity* byggs upp av scener som innehåller hierarkier av objekt. Dessa objekt kan tilldelas olika komponenter som exempelvis grafisk textur, ljud, animation, programmerad spellogik och så vidare.

För att få med bakgrundsberättelsen och ge karaktärerna liv gjordes även röstinspelningar. Ljudfilerna spelades in separat och parades sedan ihop med animationerna i *Unity*.

Testning

Då en uppföljare till Magical Garden som till viss del har gått ifrån det visuella temat från det föregående spelet har utvecklats var det nödvändigt att utföra en användartestning i utvecklingsprocessen. Tanken är att spelaren ska känna igen sig i det nya spelet så att övergången är enkel, men att det samtidigt ska vara lite nytt och intresseväckande. Det är även ny matematisk logik i spelet då nya begrepp presenteras. Detta behövde givetvis också testas för att se om det ligger på rätt nivå samt se att det visualiseras på ett lättförståeligt sätt.

Tidsramen för projektet inte var tyvärr tillräckligt stor för att göra omfattande tester med rätt målgrupp. Istället togs beslut om att ha två faser i testprocessen. En fas med en äldre åldersgrupp för att hitta grundläggande problem och en fas med den tänkta målgruppen för att utvärdera spelet i helhet.

Målsättningen med testningen i helhet var att undersöka ifall lekmoment som läggs in i undervisningen är motiverande eller distraherande. Den syftade också till att ta reda på om användarna förstår hur man ska interagera med spelet utan för mycket ledning samt om de fastnar för karaktärerna som i med.

Den första fasens testning hade också syftet att göra den andra fasen så meningsfull som möjligt. I den första omgången utfördes användartest med en målgrupp i 20- till 30-årsåldern för att se om logiken i de delspel som utvecklats var lättförståelig och om designen upplevdes som intuitiv. Testen i den första fasen utfördes med en logi-prototyp som var tecknade bilder på papper som skulle efterlikna spelets design i så stor utsträckning som möjligt. I den här första

delen behandlades frågor som:

- Förstår användarna hur de ska interagera med prototypen?
- Vad är deras intryck av karaktärerna?
- Känns spelet sammanhängande?

Den andra fasens testning var den del som var relevant för avgörandet av hela projektets resultat. I denna fas utfördes användartest med barn i den ålder som spelet är avsett för. Även här var det självfallet viktigt att barnen förstår logiken. I denna fas tillkom det dock även andra frågeställningar som till exempel ifall de verkligen lärde sig matematiken som ligger som grund för spelet. I fas två testades inte längre en lofi-prototyp utan här låg ett färdigimplementerat spel på touchplatta till grund. Detta för att barn troligen tycker det är roligare att spela ett spel som verkar färdigt och inte avfärdar det för att de tycker det är fult ritat eller har svårt att relatera till att det egentligen ska vara en touchskärm.

Det tillkom även några ytterligare frågor till den här delen av testerna som inte var lika relevant när spelet testades på en äldre spelarbas:

- Är lekmomenten distraherande eller motiverande?
- Lär sig barnen matematiken?
- Är matematiken på rätt nivå för den målgrupp som testar?
- Är logiken i utlärningsmetoderna på en bra nivå för att barnen ska förstå vad som händer?

För att utvärdera huruvida lekmomenten motiverar lärandet var det tänkt att se hur många av de som spelar som vill spela igen. Grunden för bedömningen av spelarnas engagemang låg i observationer från testledaren men det användes också en *debriefing* för att få svar på konkreta frågor som stod som stöd för de kvantitativa resultaten.

4 Resultat

Då projektet innehåller en fullständig designprocess med samtliga steg från brainstorming till testutvärdering och implementation så tillkommer även olika resultattyper. De mest väsentliga typerna är design-, projekt- och testresultat. Designresultatet beskriver hur den slutgiltiga idén uppkommit samt övriga temaförslag som diskuterats under processens gång. Projektresultatet beskriver hur den slutgiltiga designen blev, medan testresultaten speglar den feedback som införskaffats från testpersonerna under respektive testomgång.

Designresultat

Under designfasen framkom många spretande idéer om olika teman och delspel där de flesta fungerade bra på egen hand, men inte alltid så bra ihop med andra. Efter brainstormingfasen var det tre huvudteman som var klart överlägsna de andra. Ett rymdtema, ett tropiskt tema och det sista alternativet var att fortsätta i Magical Gardens tema och endast ut-

forma nya delspel. Då målgruppen som projektet är utformat för är några år äldre än Magical Gardens ursprungsmålgrupp blev valet att gå ifrån det befintliga temat för att nå en mognare publik. Konceptet var tänkt att likna Magical Gardens koncept med delspel, hopsamling av belöningar och till sist någon form av utlopp för den ihopsamlade belöningen.

Brainstormingfasen var inte helt fokuserad på att först utforma ett tema och sedan utforma delspel, utan dessa olika idéer uppkom överlappande under processens gång. Det ledde till att många av delspelsidéerna hölls öppna för att kunna anpassas till flera olika teman. Eftersom det var viktigt att hålla ett logiskt sammanhang i spelet så kretsade alla teman runt att försöka ha en överblicksvy där det var lätt för spelaren att se att alla delspelen hörde ihop med temat. Det gjorde att rymdtemat var likt ett solsystem med olika planeter för varje delspel, det tropiska temat var med olika öar, och ifall Magical Gardens tema skulle behållas var det tänkt att utforma det likt olika hus längs en gata med ett delspel vid varje hus. Dock lämnades Magical Gardens originaltema ganska snabbt vilket gjorde att många av idéerna endast användades för rymd- eller ö-temat.

Utöver att rymdtemat gick ut på att varje delspel skulle ske på en egen planet, skulle spelaren med hjälp av en rymdraket och matematikkunskaper sedan ta sig mellan de olika planeter. Där skulle spelaren spela de olika delspelen för att samla ihop någon form av belöning, likt vattendropparna i det ursprungliga Magical Garden. Det var tänkt att raketen skulle skjutas mellan de olika planeter med en slangbells istället för att flyga för egen maskin, då den skulle vara trasig och behöva lagas innan den skulle kunna flyga själv. Inspiration kom från det populära spelet *Angry Birds* (Rovio Entertainment, 2009) och matematiken skulle komma in genom att bestämma hur många av invånarna från planeten i fråga som skulle hjälpa till att spänna slangbellan för att slunga iväg raketen med rätt kraft. Idén om slangbellan resulterade i ett lite mer lekfullt mellanspel mellan varje delspel som även hjälper till att knyta ihop logiken bakom spelets design. Detta sker genom att spelaren får se hur karaktärerna tar sig mellan olika delspel

Det tema som i slutändan fastställdes var det tropiska ö-temat där varje delspel utspelades på varsin ö. Mycket av det som var tänkt till rymdtemat kunde direkt översättas. Exempelvis blev rymdraket ett flygplan och utomjordingarna blev tropiska apor. Istället för vattendroppar skulle nu flygplansdelar samlas ihop genom att göra de olika delspelen och istället för att vattna sin trädgård skulle spelaren nu laga sitt trasiga flygplan med de ihopsamlade delarna.

Projektresultat

Efter flertalet resor fram och tillbaka med bakslag och nya idéer fastslogs det som tidigare nämnts att projektet skulle resultera i en fortsättning på Magical Garden som fick gå under namnet Magical Islands. Temat blev ett tropiskt ö-landskap med bakgrundshistorien att Panders och hans vänner ska åka på semester efter att ha blivit utmattade av sitt trädgårdsarbete. Under resans gång börjar flygplanet som de färdas i plötsligt att gå sönder. Det ryker, vinglar och tappar delar som sprids över de öar som passeras för att slutligen tvingas nödlända. Nu måste Panders och hans vänner

besöka öarna, samla ihop alla flygplansdelar och bygga upp sin farkost för att kunna komma hem igen.

Pappersprototypen bestod utav två delspel och ett så kallat mellanspel. I början av en ny spelomgång får användaren se planet flyga över öarna för att sedan krascha (bilaga 1). Ett beskrivande intro ska komma att visas här för att förklara scenariot och sätta stämningen. Användaren får sedan reda på att den ska besöka en viss ö och får med hjälp av en storleks- eller avståndsbeskrivning välja den avsedda ön. Skulle fel ö väljas visas ett videostycke innehållande tät djungel och farliga djur samtidigt som berättarrösten förklarar varför den valda ön inte var korrekt, exempelvis genom att säga att ön var för stor om användaren uppmanats välja den minsta ön.

Då rätt val har gjorts kvarstår ett annat problem nämligen att flygplanet är allt för trasigt för att kunna flyga. Istället används en slangbella vars gummiband spänns med hjälp av de apor som bor på ön (bilaga 2). Detta är det så kallade mellanspelet och för att kunna ta sig till rätt ö krävs en korrekt löst matematikuppgift. Användaren får reda på hur mycket ”apkraft” som krävs för att slungas till ön. Därefter ska ett antal apor, vilka har olika tal på sin mage beroende på sin styrka, placeras vid slangbellan för att tillsammans med den eller de apor som eventuellt redan står där bilda talet som krävs för att skjuta flygplanet lagom långt. Då användaren är nöjd med sitt svar zoomas bilden ut till ö-vyn och flygplanet skjuts iväg då användaren drar i slangbellan. Skulle det sammanlagda talet på apornas magar vara korrekt hamnar flygplanet och dess passagerare på rätt ö och äventyret fortsätter. I annat fall flyger det för långt alternativt för kort, beroende på apornas tal, plumsar i vattnet och passagerarna får paddla tillbaka för att försöka ännu en gång.

De två delspel som har tagits fram hittas på varsin ö. På den första ön träffar användaren på en apa som är ledsen (bilaga 3). Han är nämligen så väldigt törstig eftersom en ilsken haj har bosatt sig i det vattenhål där han brukar släcka sin törst (bilaga 4). För att göra hajen mer tillmötesgående vill apan mata honom med kakor, eftersom kakor är något som man blir snäll utav. Hajen är dock kräsen. Han accepterar endast hela kakor. Problemet är att alla apans kakor har gått sönder och han vet inte hur han ska få dem hela igen. Användaren får alltså i uppgift att laga kakorna genom att välja den bit som saknas, för att sedan mata hajen med dem. När spelaren har matat hajen med ett antal kakor kommer Panders in i spelet. Han är användarens teachable agent och vill gärna titta på för att lära sig. Efter några gånger är det sedan hans tur att försöka och användaren får guida honom rätt. Efterhand som spelet fortlöper blir hajen mindre och mindre arg för att till slut bli glad och låta apan gå till vattnet. Som tack för hjälpen får användaren en flygplansdel som apan har hittat i djungeln (bilaga 5). Delspelet är tänkt att träna tio-kompisar, alltså vilka ental som tillsammans blir tio. En hel kaka representerar tio och övriga siffror representeras av kakbitar.

På nästa ö påträffas ännu ett delspel. Här hittar användaren en övergiven stövel i djungeln. Strax kommer ett antal myror promenerandes vilka går in i stöveln med avsikt att bosätta sig (bilaga 6). Myrorna är dock inte de enda som är intresserade av fyndet. Efter ännu en liten stund kommer nämligen en piratkatt som råkar samla på just stövlar. Han

vill inte ha några myror i sitt nya samlarobjekt och ber därför användaren om hjälp med att skaka ur stöveln. Ut ramlar ett mindre antal myror än det som gick in. Detta gör piratkatten fundersam och han undrar hur många myror som kan vara kvar i stöveln (bilaga 7). Detta blir användarens uppgift att besvara. Vid rätt svar skakas samtliga myror ut och piratkatten kan nöjt fortsätta sitt sökande efter fler stövlar. Anges ett för litet antal myror plockas detta antal ur stöveln och katten försöker sedan sätta på sig stöveln. När han upptäcker att det är myror kvar tycker han att det kittlas, vilket får honom att hoppa runt i en knasig dans. Skulle istället ett för stort antal myror anges skakas de kvarvarande myrorna ut och katten får berätta för användaren att det inte fanns så många myror kvar. I detta delspel tränas pre-algebra, alltså enkla ekvationer som exempelvis fyra adderat med vadå blir sju. Även i detta delspel kommer så småningom Panders in för att först lära sig och sedan räkna själv.

En pappersprototyp enligt ovanstående beskrivning användes vid första fasens användartestning. Till den slutgiltiga prototypen har ett intro, mellanspelet samt det första delspelet implementerats. Denna prototyp användes sedan till testning fas två.

Testresultat

I den första testfasen fick nio personer i åldrarna 20-30 år utföra användartester på en prototyp i lofi-version. Fokus låg på att samla in kvalitativ data samt se om spelet ansågs intuitivt och intresseväckande. Resultaten från den första omgången var överlag positiva och mycket av den feedback som samlades in visade att prototypen var på väg i rätt riktning. Det är dock värt att poängtera att testpersoner var i samma åldersgrupp som utvecklarna och därför relativt olik den ålder som målgruppen kommer vara. Det skulle kunna vara så att spelet är intuitivt och lättförståeligt för testgruppen eftersom de har samma mentala modell av kunskapen hos barn i målgruppsåldern som utvecklarna själva har, vilken i sin tur kanske inte alls överensstämmer med verkligheten.

Den mest betydande feedbacken som gavs i denna del kan ses nedan.

- Överlag en mycket positiv inställning till spelets story och många tyckte det var ett bra och sammanhängande tema.
- I princip alla användare var beroende av berättarrösten för att förstå vad de skulle göra.
- Flera testpersoner efterfrågade ändå en funktion för att kunna hoppa eventuella intron till uppgifterna.
- Karaktärerna var mycket omtyckta. Det var även stor spridning på vilken karaktär som testpersonerna tyckte mest om. Alla karaktärer var någons favorit.
- Interaktion mellan karaktärerna efterfrågades.
- Ett problem som frekvent återkom var att gorillan stod på en plats som var på mellanhand mellan slangbellan och de apor som kunde manipuleras, vilket gjorde att den inte fick någon tillhörighet alls. Det var oklart ifall det var en apa som kunde väljas eller en som redan stod vid slangbellan och många noterade honom inte alls.

Samtliga åsikter samlades in och designen revideras innan implementering av spelet som kom att testas i fas två.

I nästa fas besöktes en förskola som hyser barn i åldrarna fyra och ett halvt till sex år. Användarstudien utfördes på tolv barn. I denna fas gjordes försök att samla in både kvalitativ och kvantitativ data. Detta för att kunna avgöra om spelet ligger på en tillräcklig nivå så att det inte blir för svårt att förstå eller alltför enkelt att genomföra. Om eleverna inte behöver använda matematiska beräkningar för att lösa uppgifterna, utan istället kan gissa med stor säkerhet, samt får för lite formativ feedback anses spelet inte uppfylla sitt syfte.

Testet var utformat med ett antal korta inledande frågor om datoranvändning och räknevana innan barnen fick testa spelet. Detta följdes av några avslutande debriefingfrågor. Det visade sig vara svårare än förmodat att få relevanta svar på de frågor som ställdes. Den kvantitativa delen av resultatet blev därför begränsad.

Det mest betydande resultatet som observerades i denna del kan ses nedan.

- De allra flesta försökspersoner var oengagerade under lyssningsdelarna. De negligerade i princip bakgrundsberättelsen totalt.
- *Trial-and-error* användes när de inte förstod hur de skulle göra.
- I det första delspelet nyttjades pusselmetaforen mer än matematiken. Få verkade notera siffrorna på kakbitarna eller att summan skulle bli tio.
- Ju längre spelet fortlöpte desto mer intresserade verkade de bli. Ville utföra uppgifterna fortare och fortare.
- Vissa elever behövde guidning gällande vilken kaka som kunde passas ihop med vilken. Det var inte tillräckligt tydligt att huvudkakan var fristående och skulle matchas med rätt kakbit.
- De användare som uppgivit att de tyckte om matematik och brukade räkna hemma med sina föräldrar klarade av logiken i spelet på ett utmärkt sätt. Övriga hade mer problem.
-

Återkommande svar på de ställda frågorna och kommentarer från testerna.

- Spelet var enkelt!
- Matematiken var mycket enkel.
- Att mata hajen var det roligaste med spelet.
- Det var speciellt roligt att hajen blev mer och mer glad allteftersom.
- Det var tråkigt att behöva mata hajen så många gånger.

Avslutningsvis fick även två förskolelärare testa spelet. De gav följande kommentarer.

- Det känns för enkelt att gissa sig fram i kakspelet. Många barn kommer troligen att pussla mer än räkna.
- Det vore bra om hajen såg gladare ut för varje kaka man ger honom. Det hade gett bra progressionsfeedback.
- Lite väl mycket berättande. Barnen har inte så mycket tålamod.
- Trevligt att det är lite långsammare än många andra spel på marknaden idag. Det blir inte lika mycket tävling.



Diagram 1. Diagrammet visar fördelningen över barnens favoritkaraktärer.



Diagram 2. Diagrammet visar huruvida barnen önskade spela Magical Islands igen.

5 Diskussion

Matematikuppgifter

Huvudsyftet med *Magical Islands* är självfallet att barnen ska lära sig matematik, vilket ställer vissa krav på uppgifterna som spelet innehåller. De spel vi har utvecklat är tänkta att kunna variera i svårighetsgrad samt träna ett antal olika begrepp. För att veta att uppgifterna ligger på en lagom svårighetsnivå har vi utgått ifrån läroplanen för årskurs ett och två. Detta är på gränsen till att vara i överkant av den tänkta målgruppen, men efter att ha fått höra från en lågstadielärare att läroplanens nivå är lägre än den som lärs ut i skolor idag valde vi att göra på detta vis.

Trots detta har vi under användartesterna fått förfrågningar om spelen är på en lagom nivå. Det var speciellt de vuxna med tidigare erfarenhet av barn som ifrågasatte spelet och ansåg att det skulle vara för enkelt för ett barn i den åldern. Enligt testresultaten kan man dra slutsatsen att detta inte är fallet. Vissa elever klarade av uppgifterna på ett enkelt sätt och hade säkerligen kunnat lösa även mer komplicerade sådana, men dessa elever var i klar minoritet. Under testningen användes den enklaste av våra tänkta scenarion, vilket indikerar att vi ligger på en lagom nivå för vår målgrupp. För att säkerställa detta resultat är det dock nödvändigt att utföra avsevärt många fler användartester. Det skulle även vara intressant att höra fler lärares åsikter i frågan.

Ett av de begrepp som vi fått bekräftat som nyttigt av tidigare nämnda lärare är tiokompisar. Detta innebär att eleven övar på att se vilka två tal som tillsammans bildar tio. Detta är mycket användbart och bra att lära sig tidigt i sin utbildning. Detta är främst implementerat i kakspelet, då en hel kaka antar värdet tio, men det kan även användas i mellanspelet med apor och slangbella. Mellanspelet är ett moment som återkommer mellan varje delspel och det är därför viktigt att det kan alterneras för optimal inläring. Barnen ska inte kunna lära sig det korrekta svaret utantill utan förståelse för uppgiften. Detta ska uppnås genom att det sökta talet varierar samt att aporna kan ha olika styrka, värden. I det enklast möjliga scenariot kan alla apor vara värda ett, det sökta talet är åtta och på plats står redan en apa med värdet sju. Detta kan sedan utökas till att användaren får placera två eller tre apor med olika värden för att uppnå det sökta talet, med eller utan en redan ditplacerad apa.

Scaffolding

Scaffolding är ett begrepp som genomgående har varit centralt i detta projekt. Det används inom både del- och mellanspel på ett sätt som vi hoppas kommer stötta användarna i deras läroprocess på ett bra sätt. I mellanspelet varierar apornas storlek med sin styrka, alltså sitt värde. Man kan därmed enkelt se att en större apa är starkare än en mindre, i detta fall har ett högre värde. Detta kan ses som en tämligen naturlig hjälp som är enkel att förhålla sig till.

Ett annat område där scaffolding kom in på ett okonstlat sätt var i kakspelet. Här har varje kakbit till en början en storlek som motsvarar dess värde, vilket resulterar i att den sökta kakbiten har samma storlek som den saknade biten i huvudkakan. Paralleller kan i och med detta dras till pusselläggande, vilket ger en utökad chans att lösa uppgiften. Det är även, precis som i föregående exempel, tydligt att storlek ökar linjärt med värdet. Det visade sig under testningen att detta var till stor hjälp för många användare.

Till det sista delspelet var det något svårare att hitta ett sätt att tydligt få in scaffolding. Vår lösning blev till slut att använda färger på myrorna i piratkattens tankebubbla (bilaga 7). Tankebubblan visar en uppställning som är tänkt att förenkla för minnet genom att visa en slags före- respektive efterversion av händelsen. Genom att i ett första skede införa färgkodning för de myror som stannar kvar i stöveln kan dessa skiljas från de som trillat ut. För att förtydliga kan det antas att fyra myror äntrar stöveln, två av dem är röda och två är svarta. De svarta trillar ut, medan de röda är kvar. I

tankebubblans uppställning kan användaren då enkelt se att alla de svarta myrorna har lämnat stöveln och kan således räkna de röda myrorna, då det är dessa som är kvar.

I samtliga av ovanstående fall ska dessa hjälpmedel efterhand tas bort i enhet med scaffoldingmetoden. När det gäller aporna och kakbitarna så kommer deras storlek efterhand normaliseras och slutligen kommer alla vara lika stora, oavsett om de är värda två eller sju. I det andra spelet kommer alla myrorna ha samma färg, vilket tvingar användaren att helt förlita sig på de tal och den matematik som uppgiften erbjuder. Tyvärr hann vi inte implementera hur denna scaffolding gradvis ska försvinna, varför detta återstår att testa.

Formativ feedback

I mellanspelet där rätt ö ska väljas finns en risk att användaren skulle kunna använda trial-and-error för att komma vidare. Detta måste motverkas och i dagsläget är vår lösning att om användaren väljer fel ö så spelas det upp en scen där vyn zoomar in på ön och användaren får se tät, ogenomtränglig djungel och arga djur. Detta skulle göra trial-and-error en mycket långsam metod för att hitta rätt svar. Feedback ska även ges genom att berätta att den valda ön är "för liten/stor" eller "för nära/långt bort" etc.

I mellanspelet där användaren väljer ett antal apor för att skjuta iväg flygplanet till rätt ö ges feedback genom att flygplanet skjuts olika långt beroende på vilka apor som valts. För starka apor gör att det hamnar för långt bort, medan för svaga hindrar planet från att nå fram till rätt ö. Sköldpaddaskaraktern kan även säga något i stil med "Ojdå, det blev visst för mycket/för lite kraft". Barnen verkade inte reflektera speciellt mycket över denna feedback när de gjorde fel. Det var dock ingen som gjorde fel mer än en gång, vilket gör att de aldrig fick se samtliga scenarion. Detta bör dock ses som en vinst med tanke på att de utan denna hjälp kunde förstå spelets avsikt.

I delspelet där myror kryper in i en stövel är vår tänkta feedbacklösning just nu att låta katten ta på sig stöveln och börja dansa om ett för litet tal angivets. Det blir en tydlig konsekvens av matematiken; det är fler myror kvar i stöveln, vilka kittlas när man tar på den. Något som däremot är betydligt svårare att lösa är om ett för stort tal angetts. Ges ett för stort tal kommer ju katten plocka ut alla myror och sen blir problemet att det inte finns några kvar, även om svaret var fel. Att ange ett för stort tal skulle helt enkelt lösa problemet. Detta är något vi ha diskuterat mycket under våra möten. Vår nuvarande lösning att låta katten bli mycket förvirrad om det tar slut på myror innan han har plockat ut så många som användaren angett. En annan lösning skulle vara att katten blir så tokig så han slänger iväg stöveln, men detta känns lite väl dramatiskt. Vi är inte helt nöjda med lösningen i dagsläget och ser gärna att någon går vidare i ämnet genom ytterligare testning och utforskning av konceptet.

Lekmoment

Magical Garden är framförallt ett läromedel och i andra hand ett spel. Interaktionen med spelvärlden är ganska begränsad, då det mesta handlar det om att trycka på knappar. Vi tycker att detta känns lite tråkigt och redan tidigt i vår process bör-

jade vi diskutera att det kanske vore klokt att satsa på lite mer och lite roligare interaktion. Lite mer spel helt enkelt. Det finns tydliga bevis på att spel har en motiverande kraft, men det är oklart hur man ska fånga detta i ett lärospel (Garris et al. 2002). Räcker det att spelet innehåller några få lekmoment så som en växande trädgård där användaren får se sina framsteg visualiserad? Mer interaktion borde leda till ett roligare spel, men samtidigt får det inte påverka lärandet negativt. Våra lösningar i de olika delspelen bygger på tanken att när eleven löst en uppgift får hen göra något interaktivt.

I det första delspelet, där flygplanet ska slungas iväg, får användaren först räkna ut rätt antal apor och dra dessa till slangbellan. Användaren belönas sedan med ett litet roligt *Angry Birds*-liknande (Rovio Entertainment, 2009) spel där slangbellan måste dras upp för att skjuta iväg flygplanet. Flygplanet flyger sedan iväg och hur eleven än har siktat med slangbellan så åker planet ändå så långt som angavs när aporna drogs till slangbellan. Här finns dock ett problem, om användaren skulle tillåtas skjuta iväg flygplanet även om fel antal apor angetts skulle detta potentiellt kunna leda till att användaren glömmar matematiken och gör fel med flit för att få skjuta slangbellan mer eftersom det är roligare. För att motverka detta kan skjutandet begränsas till endast när rätt svar angivits, men då tappas den feedback som ges när planet åker för långt/kort. Här behövs mer testning för att vi ska kunna ta ett beslut. Tyvärr kunde funktionen inte testas under våra användartester, då implementationen inte hann färdigställas. Detta är alltså något som kvarstår till eventuellt kommande arbete.

Något annat som spel är bra på är att fånga användaren med en berättelse. Vi är dock inte helt säkra på om detta är något som tilltalar målgruppen. Från egen erfarenhet var ett spels berättelse inget som påverkade väldigt mycket när vi själva spelade spel i vår barndom. Själva interaktionen och progressionen är något som vi anser spelar större roll. I detta projekt har vi trots detta fört in lite mer spänning i storyn för att undersöka om det kan göra spelupplevelsen bättre. Vi ville se om en sammanhängande berättelse kunde ge barnen en motiverande anledning att lösa uppgifterna. I den första testomgången sågs berättelsen som något mycket positivt och testpersonerna önskade mer interaktion mellan karaktärerna. Detta lades till inför nästa test, men nådde inte riktigt de resultat vi önskade. I de nästkommande testen såg vi nämligen snabbt att vi hade gått till överdrift gällande historian. Den var alldeles för ingående och krävde koncentration av åhöraren, trots att vi försökt göra den så barnvänlig som möjligt. Att personerna i de olika testomgångarna såg så olika på detta beror troligen på att flera aspekter. För det första gjordes de testen i fas ett muntligt. Försökspersonerna fick alltså berättat för sig vad spelet gick ut på, hur bakgrundshistoria såg ut och vad de skulle göra i spelet. I det andra testet vävdes allt detta in i förinspelade ljudfiler som framfördes av karaktärerna. Förinspelade röster kan vara svårare att följa och det krävs att den inspelade rösten pratar mycket långsamt för att behålla tydligheten. Detta gör att det tar längre tid och koncentrationen hos spelaren kan gå förlorad. En annan aspekt som är väl värd att nämna är den skilda åldern mellan de två grupperna. Det som anses som enkelt, medryckande och lekfullt för en 25-åring kan ses som långtråkigt och svårt av en 5-åring. Detta bör förbättras genom nerbantade intron och

korta, kärnfulla dialoger.

En del som ännu inte implementerats och därmed inte testats är vår idé om ett lekmoment utan matematik. Här skulle man ta de delar som samlats ihop efter varje avklarad uppgift och sedan bygga ihop sitt egendesignade flygplan. Detta moment skulle fungera som en slags belöning efter att eleven presterat i de räkne baserade delspelen. Här finns ännu ett spel vi tagit inspiration från, *Bygg bilar med Mulle Merck* (Levande Böcker, 1997). Detta gamla spel fångade användaren genom att låta spelarens framsteg visualiseras genom att tillgång ges till nya häftiga bildelar. Användaren får sedan själv bygga ihop allt till ett unikt fordon. Detta fordon hänger sedan med i övriga spelet och gör sig hela tiden påmind. Fordonet blir både en bra morot och ett kul lekmoment när spelaren själv får bygga. Vår tanke är att låta användaren bygga lite på sitt flygplan i pauser som läraren kan styra. Skulle eleven själv kunna styra detta hade nog matematikuppgifterna snabbt glömts bort och byggandet hamnat i fokus. Vi tror att byggandet kan bidra till att användaren vill komma tillbaka och spela mer. Hur vi ska låta byggandet sättas igång är ännu inte helt bestämt, men att låta läraren vara i kontroll känns som en rimlig idé eftersom själva matematiklärandet ska vara i fokus. Man kan även tänka sig att byggandet kan sättas igång efter varje delspel och vara tidsbegränsat så att eleven inte fastnar.

Kritik och förbättringar

Gorillans placering i mellanspelet uppdagades vara bristfälligt i lofi-skissen då många inte kunde placera honom varken vid slangbellan eller bland de apor man har att välja på. Detta är givetvis något som är mycket viktigt och det har på andra ställen visat sig att föremål är ottydligt placerade. Samma problem fanns med kakan som skulle lagas i kakspelet. Där var problemet att det inte var uppenbart vilken kaka som skulle lagas och testarna försökte då istället para ihop alla delarna för att få ihop så många kakor som möjligt. Apornas placering var något som reviderades till andra testomgången med gott resultat. Ingen av våra testpersoner hade svårigheter med gorillan i denna omgång. Gällande kakorna så bör huvudkakan flyttas längre från de övriga och samtliga kakbitar bör göras större för att öka tydligheten.

I det första testet vad det flera som kommenterade att det blev lite osammanhängande när man kastades direkt in i mellanspelet och vissa hade även lite svårt att se kopplingen till delspelen. Många ville veta vem som fanns i planet från första början och tyckte att det hade varit bra för sammanhanget. Vid dessa test försökte vi som testledare att återge en tillräcklig del av intro och mellananimationer för att testpersonen skulle kunna ta sig igenom spelet utan problem. Efter den första delen användartester blev det klart att det hade varit bra med mer ingående animationer mellan delspelen för att få logiska övergångar och en mer flytande känsla i spelet. Testpersonerna vill också gärna ha ett förklarande intro som gav kontext till spelet. Detta löstes genom att implementera just övergångsanimationer och intro till spelet för att göra det mer sammanhängande. Efter att ha testat det på barnen visade det sig dock som tidigare nämnt att långa intron och väntetider leder till otålighet hos spelaren. En lösning kan då vara att ha med animationerna men att ge spelaren en möjlighet att skip-

pa de momenten.

Många elever som testade spelet använde sig av trial-and-error-tekniken när de inte förstod matematiken i en uppgift. Detta är ständigt ett problem när det kommer till självvärterande interaktionsmedel; det är enklare att prova de olika alternativen om den enda konsekvensen är att uppgiften måste göras om igen, utan väntetid. Det finns olika sätt att lösa detta på. Det kan exempelvis läggas in ett system att användaren bara får göra fel ett visst antal gånger eller kanske att användaren får minuspoäng om den gör fel. Vi valde en annan metod. Eftersom barnen i testerna inte alls tyckte om att vänta mellan att utföra uppgifter så la vi in att efter varje gång eleven gjort fel visades en sekvens om på vilket sätt detta var fel. Sekvensens syfte är både att informera eleven om på vilket sätt den gjorde fel, och ge formativ feedback, samt att vara tillräckligt lång för att den ska vara störande. Eleven ska alltså inte vilja göra fel för att den då får vänta en alldeles för lång tid innan den får prova igen. Det visade sig tydligt i testen att denna metod skulle behöva appliceras på delspelet för att förhindra att barnen försökte med samtliga kakbitar innan den rätta hittades. I dagsläget spelas bara ett ljudklipp upp, vilket inte var tillräckligt avhållande. Det är dock viktigt att tänka på att det är en balansgång mellan att vara tråkigt att göra fel och att spelet blir tråkigt i helhet. Det gäller att hålla sig på rätt sida av den linjen för att bevara barnens intresse och motivera dem till att räkna och lära sig.

Användartest med barn

Användartestning är något som vi i projektgruppen är mycket förtrogna med. Vi visste dock inte att det var så pass stor skillnad att utföra tester med barn som användare. Detta fick vi snart erfara. Kontakten med förskolan gick mycket enkelt. Vi skickade mail med information (bilaga 8) och fick snabbt svar. Därefter skickade vi ut ytterligare information till föräldrar (bilaga 9). Redan här skiljde sig tillvägagångssättet då kontakten inte kunde ske direkt med våra användare. När vi anlände till förskolan var många barn lite blyga och hade svårt att svara på våra inledande frågor, trots att vi försökte mjukstarta med lite avslappnat prat om namn, ålder, födelsedagar etc. Vi var även noga med att poängtera att vi ville att de skulle testa spelet eftersom de var experter och visste mycket mer om vad som var bra respektive dåligt.

De flesta barnen fick testa spelet i mindre grupper, då alla skulle få tid att försöka. Det var då tydligt vilka personer som var mer dominerande och fick ofta sina kamrater att hålla med om sina åsikter. Överlag gav svaren på de ställda frågorna mycket lite användbar information. Ett exempel som är mycket tydligt är att samtliga försökspersoner angav att matematiken var enkel, men endast ett fåtal klarade spelet utan att använda sig av trial-and-error. Enligt våra observationer verkade alltså inte räkneuppgifterna vara lika enkla som barnen uppgett. Ju längre testet fortskred desto mer förtroende verkade barnen få för oss, vilket märktes tydligt under debriefingen även om svaren inte alltid var helt sanningensligna. Efter testen fick varje barn ett diplom för deltagande (bilaga 10), vilket var mycket uppskattat.

Hade vi gjort fått möjlighet att göra om användartesten idag hade det funnits en del saker som vi hade förändrat. För

det första hade vi åkt dit en dag i förväg för att presentera oss, så att barnen sett våra ansikten en gång förut. Vi hade även sett till att vi hade haft fler inledande frågor så att barnen hade börjat öppna upp sig och släppt blygheten innan de relevanta frågorna hade kommit fram. Vidare hade det varit bra att tillsammans med förskolelärarna komma fram till en uppdelning i lämpliga kamratpar, så att alla personer hade fått möjlighet att visa vad de kunde och tyckte. I övrigt fungerade testen bra och vi fick ut nyttig information som bekräftade vissa av våra misstankar och gav ny input.

Unity

Beslutet att inte använda oss av den befintliga implementationen av Magical Garden kan tyckas något förvånande. Normalt sett brukar det vara en fördel att utgå från färdig, fungerande kod. I vårt fall var det dock ett enkelt beslut. Eftersom Magical Garden är byggt i *HTML5* och *Javascript*, vilka ingen av oss har mycket tidigare erfarenhet inom, var det mycket krångligt att komma igång med. Våra spelidéer har även en del fundamentala skillnader interaktionsmässigt mot Magical Garden, exempelvis *drag-and-drop* och den *Angry Birds*-inspirerade slangbellan. Dessa problem gjorde det omöjligt för oss att med våra resurser kunna implementera vår vision av Magical Islands. Istället togs Unity upp som ett alternativ. Samtliga i gruppen hade tidigare arbetat i Unity och vi hade därför en ganska tydlig bild av dess möjligheter. Efter en kort analys visade det sig att Unity nu har bra stöd för 2D-spel, vilket kändes lovande. Efter beslut om att använda denna spelmotor kunde vi på kort tid bygga vår första scen. Unitys tydliga uppdelning i grafisk presentation och scriptlogik gjorde det även enkelt att dela upp arbetet i två kategorier. En grupp byggde animationer och den andra programmerade den sammanbindande logiken.

Arbetet med Unity började bra och kändes mycket lovande. Det visade sig dock senare uppkomma en del problem. Vissa av dessa berodde på misstag baserade på gruppens, trots tidigare arbeten, relativa oerfarenhet av Unity. Ett exempel på ett sådant var att vi till en början använde komponenten *Animation* i de objekt som skulle animeras. Detta medförde en del svårigheter och efter ett tag upptäcktes det att Unity numera har ett *Animator*-verktyg som fungerar betydligt bättre. Vi fick därför börja om med alla animationer och utföra dem i *Animator* istället.

Versionshantering blev också ett problem. Vi ansåg det krångligt att sätta upp ett delat projekt och arbetade istället parallellt med två lokala projekt som vi med jämna mellanrum bakade ihop. Detta gick bra till en början, för att sedan inte fungera alls. Filerna ville inte samsas och Unity kraschade. Detta ledde till att vi fick bygga om alla animationer än en gång. Detta kunde ha undvikits om vi tidigt hade tagit tiden att sätta upp ett gemensamt delat projekt varifrån animationerna kunnat återhämtas. I efterhand kan det tyckas naivt att vi inte insåg att detta skulle hända, men vid tillfället ansåg vi att det skulle ta för mycket resurser att få ordning på ett delat projekt.

Spellogiken visade sig även vara mycket svårare att få ordning på än väntat. Enligt utvecklingsplanen hade vi tänkt implementera mellanspelet med slangbellan samt det delspel där hajen matas med kakor. I slutändan blev det dock svårt

nog att få rätt på ett delspel, varför vissa detaljer i melanspelet aldrig blev färdiga. Från början skrevs alla scripts i *Unityscript*, vilket liknar *Javascript*, för att kunna använda delar av koden från Magical Garden. Detta var inte så klokt eftersom vi ju hade gått över till Unity för att just komma ifrån den gamla koden, då vi inte hade någon erfarenhet av *Javascript*. Detta gjorde att vi efter mycket frustration beslutade att skriva om allt i C# istället. Efter detta gick arbetet mycket snabbare och smidigare, men för mycket tid hade runnit iväg för att hinna bygga allt innan testerna skulle genomföras.

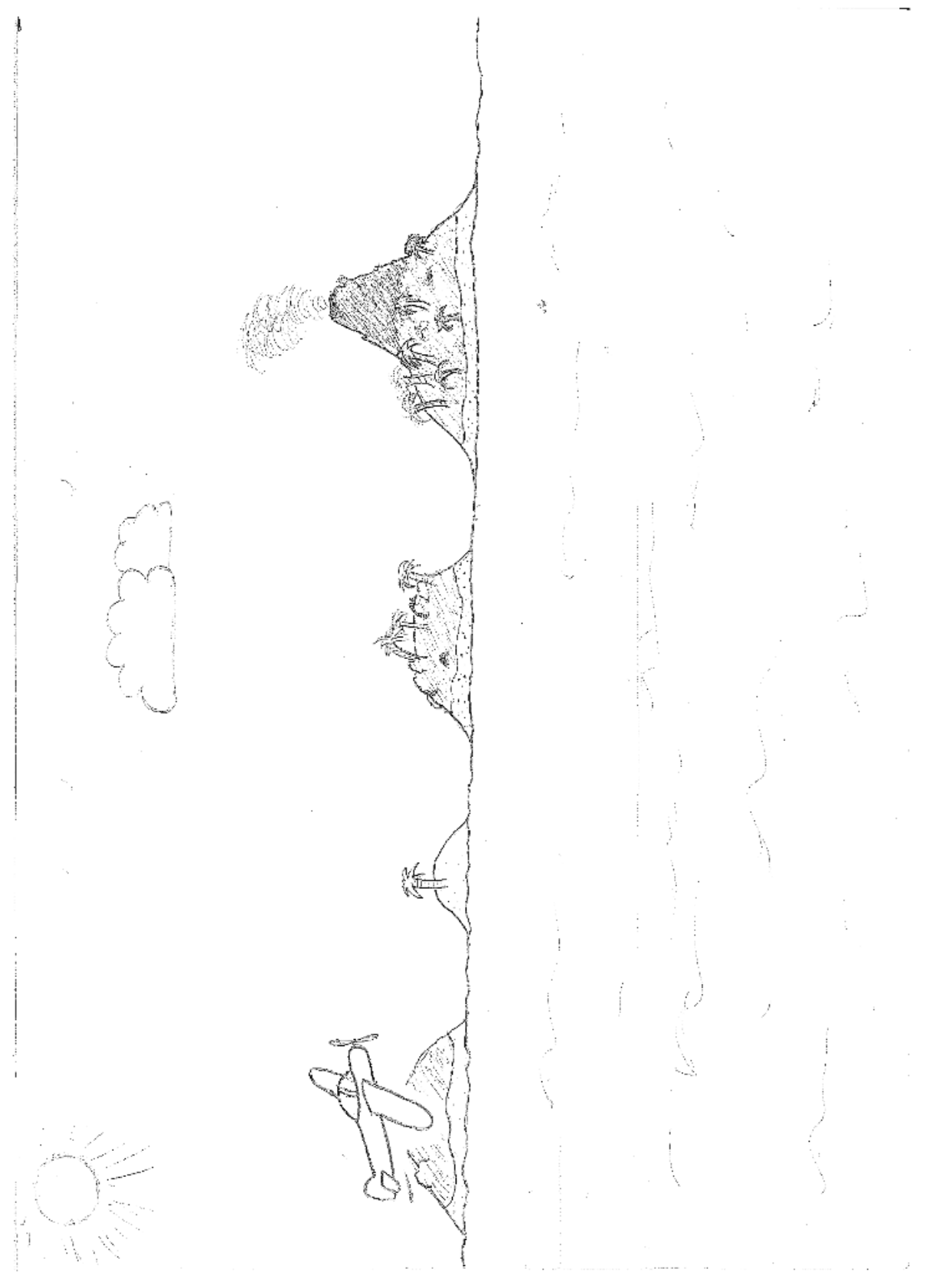
I slutändan tror vi ändå att vi gjorde rätt val att gå över till Unity, men oförutsedda problem och en något naiv tilltro till programmet gjorde att vi inte hann komma så långt i utvecklingen av spelet som vi skulle ha velat.

6 Slutsats

Projektets omfattning var från början inte bestämd vilket ledde till att ambitionsnivån blev alldeles för hög för den tidsram projektet hade. Problem så som manfall, tekniska bekymmer och tidsbrist hjälpte inte heller fallet. Trots detta har projektet nått en punkt i utvecklingen som är lovande. Projektet har resulterat i ett spel med ny design och nytänkande inom lärospel och lagt en grund för vidareutveckling inom området. Då det grundläggande konceptet är fastställt enligt rapporten ovan är det enkelt att bygga på idén med fler delspel och mer lekmoment. Kommentarer från förskolelärare sa att de tyckte att det var bra att spelet var lite långsammare än många andra som finns på marknaden idag, då lärande stimuleras av en lugn miljö. Att avläsa från barnens kommentarer så ville tio av tolv barn spela spelet igen, vilket är ett bra resultat då spelet är anpassat för just dem.

Referenser

- Bailenson, J., Yee, N., Ducheneaut, N. (2009) The Proteus Effect Implications of Transformed Digital Self-Representation on Online and Offline Behavior, *Communication Research* 36:2, 285-312.
- Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Greening, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J., Thomas, D. (2001), Manifesto for Agile Software Development, Agile Alliance
- Garris, R., Ahlers, R., Driskell, J. (2002) Games, motivation, and learning: A research and practice model, *Simulation & Gaming*, 33:4, December 2002, 441-467.
- Gulz, A. & Haake, M. (2014). Forskning och utveckling om digitala läromedel för att gynna lärande och lärare. In A. Persson & R. Johansson (Eds.), *Forskarperspektiv på kunskap, utbildning och skola – utbildningsvetenskaplig forskning vid Lund universitet*. Lund, SE: Dept. of Educational Sciences, Lund University. 3, 45-64.
- Levande Böcker (1997), *Bygg bilar med Mulle Meck*, SE: Levande Böcker.
- Linderöth, J. (2012). Why gamers don't learn more: An ecological approach to games as learning environments, *Journal of Gaming and Virtual Worlds* 4:1, 45-62.
- Linderöth, J. (i tryck). Spel i skolan - Det regelstyrda lärandets möjligheter. In *Lärare i en uppkopplad tid: Möjligheter och Utmaningar*, A. Lantz-Andersson & R. Säljö (Red.). Stockholm, SE: Natur och Kultur. 8, 171-193.
- Pilner Blair, K. (2013) Feedback in Critter Corral: The Effectiveness of Implication Versus Corrective Feedback in a Math Learning Game, *Early Education and Technology for Children – April 2013*, Salt Lake City, Utah.
- Reeves, B. and Nass, C. (1996) *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Chicago, IL, US: Center for the Study of Language and Information; New York, NY, US: Cambridge University Press.
- Rovio Entertainment (2009), *Angry Birds*, Espoo, FI: Rovio Entertainment.
- Sampayo-Vargas S., Chris J.C., Zhen H., Graeme J.B. (2013) The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game, *Computers & Education*, 69, November 2013, 452-462 Elsevier Science Ltd. Oxford, UK.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner & E. Souberman., Eds.) (A. R. Luria, M. Lopez-Morillas & M. Cole [with J. V. Wertsch], Trans.) Cambridge, Mass.: Harvard University Press. (Original manuscripts [ca. 1930-1934]), 86
- Nielsen J. (2010) *Children's Websites: Usability Issues in Designing for Kids* (Elektronisk) Nielsen Norman Group. Tillgänglig: <http://www.nngroup.com/articles/childrens-websites-usability-issues/> (2014-12-20)
- Hanna, L., Ridsden, K., Alexander, K. J. (1997) *Guidelines for Usability Testing with Children*, *Magazine Interactions*, 4:5, 9-14.
- Girard, S. (2011) *Case studies illustrations of experimental settings issues in classrooms* (Elektronisk) Bath, UK: HCI Research lab, Dep. of Computer Science, University of Bath. Tillgänglig: http://www.researchgate.net/publication/260959966_Case_studies_illustrations_of_experimental_settings_issues_in_classrooms (2014-12-20)







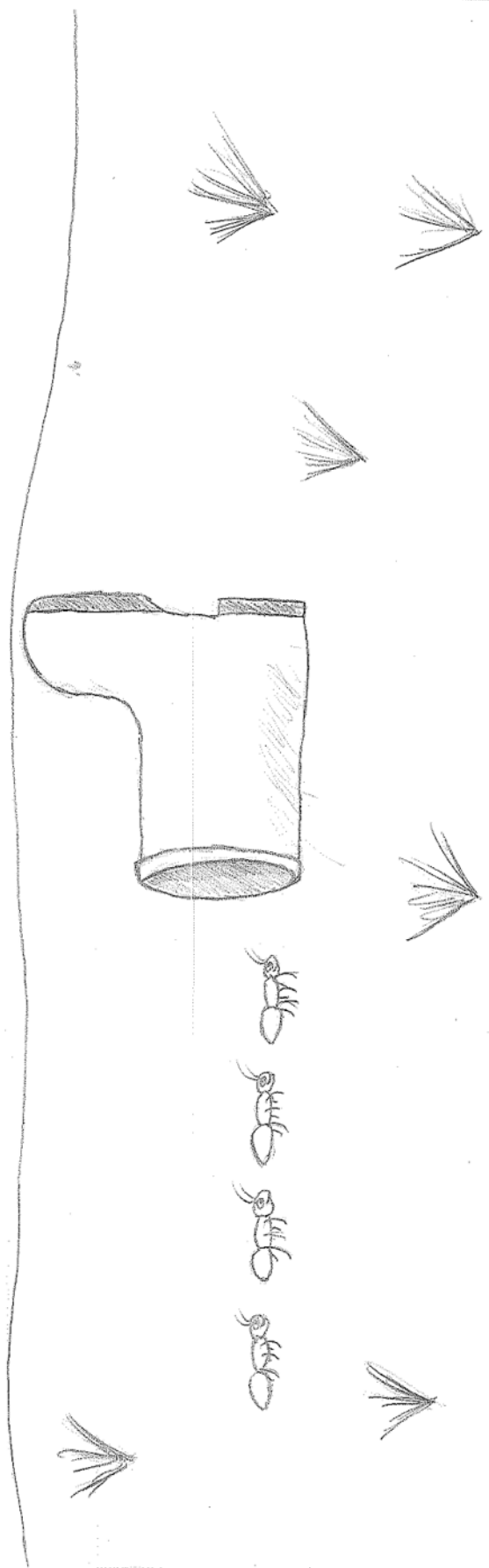
Bilaga 3



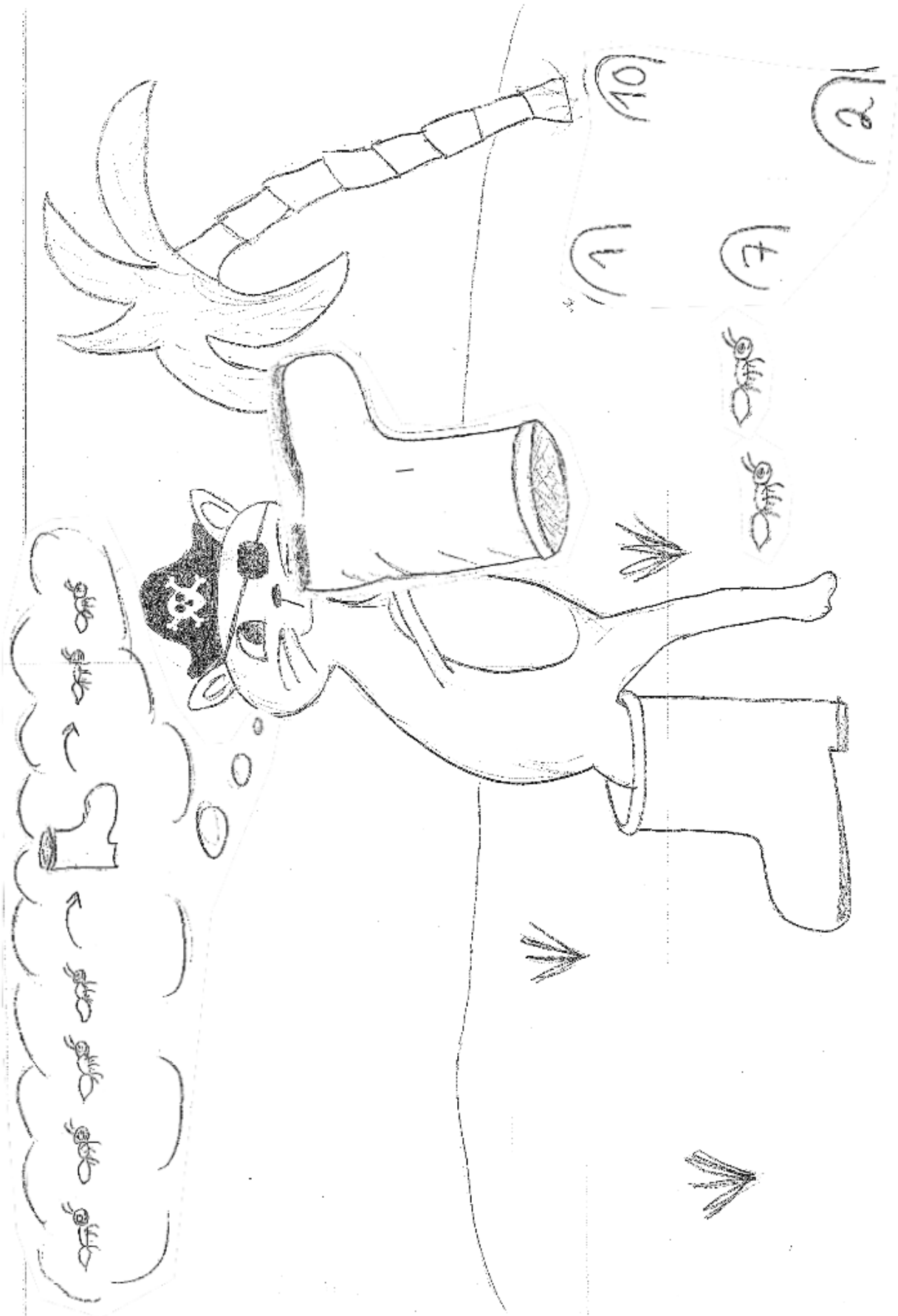
Bilaga 4



Bilaga 5



Bilaga 6

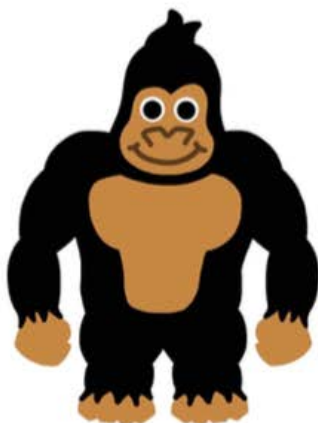


Bilaga 7
Bilaga 8

Hej,

vi är tre studenter som studerar sista året på civilingenjörsprogrammet på Lunds Tekniska Högskola. Vi läser just nu en kurs där vi utvecklar ett pedagogiskt lärospel för barn i åldrarna 5-7 år. För att utvärdera vårt arbete söker vi nu personer i vår målgrupp som kan utföra användartester på produkten. Spelet är inriktat på matematik och är utvecklat enligt läroplanen för barn i målgruppsåldern. Spelet är webbaserat och designat för att spelas på en touchplatta, exempelvis en iPad. Användartesterna är tänkta att utföras på ca 10-15 elever

och är beräknat att ta ca 10 minuter per elev. Är detta något som låter intressant för Er förskola? Tveka inte att höra av Er med frågor och funderingar. Vi ser fram emot ett eventuellt samarbete.



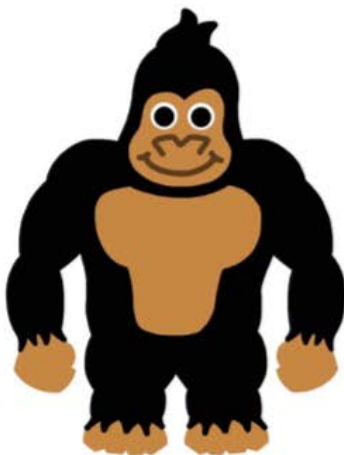
Med vänliga hälsningar,
Celie Gunnarsson
Johan Mattsson
Aron Söderling
Informations- & kommunikationsteknik, LTH

adi10cgu@student.lu.se

070-722 94 99

Hej,

Fredagen den 12 december kommer det att genomföras användartester på Backens förskola avdelning Stenen. Vi som genomför testen är tre studenter som läser sista året på civilingenjörsprogrammet på Lunds Tekniska Högskola. Vi jobbar just nu med ett projekt där vi utvecklar ett pedagogiskt lärospel för barn i åldrarna 5-7 år. Användartesterna är ett sätt för oss att utvärdera vårt arbete, undersöka produktens kvalitet och nivån lämplighet för målgruppen. Vi vill särskilt poängtera att testerna endast är till för oss att få feedback på produkten och är inte på något vis avsedda att testa Era barns förmåga. Spelet är inriktat på matematik och är utvecklat enligt läroplanen för barn i målgruppsåldern. Testerna kommer utföras på touchplatta, exempelvis en iPad, och beräknas ta ca 10 minuter per elev.



Har ni frågor eller vill ha mer information, kontakta oss gärna.

Med vänliga hälsningar,
Celie Gunnarsson
Johan Mattsson
Aron Söderling
Informations- & kommunikationsteknik,
LTH

adi10cgu@student.lu.se

070-722 94 99

Diplom

för deltagande i användartestning
av lärospelet *Magical Islands*.



Tack för hjälpen!

Diplom

för deltagande i användartestning
av lärospelet *Magical Islands*.

Tack för hjälpen!



Diplom

för deltagande i användartestning
av lärospelet *Magical Islands*.



Tack för hjälpen!

Diplom

för deltagande i användartestning
av lärospelet *Magical Islands*.

Tack för hjälpen!



Virtuell miljö vs. Verkligheten

Choice Blindness-experiment utformat för en virtuell miljö med en virtuell agent.

Emil Boman, Martin Lingonblad, Ludvig Londos, Arvid Nilsson

Påverkas människors tillit eller uppmärksamhet inför en virtuell agent av hur inlevelserik och grafiskt detaljerad agenten och dess omgivning är? Ett choice blindness (CB) experiment utfördes för att undersöka detta. Vi försökte med hjälp av en hög- respektive lågdetaljerad virtuell försöksmiljö testa hypotesen att den lägre detaljerade virtuella miljön (låg immersion) skulle öka upptäcksgraden vid en CB manipulationen. 38 försökspersoner deltog och delades in i två grupper (hög och låg immersion) och testades. I båda villkoren agerade en virtuell agent försöksledare till CB-experimentet, 16 par porträttbilder visades två åt gången för försökspersonerna och de valde vilket ansikte de tyckte var mest attraktivt. I åtta av paren fick försökspersonen motivera sitt val av ansikte, fyra av dessa var manipulerade och om försökspersonen uttalade att manipulation hade skett noterades det som en concurrent upptäckt. I miljön med hög immersion var den totala upptäcksgraden 31%, och i den med låg immersion var 43%. Vi kunde därför inte visa att den totala upptäcksgraden i låg immersion skilde sig signifikant från den i hög, däremot fanns en skillnad i när de flesta upptäckter gjordes, nämligen tidigare i låg jämfört med i hög immersion. Framtida forskning kan involvera experiment med en högre immersion grad än den vi använt genom att använda till exempel Oculus Rift eller större skärmstorlek.

1 Introduktion

Digitala experiment blir allt vanligare. Det har blivit enklare att programmera och skapa experiment digitalt vilket lett till att fler och fler använder detta medium för att skapa experiment. I många fall är detta ett fruktbart sätt att göra studier för att få bra data och jämförbara resultat, eftersom själva undersökningen kommer att vara likadan för varje försöksperson. Å andra sidan blir det, i takt med att man vill göra mer och mer avancerade undersökningar med datorn, allt fler parametrar man vill inkorporera och redogöra för.

Inom exempelvis beteendeforskning eller socialpsykologi måste man ofta göra ett val mellan naturlig och kontrollerad "setting" (miljö). En naturlig miljö inbjuder till naturligt beteende men gör det även problematiskt att kunna replikera experimentet då det finns nästan oändligt många obestämda variabler. En kontrollerad miljö – likt en experimentmiljö – kan redogöra för många av dessa variabler men kanske inte undersöker det naturliga beteende man vill un-

dersöka. Frågan är om vi kan, med hjälp av datorn, skapa kontrollerade naturliga miljöer som kan efterlikna vår naturliga miljö och på så sätt kunna undersöka naturligt beteende på ett sätt som är replikerbart.

En virtuell värld ger forskare möjligheten att ha kontroll över miljön likt en undersökning i ett labb med fördelen att miljön kan efterlikna en verkligare miljö (Fielder & Haruvy, 2009: 716). Den virtuella miljön medför också att man kan utföra experiment genom att programmera virtuella människor – virtuella agenter – till att representera människor som behövs i experimentet. De virtuella agenterna kommer då att bete sig på samma sätt för varje försöksperson i varje experiment, något som annars kan vara problematiskt att kontrollera hos riktiga människor (Blascovich et al., 2002: 112).

Syftet med denna undersökning var att replikera ett tidigare Choice Blindness-experiment (Haake & Gulz) virtuellt – mer om Choice Blindness (CB) nedan. CB-experimentet kommer vi att använda som ett mätvärde för hur väl den virtuella settingen är framställd. Genom att manipulera den virtuella settingen (hög immersion/låg immersion) vill vi kontrollera om detta påverkar upptäcksgraden av manipulationer hos försökspersonerna. Det vi vill undersöka är huruvida den virtuella settingen bidrar till en väsentlig skillnad i upptäcksgrad hos försökspersonerna. Det finns även möjlighet att jämföra upptäcksgraden i detta experiment med tidigare CB-experiment i naturlig miljö.

2 Bakgrund

Undersökningar i virtuella rum (Virtual environments, VE) kan hjälpa oss skapa och göra undersökningar i en naturlig virtuell setting. En fråga man kan ställa sig är hur pålitliga experiment man kan göra i denna setting. Litar vi exempelvis på datorer i samma utsträckning som vi litar på människor?

Reeves & Nass (1996) hävdar att vi – i stor utsträckning – behandlar datorer som om de vore människor. Exempelvis ger man mindre kritik till ett datorprogram om vi ombeds att utvärdera programmet på samma dator som man körde programmet (Reeves & Nass, 1996: 23f). Detta är en parallell till att det är mycket svårare – man får mindre kritik – om man ombeds att utvärdera en person när personen är närvarande.

I denna undersökning kommer vi som sagt att använda

oss av CB för att undersöka hur försöksdeltagare påverkas av en virtuell setting.

“Choice Blindness is a research paradigm originally inspired by techniques from the domain of close-up card-magic, which permits us to manipulate the relationship between what people choose, and what they actually get.”

- <http://www.lucs.lu.se/choice-blindness-group/>

CB-experimentet i sitt grundutförande genomförs med en försöksledare och en försöksdeltagare. Försöksledaren visar upp ett flertal bilder av människo-ansikten. Dessa bilder visas upp parvis i ett kort intervall, mellan 2-5 sekunder. Efter att korten visats vänds de upp-och-ned på bordet varpå försöksledaren frågar försöksdeltagaren vilket av ansiktena denna föredrog. Försöksdeltagaren väljer därpå höger eller vänster kort beroende på dennes preferens. När valet har gjorts skjuts detta kort fram till försöksdeltagaren som antingen får lägga kortet åt sidan eller ta upp och kolla på kortet. I vissa fall då deltagaren tar upp kortet så ska denna motivera varför hen valde detta kortet över det andra. I några av fallen då deltagaren ska förklara sitt val så har försöksledaren utfört ett magitrick - där hen har haft båda korten i paret i båda händerna och utfört ett byte. Försöksdeltagaren tvingas i detta fall att motivera det valet hen inte utförde. Oftast uppmärksammas inte manipulationen av korten och det är detta som kallas Choice Blindness.

Upptäcktsgraden av manipulationer har visat sig stiga när man för över CB till en virtuell setting (bland annat föregående ej publicerad undersökning av Haake & Gulz samt Johansson et al., 2007). Bland annat såg man här, att när tidigare genomförda CB undersökningar med en fysisk försöksledare och verkliga kort nu gjordes med en fysisk försöksledare men med digitala kort ökade upptäcktsgraden av manipulationer (Haake & Gulz). Vad beror det på? Vi ser fyra möjliga alternativa förklaringar till att man som försöksdeltagare missar en manipulation:

- Naturlig miljö som stör ens uppmärksamhet.
- Tillit till försöksledaren: ”hen luras inte”.
- Tillit till den naturliga miljön; foton förändras vanligtvis inte.
- Någon annan okänd faktor.

Är det vår naturliga miljö med all dess stimuli och dess inverkan på vår uppmärksamhet som gör att vi inte ser manipulationer, är det vår tillit till försöksledaren – att hen inte luras – som gör att vi missar manipulationer eller är det vår tillit till att har man gjort ett val så kommer detta val inte att förändras?

Om vi återskapar ett naturligt CB-experiment virtuellt kan vi försöka komma åt detta. Till vår hjälp kommer vi använda oss av termerna immersion och presence, vilka är sätt att mäta hur personer upplever den virtuella miljön – mer om detta senare. I den virtuella miljön finns möjligheten att manipulera egenskaper för att se om det har en effekt på hur ofta människor upptäcker en CB-effekt. Påverkas man av att en virtuell försöksledare har ett mindre mänskligt beteende, och möjligen inbjuder till en mindre

social respons? Påverkas man av en karg miljö utan irrelevanta stimuli? Påverkas man av en miljö som inte till fullo matchar med verklighetens fysiska lagar? Alla dessa faktorer kan diskuteras i former av immersion.

Om man manipulerar ett par eller alla av dessa egenskaper, påverkar detta upptäcktsgraden av manipulationer eller är den största faktorn som ökar upptäcktsgraden själva övergången från det faktiska till det digitala?

Immersion och Presence

Vår förhoppning är att kunna påvisa att vilken grad man avspeglar verkligheten spelar roll för resultatet av ett experiment. I “Virtual reality”-paradigmet används ofta termerna “immersion” och “presence” för att bestämma denna grad av reflektion av verkligheten. Immersion och presence används även exempelvis i spelindustrin där de dock brukar beskriva samma sak. Här kommer vi att använda immersion för att förklara det som brukar kallas “system immersion” (Slater, 1999; Slater & Wilbur, 1997), det vill säga de tekniska egenskaper som den virtuella settingen byggs upp av. Presence används däremot som ett subjektivt mått på hur mycket man “är i den digitala världen” (present in the virtual world).

Genom immersion kan man påverka persence. Bland annat så rapporterade Welch et al. (1996; refererad i Slater & Wilbur, 1997) en signifikant skillnad i hur hög och låg grafisk fidelitet påverkade hur försöksdeltagare upplevde presence när de fick utföra ett experiment med en körsimulator.

Presence

Presence är en viktig aspekt av att skapa en virtuell setting. Kan man skapa en hög presence finns det stor anledning att tro att deltagarna kommer att i större mån ha ett naturligt beteende, det vill säga bete sig som man gör i en naturlig setting (Slater & Wilbur, 1997:8).

”The fundamental idea is that participants who are highly present should experience the VE as more the engaging reality than the surrounding physical world, and consider the environment specified by the displays as places visited rather than as images seen (ibid: 4).

Hög immersion leder dock inte med nödvändighet till en hög presence. Viktig är även förmågan hos deltagaren att övertyga sig själv om att den virtuella värld hon upplever är verklig. Det brukar kallas “suspension of disbelief”. Ett utmärkt exempel på låg immersion men hög presence är boken.

Presence kan även delas upp i olika delar, främst environmental presence och copresence eller social presence. Med denna uppdelning kan man urskilja hur personer upplever att de är i miljön (environmental presence) och att de är tillsammans med en annan person (social presence). Båda tror vi är viktiga att undersöka då – som vi nämnt ovan – de kan vara viktiga komponenter för att återskapa ett CB-experiment med ett resultat som liknar det i en verklig setting.

Immersion

Immersion behandlar de tekniska delarna av en virtuell setting – hur informationen visas, de kvantifierbara delarna av programmet (Slater & Wilbur, 1997:3). Immersion kan även

det delas upp i olika delar, Slater & Wilbur (1997) delar upp immersion i fyra kategorier: Vivid, inclusive, extensive och surrounding (s. 3). Dessa är en hjälp för att bygga upp en environmental presence. Vivid beskriver de tekniska specifikationerna såsom upplösning och fidelitet. Inclusive är hur mycket av verkligheten som sluts ute. Extensive är hur många sensoriska modaliteter som är inkluderade. Surrounding är hur väl ens synfält representeras i den virtuella världen.

Ett exempel på hög immersion är användandet av en Oculus Rift eller liknande "head-mounted display". Denna både stänger ute omvärlden, ger ett synfält liknande ens eget samt möjligheten att fritt röra på huvudet. Denna används i ett virtuellt rum med en hög grafisk fidelitet där man på olika sätt kan interagera med (exempelvis både genom tal och gester) och röra sig i omgivningen.

Utöver dessa egenskaper av immersion som främst är riktade till att skapa en hög environmental presence tillkommer det även interaktion mellan en virtuell agent och deltagaren – s.k. social presence – i denna undersökning en interaktion med en virtuell agent som försöksledare. En bra animerad agent kan leda till att människor uppfattar agenten som mer "human-like, engaging and motivating" (Dehn & Van Mulken, 2000: 2). Att en agent beter sig mänskligt kallar Blascovich et al. (2002) för behavioral realism. Det är helt enkelt enklare att övertygas om att den agent man möter i den virtuella miljön representerar en människa om den beter sig likt en människa och man kan kommunicera med den på samma sätt som man gör med en människa.

Blascovich et al. (2002) hävdar också att den kritiska punkten för en god behavioral realism är att den virtuella agenten beter sig eller ter sig intelligent. Detta för att vi skall uppnå en social presence (s.112). De hävdar även:

"If social presence is low (i.e., the user or participant knows the virtual other is an agent-avatar), behavioral realism must be very high, including perhaps even photographic realism, for social influence effects to occur" (Blascovich et al., 2002: 112)

Det vill säga att när en deltagare är medveten om att det

är en agent och inte en avatar (en virtuell representation av en annan människa) så krävs det högre realism, både på agentens rörelser och dess grafiska fidelitet, för att vi skall vilja interagera med den som om den vore en människa.

Sammanfattningsvis kan vi anta att: "The more inclusive, extensive, surrounding, and vivid the VE is [...] or the more similar the transformations in the VE are to those in the real world [...] the higher the presence." (Shubert, Friedmann & Regenbrecht, 2001: 267)

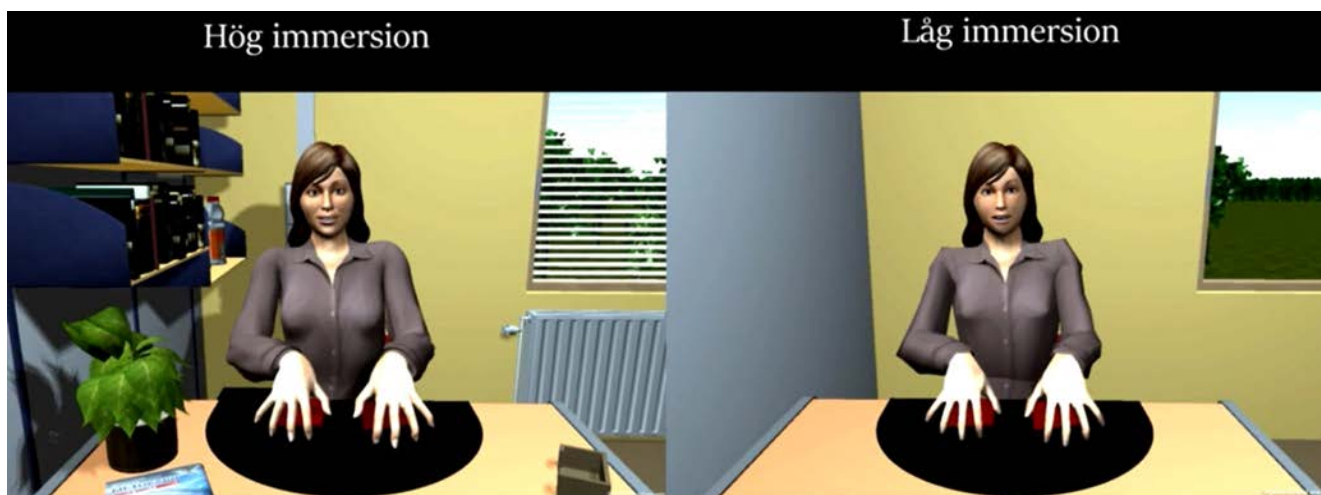
Choice blindness i en virtuell setting

Experimentet är en konceptuell replikering av Haake och Gulz (ej publicerad) CB-experiment vilka gjordes i två utföranden. Det ena med verkliga kort och en verklig försöksledare (Betty Tärning) och den andra med samma försöksledare men med virtuella kort på en datorskärm. Detta experiment blir en utökning av dessa då vi använder oss av virtuella kort och en virtuell försöksledare.

Experimentupplägget kommer att vara detsamma som de tidigare och kommer bestå av 16 stycken ansiktspår, representerade av svart-vita foton av både manliga och kvinnliga ansikten, ej mixade par. Det kommer att vara samma ansiktspår som i de tidigare experimenten och de kommer visas under samma tidsintervall (3 sekunder). Storleken på korten kommer även att anpassas så att de efterliknar experimenten som gjordes av Haake och Gulz.

Av de 16 ansiktspåren kommer deltagaren bli bedd att förklara sitt val på 8 stycken par varav fyra av dessa kommer att vara manipulerade. Dessa åtta par kommer att vara randomiserade vilket betyder att alla försöksdeltagare kommer få annorlunda ordning samt att de fyra manipulerade paren kan vara olika från försöksdeltagare till försöksdeltagare.

Deltagarna filmas för att kunna urskilja om och när en upptäckt av en manipulation sker. Utöver det har vi även en debriefing efter experimentet som efterspeglar de debriefingar som görs efter ett standard CB-experiment och ger chans för försöksdeltagaren att även göra upptäckter efter experimentets gång.



Figur 1. Till vänster en hög immersion bild av programmet och till höger är en bild av låg immersion där den virtuella miljön har ändrats enligt Blascovichs.

De två parametrarna vi kommer att undersöka är hög samt låg immersion. I hög immersion eftersträvades att replikera experimentet (med verklig försöksdeltagare och verkliga kort) i så hög grad som möjligt. En egenskap är att deltagaren kan prata naturligt med försöksledaren via en mikrofon. För att separera två grader av immersion valde vi att endast sänka eller ta bort saker i den med låg immersion, med hög immersion som utgångspunkt. Alternativet skulle vara att försöka öka den höga genom exempelvis graden av skala (Johnsen et al., 2010) eller inklusion (t.ex. med hjälp av Oculus rift).

Fyra aspekter av environmental presence valdes ut att manipulera. De valdes både med tanke på upplevd presence och möjlig påverkan på effekter som kan påverka CB (se diskussion ovan) och var som följer: Spatial presence (Slaters extensive), involvement (inclusive), upplevd realism (Slaters vivid) och social presence.

I versionen med låg immersion togs all överflödig möblering bort (se Appendix X för noggrannare beskrivning), för att sänka variabeln surrounding. Detta liknar det som gjordes i Welchs studie (Welch et al., 1996) där de också sänkte den grafiska fideliteten. Även handen som visade vilket kort som valdes togs bort, försöksledaren Betty fick sin röst datoriserad och hennes ansiktsanimeringar minskades i komplexitet, för att påverka variabeln extensive. I Blascovichs studie (Blascovich et al. (2002) gjorde man liknande förändringar.

Även skuggorna som fanns i versionen med hög immersion togs bort, detta testades i en gammal studie (Slater, Usoh & Chrysanthou, 1995). Även programmets grafiska inställningar sänktes för versionen med låg immersion för att sänka variabeln vivid. Målet med låg immersion var att sänka upplevd presence men inte påverka programmet negativt eller störa deltagarens uppmärksamhet, det vill säga att det varken stör eller drar uppmärksamhet ifrån CB-uppgiften. Målet var att sänka upplevd presence jämnt för de fyra aspekterna för att inte en skulle sticka ut och störa försöksdeltagaren.

Tabell 1. Visar medelvärden hos de två grupperna i de vardera fem aspekterna av presence som studerades.

Grupp	Spatial	Involvement	Exp. realism	Social
Hög immersion	2,28	2,05	1,65	2,52
Låg immersion	1,84	1,3	1,15	2,16

3 Pilotstudien

Design

En pilotstudie utfördes innan huvudexperiment. Syftet med pilotstudien var att fastställa att våra två grader av immersion faktiskt speglar en upplevd skillnad i presence samt att de inte verkar påverka CB-uppgiften negativt. Studien utfördes genom att vi först lät försöksdeltagaren göra en Immersive Tendencies Questionnaire (ITQ) (se appendix III).

Detta frågeformulär var till för att få en insikt i vilken utsträckning försöksdeltagarna tenderar att involvera sig i en virtuell (eller textbeskriven - bok) värld. Efter experimentet fick deltagarna även besvara ett frågeformulär om upplevd presence, "The Igroup Presence questionnaire" (IPQ) (se appendix IV). Detta frågeformulär behandlar fyra kategorier av presence (Spatial, involvement, exp. realism) och hur mycket presence personen kände i experimentet. Formuläret var utökat med fem tillagda frågor hämtade ur Garau et al. (2005) vilka behandlade social presence. Alla frågor besvarades på en femgradig skala.

Pilotundersökningen användes som grund för att visa att vår låg respektive hög immersion-grupp faktiskt har låg respektive hög immersion. Vi inkluderade även en standard debriefing angående CB där vi ställde frågor som stegvis blir mer och mer specificerade runt kortmanipulationerna. Ökningen av specifikation går från frågor som t.ex. om man tyckte att experimentet gick bra, till att man avslöjar att manipulationer har skett och ifall försöksdeltagaren upptäckte dessa. Vår hypotes är att det är en skillnad i presence mellan den hög och låg virtuella miljön.

Metod och procedur

Tio stycken studenter rekryterades på Ingvar Kamprads Designcentrum i Lund för att delta i pilotundersökningen. De fick ingen ersättning för sitt deltagande. Alla besvarade frågeformuläret om "immersion tendencies" innan de utförde undersökningen. De randomiserades in i två grupper om fem, som vardera fick hög respektive låg immersion-villkor. Direkt efter experimentet utfördes presence-formuläret.

Målet med pilotundersökningen var som tidigare nämnt att fastställa att de två villkoren skiljde sig åt i hänseende till upplevd presence och att de parametrar vi manipulerar inte störde utförandet av CB-uppgiften. Efter presence-formuläret gjordes även en debriefing för upptäckt av manipulationer men eftersom detta inte var pilotundersökningens huvudfokus analyserades inte detta utan användes främst för att testa experimentutförandet.

Resultat

Vi fann en skillnad på de båda grupperna i upplevd presence, vilka alla visade en viss försämrad upplevelse av presence jämfört mellan grupperna. I efterintervjun samt i en yttlig sammanfattning av concurrent detection fann vi även att det var en mindre upptäcksgrad av manipulationer i hög immersion jämfört med den i låg immersion. ITQ-resultaten hade ett medelvärde på 0,62 i hög immersion och 0,65 i låg immersion.

Diskussion

Resultatet motsvarade vår hypotes såtillvida att det visade på en tendens åt att det finns en skillnad. Låg immersion verkar visa på en allmänt mindre presence-känsla. Detta betydde att vi kunde gå vidare med experimentet på en större skala och testa upptäcksgraden då vi med pilotundersökningen menade oss ha visat att de parametrar vi manipulerat i låg immersion är av skild karaktär från den i hög immersion samt att de inte verkade störa deltagarna under CB uppgiften.

4 Choice Blindness Experimentet

Deltagare

41 studenter (medelålder: 24, $sd=2,8$. 21 st kvinnor) vid Lunds Universitet samt Lunds Tekniska Högskola rekryterades på Ingvar Kamprads Designcentrum i Lund. Som tack för sin medverkan kompengserades de med ett kaffekort värt 25 kronor. 3 deltagare exkluderades ur analysen på grund av att de antingen inte hade fullföljt uppgiften eller hade fått ett felaktigt experimentprogram. Totalt analyserades data från 38 deltagare, 19 i respektive grupp med medelålder i hög: 24,2 ($sd=3$, 10 st. kvinnor); låg: 23,8 ($sd=2,6$, 11 st. kvinnor).

Design och procedur

Undersökningen delades upp i tre delar. Ett inledande frågeformulär (ITQ), själva undersökningen samt en avslutande intervju. Frågeformuläret och intervjun gjordes i samma rum medan undersökningen gjordes i ett annat. Innan undersökningens början informerades alla försöksdeltagare (fd) om att målet med undersökningen var att interagera med en virtuell agent med hjälp av röstigenkänning. De blev också informerade om att den virtuella agenten skulle komma att instruera dem att utföra en uppgift.

Frågeformulärets syfte var att säkerställa att de två grupperna var likställda, både hur de uppfattar sin förmåga att leva sig in i en virtuell värld och deras datorvana. Själva undersökningen (se ovan för specifikationer angående antal par, antal manipulationer och så vidare) utfördes genom att använda en s.k. Wizard of Oz-teknik för att styra programmet av en person i samma rum som fd, dock bakom en skyddande skärm så att den ej skulle bli sedd. Det verkade inte som att någon deltagare upptäckte denna styrning, utan de agerade som om programmet reagerade på deras röststyrning. Personen som styrde programmet försökte i så god mån som möjligt att emulera hur ett program borde reagera mot röststyrning och gjorde detta på samma sätt oavsett hög eller låg immersionegrad. Den efterföljande intervjun gjordes för att ge deltagarna ännu en chans att uppvisa att de upptäckt manipulation och följde en standardmall för CB-briefing. (Se appendix X för en fullständig teknisk dokumentation om hur programmet var uppbyggt och vilka resurser som användes.)

Resultat

Concurrent detection (CD) kodades av två personer i gruppen vilka var blinda för immersionegraden hos försöksdeltagaren. CD kodades individuellt av de två för att sedan jämföras. I de fall då de inte var eniga diskuterades dessa fall innan kodarna hade vetskap om immersionegrad eller resultatet från efterintervjun.

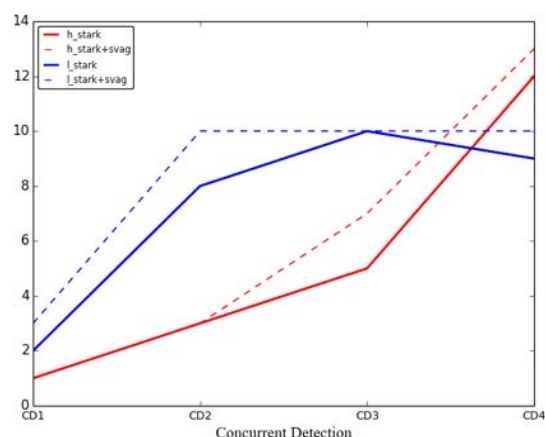
CD kodades i två grader. Hård vilken definierades som någon form av verbal upptäckt, antingen i form av att tydligt säga att det inte var det kortet de valde eller något vagare, men fortfarande verbalt, exempelvis: "var det verkligen det kortet jag valde?" Det kodades även en svagare form av CD. Denna kategoriserades av att försöksdeltagaren antingen rynkade på pannan, såg frågande ut eller att försöksdeltagaren svarade på ett sätt olik de andra paren där hen blev ombedd att förklara sitt val. Denna svagare

form av upptäckt är inte en lika tydlig upptäckt då det kan vara så att försöksdeltagaren svarar annorlunda eller beter sig annorlunda men egentligen inte blir medveten manipulationen. Det är också möjligt att försöksdeltagaren blev medveten om manipulationen men valde att inte verbalt reagera. (Se graf 1.)

Total upptäcktsgrad av CD var i hög immersion 27% (inkl. svag upptäckt = 31%). I låg immersion var den 38% (inkl. svag upptäckt = 43%). Ett t-test gjordes på total upptäcktsgrad i både stark och svag CD ($t(150)=1,4824$, 95% CI [-0,07; 0,5], p-värde = 0,14) utan att visa på någon stark skillnad på total upptäcktsgrad mellan grupperna.

Det syns också tydligt i graf 1 att upptäckten av manipulation skedde i högre grad tidigare i låg immersion. Redan vid andra manipulerade paret fanns det en skillnad i upptäcktsgrad gentemot hög immersion där främsta upptäckt skedde först vid det fjärde manipulerade paret. Även ett t-test mellan CD2 i båda grupperna visade på en statistisk skillnad dem emellan.

ITQ användes, som tidigare nämnt, för att fastställa att det inte var en snedfördelning i hur stor utsträckning försöksdeltagarna tenderar att involvera sig i en virtuell värld mellan de båda grupperna. Medelvärde på ITQ i hög immersion var 0,63 medan det i låg immersion låg på 0,59. Ett t-test visade ingen statistisk skillnad dem emellan ($p>0,05$). En linjärregressionsmodell (Total upptäckt ~ medelvärde ITQ) visade heller ingen signifikant påverkan av inlevelse tendens på upptäcktsgrad.



Graf 1. Concurrent detection i låg och hög immersion (blå respektive röd) med både stark kategorisering och stark plus svag (streckade linjerna).

5 Diskussion

Även om studien inte visade någon stor statistisk skillnad i total upptäcktsgrad mellan hög och låg verkar det ändå ske någon form av påverkan på upptäcktsgrad dem emellan. En större undersökning med fler deltagare skulle möjligtvis kunna urskilja vårt antagande att låg immersion ger en högre upptäcktsgrad. Man kan också spekulera i om den låga inte var låg nog utan de aspekter vi manipulerade kanske borde ha sänkts ännu mer.

Det var inte någon signifikant skillnad mellan hög och låg immersion i totalt antal CD däremot såg vi – vilket man kan se i graf 1 – att de har olika karaktär i när upptäckterna görs. Vad för olika tolkningar kan man göra om att låg immersion upptäcker tidigare än hög immersion? Vi kan spekulera utifrån de resultat vi fått: Kan det vara så att skillnaden mellan de två miljöerna beror på mängden stimuli i hög, det faktum att det helt enkelt finns mer saker att fokusera på. Då kan det vara att man upptäcker tidigare i låg för att man inte har något annat än ansiktena att fokusera på, med andra ord, det är inget annat som drar till sig uppmärksamheten?

Våra resultat verkar ligga i paritet med tidigare CB-experiment. Till exempel hade originalstudien av Johansson et al (2005) en total upptäcktsgrad på 26%. Detta skulle påvisa att det är möjligt att återskapa choice blindness effekten i en virtuell setting. Övergången från verklighet till virtuellt verkar inte ha en negativ påverkan på resultatet i allmänhet. Däremot verkar sämre datorprogram påverka resultatet om inte i total upptäckt så åtminstone på något sätt, i detta fallet i när de flesta upptäckter sker. Det kan därför vara bra att ha detta i åtanke vid skapandet av den virtuella miljön och interaktionen med virtuella agenter, samt vid sammanställning av resultatet från en sådan undersökning. Beroende på uppgiften i undersökningen kan det vara värt att lägga ner mycket tid på att arbeta upp ett bra program som har en bra mappning med verkligheten innan man testat det på riktigt så att säga.

Framtida forskning

Vi ämnar att vidarearbeta vår undersökning med en större statistisk analys samt jämföra vår hög immersion version med de tidigare undersökningarna utav Haake och Gulz vilka använde samma experimentparametrar (se ovan). Detta skulle ge en ännu bättre jämförelse med ett verkligt CB-experiment än att exempelvis jämföra med Johansson et al. (2005) där andra parametrar, exempelvis andra ansiktspar, användes.

Vår undersökning möjliggör även alternativ att antingen försöka öka presence och på så sätt möjligtvis även sänka upptäcktsgraden ännu ett steg, även om det öppnar upp frågan om det är relevant att försöka sänka upptäcktsgrad överhuvudtaget. Exempelvis så lär användandet av Oculus Rift öka graden av upplevd presence men frågan blir om det sänker upptäcktsgraden. Att be försöksdeltagare att göra ett "enkelt" ansiktspreferenstest och samtidigt be dem använda en "head mounted display" kan öka misstanken för experimentet och göra att man faktiskt ökar upptäcktsgraden.

Åt andra hållet så kan man även vidareutveckla vår låg immersion version genom att antingen sänka alla aspekter ännu mer eller fokusera på en del. Till exempel kan man fokusera på den virtuella agenten, för att genom att endast manipulera en aspekt möjligtvis kunna urskilja olika aspekter som skapar CB effekten individuellt.

Referenser

- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2), 103-124.
- Dehn, D. M., & Van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International journal of human-computer studies*, 52(1), 1-22.
- Fiedler, M., & Haruvy, E. (2009). The lab versus the virtual lab and virtual field—An experimental investigation of trust games with communication. *Journal of Economic Behavior & Organization*, 72(2), 716-724.
- Garau, M., Slater, M., Pertaub, D. P., & Razzaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 14(1), 104-116.
- Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, 310(5745), 116-119.
- Johansson, P., Hall, L., Gulz, A., Haake, M., & Watanabe, K. (2007). Choice blindness and trust in the virtual world. Technical report of IEICE: HIP, 107(60), 83-86.
- Johnsen, K., Rossen, B., Beck, D., Lok, B., & Lind, D. S. (2010, March). Show some respect! The impact of technological factors on the treatment of virtual humans in conversational training systems. In *VR* (pp. 275-276).
- Reeves, B., & Nass, C. (1996). How people treat computers, television, and new media like real people and places. CSLI Publications and Cambridge university press.
- Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence*, 10(3), 266-281.
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 6(6), 603-616.
- Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5), 560-565.
- Slater, M., Usoh, M., & Chrysanthou, Y. (1995). The influence of dynamic shadows on presence in immersive virtual environments (pp. 8-21). Springer Vienna.
- Welsh, R.B., Blackman, T.T., Liu, A., Mellers, B.A., Stark, L.W. (1996) The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence, *Presence: Teleoperators and Virtual Environments*, MIT Press, 5(3), 263 - 273.

Appendix I: Manus

Välkommen/Intro.

-Hej och tack för att ni vill medverka i denna undersökning. Jag är en virtuell agent och kommer vara undersökningsledare i denna undersökning som är ett ansiktspreferenstest.

Själva undersökningen kommer att ta ungefär 20 minuter därpå kommer det att följa en kort intervju, hela undersökningen kommer max att ta en timme

(Undersökningen är varken fysiskt eller psykiskt farlig,) Om ni känner er obekväma eller vill av någon annan anledning vill avbryta så är det bara att säga till. Det kommer det inte att påverka eventuella senare undersökningar.

-Undersökningen kommer ske helt via skärmen och ni kommer att kommunicera med mig genom mikrofonen. Med hjälp av röstigenkänning kommer jag att kunna höra vad ni säger så försök tala så naturligt som möjligt.

-Under undersökningen kommer jag visa totalt 16 ansiktspär. Jag kommer hålla upp dem två åt gången i tre sekunder, sedan kommer jag lägga ner dem, er uppgift blir då att välja det ni tyckte var mest attraktiv genom att säga antingen höger eller vänster. Tala högt och tydligt. Under vissa par kommer jag även att ställa några följdfrågor, men det är inga svåra frågor så det skall nog gå bra.

Vill ni höra instruktionerna igen eller är ni redo att starta undersökningen?

Undersökning (Ströfraser och följdfrågor att lägga in under undersökningen)

- Vad gjorde att ni valde detta ansikte framför det andra?
- Vad var det som stack ut som gjorde att ni valde detta ansikte?
- Varför valde ni detta?
- Fanns det något som gjorde att ni valde detta ansikte framför det andra?
- Var det något speciellt som gjorde att ni föredrog detta ansikte?
- Något annat som gjorde att ni valde detta ansikte?
- Var det någon annan del som gjorde att ni föredrog detta ansikte?
- Okej, något annat?

Okej, då tar vi nästa par.

- Tackar, Okej, då kör vi nästa par,
 - Nu har vi gjort hälften
 - Sista paret nu, (säga vilket parnummer vi är på?)
- Slut

Appendix II: Virtuellt Setting

Denna del är uppdelad i två delar. Först redovisar och motiverar vi hur vi byggt upp den virtuella miljön och den virtuella karaktären. Sedan kommer vi redovisa och motivera hur experimentet kommer utföras.

Vi har byggt upp den virtuella miljön efter fotografier av experimentrummet som användes i Haake och Gulz tidigare experiment. Vi har använt plattformen Unity för att replikera och bygga upp experimentrummet se figur 1 och figur 2 i appendix. I den virtuella miljön har vi försökt implementera så mycket av experimentrummet som möjligt tex bord, stol, fönster, persienner, element, bokhylla med tillhörande böcker och ett svart skrivbordsunderlägg. Väggarna har vi byggt genom att manipulera kuber som är ett sk geometri objekt i Unity. Man kan applicera ett material med specifik textur och färg på ett geometri objekt. Texturerna till materialen har vi byggt i GNU Image Manipulation Program (Gimp) och exporterat till Unity. Genom att modifiera och manipulera geometrin objekten har vi kunnat skapa det som finns i rummet till exempel bokhyllan, skrivbordsunderlägget och fönstret. Till fönstret har vi byggt en fönsterkarm, handtag och persienner genom att manipulera geometri objekt i Unity. Bokhyllan, skrivbordsunderlägget och lite annan geometri har vi byggt upp i Autodesk 3Dstudio Max. Andra objekt i den virtuella miljön som böckerna, elementet, stolen, bordet har vi importerat från gratis databasen Archive3D. Vi har modifierat några av dessa importerade objekten tex elementet har vi lagt till ett metallrör och flyttat på ventilen, som vi byggt i Autodesk 3Dstudio Max. I Unity har vi lagt till en himmel och byggt skog utanför fönstret av det virtuella försöksrummet. Vi har försökt replikera rummet i så stor grad som möjligt för att försöka skapa samma känsla i miljön för försökspersonerna i den virtuella miljön som den miljön som användes i tidigare experiment. Att rummet i den virtuella miljö liknar det verkliga rummet är väldigt viktigt för att skapa en så hög immersiv miljö som möjligt. Det var viktigt att vi med omsorg valde texturer och geometri objekt som har bra upplösning och liknar riktiga objekt.

Eftersom det behövs en virtuell försöksledare i den virtuella miljön har vi valt att replikera försöksledaren Betty Tärning från det tidigare studien. I AutoDesk CharacterGenerator skapade vi ett ansikte och en kropp till den virtuella karaktären. Detta gjorde vi genom att slå ihop olika ansiktsdrag/kroppsdelar från färdiggenererade ansikten/kroppar för att modellera fram ett ansikte och kropp som liknar Betty. För att animera Betty använder vi Bezier kurvor, dessa består av fyra 3D orienterade punkter. Det betyder att vi skapar en start och slut punkt där vi vill att kroppsdelen ska gå igenom och två punkter vilka kroppsdelen nödvändigtvis inte behöver gå igenom. Utifrån de fyra punkterna så skapas en kurva som beskriver kroppsdelen rörelse, genom att manipulera de två punkterna "mellanliggande" punkterna kan vi ändra kurvans form och därmed rörelsen. I Unity har vi programmerat i programmeringsspråket C#. För att knyta ihop Bezier kurvorna till de olika rörelser tex att visa ansiktsparen, lämna fram ett kort och ta tillbaka händerna behövde vi använda Inverse Kinematics. Med Inverse Kinematics kan vi låsa en led i den virtuella agenten och på så vis räkna ut var tex armbågen skall var i förhållande till den låsta leden. Detta behöver vi implementera för att kroppsdelen hos den virtuella agenten skall hamna på rätt ställe vid rätt tidpunkt. Vi har fått MotionCapture (MoCap) filmer från Magnus som var inspelade med rörelser från det tidigare experimentet. Vi har valt ut delar, med lite flimmer, av det inspelade MoCap materialet i programmet Autodesk Motion Builder. I Motion Builder kopplade vi sedan 3D punkt molnet till ett skelett för att kunna exportera en animation till Unity. Animationer vi importerat till Unity har vi använt för att lägga ovan på de programmerade animationerna. Detta för att rörelserna skall se mer naturliga ut. Vi har med en ansiktskamera Asus Xtion pro live spelat in ansiktsrörelserna i programmet Faceshift. I Faceshift har vi först tränat upp en modell och sedan spelat in ansiktsrörelserna och ljudet. Vi använde en skådespelerska för att läsa upp ett manus (se Manus i appendix). Dessa ansiktsrörelser kommer vi exportera till Unity och lägga på den virtuella agenten så att när programmet spelas upp kommer den virtuella agentens ansikte följa de inspelade ansiktsrörelserna. Eftersom ljudet spelades in samtidigt som ansiktsrörelserna med en mikrofon "Zoom H4". Vi har använt AudioCity för att ta bort brus och för att editera klippen till rätt längd. Längden på klippen behövde vi editera för att de skulle synkronisera sig till ansiktsrörelserna.

Appendix III: Immersive tendencies questionnaire

Immersive Tendencies Questionnaire: (http://nuweb9.neu.edu/gurmethods/?page_id=61)

1. Hur lätt har du för att känna dig djupt involverad i filmer eller TV-serier?

| | | | |
INTE ALLS NÅGOT VÄLDIGT ENKELT

2. Har du någon gång blivit så involverad i ett TV-program eller bok så att folk omkring dig har haft svårt att få din uppmärksamhet?

| | | | |
ALDRIG IBLAND OFTA

3. Hur uppmärksam känner du dig just nu?

| | | | |
INTE ALLS NÅGOT MYCKET

4. Har du någon gång blivit såpass involverad i en film så att du har missat saker som skett i din omgivning?

| | | | |
ALDRIG IBLAND OFTA

5. Hur ofta händer det att du känner att du kan identifiera dig med karaktärerna i en berättelse?

| | | | |
ALDRIG IBLAND OFTA

6. Har du någon gång blivit så involverad i ett dator- eller TV-spel att det känns som om du vore i spelet istället för att gå omkring med en handkontroll och titta på en skärm?

| | | | |
ALDRIG IBLAND OFTA

7. Hur kry/vaken känner du dig idag?

| | | | |
INTE ALLS NÅGORLUNDA BRA

8. Hur bra är du på att ignorera distraktioner från omgivningen när du är engagerad i något?

| | | | |
INTE ALLS NÅGORLUNDA BRA

9. Om du ser på sport, händer det då att du blir såpass engagerad i matchen att du reagerar som om du var en av spelarna?

| | | | |
ALDRIG IBLAND OFTA

10. Har du någonsin blivit så inne i en dagdröm att du missar saker som sker i din omgivning?

ALDRIG IBLAND OFTA

11. Har du någon gång haft drömmar som känts så verkliga att du känner dig desorienterad när du vaknar?

ALDRIG IBLAND OFTA

12. Tappar du någonsin bort tiden när du engagerar dig i någon form av sportaktivitet?

INTE ALLS NÅGOT HELT

13. Hur bra är du på att koncentrera dig på saker/aktiviteter som tilltalar dig?

INTE ALLS NÅGORLUNDA BRA

14. Hur ofta spelar du dator- eller TV-spel?

ALDRIG IBLAND OFTA

15. Har du någonsin blivit uppgjagd av en biljakt eller dylikt på film eller i TV-serier?

ALDRIG IBLAND OFTA

16. Har du någon gång blivit rädd för något som händer i en film eller i en TV-serie?

ALDRIG IBLAND OFTA

17. Har du någonsin förblivit rädd eller orolig en lång stund efter att du sett en läskig film?

ALDRIG IBLAND OFTA

18. Blir du så inne i att göra något att du tappar känslan för tid?

ALDRIG IBLAND OFTA

19. Hur många timmar om dagen, i genomsnitt, spenderar du framför datorn?

0-1 4 7-8+

Appendix IV: Igroup Presence questionnaire och social presence

Igroup Presence Questionnaire: (<http://www.igroup.org/pq/ipq/index.php>)

Social Presence: Garau, M., Slater, M., Pertaub, D. P., & Razaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 14(1), 104-116.

- I den datorgenererade världen upplevde jag en känsla av att "vara där".

Inte alls				Väldigt mycket

- På något sätt kände jag mig omsluten av den virtuella världen.

Instämmer inte alls				Instämmer helt

- Det kändes som att jag bara betraktade bilder.

Instämmer inte alls				Instämmer helt

- Jag kände mig inte närvarande i den virtuella miljön.

Jag kände mig inte närvarande				Jag kände mig närvarande

- Jag hade en känsla av att jag gjorde saker inne i den virtuella världen snarare än att jag kontrollerade något utifrån.

Instämmer inte alls				Instämmer helt

- Jag kände mig närvarande i den virtuella världen.

Instämmer inte alls				Instämmer helt

- Hur medveten var du om omvärlden medan du befann dig i den virtuella världen? (t ex ljud, rumstemperaturen, andra personer etc.)

Extremt medveten			Totalt omedveten	

- Jag var inte medveten om min omgivning.

Instämmer inte alls			Instämmer helt	

- Jag var fortfarande medveten om den fysiska omgivningen.

Instämmer inte alls			Instämmer helt	

- Jag var fullständigt uppslukad av den virtuella världen.

Instämmer inte alls			Instämmer helt	

- Hur verklig upplevde du att den virtuella världen var?

Fullständigt verklig			Inte verklig alls	

- Hur stor överensstämmelse upplevde du att det fanns mellan den virtuella världen och den verkliga världen?

Ingen överensstämmelse			Stor överensstämmelse	

- Hur verklig upplevde du att den virtuella världen var?

--	--	--	--	--

Lika verklig
som en fanta-
sivärld

Oskiljbar från
den verkliga världen

- Den virtuella världen framstod som mer realistisk än den verkliga världen.

--	--	--	--	--

Instämmer inte
alls

Instämmer helt

- Kändes det som om du var ensam i rummet eller i ett rum med en annan människa?

--	--	--	--	--

Helt ensam

Med en annan
person

- Hur ofta talade du till försöksledaren som om hon vore en människa?

--	--	--	--	--

Aldrig

Hela tiden

- I vilken grad kändes det som om du och försöksledaren var i samma rum?

--	--	--	--	--

Aldrig

Hela tiden

- Under hela experimentet, hur ofta kändes det som om försöksdeltagaren förstod dig?

--	--	--	--	--

Aldrig

Hela tiden

- Under hela experimentet, hur ofta kändes det som om du pratade med en annan människa istället för en datogenererad karaktär?

--	--	--	--	--

Aldrig

Hela tiden

Cognition, Feedback and Educational Software Mathematics

Ryo Yamazaki, Andreas Stephens, Hannes Johansson, Johanna Petersson

yamazakish@hotmail.com, andreasstephens@gmail.com, hannes.johansson91@gmail.com, min.mail@jonannap.se

By focusing on feedback and the embodied natural mathematical competences of human beings it is possible to provide more efficient educational software that is grounded in cognitive scientific theory. Using conceptual metaphors such as ‘object collection’, ‘points on a line’ and ‘containers’ offers a conceptual base that is in line with these natural competences. The following five principles are proposed to provide a scientifically grounded base for educational software design that can optimize the learning experience for students: problem solving; object representation; world feedback; interactive workspace, and; gradual transferal. Lo-fi, Mid-fi and Hi-fi prototypes were developed during the project and compared to an educational software on the market, with a focus on these principles. Three exploratory pilot studies highlighting central aspects of the Hi-fi prototype were conducted. A transferal test was conducted to investigate our hypothesis; that our prototype would invoke successful strategy transferal to a high degree, with a high level of implicit understanding of the solution. The results from the transferal test were elucidating even if inconclusive with respect to our hypothesis.

1 Introduction

With the increasing availability of technology in today’s classroom, in terms of computers, tablets and smartphones, contemporary mathematics education ought to incorporate these assets to facilitate learning in the most efficient way possible with the help of educational software (Ginsburg, Jamalain & Creighan, 2013). Compared to conventional modes of teaching, such as “the teacher speaking in classroom”-methods, educational software has the benefit of being possible to be personalized to match individual students’ levels and development. It is also cheaper and less dependent on the quality of the teacher. Furthermore, technology can have added interactivity, and offer wider availability for feedback and multi-modal learning.

This paper will investigate educational software design, with a special focus on feedback, since it is through feedback that students see the results of their actions or inputs. However, a number of the educational software for elementary school mathematics on the market seem to only translate traditional math-books into digital form. Recent investigations of the educational softwares available in commercial application stores have shown that many are insufficient in their content and design (see, e.g., Ginsburg, Jamalain & Creighan, 2013). While simply translating books to digital

format is indeed one way to go, the benefits of digital learning might be severely undermined by limiting educational software’s to such formats.

From the stance of embodied cognition, we will investigate what good educational software should incorporate to go beyond the digital-book learning experience, and how this is most fruitfully presented to the student as feedback. We will then analyze the educational software “Qnoddarnas Värld”, a typical teaching app for children age 7 to 9, and propose potential areas in which it can be improved. These changes are discussed in relation to a number of Lo-fi and Mid-fi prototypes and their implementation in a Hi-fi prototype. Lastly our created prototype is tested in an informal exploratory pilot study, to judge if the theoretical implementations are indeed successful, as well as to indicate strengths and weaknesses in our design.

2 Theoretical Discussion

2.1 What is Good Educational Software?

In defining what successful educational software should incorporate it is necessary to analyze two factors that define the value of educational software; the limitation of non-digital mathematical learning, and the potential of digital learning. In other words, if there is something that educational software can do that would be beneficial for the learning process but is not possible in textbook form, those are the things that should be implemented.

While there are several things a tablet or computer can do that a book cannot, we will limit our focus on three elements. First comes the ability of educational software to react to a student’s action, thus allowing interactivity and a means to give feedback. While this can be provided to a student by a teacher, an educational software can follow each step in a learning process for *each student* during the same period of time, something that is impossible for a teacher. Second comes the possibility to show visual aids continuously and if needed in animated form. Lastly there is the possibility to provide a simulated world to the user through the calculatory capability of the hardware. These three elements, each based on the continuous feedback relationship between software and student, are not necessarily clearly distinct from one another. For example, touching an object on a screen and then moving it, uses all of the elements at once. They will however suffice as a general example of what educational software can do.

The question of what conventional textbooks are lacking is another issue that does not have an absolute answer. To properly answer that question a comprehensive theory of the most ideal mathematical education is needed. The difficulty of that is that an educational setting is rarely uniform. Even when defining how classroom mathematics should be taught there are unpredictable factors such as student motivation, specific learning goals and strategies, interactions between the student and the learning environment comprised of, among other things, teachers and other students (Hierbert & Grouws, 2007). Since all of these factors intertwine, experimental results and previous successful approaches may be hard to replicate. It is then necessary that a comprehensive model of what constitutes good mathematical education needs to take it all into consideration and also explain a valid approach to each and every setting. Such an all-encompassing explanation of mathematical education is not plausible to obtain at this point. We will instead be forced to approach the issue of what mathematical educational software should do by assuming certain general theoretical standpoints.

2.2 *The Embodied Mathematical Approach*

As the theoretical framework for defining a good educational software learning material we chose to focus on the concept of embodied mathematics, as expressed by Lakoff and Núñez (2000). The key point of the embodied mathematical approach is that human mathematical abilities are grounded in our real perception of real life objects and aspects, not in an abstract plane.

Humans, as well as various other animals, are born with innate mathematical competences, such as basic numerical capabilities (Lakoff & Núñez, 2000). This is also discussed by, for example, Plotkin (1993) who argues that humans, as well as most other animals, have constrained and directed abilities to learn and develop their intelligence, and hence are not born tabula rasa (Plotkin, 1993: 131–2). Babies can, for example, at a very early age discriminate between small amounts up to about four. To a large extent these capabilities are unconscious. From these innate simpler capacities we build more abstract mathematical concepts using conceptual schemas and metaphors. With this in mind, arithmetic can be described, and understood, using the metaphor ‘object collection’ (Lakoff & Núñez, 2000: 54–60). The neural connections that are built up by the basic number capabilities and experiences of objects being added and subtracted constitute conceptual metaphors. To be able to grasp arithmetic, a student needs both the innate capabilities and conceptual metaphors (Lakoff & Núñez, 2000: 55). This means that educational software should take these capabilities and metaphors into account when constructing exercises to make a good educational tool and the feedback it provides.

Object collection is a good metaphor for arithmetic exercises and Lakoff and Núñez argue that the different fields and ideas in mathematics all can be described in similar embodied ways.

Assuming that real world objects are indeed the simplest and most natural form of mathematics, what an educational software should utilize is the intuitive sense that concrete representations gives.

There have been well-documented cases of transferability problems in people who can solve everyday mathematical problems, but who are not able to do the same task presented

as an abstract mathematical exercise (Lave, 1988). While the more extreme claims of situated cognition and the impossibility of task-transferability are overstatements (Anderson, Reder & Simon, 1996), there is nonetheless a value in observing the relationship between concrete task solving and abstract understanding. If we assume that concrete problems are more natural than abstract ones as well as useful as a basis to obtain abstract mathematical understanding from, then this should be used when learning mathematics. Many textbooks do use concrete object representation to illustrate mathematical problems in the form of pictures and example. With educational software there is the added possibility of making these objects interactive and possible to manipulate and thereby make them a stronger learning support.

A key difference between mathematics education and other fields such as social science is that students to a lesser extent can relate what they're learning to actual world interest (Stodolki, Salk & Glaessner, 1991). If mathematics, like any other science, has its base in experience and perception, the primary focus on the abstract can explain the widespread dislike and feeling of alienation many students feel when confronted with very abstract mathematics (Ashcraft, 2002). By using an embodied cognitive framework in mathematical education there is a possibility to help make more sense of mathematics to those who didn't feel confident with it previously.

What we then need to do is identify how the above-mentioned theoretical framework could translate into effective educational software. If mathematics requires either a real or mental world to be meaningful, educational software could provide a digital world for the student.

2.3 *Five Principles Concerning Embodied Mathematics and Feedback*

The basis of all five principles to be presented is that we learn concepts through seeing them in the world. At the very ground of our approach to the education of mathematics is the continuous interaction between the user and the world, where the student acts upon and gets feedback in response. In this sense, many of our principles will be about feedback. The first principle will explain how a world can be created in which the student feels there is a possibility and reason to interact with. The second principle will explain why representations in the world are useful in showcasing relationships and rules. The third principle will explain how the reaction of the world to the answer the user inputs can help lead to a better solution. The fourth principle will explain the possibilities of additional interactive tools in the world facilitating cognitive tasks. And the fifth principle will explain our approach in how we can take solution strategies garnered from our digital world and transfer it to abstract mathematical problems.

2.3.1 *Principle of Problem Solving*

Solving a Problem Rather than Doing Math

In harnessing the intrinsic mathematical capability of the student it is important to note the distinction between mathematical problem solving and mathematical school exercises. While mathematical problem solving is nearly universally available for humans, math-avoidance and math-anxiety is

frequent in the adult population (Betz, 1978). Such fear seems perplexing if we assume that mathematical cognition is natural to us. A potential cause for the anxiety is thought to lie in the nature of how we teach mathematics. Mathematical education, in contrast to other subjects, rarely promotes experimentation, assumptions or trial and error. Instead a binary scale of right or wrong judges the student. This means that when a student commits an error there is no help to gain about what did go wrong or how to fix it (Furner & Berner, 2003). From the point of the embodied mathematical approach such a method of teaching is flawed. In real life problem solving, it is unlikely that one is confronted by a judgment of “wrong” when committing an error. Indeed the type of error and the context, in which it occurs all give different responses, with some having a potential negative consequence for the solver, while others might have no consequence at all, provided the solver reacts and manages to rectify the error. Real life use of mathematical abilities rarely follow an “answer and be judged”-format, but rather a “try and solve”-format.

The discussion above leads us to a new approach in designing mathematical exercises. Instead of focusing on the mathematical task as something that requires one answer that is either right or wrong, it can be useful to provide mathematical tasks that do not involve judgment. By returning to the concept of simulated reality in educational software it is possible to design mathematical problems as a scenario that can only be solved through mathematics. This type of task is sometimes called *anchored instruction*, and has been shown to have potential benefits above regular instructions (Bottge, Heinrichs, Mehta & Hung, 2002). Implementing such a task in a digital world could have the benefit of potentially suppressing the anxiety related to “test condition math”. Furthermore it can escape the paradigm that mathematics requires testing from a teaching authority, when instead the success of the solution is given so that the student him- or herself can see it from the reaction, or feedback, of the world. The knowledge that all of this is done in a simulated and digital form may further aid students who are cautious about mathematics to want to experiment and feel less fear of making a mistake.

The Relation Between Task and Learning

‘Learning by doing’ is a concept that is applicable to mathematical education. Studies have shown that students only need to work out examples and problems without explicit instructions to fully learn to solve and understand mathematical solution strategies (Zhu & Simon, 1987). With that said there are certain factors worth observing in how to design a specific task for a certain goal. Above we noted that a mathematical problem could be expressed as semi-realistic anchored scenarios of actual problem solving. It is important that the method of solving that scenario does not diverge from the actual mathematical strategy to be learned. Mathematical problems are at a simpler level well-defined convergent types of problems – meaning that all elements of the problem are given and that there is one goal state to be achieved (Jonassen, 1997). The general solution in solving a convergent problem is through manipulating one or several of the elements through certain rules that are allowed, and through one or several steps leading it to the goal. Two mathematical problems are equivalent to each other in terms of problem solving strategy if all the factors that define them

are identical, namely the goal, the elements, and the allowed moves. Thus a valid transition of a mathematical textbook problem into a digital scenario should assure that these three factors are identical, or at least not significantly different.

When learning to solve a mathematical problem, the relevant knowledge is to understand how to move elements so that the problem approaches a goal state. The problem should provide an opportunity for the student to test possible manipulations and compare them to the desired result. The goal state and elements of the problem solving can be explained through different feedback, for example either visual or verbal information, in educational software. Allowed movement arguably needs less instructions in an educational software compared to solving mathematics on paper, since restriction in the program along with feedback can lead the student in what moves can be done. Additional aid in how to solve the problem step by step can be given through visually worked out solutions. Through seeing how one might solve the exercise the student can focus on understanding the solution strategy, and minimizing the cognitive load, that is not relevant to the learning task at hand (Paas, Renkl & Sweller, 2003).

Promoting Intrinsic Motivation to Task

The benefit of educational software compared to conventional learning material is often presented in terms of educational software being more entertaining than conventional school education. There is often an assumption that digital education material can harness the motivating and addictive characteristics of video games. Indeed game-based learning has been shown to be more effective than traditional forms of learning (see, e.g., Tüzün et al., 2009; Kebritchi, Hirumi & Bai, 2010). Furthermore, children are to a larger extent familiarized with technological material (Divjak & Tomić, 2011: 16). However in implementing game-like “fun” elements it is important not to focus on the surface properties. Many game-like educational softwares suffer from an awkward discrepancy between the fun game-part and the boring learning-part (Ginsburg, Jamalain & Creighan, 2013: 107). Such an implementation will at best bribe the user into complying with the learning for the benefit of enjoying the rewarding game-part. This is an example of so called external reward (Habgood & Ainsworth, 2011). Previous research has shown that external reward might have a short-term benefit in motivating the user to perform a specific task, however it may also lower the actual motivation for users to continuously perform (Deci, 1971). While recent studies are not entirely consistent with these ideas, there is nonetheless evidence indicating that an approach that does not utilize external reward but rather encourages intrinsic reward – that is finding a pleasure in the learning task itself – has an increased benefit in aiding learning on short- to mid-span, with a better transferability of the ability gained through such a method also observed (Habgood & Ainsworth, 2011).

To utilize intrinsic motivation in educational software, what is necessary is to incorporate the game-like elements in the tasks themselves, not in a side-dish manner that disrupts the flow of learning. To do so can mean that the interactivity and the strategy development for the tasks can be done in a game-like setting where students are given a natural narrative that relates to content of the task. Game-like element such as interaction and a rewarding visual or auditive feedback effect

for solving the level can also be used, so to not seem out of place in relation to the task performed.

Another important factor for motivating the student is in how educational software challenges the user. While the idea of challenges is naturally engaging in games, a learning challenge can have a negative connotation. This is because the failure to complete the challenge in learning can have negative consequences and furthermore the challenges are often presented in an arbitrary and one-after-the-other type of fashion in mathematics education. Gee (2003), argues that competent video-game developers have construed an effective way to make the challenges entertaining rather than frustrating in their games. Gee notes that games usually have a sort of structure, in that there is a tutorial-like initial stage where one new type of knowledge is given, and the following stage will be about mastering them, until a boss eventually comes up where many of previously gained skills need to be utilized. In presenting new knowledge in a meaningful and gradual manner the challenges are never too hard for the user. Arguably a similar layout of problems is already present in conventional mathematics education. Yet one missing aspect in early-level math is that early knowledge is seldom tested. For example, one chapter usually consist of one type of question, and when it is done the other starts, where again only the new type of question is tested, the previous learning seeming to be forgotten. The user gets no sense of progression in learning, since the efficacy from previous learning is virtually not relevant in the newer chapters. To avoid this we suggest that newer more puzzling math questions in the educational software should only be solved through the skills gained several chapters previously. In doing so the user can get a sense that they are getting gradually more competent, and feel a pleasure in the challenges.

To summarize this principle the main point is that digital educational software can distance itself from abstract math learning. Instead of giving a task that is merely "Add these numbers" it can give a scenario in which the user needs to be able to add the same numbers to solve. By doing this the user will not need to interpret the meaning of what he or she is doing. Instead the focus will be shifted to how to solve the problem.

2.3.2 Principle of Object Representation

Working Memory Load and Representation

Within the framework of embodied mathematical cognition it is possible to incorporate the idea of working memory. Human cognition is limited in the capacity of processing (Baddeley, 1992). This implies that having to tend to simultaneous tasks can mutually hinder the performance of each. In terms of learning it is also rarely the case that the student can be immersed in exactly the content of information that he or she wants to learn (Paas, Renkl & Sweller, 2003). Learning is thought to be a process of solidifying mental processing in working memory to a recallable long-term memory schemata (Paas, Renkl & Sweller, 2003). Certain cognitive tasks involved in learning can be of little to no value in the process of creating long-term memory schemata. A textbook can be deemed demanding when the content structure is hard to understand and when it requires additional cognitive load by the reader to make sense of text. Similarly an instruction that requires extraneous working memory load by the user to

simply understand it can hinder the acquisition of schemata through straining the limited cognitive capacity (Paas, Renkl & Sweller, 2003).

The theory of working memory load is applicable in a more general sense when it comes to mathematical education. Firstly, it will mean that it is necessary to allow users to focus specifically on the learning, and not on the irrelevant aspects of the learning material. An ill-designed educational software requires of the user to follow arbitrary rules or systems that provide little or nothing in terms of the building block of mathematics. Instead it strains the cognitive capacity of the user, and perhaps even worse, gives the false impression of that the arbitrary rules are important to be remembered. Furthermore, reducing the time spent on understanding the system will also create more time for the actual learning experience. To reduce such unnecessary load for the student it is important to design educational software so that its rules are clear and intuitive, and that it is flexible in what input it accepts. 'Intuitive', in this context, would mean that the 'laws' or rules of the educational software are understandable from that of real world laws and understanding. In other words, the system image must make the user understand the designer's model (Norman, 1988: 16). Norman argues that this is efficiently done by affordances and constraints that are built into the relevant system (Norman, 1988: 9–11, 82–92). In the following sub-sections these and other aspects of learning and educational software will be discussed in more detail. An effective strategy is to clearly indicate what aspects of the screen are relevant at any given point, and what parts of it allow interaction. Such attention bringing to relevant features can be done visually, verbally and through the agent's internal understanding of the rules of the educational software.

Secondly, the cognitive load of mathematics can be assumed to be larger than the sum of understanding the rules and learning the content. Indeed our original assumption of embodied mathematics assumes that humans use mental representation in solving mathematical problems (Lakoff & Nuñez, 2000). Mental representations can be manipulated to solve problems, but in doing so it may require additional cognitive load. Keeping mental representation in the head is a cognitively straining task, called representational holding (Moreno & Mayer, 2007). In so called mental rotation tasks a person is required to rotate the representation of an object that he holds in his or her mind. Doing so intervenes with simultaneous visual cognitive tasks, indicating that they both utilize the same capacity (Hyun & Luck, 2007). Another famous example is the Tetris rotation task, in which most Tetris players found it easier to manipulate the digital Tetris block by pressing a button, even though they were fully capable of rotating the block in their mind and solving the problem without additional physical activity (Kirsh & Maglio, 1994). By off-loading representation to the real world we can easily through our senses obtain the relevant aspects of the representation, without it needing to be constantly imagined in the head (Zhang & Norman, 1994).

Object representation is then something that takes effort, but can be optimized through creating the representation in the world itself. Much like the rotation and manipulation of a digital object can be easier to grasp than the mental rotation and manipulation of the same object, mathematical manipulation of representation could be improved by having them

externalized and offloaded into the software. The concept itself might not be new, considering that traditional mathematical education utilizes pen and paper to externalize certain aspects of problem solving. Still, the interactivity and continuous change of visual feedback available in educational software's allow for easier and more meaningful representations to be externalized than in paper form.

Natural Metaphors and Choice of Representation Type

The importance of developing models that most efficiently represents the abstract ideas of mathematics is discussed by for example Ginsburg, Jamalian and Creighan (2013: 91). What can be expressed through object representations are mathematical entities – at the most basic level, numbers. Any natural number can be expressed through general measures such as length, weight, volume, position or amount. Furthermore different mathematical terms can be expressed in a consistent system using several measures. For example, an interval can be expressed as the length of a line, while a number can be expressed as a point on that line (Lakoff & Núñez, 2000). Some combinations can be incoherent and unnatural. For example, representing a number through length in one instance and through weight in the other will render the aspect of comparison meaningless. Some choices of representation can suffice, but might nonetheless be less than optimal, in illustrating a mathematical entity and a problem. The representation of an amount of objects for example is intuitive and graspable when the number is small, and the task is to add several stacks of them together. However, when the number gets large, such as in the hundreds, it is less apparent to see how the representation can facilitate the addition. Each representation thus has different characteristics, which is understandably so considering that mathematics is used to solve a wide range of world problems, where in each case the number might mean something different. The choice of representation should be decided on the base of the nature of the task and the types of mathematical entities involved.

While the trait of objects can be used as representation of mathematical entities it is also relevant to consider the type of object to choose. Different types of objects can have different implication and meaning depending on cultural cues. For example, it is hard to accurately represent the number 2301 in number of beads, but it can easily be done when instead it is imagined in amount of bills of money. Choosing money as a way to represent something utilizes the student's worldly knowledge of how a 100 bill is ten times as valuable as a 10 bill. Thus the need for the student to learn a new complex system disappears.

In all cases it is important that the total problem space utilizes a consistent model of representation, so that the result of the mathematical task can be interpreted through the result of the representational aid and vice versa. An effective model can aid the student in understanding relationship between mathematical entities (Ginsburg et al., 2013), and thus apply the concrete knowledge of what is going in the digital world, to understand mathematical rules.

To summarize this principle, there are several benefits of making mathematical concepts visible digitally in the software. Coupled with the principle of problem solving this leads to a coherent world in the software, which the user can attempt to solve the task. If the user gets faced with a problem “You need to solve scenario A with objects B”, then he

can form real hypothesis about it from real world knowledge. This approach thus distances itself from learning through memorization but approaches a more natural learning through world. How this element of learning can effectively be added is discussed in the next principle.

2.3.3 Principle of World Feedback

Types and Levels of Feedback

A fundamental point in the “learning by solving”-approach is that you can gradually come closer to a solution through the mistakes you make. For if you did not make mistakes from the beginning, it would mean you did not need the additional learning, and if you did commit errors without learning from it, the process would be more about guessing and hoping rather than learning. It is thus important to construct a creative environment for efficient learning where feedback plays a natural role (Kinzer, Bradley & Morandi, 2013).

There are numerous ways to categorize feedback processes. Positive feedback can be understood as an indication that the student has answered correctly, while negative feedback is an indication of an error. Furthermore, feedback can be either motivational, evaluative or descriptive (Kinzer, Bradley & Morandi, 2013). Different categories might be variously effective due to contextual aspects (Hattie & Timperley, 2007). In many cases corrective feedback – right or wrong feedback – that gives the student the “result” of her chosen action or answer, seem to be a natural “first choice” in educational software's, most likely since it is relatively easily implemented. This can be done in multiple ways. Sounds, colors and symbols – as well as other implementations – can be used to indicate the corrective result. These more typical feedback signals use the culturally learned background knowledge that the students are thought to have in common. Classical examples of this kind could be sounds of different pitch with a positive or negative connotation, or green check symbols for right and red cross symbols for wrong. Corrective feedback however only provides surface level information of the error. A negative corrective feedback can make the student check for any conceivable errors, or if such are not found it can make him or her reconsider his or her solution strategy (Hattie & Timperley, 2007). Being told you did wrong only helps if the student is in the position to be able to by him or herself to go back and find the error. If however the level of understanding is not sufficient, simply being confronted with “wrong” as feedback is of little help. Clearly effective feedback should go beyond the surface level, and give cues to help understanding of the error and the level of progress, so that the student can form a solution strategy from the information given through the feedback.

Formative Feedback

A more improvement oriented feedback is what is called formative feedback. Unlike corrective or summative feedback which has its main intention to evaluate and indicate the correctness of an answer, formative feedback instead focuses on helping the user understand his answer and action, and thus to set the user on the right trail towards a better solution (Shute, 2008). Formative feedback is in its definition a more information bearing type of feedback than that of merely summative feedback.

We argue that an effective way of providing formative feedback is through implication – consequence – feedback. Above mere indication of the correctness of an answer, implication feedback presents the consequences, e.g. the physical implications, of what that answer would mean. With this type of feedback the outcome of the chosen action leads to concrete results that are shown to the student. This causal indication of an error or correct answer is closer to problem solving scenarios encountered in real life. We face a problem, attempt a solution, and see how the world reacts to the solution. The added benefit of implication feedback is that it can illustrate the specific type of error and the magnitude of such an error. If a student gets the task to measure the volume of a glass, so that he or she can fill it with the right amount of water, it is easy to illustrate the correctness of an answer through implication feedback. If the amount chosen is too high, there will be too much water and the glass will be overflowing. If the amount is too small, the glass will only partly be filled. In each case, the student is informed with the type of error (too high or too low value), but also with the size of the error. Even though this way of feedback does not explicitly tell which part of the solution strategy that caused the error, it nonetheless provides a tool for systematically investigating it. The student can change the parameters, experimenting with the program and see how each strategy, or input, affect the error. In essence implication feedback grants a way for the student to experiment, much like in real life. When finally the correct solution is realized the student will hopefully not only have knowledge of how an answer is correct, but also why all other answers are incorrect.

On Demand Help

However, to expect all students to be able to experiment their way into a general solution to a mathematical problem is perhaps asking too much. A metacognitive strategy of how to generally find a solution might not be present a priori in each student. However the assumption of embodied mathematics implies that even though such a strategy is not being formed and used by all or even most students, the *ability* to form and use it is near universal. Clearly what is desired then is knowledge about how to support students into forming these strategies. In real life problem solving there is an additional aspect of finding a solution, namely the possibility to ask for help. While help usually is provided by the teacher, an educational software too may provide a help-function to lead the student to the correct answer in case a strategy cannot be found by the student herself. These types of on-demand help can aid learning, but are not always used effectively by students that could benefit from them (Aleven, Stohl, Schworm, Fischer & Wallace, 2003). To make the help-function useful, our approach is to make the way help is given mimic that of real life help seeking from tutors. Help will only be given when the user feels the need for it, since excessive help can deter from the focus on the task (Hierbert et al., 1996). In our approach the student needs to be able to pinpoint the type of information that he or she is in need of through reflection of his or her own understanding of the problem structure. The different types of help to each task can be expressed in a list, each having natural language question as a base, e.g., “How do I go further from this point?” If such a system is successful the student will hopefully gain a meta-strategy to seek

effective help through relevant questions, relating to his or her own level of progression.

In short the principle of world feedback gives a tool for the user in the digital software problem space to approach an answer through her own actions. The world reacts to one answer and provides signals to allow the user to infer what went wrong and how. The information in this signal, needs to be qualitatively sufficient for the user to be able to use it as a tool in the next attempt at solving the problem. To achieve this it is necessary to imagine the scenario of task the user is attempting to solve, and to use his or the program’s representation of the problem solving strategy to indicate what was different from the goal state.

2.3.4 Principle of Interactive Workspace

Epistemic Action

When presenting the principle of object representation we briefly mentioned how Tetris players manipulate the digital world, to off-load the cognitive load posed on internal processing (Kirsh & Maglio, 1994). Having representation in the world decreases cognitive demand, but it also allows for a specific type of action, called epistemic action. In traditional cognitivist view actors have certain goals and intentions, and through internal cognition they define how to achieve this, upon which the solution is acted onto the world. Epistemic action is different from this pragmatic sort of action, in the sense that it does not have the aim to change the state of the world through an action, but rather to change the cognitive state of the agent her- or himself through the action (Kirsh & Maglio, 1994). Epistemic action can use objects in the world to simplify the computational task that the user is in and reduce the risk for error.

In real life, we often employ strategies to simplify the computational demand through making the relevant sensory aspect tangible. If asked to compare the length of two objects, we can usually judge the rough relationship just by looking. But when the difference is small, and there is reason to be careful, we can pick the two objects and place them next to each other, thus making the difference visible at a single glance. This is a case of epistemic action, in which we place objects in the world to make the answer to a question visible in the world itself. Off-loading the cognitive demand and task into the world itself is therefore a natural solution strategy we employ, and in constructing educational software that replicates real life problem solving scenarios this is something that should be taken into consideration.

Understanding by Using

The concept of learning by interacting with the world has been applied to mathematics through what is called concrete manipulatives, where users can manipulate something concrete to gain an understanding of the abstract ideas behind. This method has, in a number of studies, for example presented in a meta-analysis by Carbonneau, Marley and Selig (2013), provided an effect on learning, albeit small- to medium-sized. The concept has been successfully implemented into educational software’s, in so-called virtual manipulatives. Students who learned fraction through virtual manipulatives gained understanding through experimenting and comparing the representation with the fraction, and exhibited

less tendency of common fraction addition error (Suh & Heo, 2005). Concrete or virtual manipulatives differ from static representations in the regard that there is an element of interactivity and possibility of experimentation (Moyer, Bolyard & Spikell, 2002). When implementing this to an educational software it is therefore necessary to reflect upon the nature of the concept that gets represented.

Digital Aid as Manipulative and Epistemic Action

Epistemic action uses real world relationships and sensory information to facilitate and reduce error in complex cognitive tasks. Virtual manipulatives allow students to interact with a parameter or tool to understand the underlying relationship. We argue that the benefits of both can be implemented into an educational software through creating a digital interactive workspace which abide certain mathematical rules. The idea is to create virtual manipulatives within a problem that can be used as a mean to transform the workspace so it offloads certain cognitive demands to the world. A very basic implementation of this would be to take examples of epistemic action that is the norm in traditional mathematical education and make them automatized in the digital world. An example would be the restructuring of the numbers used in paper arithmetics. Most children learn in school that the addition “325+452” can be solved on paper through columnar addition, where the two addends get aligned so that one of them is placed on top of the other. Then the complex task of adding two three-digit addends becomes reduced to step-by-step adding each column to each other. The strategy involves taking one problem and separating it into several sub-problems, which is then facilitated through the layout of the numbers on the paper. The latter part of this strategy can be said to be a very common type of epistemic action, since the solution strategy “take the first digit of each addend and count the sum of it, then take the second digit of each and count the sum of it...” is essentially doable without actually columnally aligning the two addends. The replacement is done just to facilitate the actual cognitive task and reduce the risk of error. But to be able to do this restructuring the student needs to be in position to understand that the addition “325+452” written in one line is equivalent to the same two addends written columnally on top of each other. The equivalence of an addition, can be expressed within the software as a manipulative to be used. Indeed through implementing a manipulative as a help tool within a problem, a simple equivalence relationship, which is hardly of very interest were it to be implemented independently, gains substantial use. Letting the software interactively restructure the problem elements is then a way of making students epistemically act according to a rule within the world expressed as a manipulative.

This notion can be taken further, so that some instances of manipulatives can be available in certain problems, and that others can be thought of as a more universal tool kits. Arithmetic abilities such as addition and subtraction are integral parts of solving many mathematical problems but nonetheless distinct from the ability to form and understand solution strategies. There might be a benefit in off-loading this task to the world, so the student can focus on learning and understanding the solution strategy at hand. A meta-analysis of studies analyzing the effect of calculators in precollege math education show substantial benefits in conceptual and opera-

tional mathematical skill as a result of testing and learning with calculators (Ellington, 2003). Although effects are less certain when it comes to specifically elementary school level mathematics, this may also be due to the fact that studies regarding the effect of calculator integral education at an early level is scarce, and that learning to effectively use calculators at an early age might be a difficult task in itself. With today’s technology however, the control of simple arithmetic can be done intuitively and without much prior knowledge. An example would be to integrate “addition by touch” in which case all the student needs to do is to drag two number on a tablet screen together with his fingers so that it forms a sum. Dragging two numbers together to form a sum can then be thought to be a specific epistemic action in a particular virtual world to manipulate the problem space.

To conclude, the principle of interactive workspace is aimed at providing the user a space in which she can perform actions to epistemically test herself to a solution strategy. Unlike the principle of feedback an interactive workspace does not need to be limited as reaction to an answer of the user, but instead can be the world reacting to a purely epistemic action. Again when designing appropriate manipulatives it is necessary to take into account the scenario, type of representation of the task. The relevant relationships illustrated by manipulatives should be highlighted in the scenario and naturally graspable through the type of representation in the problem space.

2.3.5 Principle of Gradual Transferral

Understanding the Similarities of Solutions

All previous principles have stemmed from the idea of embodied mathematics implying that early level mathematical education can benefit from a reliance on knowledge regarding human cognitive capacities. Mathematical problem solving should hence be grounded in, and take into account, real world perception and representation. Yet one must keep in mind that the final goal of grade school mathematics is that the student gets a conceptual understanding not only limited to certain concrete scenarios. By distancing the education material from this ultimate goal there is a risk that the knowledge accrued will itself be of a different kind than that offered through traditional mathematics education, and then the embodied mathematical approach can no longer be deemed a potential improvement or substitute to the former, but rather a completely different field. Indeed we have previously noted the difficulty when it comes to mathematics being context dependent (Lave, 1988). And even when contexts, in which specific tasks are given, can be similar it is still not the case that transferal of learned solution strategies can be applied to problems expressed superficially differently (Catrambone & Holyoak, 1989). Our design assumptions are, however, twofold. Firstly each solution strategy learnt is assumed to be identical in terms of deep-lying strategy to those found in math textbooks, and secondly a transfer of knowledge going beyond superficial difference will be possible as long as certain care is taken into the manner in which tasks are presented and explanations are given (Anderson, Reder & Simon, 1996; Catrambone & Holyoak, 1989).

Removing the Scaffolding, Proximal Learning Zone

We previously noted that the transferal of mathematical ability between contexts is not impossible. Focusing first on the concrete and then showing its relationship to the abstract problem has proven to be an effective strategy for those with mathematical learning disabilities when it comes to learning algebra (Witzel, Mercer & Miller, 2003). The aim is to apply the same system in our educational software. To do so we will apply Vygotsky's theory of proximal learning zones (Chaiklin, 2003). The concept of proximal learning zone states that students can perform and solve problems together with supervising assistance that is greater than he can solve by himself and further that this supervised understanding will be able to be understood without additional help in the future. While the original idea assumed the existence of a human advisor, we try to implement this notion as the aid given by the educational software. While human advisors can guide attention and provide situation specific instruction, current technology may lack in a capacity to do so. In our current situation however we assume that the student has already been able to solve the problem through software aid, and what is relevant is the transferal of this to more abstract tasks. With human supervisors the transferal can be assumed to be done rather easily, that is, through pointing out that the problem that was solved before is identical to the new problem, and you only need to replace each element with another and use the same strategy. We believe this can be implemented in educational software through forcing the user to take notice of the relationship between the past aided problem and the new unaided textbook problem.

What is necessary for educational software to show that two problems who are superficially different are indeed identical is that the element needed to be interchanged from the first problem to the second is not too distant. We will do this by gradually removing the concrete aspects of the problem, such as representation aid, interactive aid, real world scenario, and so on. Through removing each of these aspects at a time the student will hopefully not notice that the task itself has changed, but rather have only gotten slightly harder. An additional benefit of gradually removing the digital aid is that the program can pinpoint the exact moment in which the problem arises for the student in transferring the solution. If for example a student can solve a multiplication task through using representations in the programs, but fail to solve the identical problem without the representation it is obvious either the student has failed to make the connection between the number and the representation, or that he fails to perform the task without working memory aid in the world. The former can be helped through instructing the student to see that both problems are the same, and the latter can be helped by either instructing the student to imagine the representation in his mind, or that he take a piece of paper and write his own representational aid on it. Finally when the transition from the concrete problem solution to abstract problem is done the two can be shown together further illustrating the similarity between the two.

The principle of gradual transfer integrates the above four principles and attempts to make the knowledge gained in the software task space transferable to school mathematics through gradually showing the similarity between the two. By not making instant jumps and by keeping the solution strategy identical throughout each step the user can hopefully

infer that the starting, heavily aided software scenario, is identical in deep lying problem strategy as the last step, a purely abstract mathematical task.

2.3.6 Summary of principles

Through the above discussion there is reason to believe that educational software can be improved by an implementation of our principles. While the embodied mathematical framework might be neither optimal nor all-encompassing we believe it can form a suitable design alternative to many available softwares on the market that can be worth inspecting. To summarize the principles, our hope is that this approach can be applied to in general facilitate understanding of mathematical concepts by allowing users to learn from the world, and in the process removing unnecessary working memory load to the learning task. While all principles are based on the general assumption that students are able to make inferences and form strategies to find solutions, each principle itself has certain backing in theory and research that shows they might be fruitful. Thus the next step in this endeavour is to see if the framework not only is supported by existing theory and research, but if it can be used fruitfully in the practical designing process.

2.3.7 Example of design process using the framework

Before proceeding to analyzing our actual problem we will attempt to show an example of design using the five principles in order to illustrate how it can be done. The example will use "Single digit addition with regrouping" as the concept to be learned. For those not versed in the world of elementary school mathematics this would be additions like "4+8" where the sum of addends is greater than 9. This is an example of an abstract concept, since the user needs to learn that the number above 9 is not some whole new type of number but rather a combination of 1 and 0. How can one explain that the 1 in 10 is ten times as valuable as just a single digit 1? In creating a task relevant to showcase this relationship we must first try to apply the principle of problem solving and principle of object representations. Is there any analogy in the real world to digit regrouping? An example could be the concept of container, where each digit represents a box. When the value is too large it cannot be contained in one box, and thus it needs to be placed in a bigger box. Concretely imagining this one can say there's a cardboard box where you place apples where only 10 can be placed at any given time. Every time one box gets filled to maximum it is moved over to a bigger contained box, this in time able to accommodate 10 smaller boxes. Using boxes as representation it is then necessary to find an appropriate scenario where this representation makes sense. An example would be the user is working in a factory, and needs to sort the products into appropriate boxes for shipping. From this appropriate use of workspace and feedback can be formed. An addition between 5 and 6 could be expressed as "Put these five and six apples in the cartons". Feedback can be given if the user fails to put all apples in the boxes in a systematic manner by highlighting apples not placed in any box. Allowing apples to be placed by mouse or touchpad movement gives an additional element of manipulatives to it, especially if the placed apples in the boxes is represented and updated real time to the placement of apples. By dragging and putting

an apple to a small box a single digit value increases, and by placing the filled small box to a bigger box a second digit value appears. This way, the task of physically adding the apple causes an increase in number, and by showing them together they highlight the relationship that occurs when adding numbers to each other. The user will hopefully make a connection, that the 1 in 10 is in fact not an apple, but stands for 1 box containing 10 apples. The scaffolding can be gradually reduced by removing the dragging element and then the visual aid.

The example above was thought up with a couple of minutes. Given more time more elaborate or interesting scenarios may be designed, and the type of feedback can hopefully be improved. Hopefully it however illustrates how our framework can be applied. The steps required are, “Identify the concept characteristic to be learned”, “Find an appropriate analogy to that concept in real life”, “Create a problem solving scenario in which the analogy can promote interactivity and provide relevant feedback”.

3 Prototypes

To show that the embodied mathematical framework is useful we will aim to apply it to a pre-existing educational software and suggest concrete points of improvements. The educational software Qnoddarnas Värld, for the third grade, by Natur och Kultur offers mathematical – as well as Swedish language – exercises for elementary school students. Qnoddarnas Värld is used in several elementary schools across Sweden, and thus was chosen as a typical example of educational mathematical software. Analyzing the product itself will therefore highlight the difference between common strategies employed in designing these softwares to our own approach. We will not only analyze and suggest areas of possible implementation of our framework but also create a prototype that deals with some of the concepts in Qnoddarnas Värld in an alternative way. We have chosen to focus on the broad topic of “number sense” and the exercises “Description of Numbers” (in Qnoddarnas Världs “Blueberry 1”) and “Expanded Forms” (in Qnoddarnas Världs “Blueberry 2”). However, as will be mentioned later in this chapter our framework is applicable not only to limited types of mathematical learning.

To enable fruitful implementations of new ideas, we wanted to thoroughly ground the underlying conceptual design for what we wanted to achieve in theory. From these ideas we discussed and designed various Lo-fi prototypes of possible exercise-improvements, as well as completely new designs. The Lo-fi and Mid-fi prototypes were made using InDesign and Photoshop that resulted in PowerPoint presentations. These prototypes made it possible to step-by-step “walk” through an exercise and its phases. We chose to start in this medium since we had prior knowledge regarding the programs and that they allowed for very fast implementation, where we could use screen dumps from the actual educational software and cut and paste. This method made it possible to quite fast investigate several ideas and to single out those we thought most promising.

On the basis of the Lo-fi and Mid-fi prototypes we chose to focus on one idea for the Hi-fi prototype. For several reasons we decided to use HTML5/JavaScript for the implementation of the Hi-Fi prototype. One important reason is the

cross platform properties making such software possible to run in every modern web browser. This means that we can use it on both android and iOS tablets and also on PC and Mac computers. We were also using a HTML5 game engine library, Phaser www.phaser.io, that gives basic functions for moving images and detecting touches. This makes simple manipulations, such as click and drag, simple to implement and instead we can focus on designing the prototype. JavaScript also gives a great opportunity for collecting data during tests. It is rather easy to set up timers, counters and make other measurements.

3.1 Qnoddarnas Värld

Qnoddarnas Värld is an environment where students encounter animals and fictive anthropomorphic animal-like characters named “qnoddar”. Qnoddarnas Värld is presented as a story book like setting. In order to make students associate to a context of study and learning the first thing that meets a user is a study-chamber with a framed picture of a qnodd. Above mere mathematical exercises the student can enjoy game-like features such as customizing the appearance of the room where the qnodd lives by choosing from different preset designs, and reading e-books from a digital bookshelf. The roles of the student and the fictional characters, respectively, are however never clearly stated and so remain a matter of interpretation. When exiting the qnodd’s room through the one available door the student is presented with a choice between mathematics and Swedish language. If mathematics is chosen, a new frame with a fictive world map is offered, from which students can choose where to go. Each area hosts exercises, represented as berries on paths that the students can unlock as they complete the tasks presented.

3.1.1 Blueberry 1

Blueberry 1 is the first exercise out of nine focusing on number sense. The exercise is divided into three kinds of increasing difficulty, of which we have chosen the third, called “numbers”. The scene for the exercise contains two characters; a rabbit and a “qnodd” – the fictional characters of the educational software. A text states that the student is supposed to write the number the rabbits asks for – this statement is also presented as an audio message. The rabbit describes a number, mixing text and numerals, in a speech bubble, and the student is supposed to figure out the number and write it in numerals in a blank speech bubble stemming from the character. In other words, the student is implicitly asked to step in and use the qnodd character as a form of avatar (that is, a representation of herself – since she is the one who answers the rabbit’s question).

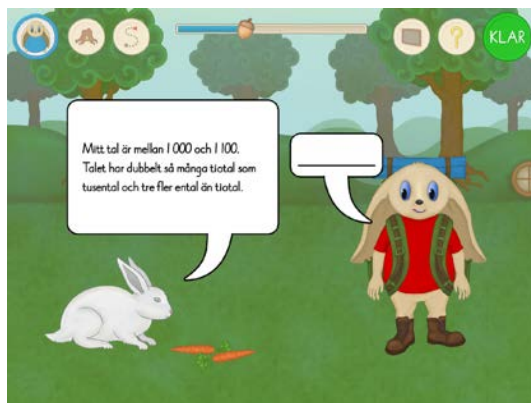


Fig. 1. The scene depicts how the questions in the exercise Blueberry 1:3 are posed.

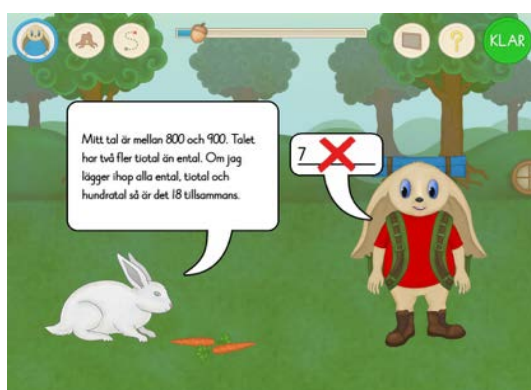


Fig. 2. The feedback indicating a wrong answer.



Fig. 3. The feedback indicating a correct answer. The star is added to the bar in the top of the screen.

When the answer provided by the student is wrong, a red cross symbol is presented as feedback over the answer, as shown in figure 2 below. Thereafter the original question reappears, and this continues until the correct answer is given.

When the student provides a correct answer a star appears, as seen in figure 3, and consequently there is an increase in a “score bar” on the top of the screen. The next question then automatically follows, until the bar is full and next exercise-step is unlocked.

Discussion

Taking the above exercise as a more or less standard type for the entire software, there are clear differences in this design and a design according to our framework. Firstly the exercise is not anchored to any actual scenario. This is by no means a flaw as such but it illustrates a different approach. With that said when a software clearly provides a gamelike setting and environment it begs the question what the relationship is between the character and the task. The fantasy setting does little more than offering visual aspects irrelevant to the task. While the exercise aims to train number sense there is no representational aid to illustrate the relevant aspects of the task. There is no interactivity or strategy to come closer to a solution. The feedback provided is only corrective and gives no clue as to how to change the answer the next time.

The restrictions on answer input is also problematic in the sense that when the student wants to write an answer, the iPad’s keyboard covers about half the screen. Since this is a built-in feature, the placement of both figures should have taken this into account but this seems not to have been the case. Not being able to read the task limits the ways in which the student can answer the task and causes unnecessary working memory load since the potential of using the display as a tool for solving is diminished. Essentially the software requires the user to first complete a comprehension phase and only when the answer is clear start to input it. This demarcation between solving and answering is not in line with the embodied mathematical approach

Implementations

We wanted to investigate how the exercise could be reconfigured to fit in the embodied mathematical framework by means of concrete changes. Regarding the problem solving principle, the overall situation can be interpreted either as if the student should provide help to the qnodd character or that the character is an avatar that represents the student herself. In any case it is unclear why the rabbit wants to know the answer to the questions it poses and similarly unclear what the student or the character has to gain from providing a correct answer. This said, there is the rather abstract element of gratification in gaining a star that raises a bar and eventually leads to promotion. As discussed above, this can be interpreted as a form of gamification. Concerning which moves are allowed and what the goal state is, the exercise is reasonably clear, even though the narrative around it is diffuse.

Our overarching goal was to ground the exercise in the embodied mathematical framework presented above. The usage of the conceptual metaphors ‘object’ and ‘collections of objects’ for numbers and sets respectively, could provide such a grounding. As mentioned when discussing the principle of object representation, objects are a natural way to represent numbers. We therefore developed a Lo-fi prototype that used balls as concrete objects. In figure 4 below we color-coded the different types of balls according to their properties.

situation and context, and the student is not provided with much of an explanation or motive for why he or she should put money on the counter. Even more problematic is the order of the representational aid given within the task. The student needs to first answer correctly to the expanded form question without any representation of money, and only after this does the money and additional instruction of putting money in the tray. This is problematic since the representational and interactive manipulation of the world are both there to facilitate learning of new concepts, but the fact that it requires the user to first answer correctly means the help only really appears when the user already knows how to solve the task. Furthermore the dragging money to the tray must be done in accordance to the answer previously put in, and it is not clear why there should be a required additional task when the question is already answered correctly.

Similarly to the previous exercise, the exercise blueberry 2 uses very limited corrective feedback. There is no indication of where the error is or the magnitude or direction of the error. The input accepted is only numerical, and while the task of expanded form requires the user to enter the number in a orderly fashion from thousands, hundreds, and so on, there is no input restriction to each field in this regard. This means that the student can enter the correct numbers, but without knowing the orders that the software requires she is essentially lost. The learning aspect of the exercise appears first when the user has understood expanded form and the way that the software requires an answer to be entered. As such, the initial screen presented to the user is closer to a password input screen than to a learning material.

Implementations

Our initial take on the exercise was to merge the two tasks and to present the money directly. The student then gets a visual aid and can externalize some of his or her conceptual model, in line with our second principle concerning the minimization of unnecessary strain on the student's working memory, as well as in line with the fourth principle of interactive workspace and epistemic actions. Mapping the more abstract numbers to the more concrete money offers a natural way to understand the concept of expanded form. With such a design the exercise with the concrete manipulatives will offer an actual chance of learning through the manipulation and not just offer a test. This said, it is not unproblematic to design the exercise since it is important to keep the learning aspect of "knowledge in the head" (Norman, 1988: 79). As the exercise proceeds, a gradual transition from a very concrete task involving strong aid and support to a more abstract task should occur.

By connecting the two categories numbers and money, and then make changes in one category that will affect the other, a natural way to understand the abstract and the concrete concepts can be enabled. This would offer immediate visual and conceptual feedback to the student and show the result of his or her actions. We hence decided to design a Hi-fi prototype based on this multi-modal concept. These changes were initially investigated through minor design-rearrangements, where the money was immediately visible and connected to the numbers. Furthermore, an added feature where the money the student present as the solution to the exercise gets pooled and compared to the correct amount

would give direct feedback in a way that lacks in the original design:



Fig. 7. A redesigned version of exercise 2:1, where the added space is used to present the students answer and then compared to the correct answer.

Although we ultimately decided to focus on a re-designed version of the exercise, we made a Mid-fi prototype investigating another concept. Here we tried to see if we could develop the exercise beyond the constraints set up by the current design and incorporate the principle of problem solving more clearly. Basing the exercise on a real life problem solving situation we used an anchored setting of a "shop", where the student was supposed to buy something from a shop-keeper. This integration of a narrative, discussed by Ginsburg, Jamalain and Creighan (2013: 106), might provide a more meaningful context and motivate students more.



Fig. 8. The shop set-up was used as a real life inspired mathematical problem solving situation.

The use of a store presents students with a situation that most of them probably are familiar with. This is likely to increase the chance of the user seeing the point of trying to solve the exercise. The mathematical skills they acquire are in this way connected to motivating factors such as a willingness to master similar situations that they might find themselves in. The task (buying something), the representation (money) can be consistently illustrated through this setting and immediately give a cue to the user what the meaning of the task and how to possibly go about it.

After an initial description of their task the student gets to choose an item of their liking. She then encounters the shop-keeper at the register:



Fig. 9. The exercise keeps the initial metaphor and use of money although provides a real life context.

In this step we wanted to link the concrete element of money to the abstract mathematical concepts. We therefore use the same conceptual metaphor as the one already present in the educational software, with the difference that we connect it to a real life situation. An additional feature is that we connect the numerals and the money, so that if one is lowered or raised the other complies. The interactive element of moving the money becomes a concrete manipulative by making the changes in money position reflect the change in value displayed. This should be made visible by affordances such as color and moving features. An example of this feature is visible in (Fig. 9) where the added coin is thought to be connected to the green numeral. When the correct amount is presented the shopkeeper indicates that everything is okay and the student receives the merchandise. When the wrong amount is presented, what and how much that is incorrect should be clearly shown so that the student can understand how a more proficient strategy could be applied. Comparing the desired number and inputted number physically close to each other allows for easier visual comparison, similar to how a user would epistemically check difference between two objects.



Fig. 10. The feedback should clearly indicate where the problem is, and how it can be solved.

3.2 Summary

The above two exercises and many more inspected but not presented here indicate that Qnoddarnas Värld uses very little of what our framework highlights. Representational aid is not always provided, and when it is, it is presented ineffectively. Anchoring to scenarios seems to not occur, as each exercise is expressed in an abstract school math task manner while the fairy tale element only seems to exist as background. While interactive elements to a certain extent exist they seldom qualify as a concrete manipulative, since they do not effectively highlight the relationship between the object used and what it is representing. In certain cases like the dragging of money in blueberry 2 the interactive element is added after the actual task, making it seem like an additional game element not related to the actual problem. The use of epistemic action is also hindered by the fact that the software requires specific input, restricted to letting the user answer the exercise. There is a lack of options, such as, to write notes or put memory aid on the screen, which is possible when doing similar exercises with pen and paper.

3.3 Hi-fi Prototype

Our Hi-fi prototype combined elements both from the original exercise and our Lo-fi and Mid-fi prototypes, and can be found online: <http://intraktion.azurewebsites.net/>

We have, primarily worked with the exercise blueberry 2:1, present under the link "Tal". The prototype that was used in our pilot studies was designed with the shop-metaphor in mind (yet in the Hi-Fi-prototype it was re-designed to look more like the original exercise in Qnoddarnas Värld). The correct amount of money should be dragged to the table, and the corresponding numbers at the top of the table are updated dynamically. We included multiple core elements in this first prototype version: object manipulation (the money movement) and provision of basic corrective feedback. This version was very much a proof of concept to be expanded upon.



Fig. 11. The start frame of the first version of the Hi-fi prototype.

Developing the High-fi prototype further, it was important to make the prototype incorporate as many elements as possible from the original exercise so that comparisons between them would be manageable. We thus used a qnodd as a shopkeeper and based the background on the original exercise, albeit with minor changes.

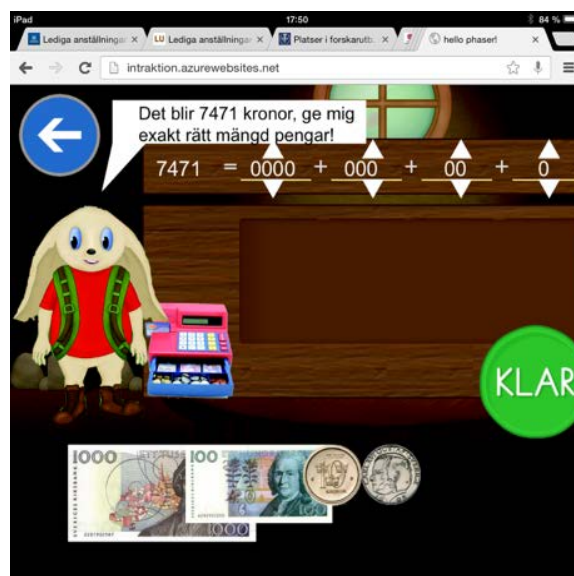


Fig. 12. The start frame of the final Hi-fi prototype.

4 Pilot Studies

Three pilot studies were conducted on children who were three, two and one years younger than the intended target group for the educational software. By this procedure it was possible to pin-point strengths and weaknesses in the prototype design as well as getting an insight into which aspects of the prototype that were possible to grasp and which aspects that was difficult to understand for students in the age-group.

4.1 Pilot Study 1

In the first exploratory pilot study on our Hi-fi prototype of exercise blueberry 2:1, we chose to use a subject that was five years old and did not have any prior formal mathematical training. The subject was thus entirely unfamiliar with the context of mathematics learning and mathematical exercises. By this choice we hoped to get an indication about the embodied mathematical design that was as free from prior influences as possible. It did however mean that the subject officially was “too young”, since the exercises target third graders, about nine years old, and that the subject would probably not be able to actually solve any of the mathematical exercises. What we observed, however, was how the subject interacted with our Hi-fi prototype. This made it possible to point out what was clear and easily understandable in our design. The information gained served to indicate strengths and weaknesses in our design. The Hi-fi prototype was presented on a tablet. One test leader, one parent and one sibling was present during the study.

4.1.1 Pilot 1 Discussion

The subject was from the start uncertain about how to proceed. He initially tried to press the “down arrows” indicating thousands and hundreds. Since this did not result in anything happening he asked for guidance before interacting any further with the software:

“Where are you suppose to push?” [00:36]

This issue is related to the fact that the prototype has constraints in that it only accepts positive numbers. A possible work around would be some feedback clearly showing that it is only possible to increase the amount from the initial stage. After some encouraging comments from the test leader, the subject tried to press the screen in what looked like a random fashion. After a short while he was able to move a hundreds bill. When he realised that it was possible to interact with the money he began to drag the bills. The location of where to position the bills seemed however unclear to him.

The question was repeated and the subject seemed to connect his task to a location. He started by dragging the money towards the figure. When the role of the money had become more clear to him, he had no trouble realizing that the coins were possible to move as well. Connecting the money to the numbers presented was not an easy task. The subject saw that a number was indicated, he then asked:

“Is it supposed to be a ‘one’?” [01:55]

This was due to him having presented a single hundred bill, and was not connected to the number that was being asked

about. The subject can thus be said to not understand the connection between the expanded form of the presentation and its linkage to the money. At a later stage he did however see the connection [02:24], which likely had to do with that he had started a new trial and therefore had a “clean screen” before him. In other words; when he could see that a number appeared as a direct consequence of him positioning a bill in the correct position and no additional noise was present he did understand the connection that previously had been too difficult to see.

The subjects attention was not possible to hold for long, but we were able to get some indications of the prototype design. One important lesson from this pilot study was that the numbers needed to be more saliently presented. In order for a subject to directly see the connection between the money and numbers there can be no room for not noticing that something new has happened with the numbers when money is added or subtracted or vice versa.

Another aspect that made the exercise difficult was that we had not been able to include an introductory demonstration and the subjects had to infer from a brief spoken description. The less prior experience with similar exercises and mathematics in general the more difficult it is to understand what is expected of you.

4.2 Pilot Study 2

In the second exploratory pilot study we wanted to get some additional feedback on our initial Hi-fi prototype. The Hi-fi prototype was presented on a laptop, whereas the Qnoddarnas Värld exercise was presented on a tablet. One test leader was present during the study and one of the subjects parents was in the next room.

The subject was initially encouraged to interact and use what was being presented to him generally, without any further specific instruction. When he got stuck, we cautiously indicated how to proceed, providing as little information as possible. The subject was six years old and thus, as the previous subject, younger than the target group.

4.1.1 Pilot 2 Discussion

From the start it was clear that the subject did not understand what was asked of him, neither with the prototype or the original exercise – even though he just had used the prototype with its similar case set-up. This result was however to be expected since he could not read the instructions himself nor understood what was being asked of him, since he was new to the context:

“I don’t even know what I’m supposed to say.” [00:17]

After encouraging general remarks to talk out loud about his thought process from the test leader the subject mentioned that what he had first seen as standing out in the scene was the money:

“[...] I didn’t mean the man; I meant the money!” [00:35]

This can be interpreted as that the object representation is reasonably chosen in that it is recognizable and “stands out” which would indeed be preferable. The subject could not

read the text explaining the exercise so the test leader read the message out loud:

“I can’t read that much text.” [00:50]

This feature should be built into the final Hi-fi prototype and be presented as an audio message, which made it reasonable to read it out loud here. The subject did however not understand what was expected of him on the basis of the message.

When explicitly encouraged to use the mouse-pad the subject’s first choice was to push the “finished”-button. This can be understood as that the button-feature was salient, and given that the subject was unable to read its label it was a reasonable choice to try. It was however not correct given the exercise. From these initial observations it is obvious that it would be a good idea to work on making the features of the prototype as salient and obvious as possible. Even though the target group will most likely be more used to similar exercises there should be no uncertainty regarding what is possible to perform – in line with the principle of object representation – given the importance to minimize working memory load. Furthermore, in line with the principle of world feedback, the subject should not have any difficulty to figure out how to do something or what to do.

After some trial and error the subject noticed that the money was an element that was possible to interact with:

“Wow! How did that happen?” [01:55]

The feature that made him realize this was that the bills moved while being pressed. This is an important finding that needs to be thoroughly incorporated into the Hi-fi prototype, that when something is clicked it should move or change color to clearly indicate that manipulation is possible. It is likely a good idea to make these features be present when hovering the cursor or finger above the target.

The subject right away started to move the money, which is an indication that the feature is possible to grasp easily, although he initially moved it to a random blank space in the lower right corner of the screen. After a while the subject said that he would place the money on the table which indicates that he understands the concept of the table and that it would be a reasonable place to place the money. His first thought was to give the character in the exercise “all the money he [the subject] had”. He did not see or make a connection between the money being placed on the table and the simultaneous increase in numerals:

“I’m gonna put the money on the table” [02:23]

This means that the feature of numerals being mapped to money/objects must be much more saliently designed. Even after this feature was pointed out by the test leader the subject did not understand the connection. This is however once more most likely caused by his lack in mathematical training, since he was only able to read the numerals one at the time – not connecting them into tens or hundreds etc.

When asked how he reasoned when he placed the money on the table he indicated that he was thinking as if he was in a store:

“[...] as if it were in a store. [...] and you put it [the money] on the counter.” [03:55]

This indicates that the general metaphor and design is natural and possible to grasp. After being encouraged to explain his thoughts, the subject described how he should first give one then two then eight then one to the store clerk. Since the subject only understood the concept of one numeral at the time this was in a sense correct given his knowledge level. After understanding that he had not performed correctly, indicated by the test leader, the subject understood it as that the money on the table should be removed – which is correct.

Concerning the task in the original educational software, *Qnoddarnas Värld*, the subject did not grasp where to place the objects or what to do. Rather than to connect the objects and numerals he “filled” the containers indicative for where the money ought to be placed.

4.3 Pilot Study 3

A third exploratory pilot study was conducted to gather additional feedback on our Hi-fi prototype. The third subject was seven years old and thus too relatively unfamiliar with the context of mathematical learning and mathematical exercises. He said that he had some experience of buying things in a store by himself, which can indicate that the problem solving case of a store is reasonable considering the age group. As regarding the previous subjects, this means that the third subject also was younger than the intended target group – although in close proximity. We observed how the subject interacted with our prototype as well as the original educational software. The result from this pilot study should primarily be seen as a general informal indication of strengths and weaknesses in our design. Both our Hi-fi prototype and the original exercise was presented on a tablet. One test leader and two observers – one filming – was present during the study as well as one of the subjects parents, who occasionally interacting with the subject.

As in the previous pilot studies, the subject was initially encouraged to interact and use what was being presented to him generally, without any further specific instruction. When he got stuck, we cautiously indicated how to proceed, providing as little information as possible.

4.3.1 Pilot 3 Discussion

The pilot study began with our Hi-fi prototype being presented to the subject on a tablet. The subject immediately started to use the money and appeared to try and “give” the money to the cashier in the software, by swiping the bills and coins towards him. The subject did not find the “ok”-button which made him unable to get to the next step. He did however present the correct amount and was able to continuously solve the exercises in a fast tempo after the button was pointed out to him. This, indicates that the button needs to be much more salient. Furthermore, the subject said that he held the amount of bills in mind rather than use the numbers presented as feedback which indicates that the prototype needs to present the numbers in a more salient way as well:

“I count it in my head.” [05:01]

He did however profess to use the numbers as support when he lost track of the amounts he was working with. When asked if he had bought something in a store by himself he said that he had, and regarding if he thought that this was similar to the exercise he answered:

“Kind of.” [06:55]

Regarding Qnoddarnas Värld the subject, after some initial trial-and-error, understood where to press on the tablet to be able to enter numbers. After the first hesitation the subject intuitively grasped that he ought to change the cursor position to fill all the blank spaces.

The subject's first attempted solution expanded the number, although not in the intended order. The exercise did only accept thousands, hundreds, tens and ones – in that order. All other answers were indicated as wrong with the sole feedback of a red cross. This meant that the subject got four red crosses when he pressed the ok-button, which made him perplex. When asked what he thought was wrong he said that he did not know. Since he could not understand what was wrong he asked for help and the test leader directed him to the software's help function in the form of a button with a question mark. The help consisted of a repetition of the previously stated demand to “write the number in expanded form”. The subject did not get any aha-experience from this information and still did not know what had been wrong with his answer or how he ought to go about to present the correct one. He tried again with another order and way of expanding the number, which also resulted in a correct amount albeit not in an acceptable order. This was repeated a third time. At this point the subject was told – got feedback – that a key element in the exercise was the order in which he presented the number. From this minor input he was immediately able to give a fully correct answer by himself and solved the exercise.

In the second step of the exercise the subject understood how he should use the money and where to place it. He proceeded to give a correct answer straight away. The second time he tried, he did however get two numbers wrong, and red crosses appeared. He was not able to get it right and when asked about his thought process he said that he tried to add it together in the right order. When asked about the exercises in general, the subject gave answers that may indicate that he was close to an understanding of the concept behind the exercises, although he expressed this understanding in a vague way, which leaves some uncertainty. He described how:

“[...] it's supposed to be the same in the end.” [22:19]

This can be interpreted as that he realized the importance of thousands, hundreds and so forth. He was however unable to get a “paper exercise” entirely right without the help of the educational software or the object representations, which shows that he had not gained a full understanding of the concept “expanded form”. This was of course not unexpected given the subject's age and the short amount of time he spent with our prototype and the educational software.

5 Transferral Test

Basing on the theoretical discussion above we believe that our approach to a mathematical educational software has a potential to serve as a well-grounded framework for designing mathematical education material for school and everyday use. But in order to evaluate this potential, large-scale testing in actual school settings are required, something that can not be carried out in the frame of the current project. However, as a first step we conducted a smaller-scale study to investigate whether software designed according to the framework of embodied mathematics would live up to the prediction made in our theoretical analysis regarding students interactions with the software, and their learning from it. More specifically the small-scale study focused on how solving a problem in the software relates to solving school-math test problems with similar deep lying structure. In short we aimed at testing whether a similar deep level problem-solving would lead to analogical transfer to more abstractly formulated tasks, and if this transferral would base on an understanding of the learned solution.

The following transferral test section will explain the structure of our test, followed by a result analysis of the subject's performance on paper test pre and post software learning. Lastly we present an interaction analysis of the manner in which the subject used our hi-fi prototype.

5.1 Hypothesis

By virtue of our qualitative approach we will be interested in the actual interaction between the user and software. More generally our assumption is that a decrease in cognitive demand will be present for understanding the task space in our Hi-fi prototype. Specifically our approach to learning through feedback would hopefully lead to that users *learn* through each mistake, thus not making the same sort of mistakes repeatedly. We also believe that there will be few instances of wrong inputs in the sense of inputs that have a clear intention but nonetheless fail due to the software. Further we believe our approach will cause few verbal instances of “giving up” or admitting you don't understand due to motivation or failure to understand task. We predict that our prototype will allow for flexible solution strategies that can differ between individuals but be equally valid.

As for the transfer testing we will be interested in several aspects. Firstly, will users are able to solve problems with identical deep lying structures to the ones learned? Secondly, how will the solution in the test phase differ from the solution in the learning phase in time, working memory demand and approach? Furthermore we will be interested to see if not only the solution is transferred, but the conceptual understanding of the learned term, together with implicit knowledge used in the solution. We believe that our prototype to a high degree will invoke successful strategy transferral with a high level of implicit understanding of the solution.

5.2 Method

We will focus on *far transfer* from the concrete to the abstract by removing contextual aspects and scaffolding, to investigate whether *analogical transfer* is possible, where generalizable knowledge from the original task to another is seen.

While we have certain quantifiable measures in the hypothesis these do not in themselves tell enough. The participant's interaction with the software will therefore be filmed and analyzed through the points mentioned in the hypothesis. Since it is not possible to analyze each and every action we will focus on instances where problems arise, and how the student solves these problems.

The transferal test will contain 8 questions where two of them will test transfer of strategy using identical deep level structure question as the ones trained in the software. Two questions will test transfer with additional elements but identical solution strategy information. Two questions will test reverse order strategy (starting with goal state and deducing base state). Two other questions, finally, concern mathematical rules implied or involved in the solution, rather than direct solution strategies. To succeed on the first two types of questions a student needs to see similarity in structure and apply learned strategies. The two latter types of questions will concern how participants make inferences about the learned strategy and understand the implications of what they are doing. In each question the participant are encouraged to describe not only their solution but their strategy and reasoning.

During both the training phase and testing phase, time will be recorded. Large differences found in the time concerning training or testing will be included in the analysis.

5.2.1 Participants

The participant (*fem*, *age* = 8) had started her second year in elementary school and thus had some minor prior mathematical education regarding expanded numbers. She is hence in our target group of eight to nine year olds.

5.2.2 Procedure

The participant, and her caretakers, were at the start of the transferal test presented with an overview of the test design and informed about their rights to decline continued participation at any time. In the transferal test's first step, eight mathematical questions were given in paper form. After these were answered, the participant was instructed to use our Hi-fi prototype, presented on a tablet. In the third step the participant was asked to complete eight additional paper exercises – all exercises are presented in Appendix I. Both the testing and interaction phase was filmed for later analysis.

5.2.3 Analysis

An evaluation scale of four levels was used to categorize the participant's answers to the transferal test questions; wrong answer (0), wrong answer but right strategy/presentation (1), right answer but not entirely correct strategy/presentation (2), and correct answer (3).

For the interaction analysis each instance where the user encountered a problem, either through stopping the answer input or the program returning an unsuccessful feedback was counted as an observation for our analysis. Each observation will be categorized to any of our prediction according to user action to the software and user's verbal report, if the problem does not fit in any of our predicted categories it will be counted as a "separate". Given the vague nature of this distinction the number of instances might not be an accurate measure to any underlying cognitive phenomenon caused by

the educational software, but as a user test the fact that any one individual might face a certain problem in itself might be an interesting observation. Thus we aim not to finding any proof of our hypothesis in this text in any strict manner, but we will hopefully find indications on how users usually interact with educational software, and through it gain a glimpse of how target users utilizes mathematical softwares.

5.2.4 Ethical Aspects

Since the participants, both in the transferal test and the pilot studies, are under-age there are a number of ethical aspects to take into consideration. First of all we tried to minimize the exposure of the participants, as mentioned. This said we were not able to make the filming completely anonymous since the participants sometimes moved in unexpected ways which made their identity possible to discern. All participants and their caretakers had however agreed to being filmed in advance and did not report any concerns regarding this issue.

The transferal test was non-invasive and did not involve any sensitive topics. Nevertheless we took care to explain, both for the participants and their caretakers exactly what participations involves and make sure that they knew that they could withdraw from participation at any time. (We were aware that feelings of commitment may influence the participants to feel obligated to continue.)

It was also important to avoid that a participant would feel bad if she did not succeed in the tests. Although this is a general phenomenon in all education it was pointed out to all participants before they started that the pilot studies and the transferal test was explorative and that there was no way to do *wrong* since all answers were of interest. We mentioned more than once that incorrect answers might be more interesting than correct ones in order to minimize competitive elements and potentially associated stress.

5.3 Results of Abstract Transferal Tests on Paper

In short we could not find any indication that transferal has occurred from software to school task. The subject was already capable of solving the test tasks confidently and so it was not possible to test any learning effect from the software.

5.3.1 Results of Transferal Test 1

As an answer to question one the subject wrote what kind of expanded number each number represented, indicated by arrows. She wrote; "thousand", "hundred", "10", and "ones". This is interpreted as (2). Question two was written as; "4thousands", "8hundreds", "9tens", "3ones". Once more this will be interpreted as (2). Question three was written as; "19thousands", "8hundreds", "2tens", "3ones", which will be interpreted as (2). Question four; "3hundreds", "0tens", "3ones", which will be interpreted as (2). Questions five presented the number "3", which was correct (3). Questions six presented the number "0", which was correct (3). Questions seven marked the correct number (3). Questions eight marked the correct number (3).

From the eight questions no answers were interpreted as 0 (0 %) or as 1 (0 %). Instead half (4) of the answers were interpreted as 2 (50 %) and half (4) as 3 (50 %), (*time* = 20:23 min)

5.3.2 Results of Transferal Test 2

Question one; “8thousands”, “4hundreds”, “910” [where the ‘10’ is in smaller size], “2ones” (2). Question two was written as; “1thousands”, “7hundreds”, “610” [where the ‘10’ is in smaller size], “3ones”. Once more this will be interpreted as (2). Question three was written as; “61thousands”, “8hundreds”, “910” [where the ‘10’ is in smaller size], “3ones”, which will be interpreted as (2). Question four; “2hundreds”, “010” [where the ‘10’ is in smaller size], “4ones”, which will be interpreted as (2). Questions five presented the number “6”, which was correct (3). Questions six presented the number “0”, which was correct (3). Questions seven marked the *incorrect* number (0). Questions eight marked the correct number (3).

From the eight questions one (1) answers was interpreted as 0 (12,5 %) whereas no answer was interpreted as 1 (0 %). Half (4) of the answers were interpreted as 2 (50 %) and three (3) answers as 3 (37,5 %), (*time* = 05:26 min)

5.4 Interaction Analysis

The participant self-reported that she liked mathematics, had a good teacher and was competitive by nature. She was thus highly motivated to partake in the transferal test and to solve all exercises. This made the test easy in that she gladly and quite efficiently reported her thought process at numerous occasions throughout the transferal test. Before the test started the subject got a very short auditory explanation of the concept “expanded form”. No written examples were shown. Regarding expanded form the subject said that she had some prior experience:

“We have talked a little bit about what order the numbers should be in [...]” [00:13]

5.4.1 Transferal Test 1

The subject was able to start right away and directly described her thought process:

“First, I’m thinking two thousand, and then that is the thousands.” [01:10]

The subject then indicated which number was the thousands, rather than to write the number in expanded form. She did this by writing the word “thousand” and drawing an arrow towards the correct number. She then continued by reading the entire number out loud, followed by pointing out the hundreds and tens. She indicated these numbers in the same way writing “hundred” and “10”; drawing arrows to the correct place. At this stage it is interesting to point out that she switched between letters and numbers, since she wrote tens in numerals rather than with letters as she did with the other numbers. This was something she did continuously for tens.

When answering the second she did switch the way she reported her answers. Instead of drawing arrows, she wrote out her answers as follows; “4thousand”, “8hundred”, “910s”, “3ones”. She did not comment on why she switched or seemed to reflect on it at all. Throughout the rest of the test she stuck with this way of reporting her answers for relevant questions.

The third question asked for the expanded form of 19823. At once the subject was a bit uncertain, stopped and asked:

“I don’t know if it’s nineteen thousand or [...] some strange number that I have not learned.” [03:27]

It seems as if she had no prior experience of numbers this high and was thus flummoxed about how to proceed. She eventually wrote down “19thousand” which in a sense is correct although points towards that the concept of expanded form was not generalizable. It is possible this answer stems from a way of learning expanded form as separating the numbers into sub-number types all noted with one word. In both Swedish and English the number “ten thousands” has no unique name, but rather is expressed as a number of thousands. The system of verbally expressing the number thus changes from the thousands, and it is possible the subject learned expanded form intimately connected to the verbal expression. She did not see the pattern to how the numbers were presented and what an correct answer ought to entail.

The other numbers were presented as in the second question although the tens stood out again – being reported as “210s” rather than “2tens”. She was keen on getting the answers correct and said:

“I’m gonna take a chance – I hope it’s correct.” [04:35]

As previously stated the subject wanted to present the correct answers, and wanted confirmation. The test leader at this point said that all answers was of interest, trying to encourage her to continue without interfering too much in the process.

The fourth question did not pose any problems for the subject, who solved it quite rapidly without saying anything. This can be interpreted as reasonable since it consisted of a smaller number of only hundreds tens and ones.

The subject thought that the exercises were over after the first four questions. This is likely since the first four questions were formulated in the same way whereas the following four differed somewhat. When solving the fifth question the subject started with a very elaborate counting procedure, where she added different numbers together and tried to “build up” the correct answer. Eventually she was able to solve the question and it then became clear that she had “over-thought” the question severely. Midway through the process she lost hope saying:

“I don’t get it.” [11:32]

The test leader then said that since she had tried hard, she might proceed to the next question. She then said:

“No. I *will* solve it!” [11:44]

She continued to struggle and was eventually able to reach a conclusion. The test leader pointed out that there was supposed to be equal amounts on both sides of the equal sign. This was something the subject already knew but nevertheless led to an *aha-experience*:

“Aha! Now I get it! Nine something, eight something *plus* something is equal to... So it is that which I’m supposed to get; if you add everything up.” [11:37]

Such an interpretation might follow from that the aha-moment indicates that the subject saw the relevant pattern which the question was based on. For example Gärdenfors and Lindström (2008; see also Gärdenfors, 2008; Lindström, 2011) argues that *understanding* can be seen as seeing the relevant pattern or structure of a task.

She still had to struggle a bit with the following question that had the similar structure, although this time she was able to get the correct answer much faster. She eventually was able to formulate a strategy that she thought worked and checked her answer of the previous question using this strategy and was able to confirm that the answer was correct.

The seventh and eight questions were solved quite fast. The subject had some minor initial difficulties of understanding what was being asked of her but then started to add the expanded numbers together correctly. An interesting aspect was that her reporting of her thought process indicated that she was on the wrong track since she seemed to add numbers together in an incorrect way, but the resulting numbers came out right in all cases nevertheless. Arguably there were steps in her thought process that she did not articulate even though they were clearly there.

5.4.2 Transferral Test 2

In the second transferral test, the subject proceeded to answer the first four in a very rapid pace. She gave correct answers although still writing the numbers in the same incorrect way, even though she had seen a correct presentation in the Hi-fi prototype.

She started by saying the entire number that was asked for, then focuses on the thousands:

“It’s eight thousand four hundred ninety two, so I take eight. And that’s the thousands.” [01:13]

The subject proceeded by pointing out all the numbers and their connected form in a correct manner. This was repeated in the second question.

In the third question she wrote “61thousand”. As in the previous transferral test she still did not break the number in two, following the correct pattern.

Regarding question five and six the subjects said:

“Now, it’s such tricky ones” [02:52]

She did however solve the two questions very rapidly reading out loud – she saw the pattern. Regarding the two last questions, she solved these in a similar rapid pace. She did however make a mistake – her only one – in question seven, where she incorrectly indicated that 5996 is larger than 7112. This most likely is due to carelessness, although she stuck with her answer when asked to explain her thought process regarding that question in specific. It is possible that the two nines in the number make it stand out as larger than it actually is.

5.4.3 Aspects Concerning the Transferral Test Questions

A consequent error in the subject’s presentation regarding the expanded form is to be seen throughout the test, where she writes “8thousand” rather than “8000”. The subject shows an understanding of the concept of expanded form and gives correct answers to principally all questions. However, her correct answers are incorrectly presented. This might be due to the lack of an initial example, showing how a correct answer would look like. This said the subject self-reported that she had some prior experience with expanded form and should thus have a minimum exposure of the correct way to write an answer.

It is interesting to notice that the subject switches her strategy as early as between the first and the second question. This indicates that she has some prior understanding to what is being asked of her, since the second strategy is much closer to the correct one than the first. Answers are then given in the same way in both of the tests.

The subject is much faster – about four times – after using the prototype. This could be taken as an example that the increased training at the task in the software transferred to faster ways of finding solution in the follow up test. However such an interpretation is not supported by any act or utterance during either training or testing phase. A significant portion of the initial paper test is spent on trying to understand the question and form an answer (unintentionally leaving the “learning” phase actually in the testing phase itself), thus it is very likely that the decrease in time is caused by the lack of additional solution finding effort second time around.

5.4.4 Prototype Interaction

We could not observe any indication of the subject acting in accordance with our hypothesis. This is due to the fact our hypothesis mostly assume our Hi-fi prototype will have fewer instances of some action, but since we do not have anything as a baseline to compare with we cannot say anything about it. However it is possible to analyze the general interaction of the subject with the software still. We could not observe any situation where the subject seemingly learned from her mistake, indeed we could hardly detect any error at all. The few errors occurred seemed to have happened as result of flaw in the program or unintentionally pressing the screen, none of which was sufficiently strong to make the subject rethink her solution strategy used. While this general success of the subject can be taken as an indication that our software is sufficient in providing an intuitive exercise to the user, this data will at best suggest that a user with an already decent grasp of the solution strategy will be able to apply it. It is therefore worth noting a subject with less prior confidence in the area may react in a wholly different manner.

In the second step of the transferral test the subject completed eight exercises with the Hi-fi prototype. The subject started to press both the arrows, indicating addition and subtraction, *and* dragged the money onto the desk. It is possible the subject had a decent assumption of the upcoming task, and indeed seemed to understand the desired goal state of the task immediately. This can be taken as an indication that there was successful transfer from the abstract mathematical concept of expanded form to a digital problem solving scenario.

rio. This was done before paying any attention to the specific amount that was asked for:

“So I’m supposed to bring forth, so it becomes as that?[pointing at the number in the question on the screen]” [00:27]

After the initial test of both the arrows and the money, the subject preferred to only use the arrows. Most likely this was because she immediately understood the underlying concept and choose the fastest way to present the correct solution. The object manipulation was thus not necessary for her to feel comfortable with in providing a solution. This interpretation is supported by the fact that she did use both money and the arrows and thus knew that it was possible to use both.

The subject is able to use the arrows to indicate the correct amount for the first question and reads the number out loud. She then directly presses the ok-button which means that she understands its role.

She starts the next question in silence, presenting the correct amount for each part of the expanded form. While doing so she pauses and says the “form” out loud:

“Thousand, two hundred, eleven crowns”. [02:00]

Since the amount asked for – which she correctly presents – is 7211, her formulation indicates that she first only acknowledges that she has presented the thousands, rather than the specific amount. She then switches and says the correct amount, followed by presenting the two last forms together.

For the next question she reads the entire question out loud. At this point the subject accidentally pressed the ok-button which resulted in that she got feedback that there was not enough money [02:51]. This made her stop, but she was later able to continue to present the correct amount, pointing out:

“I didn’t mean to press.” [03:01]

This only happened once which makes it seem less problematic, than if it had happened more often in which case the prototype reasonably should be re-designed to avoid similar mishaps. As it were a single mistake does not seem to be enough to indicate a genuine problem, especially since the usage of the prototype, in this matter, went fine during the rest of the test.

The subject reads the following question – not out loud –, presents the correct amount of thousands using the arrows. She then says the amount and the following out loud. This pattern is then repeated for the following questions.

5.4.5 Aspects Concerning the Prototype

The subject does not change her strategy regarding her way to present answers, after using the prototype. This can be interpreted as that the prototype visualizes the numbers in an insufficient manner. The subject ought to have no difficulty in connecting the numbers with the correct way to present an answer in expanded form. Future investigations should experiment with how this is done in the most efficient way. The saliency of the numbers may be insufficient. Since the subjects immediately start to work with the numbers, it is relatively clear that she understands what the numbers indicates

and that it is with the numbers she is expected to provide an answer. However, the *form* of the numbers might not be highlighted enough, or stand out in a way, that makes a subject connect the importance of this aspect. Since the subject was able to efficiently use the prototype to give correct answers, this lack of transferal from the prototype to the second transferal test is to be seen as a failure in our prototype design, either in our lack of a sufficient introduction and/or in the saliency of the numbers in the prototype.

Other aspects, such as the money and the arrows, indicating the possibility to increase and decrease the amount, was directly grasped and thus probably salient enough. In a similar manner the ok-button was understood without any seeming problems.

5.4.6 Contextual Aspects

The subject did the transferal test in her home, which is an environment that ought to put the subject at ease. She could take her time and focus on the test without being in any artificial environment. This might however lead both to that the subject answers “better” since she is comfortable, or that she, given the lax environment does not make the same effort as she would in a school or laboratory setting. In other words a sterner context might drive the governing epistemic standard to a higher level where the subject would feel an increasing pressure to do good (see, e.g., DeRose, 2009). This said the subject, as previously mentioned, self-reported that she was highly motivated, which can be interpreted that contextual factors are less important in the way she views the situation.

One factor to be mentioned is that the subjects younger sibling was present in the beginning of the test and offered some extra “cognitive load” due to sounds and minor interactions. This did not seem to affect the subject in any relevant way, but ought to be mentioned.

5.4.7 Weaknesses and Strengths in the Test Design

The difficulty in obtaining test subjects proved to be a crucial limitation for our test. While we wanted to investigate the relationship of how strategies learned in software transfer to other case, we were not illuminated at all on this point, since the subjects were (from the perspective of our study) at a too high level of understanding prior to the experiment. Ideally we would have had several more participants and at least one that lacked knowledge of the concept expanded form, but nonetheless understood all the pre-requisite mathematical concepts and had reading ability to be able to learn it. Such an ideal subject is hard to find, and indeed would be best found by cooperation with teachers who are knowledgeable about their pupils’ math development.

A weakness in our test design is that by only focusing on our Hi-fi prototype it is impossible to compare it with Qnoddarnas Värld in any definite way. In an alternative test design one group should do the test as in our present study while another group did it but with Qnoddarnas Värld instead of our Hi-fi prototype. While inter-software comparison can be done even on the same subject there is also a factor of in which order the software is presented to the subject influencing the result. Furthermore such a within individual comparison would pose several difficulties in comparing transfer test, since it is not possible to determine which strategy from which domain get transferred to which task. Again this pro-

blem arises from the lack of total number of potential participants we could acquire. For future studies of effect of software it is necessary to have a pool large enough to be able to confidently compare differences between individual.

A potential source of problem for the analysis is the informal test setting. While the subject's home was chosen as to recreate a natural environment this also lead to a number of potentially influencing factors. The subject often sought confirmation from the test leader, and thus while no explicit information was given about her answer it might have nonetheless influenced her own judgment on if she was going the right way in the process by implicit cues given by the test leader.

Another influencing factor to the experiment was that our Hi-fi prototype was lacking in several aspects mentioned in the theory. We did not manage to implement a more thorough visual feedback system like we wanted, but instead we were restricted to a only slightly modified corrective feedback. The scaffolding removal was not present to the extent desired and thus was never called into action during the testing phase. As such many of our theoretical predictions should not have appeared even if our assumptions were true and our testing sufficient. Furthermore it would have been ideal to test not just one type of prototype since the problems arising would expected to be different if task difference or type was different.

A strength with our test design, and the focus on one subject, was that it made it possible to analyze the interaction in depth.

Concluding Remarks

The “embodied mathematics”-approach has enabled us to investigate the cognitive base surrounding ‘feedback’ in interesting ways. Furthermore, the analysis of the existing educational software Qnoddarnas Värld has made it possible to identify many key issues to work with, and from these we have implement various improvements in prototypes. The Lo-fi and Mid-fi prototypes are to be considered as preludes to our Hi-fi prototype, which we tested in pilot studies. The main result from the pilot studies is that strengths and weaknesses in our design were highlighted, which we aimed to improve upon.

The final test done showed no indication either in support or against our framework and its benefits. To do so, we will need a larger number of participants at the appropriate level of math knowledge and skill. With that said our framework bases on previous research on mathematical education and often specifically on digital softwares. That the implementation of all of those would yield no difference in the quality of the product would be counter-intuitive. Pilot study 2 also suggest educational software through ill-conceived design can limit the user and hinder the learning experience. Further studies should be conducted about the actual use of mathematical software used in schools to seek valid design strategies that can be used to improve existing products.

References

Aleven, V., Stahl, E., Schworm, S., Fischer, F. & Wallace, R. (2003). Help Seeking and Help Design in Interactive Le-

- arning Environments. *Review of Educational Research*, 73: 277–320.
- Anderson, J. R., Reder, L. M. & Simon, H. A. (1996). Situated Learning and Education. *Educational Researcher*, 25(4): 5–11.
- Ashcraft, M. H. (2002). Math Anxiety: Personal, Educational, and Cognitive Consequences. *Current Directions in Psychological Science*, 11(5): 181–185.
- Baddeley, A. (1992). Working memory. *Science*, 255(5044): 556–559.
- Betz, N. E. (1978). Prevalence, Distribution, and Correlates of Math Anxiety in College Students. *Journal of Counseling Psychology*, 25(5): 441.
- Botte, B. A., Heinrichs, M., Mehta, Z. D. & Hung, Y. H. (2002). Weighing the Benefits of Anchored Math Instruction for Students with Disabilities in General Education Classes. *The Journal of Special Education*, 35(4): 186–200.
- Catrambone, R. & Holyoak, K. J. (1989). Overcoming Contextual Limitations on Problem-Solving Transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6): 1147.
- Carbonneau, K. J., Marley, S. C. & Selig, J. P. (2013). A Meta-Analysis of the Efficacy of Teaching Mathematics With Concrete Manipulatives. *Journal of Educational Psychology*, 105(2): 380–400.
- Chaiklin, S. (2003). The Zone of Proximal Development in Vygotsky's Analysis of Learning and Instruction. In A. Kozulin, B. Gindis, V. Ageyev & S. Miller (Eds.), *Vygotsky's Educational Theory in Cultural Context* (pp. 39–64). Cambridge: Cambridge University Press.
- Cowan, R. & Powell, D. (2014). The Contributions of Domain-General and Numerical Factors to Third-Grade Arithmetic Skills and Mathematical Learning Disability. *Journal of Educational Psychology*, 106(1): 214–229.
- Deci, E. L. (1971). Effects of Externally Mediated Rewards on Intrinsic Motivation. *Journal of Personality and Social Psychology*, 18: 105–115.
- DeRose, K. (2009). *The Case for Contextualism: Knowledge, Skepticism, and Context, Vol. 1*. Oxford University Press: Oxford.
- Divjak, B. & Tomić, D. (2011). The Impact of Game-Based Learning on the Achievement of Learning Goals and Motivation for Learning Mathematics – Literature Review. *Journal of Information and Organizational Sciences*, 35(1): 15–30.
- Ellington, A. J. (2003). A Meta-Analysis of the Effects of Calculators on Students' Achievement and Attitude Levels in Precollege Mathematics Classes. *Journal for Research in Mathematics Education*, 433–463.
- Furner, J. M. & Berman, B. T. (2003). Review of Research: Math Anxiety: Overcoming a Major Obstacle to the Improvement of Student Math Performance. *Childhood Education*, 79(3): 170–174.
- Gee, J. P. (2003). *What Video Games have to Teach Us about Learning and Literacy*. New York: Palgrave Macmillan.
- Ginsburg, H. P., Jamalian, A. & Creighan, S. (2013). Cognitive Guidelines for the Design and Evaluation of Early Mathematics Software: The Example of *MathemAntics*. In L. D. English & J. T. Mulligan (Eds), *Reconceptualizing Early Mathematics Learning*. Springer Netherlands. 83–120.

- Gärdenfors, P. (2008). The Role of Understanding in Human Nature, presentation paper at *ESSSAT*, Sigtuna.
- Gärdenfors, P. & Lindström, P. (2008). Understanding by Experiencing Patterns, in P. Gärdenfors & A. Wallin (Eds.), *A Smorgasbord of Cognitive Science*. Bokförlaget Nya Doxa: Riga. 149–182.
- Habgood, J.M.P. & Ainsworth, S.E. (2011). Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. *Journal of the Learning Sciences*, 20: 169–206.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1): 81–112.
- Hyun, J. S. & Luck, S. J. (2007). Visual Working Memory as the Substrate for Mental Rotation. *Psychonomic Bulletin & Review*, 14(1): 154–158.
- Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., ... & Wearne, D. (1996). Problem Solving as a Basis for Reform in Curriculum and Instruction: The Case of Mathematics. *Educational Researcher*, 25(4): 12–21.
- Hiebert, J. & Grouws, D. A. (2007). The Effects of Classroom Mathematics Teaching on Students' Learning, In *Second Handbook of Research on Mathematics Teaching and Learning*, 1: 371–404.
- Hyun, J. S. & Luck, S. J. (2007). Visual Working Memory as the Substrate for Mental Rotation. *Psychon. B. Rev.* 14(1): 154–158.
- Jonassen, D. H. (1997). Instructional Design Models for Well-Structured and III-Structured Problem-Solving Learning Outcomes. *Educational Technology Research and Development*, 45(1): 65–94.
- Kebritchi, M., Hirumi, A. & Bai, H. (2010). The Effects of Modern Mathematics Computer Games on Mathematics Achievement and Class Motivation. *Computers & Education*, 55(2): 427–443.
- Kinzer, C., Bradley, J. & Morandi, P. (2013). Feedback to Support Learning in the Leadership Institute for Teachers. *The Mathematics Enthusiast*, 10(3): 563–582.
- Kirsh, D. & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, 18(4): 513–549.
- Lakoff, G. & Núñez, R. E. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics Into Being*. Basic Books: New York.
- Lave, J. (1988). *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Cambridge University Press.
- Lindström, P. (2011). *Understanding as Experiencing a Pattern*. Ph.D. Thesis, Cognitive Science, Department of Philosophy, Lund University: Lund.
- Moreno, R. & Mayer, R. (2007). Interactive Multimodal Learning Environments. *Educational Psychology Review*, 19(3): 309–326.
- Moyer, P. S., Bolyard, J. J. & Spikell, M. A. (2002). What are Virtual Manipulatives?. *Teaching Children Mathematics*, 8(6): 372–377.
- Norman, D. A. (1988). *The Design of Everyday Things*. Basic Books: New York.
- Paas, F., Renkl, A. & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, 38(1): 1–4.
- Plotkin, H. C. (1993). *Darwin Machines and the Nature of Knowledge*. Harvard University Press: Cambridge, MA.
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1): 153–189.
- Stodolsky, S. S., Salk, S. & Glaessner, B. (1991). Student Views About Learning Math and Social Studies. *American Educational Research Journal*, 28(1): 89–116.
- Suh, J. & Heo, H. J. (2005). Examining Technology Users in the Classroom: Developing Fraction Sense using Virtual Manipulatives Concept Tutorials. *Journal of Interactive Online Learning*, 3, 4: 1–21.
- Tüzün, H. et al. (2009). The Effects of Computer Games on Primary School Students' Achievement and Motivation in Geography Learning. *Computers & Education*, 52(1): 68–77.
- Witzel, B. S., Mercer, C. D. & Miller, M. D. (2003). Teaching Algebra to Students with Learning Difficulties: An Investigation of an Explicit Instruction Model. *Learning Disabilities Research and Practice*, 18: 121–131.
- Zhang, J. & Norman, D. A. (1994). Representations in Distributed Cognitive Tasks. *Cognitive Science*, 18(1): 87–122.
- Zhu, X. & Simon, H. (1987). Learning Mathematics from Examples and by Doing. *Cognit. Instr.* 4: 137–166.

Appendix I - Transferal Test Questions

Transferal Test Questions 1:

1) Write the number in expanded form:
 $2345 =$

2) Write the number in expanded form:
 $4893 =$

3) Write the number in expanded form:
 $19823 =$

4) Write the number in expanded form:
 $303 =$

5) Fill in the correct number:
 $9_23 = 9000 + 300 + 20 + 3$

6) Fill in the correct number:
 $98_7 = 9000 + 800 + 7$

7) Which expanded form is greater of “ $2000 + 700 + 90 + 8$ ” and “ $4000 + 100 + 10 + 2$ ”?

8) Which expanded form is greater of “ $3 + 70 + 900 + 8000$ ” and “ $5 + 10 + 100 + 2000$ ”?

Transferal Test Questions 2:

1) Write the number in expanded form:
 $8492 =$

2) Write the number in expanded form:
 $1763 =$

3) Write the number in expanded form:
 $61893 =$

4) Write the number in expanded form:
 $204 =$

5) Fill in the correct number:
 $4_89 = 4000 + 600 + 80 + 9$

6) Fill in the correct number:
 $33_2 = 3000 + 300 + 2$

7) Which expanded form is greater of “ $5000 + 900 + 90 + 6$ ” and “ $7000 + 100 + 10 + 2$ ”?

8) Which expanded form is greater of “ $2 + 40 + 900 + 3000$ ” and “ $7 + 30 + 600 + 2000$ ”?

Oculus Rift: Virtual green environments for elderly

Alexander Badju, Elise Ellerstedt, John Larsson, Joel Olofsson

How would elderly people want a virtual environment to look like? In the work that has lead to this article, we have tried to map how elderly people should interact with a computer and virtual reality, as well as investigate what they would like to experience in a virtual world, aesthetically as well as activity wise.

Our early user tests and interviews showed us that the people, who wanted to try Oculus Rift, to a large extent, also had previous computer knowledge and they were also curious to see and test virtual reality. The tests and interviews concluded that test persons liked nature and the ocean, furthermore they wanted to have an activity to do in the virtual world.

The user tests that have been conducted have served the purpose of mapping what elderly people prefer in virtual reality, green environments or urban environments. To receive a result in this field, activity contra passiveness has been explored. In the activity, the test persons have performed a quiz walk. The whole test was done with a 2x2 test model. Passive green environment, active green environment, passive urban environment and active urban environment are the four alternatives that have been tested and evaluated. This test model was chosen through consolidation and communication with Mattias Wallergård and Susanne Frennert. The decision to have the same activity in both the active environments was made when meeting with Matilda Van Den Bosch at The Agricultural University of Sweden.

The concepts that were used to measure which environment is the most preferable are called awareness and cognitive absorption, where awareness is the main concept. This concept is a measurement of to what extent a human is aware of "what" he/she is doing and "where" he/she is as well as how prone the user is to use the system again. This is one of the most important aspects since the point is to have the elderly be in the virtual environment maybe a few times a week to pass time and to activate themselves.

1 Introduction

Because of the medical prosperities of today and the near future, the Earth's population is getting even older. As a result, the shortage of personnel at homes for the elderly is increasing. It is no longer self-evident with walks in the park and time to spend in the garden. Many of the seniors get isolated in their homes. Virtual green environments presented with *Oculus Rift* are an alternative for the senior citizens

who lack interaction with nature in their every-day life. The revolutionizing technology is built with a small high-resolution screen, which is placed on your head like a pair of ski-goggles. It is also known as a *Head-mounted display (HMD)*.

This generation with minimal computer experience is however hard to motivate, that is at least what our early pilot tests revealed. Therefore it is important that the environments and the activities captivate the users. In this article a study is presented of what type of environments appeal to elderly people, in what degree they are aware of their surroundings when using *Oculus Rift* as well as how they experience wearing the equipment. However, what is most important for devoted and committed users, green environments or activity?

1.1 Objective

The purpose of this project is to develop virtual environments for *Oculus Rift* and investigate how elderly respond to them. In earlier studies done by e.g. *The Swedish University of Agricultural Sciences*, it has been proven that green rehabilitation and garden therapy have a positive impact on seniors' health and wellbeing. Unfortunately many elders have difficulties moving and therefore spend most parts of the year indoors. In the future, green virtual environments will hopefully become complementary for the senior citizens who are not mobile enough to move outside without assistance. An important factor is to make the environments attractive enough for people to return to them, for a person who is not used to computing and these types of devices, a sort of personality change or openness to technology is needed, which could prove to be a challenge. In this article the aspects that needs to be considered when developing environments aimed for seniors are presented.

1.2 Target group

The target group for our tests are older people around the age of 55. They also need to have some computer experience to be able to take part in the experiment. For future use however, as explained in *1.1 Objective*, the target group for end consumers are people who are not able to be outdoors and activate themselves for different reasons. This technology will be developed to a higher level of sophistication by the

time today's 55 year olds are older, retired and less mobile.

1.3 Restrictions

The project was limited to a few weeks and because of that reason, the scope of development needed to be restricted. Therefore, complete 3D environments were downloaded from *Unity's Asset Store* (Unity is the program and development environment we used) and additional functions were then implemented by our team to fit the requirements of this study.

After pilot-testings, performed in an early phase of the project, at a rallying point for elderly it became evident that today's seniors with the average age of 75-80 had difficulties navigating with computers, especially keyboards. Consequently, the interaction possibilities in the environments in this study are limited. This is considered to enhance their experience, as they do not need to waste energy and become frustrated when trying to move around single-handedly in the virtual environment.

1.4 Method

The project started with recovering information from existing articles about research that was relevant to our study. It gave us a steady ground and insight in the aspects that needed to be considered when producing a study like this. At an earlier stage of the project we visited *The Swedish University of Agricultural Sciences*, situated in *Alnarp*, to meet with *Matilda Van Den Bosch* who is a researcher in environmental psychology. She gave us feedback and tips that were valuable for the continued development.

An initial user test was performed at a rallying point for seniors where the users tested three different environments with *Oculus Rift*. It became evident that the elderly had difficulties navigating with computers and therefore a younger target group was chosen. At this point we also decided to limit the user's interaction possibilities as explained under *1.3 Restrictions*.

After deciding the dependent variable *awareness* we defined the qualifications to be tested. A 2x2 test design, consisting of four different environments, was employed and later used in the user tests with the purpose of pitting the environments against each other. The environments were downloaded and developed in *Unity* to satisfy our requirements on functionalities. The user tests were then performed with twelve participants.

1.5 Scales and measuring method

We have chosen to use a six-point scale when grading our interview questions. This is because we want the scale to lean one way or another so when a participant chooses an alternative, he or she is forced to say more than if the scale would have been five-point scale. (Shaughnessy et al., 2010)

For analysing the data we have chosen mean instead of mode and median. This is first and foremost because it is a six-point scale system where mean is preferable before mode and even median when the standard deviation is low. (Shaughnessy et al., 2010)

2 Theories

2.1 Oculus rift and human eyesight

The most common way for humans to perceive and understand the world around them is by observing it through visual perception.

There is a term called *Binocular depth cues*, which defines different types of clues the human eye uses when observing the observable world. These clues could be the depth or distance in a picture for example. Since our eyes are located approximately six centimetres from each other, an object that is closer than 30 meters from our eyes will appear differently for the right and left eye respectively. E.g. hold an object in front of your eyes and, close one eye at a time, then do it in quick succession, you can see that it seems as if the object shifts position, this is because the right and left eye perceive the object from different angles. *Binocular depth cues* define stereopsis whilst another term called *Monocular depth cues* is the definition of a more accurate distance estimation (Wallergård, Eriksson, Alce, 2013).

The *Oculus Rift* device has a screen for each eye and the picture projected in each screen is projected in such a way that the phenomenon described earlier with the shifting object is eliminated. When looking through the *Oculus Rift* the picture is calibrated so it looks real and accurate.

2.2 Activeness vs. Passiveness in virtual reality

The majority of tests and research done on activeness contra passiveness in virtual reality has mostly been about how efficient humans learn as well as how the memory is affected through different types of experiments.



Figure 1. Front view of the Oculus Rift



Figure 2. Back view of the Oculus Rift, the two eye screens can be seen (Figure reference: http://en.wikipedia.org/wiki/Oculus_Rift)

In an experiment conducted by students at *Université René Descartes* in Paris, human episodic memory in virtual reality has been examined. Their experiments were based on two people travelling in a car into a virtual town, which none of them had seen before, the test persons had to remember and recognize e.g. landmarks, monuments etc. One person in the car was the driver (active person) and the other was a passenger (passive person). Their results were as follows, the passive passenger had a better overview of the town but the active passenger could remember details in the environment as well as occurrences more thoroughly. Their conclusion was that the test subjects remembered things and situations crucial for their given task (Gaën Plancher, Sylvain Hauptert, Valérie Gyselinck, Hubert Tardieu, Serge Nicolas and Pascale Piolino, no date).

A number of researchers at *ICREA-University in Barcelona* have used *Oculus Rift* as distraction therapy at dentist appointments. The researchers in this case examined activity contra passivity regarding how patients experienced the dentist appointment. The results showed that the test subjects who were active in the virtual environment perceived the experience as if they had more control, consciousness and more pleasurable than the test subjects who were in the passive virtual environment (Tanja-Dijkstra K, Pahl S, White P. M, Andrade J, Qian C, Bruce M, May J, Moles R. D, 2014).

2.3 Motion Sickness

Motion sickness has been a problem within the virtual reality domain. Along the lines of Richard Yao et al. (2014) best practice document there are a number of actions you can take to prevent motion sickness. One of these is to implement an avatar for the player that casts a shadow. This helps the player to get a sense of self-reflection and in that case reduce motion sickness. Furthermore you should make the speeds and acceleration at a pace that is not too fast, since it might cause motion sickness for some people. Another thing to think about when designing virtual worlds is heights. The people that experience nausea when standing in high places in the real world are likely to have the same problem in virtual reality and that factor should therefore be taken in consideration during development of virtual realities.

To avoid motion sickness it is also important that the *Oculus Rift* is in good condition as well. Malfunctioning gear is one common problem when working with virtual reality. When the gear is not working properly, latency issues and distortion could cause motion sickness. At last it is important to acknowledge experience as an important factor for any virtual reality project. This factor can be divided in two aspects, first the more experience a player has with virtual reality the less is the risk that he or she experiences motion sickness. This is the same as in general terms of experience and the expectation of troubles. The second point of view of experience is that too much is never good. If a player sits with the *HMD* on for hours it is not a surprise that motion sickness will be the end result. (Richard Yao et al. 2014)

2.4 Awareness

A great challenge for virtual reality developers is to attract the user/consumer back to those environments. *Goel et al.*

(2011) proposes a model that builds on predicting the users intentions to return to a virtual world. The idea is that the user perceives virtual reality as a place where he or she had useful experiences (according to *Theory of place Attachment*) and the quality of these experiences will influence the users *intention to return*. The intentions to return to a virtual reality are based on a state of deep commitment (*cognitive absorption*) that the user experiences when they perform an activity and in this immersed condition lose track of time.

Goel et al. (2011) developed a theory on *cognitive absorption* and how this state can be obtained, which properties in virtual reality that can contribute to increase the experience of *awareness*.

The users propensity to return to virtual reality is based upon their emotional experience of the place they are in. The users creates a “site sense” with the following questions “where am I?”, “what am I supposed to do?”.

Goel et al. (2011) assumes from *Interactionist theory of place attachment (ITPA)* two related concepts underpinning the theory is *place* and *space*. The *space* is an abstract phenomenon without any personal experiences attached. When the person has explored the *space* and realized what it contains and what to do, the person then creates a value of the *space* and then that becomes a *place*. A *place* is a *space* that contains value, and the *space* transforms into a *place* when the person has interacted with it. According to *ITPA*, virtual reality is a mediated *place* that develops into a *space* through the users sensory experience. Just like in the real world VR (Virtual Reality) can create experiences that are equivalent to the experiences emerged in the real world, with equally high intensity. Studies in the field of *Cognitive Neuroscience* have shown that the image of a chocolate bar activates the same regions in the brain, as it was a real object. A place represented in VR can provide both positive and negative experiences, like in the real world (Goel et. Al., 2011).

Awareness is a condition where a person processes vague, general information received through sensory perception. Information about “what” and “where” as the person becomes aware of consolidates, forming *awareness*. When a meaning gets ascribed to such information, it can result in both positive and negative experiences attributed to that information. There are three types of *awareness* Goel et. al. (2011) takes into account, which determine the degree of *cognitive absorption*.

Social awareness: The user’s perception of how he or she should understand and interact with other characters in VR. *Social awareness* is important because the meaningfulness of activities considered to be socially constructed.

Task awareness: concerns the interaction with object in VR and the user’s perception of what should be done in VR to achieve a given goal. *Task awareness* is defined as the user’s perception of what has to be done in VR, based on instructions, tools and actions of other characters. It may involve series of activities to achieve a given goal.

Location awareness: concerns navigation in VR and by definition the user’s perception of where the person is in context of object in the VR and what activities are supposed to be done there.

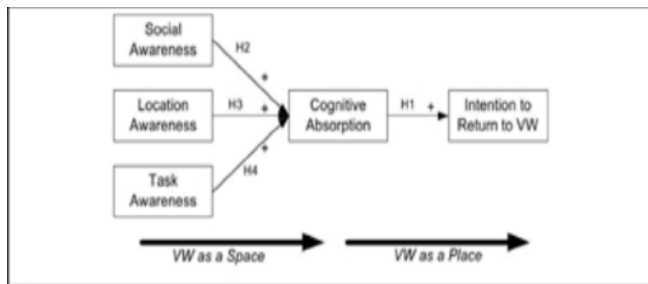


Figure 3. Goel et al. (2011) p. 758

Agarwal and Karahanna (2000) have noted conceptual similarities between the state of *cognitive absorption* and the experience of *flow* from earlier studies. Both include absorbed attention during activities where users appear to enjoy themselves intensely and are so involved in an activity like nothing else seems to matter. Dimensions of *flow* include intense concentration, a sense of being in control, loss of self-consciousness and a transformation of time.

In our study we are going to focus on the user's time perception and the experience of control besides *Location awareness*, *task awareness*, *Cognitive absorption* and *Intention to return* since they seem to matter in relation to experience of *cognitive absorption* according to the framework presented by Agarwal and Karahanna (2000).

2.4.1 Design guidelines to encourage location awareness and task awareness in VR

Users in Virtual Reality get visual clues from their surroundings. The users attribute meaning to those visual clues based on memories of interaction in similar environments. If a user gets a clear idea of where he or she is located in the virtual environment and can avoid redundant searches for places and objects, it will have a positive effect on the experience of immersion. To mark strategically important locations, such as the entrance to important places will encourage the experience of *location awareness* and have a positive effect on the experience of immersion.

The successful interaction in VR depends upon how well the users perceive the visual clues that are sent by the objects. To achieve and facilitate *task awareness* the designers can highlight certain aspects of the interaction, such as putting up and aura around objects in VR that would provide a strong indication of how it should be used.

2.5 Virtual reality with Oculus Rift and pain

Earlier in this article it was mentioned that scientists at the ICREA-university in Barcelona used *Oculus Rift* as distraction therapy at the dentist in purpose to decrease the pain for the patients. *Oculus Rift* was used as distraction tool in this case. The patients shared their experience of reduced pain through interviews and their stories were confirmed with the heartbeat and blood pressure of the patients. This research also showed that interactivity had a role along with a high

level of consciousness in the virtual world, so called *presence*. The scientists came to the conclusion that not only did *Oculus Rift* help momentarily for the patients but it also contributed to a positive memory of the treatment. It was especially noticeable on the persons who had fear of dentists (Tanja-Dijkstra K, Pahl S, White P. M, Andrade J, Qian C, Bruce M, May J, Moles R. D, 2014).

2.6 Sound effects in virtual reality

Research examining the meaning and importance of adding sound effects in virtual reality has shown that the experience of realism can be increased through adding auditory modalities to the virtual worlds (Annerstedt et. al., 2013). Other studies show that nature sounds were recognized as pleasant and sounds from e.g. a computer are recognized as unpleasant. Bird song and the sound of pouring water were recognized as the most pleasant sounds. Those sounds have been used in stressful environments like surgical operations and have demonstrated a relief of stress on the autonomy nerve system (Annerstedt et al., 2013).

Annerstedt et al. (2013) has studied sound effects impact in relation to stress reduction. The study had three different conditions: virtual forest that incorporated sounds characteristic for a forest, a virtual forest without any sound effects and a control group without both. Virtual forest with sound effects was the favourite and was taken for having the most positive effects on stress reduction amongst the contenders. The hypothesis was partly confirmed and it stated that sound effects have increased the experience of realism.

2.7 Change the behaviour in people

Physical activity has shown to have great positive effects on a person's health. There are studies that have been done to find ways to activate elderly people, since most of the elderly population is stationary. One of the studies is called the *Steps-to health*-system. The system uses a computer-animated coach to encourage the elders to exercise through walking. For the test, the seniors received a pedometer that measured the covered walked distance. The people in the control group got a laptop with a touchscreen to use during the time period of two months. The whole test group was followed for the study during the duration of ten months. The control group received the task to go online and discuss walking and exercise with the virtual coach. The result showed that the seniors were doing more exercise the first two months but after that, they got more and more lazy and their activity went down. This implementation was designed to increase the activity amongst the elderly but it required more structural, personality and mind-set changes to increase the exercising even more. (Paasche-Orlow, M., Silliman, R., Winter, M., Cheng, D., Henault, L., Bickmore, T. 2012)

2.8 Garden therapy

The first studies about garden therapy were presented in 1984. Roger Ulrich at the university of Texas in the US made a study where patients who were recently operated got a view over a park or a brick wall. The patients with the view over a park healed quicker than the patients who only

watched a brick wall. The patients did not receive any physical activity or fresh air, just watching the nature made them heal faster. The patients only viewing the brick wall did not feel as if they were located in a calm resting place.

Table 1.

1. Peacefulness	<i>Calmness, silence and signs of care. The sound of wind, water, birds, insects. No litter, no weeds, lack of disturbing people. In buildings a peaceful feeling can be experienced in churches.</i>
2. Nature	<i>Fascination of wild nature. Plants appear to have been growing naturally. Stones have natural moss growing on top of them. They are not created by any human but by a higher power, so to speak.</i>
3. Richness of species	<i>In this dimension you can experience the richness of different species, animal life as well as plant life.</i>
4. Space	<i>The place/space offers a calm feeling and a feeling of entering a new and different world.</i>
5. Openness	<i>A quality, which offers a view as well as spontaneous activities. E.g. a green field, which invites people to sit down.</i>
6. Protection	<i>A safe and sound place, a shelter of sorts. A secret oasis where you can be yourself but at the same time experiment and have fun.</i>
7. Social	<i>A social arena, a meeting point for pleasant hangouts, such as parties.</i>
8. Culture	<i>A place that offers fascination of people's values, sacrifices, hard work and experiences. This dimension should also include historical elements, such as an ability to reminisce about old times.</i>

Research showed that people feel calm and safe when sitting at a small height with view over a bright nature site with sparse tree vegetation and water in the background. Ulrich's (1984) theory goes back to the Stone Age. It was then preferable to live and have a residence near food and water with a view over the surroundings. (Grahns P, Ottosson Å, 2010, page 58)

2.8.1 Grahns eight perceived dimensions

Per Grahns has conducted research on green areas in cities and has done a scientific study together with Ulrika Stigsdotter on 2000 randomly chosen swedes. As a conclusion Grahns stated "Each individual in a green environment has a

need to find meaning and purpose to help them in their situations and this occurs in an interaction between human and nature". Grahns found that people gravitated towards green environments to experience some qualities or dimensions only a green environment could offer. Here follows Grahns's eight dimensions (Grahns P, Ottosson Å, 2010, page 67, 68).

3 User tests

Experimental design

In our study, 2x2 design is applied with four different conditions. The dependent variable is the experience of awareness, which in turn is divided into subcategories such as, location awareness task awareness, the control experience, cognitive absorption, intention to return and motion sickness that we believe has substantial effect on the other subcategories and the holistic experience of Cognitive absorption.

The measurable four conditions are passive urban environment, active urban environment, passive green environment and active green environment.

An important factor that has influenced the design of our experiment is the experience of motion sickness and dizziness that are common consequences when using Oculus Rift. For that reason, we chose to not present all four virtual environments to each one of our participants since the risk of motion sickness and other discomfort increases the longer Oculus Rift is worn by the participants. Instead of within subject design where participants go through all the conditions we will choose between subjects design. Each participant will go through two of the total four virtual worlds.

The formations of Experiment design:

The dependent variable: Awareness

Location awareness, task awareness, the control experience, cognitive absorption and Intention to return were subcategories to awareness that were measured with a 6 point scale (where 1 represented "totally disagree" and 6 "totally agree"). Those subcategories were taken from the studies of Goel et. al. (2011) and Agarwal & Karahanna (2000) and compose a mixture from the two studies, to get as clear picture of Cognitive absorption as possible.

Another important feature in the context of user experience was measured, the experience of motion sickness, also with a six point scale.

Each subcategory had 2-3 question related to the participant's experience of the certain aspect of the virtual environment. Questions aimed to answer location awareness for example contained the following questions: "I was conscious of elements I passed by in the virtual environment", "I was curious about objects in the virtual environment". Questions aimed to analyse cognitive absorption contained following questions: "Time appeared to go by quickly when I was passing through the virtual environment" "I lost track of time when I was going through the virtual environment".

The full form is available as an attached document. Questions 1-3 are aimed to analyse Location awareness, 4-5 Task awareness, 6-7 The Control experience, 8-9 Cognitive absorption, 10-12 Intention to return, 13-15 Motion sickness.

The independent variables: **A:** Urban passive environment, **B:** Urban active environment, **C:** Passive green environment, **D:** Active green environment

The design contains independent variables with the opportunity to obtain following results according to *Borg and Westerlund* (2006, pp 281-286):

- We looked for a main effect of the green environment / urban environment. The degree of perceived *awareness* has to do with what kind of environment participants encounter (the expected result is that the green environment would be preferred by participants).

- We looked for a main effect of passivity / activity. The degree of perceived *awareness* has to do with whether there is an activity in the virtual environment.

- We looked for an interaction effect between the environment and the activity / passivity (The expected result was that active green environment would be outstanding in relation to experienced *awareness* in contrary to the three other tested environments).

3.1 Test structure

The test was performed in three subsequent phases:

1) Participants received a presentation of *Oculus Rift* and got informed that the *University of Lund* was making a study to evaluate if the technology can be used by people aged 55+ in the future (and not just people under 30 who are the current users). The goal of the study was to evaluate how participants react to the new technology and identify their needs and preferences. The participants were withheld from the actual purpose of the experiment in order to provide as objective answers as possible. Participants were informed that their participation was entirely voluntary and that they could interrupt at any time if they experienced discomfort in any way. Participants were informed in advance what kind of environments they were going to encounter and how they could interact with the virtual environment (by pushing any button on the keyboard and looking around).

2) Participants encountered a virtual environment and answered a questionnaire after each tested environment. When they had participated in both virtual environments, participants answered a questionnaire designed to compare the both environments.

Table 2.

Participant 1: A+B	Participant 2: A+C
Participant 3: A+D	Participant 4: B+A
Participant 5: B+C	Participant 6: B +D
Participant 7: C+A	Participant 8: C+B
Participant 9: C+D	Participant 10: D+A
Participant 11: D+B	Participant 12: D+C

3) The participants got debriefed, thanked for their participation and informed that they can get information about the result of the study in the future if they wish.

A semi-structured interview method was used in the study, and the questions were read to the participants by test-leaders. We had chosen to use a six-point grading scale for our interview questions in order to get the answer that leans to one side or the other.

The semi-structured interview method included follow-up questions where participants could elaborate their answers and not be bound to a 6-point scale if they wished to comment on any aspects of a tested virtual environment.

The tests took place in a small soundproof room with only the participant and test-leaders present. The participant's answers have been written down by test-leaders, and the participants were filmed while they encountered the virtual environments.

Besides questionnaires we also applied an observation template, which aimed to track the participant's head movements in order to see if they grasped the *Oculus Rift* technology or if they would perceive it as a regular screen and not look around.

3.1.1 Predictions

Our predictions for this study is accordingly to earlier stated theories from Tanja-Dijkstra et al. (2014) about activity and passivity to think the most chosen environment should be an active environment since activity is preferable to passivity. Accordingly to *Ulrich* (1984) green environments are preferred over urban environments for recreational activity, which is something we want the elderly to do when they are experience our virtual worlds.

4 Results and analysis

4.1 Environments

4.1.1 Activeness vs. Passiveness in green-and urban virtual environments

The purpose of our initial pilot studies was to compare activeness and passiveness in a virtual environment in order to get an understanding of what the elderly in our target group would prefer. The results of the initial pilot studies showed that the people in our target group would rather have an activity in their virtual environments than being passive. As explained earlier in this article our final tests were done comparing an active green environment, a passive green environment, an active urban environment and a passive urban environment to see which of the four types the test subjects preferred. The urban environment was represented by a Japanese city and the green environment was a tropical island. The tropical island was chosen for its natural and realistic look, it fulfils some of *Grahn's* (2010) eight dimensions. *Peacefulness* through bird sounds and other natural sounds, the *Nature*-dimension through the detailed portrayed nature (trees, grass, rocks etc.). The *Space*-dimension is represented by simply entering a tropical island environment from previously being in a normal room, *Protection* is experienced since the virtual environment is an oasis of sorts and

a room and a place where you can let your fantasy run free. The *Richness of species*- is also fulfilled thanks to birds flying around as well as many different plants and trees.

4.1.2 The active environments

The activity chosen for the active virtual environments was a so-called “quiz walk”. A quiz walk is an activity where a person moves along a given path and at certain locations on the path there are quiz questions, which the person first reads and then answers. The setup of a quiz question is as follows, first there is a number indicating which question the person has arrived at. Then the question is presented, which the person first reads, when answering the question there are three alternatives to choose from. The alternatives are numbered 1., X., 2.. The person then proceeds with answering the question by marking e.g. a 2. on e.g. question number five. When the question has been answered, the person continues on the given path heading to the next question where he/she repeats the same process.

Fråga 5. (5/5): Hur många drottningar finns det normalt i en bikupa?
1. 3 st
X. 18 st
2. 1 st

Figure 4. An example of a question from the quiz walk (in Swedish).



Figure 5. The blue line is the path the user travels along in the active green environment.



Figure 6. The green box surrounding the question (in the

green environment) is the zone described in the text above.

In our active virtual environments the quiz walk is quite similar to a real one, the difference is that the people do not walk along the path themselves. Instead we programmed the environments in such a way that the person is still but moves along the predefined path automatically, the metaphor we use is that the person is sitting in a wagon on a rail. Since the user rides along the path the *location awareness*-variable may suffer, but *task awareness* is tested to a larger extent because of the quiz activity. Experience tells us that people get very involved and focused when quizzing.

When the person is moving along the path, he or she will encounter the quiz questions, which are placed within an even distance of each other. There are five questions in total for each active environment. When the person is moving along the path and a question appears, the question can be seen from a distance to avoid startling the person with an unexpected pop-up window, the person stops automatically in front of the question so that the person can read and answer. When answering the question, the test person simply said what alternative he or she believed to be correct for that particular question and we wrote it down. The question has a zone surrounding it, which recognizes when an object enters the zone and then proceeds to stop that object in its path in front of the question, within reading distance.

4.1.3 The passive environments

The environments which are passive are designed in such a way that the users travel on the same predefined path as in the active environments, the difference lies in the users not taking part in the quiz walk activity. The user can still look around in the environment and listen to the sounds surrounding him or her.

The reason for the user just being able to ride along the path and not being able to walk around by him-/herself is because it is important that the environments, as well as the interaction differences between the different environments, are as similar as possible to avoid distorted results in the final user tests.

4.1.4 Sound effects

Both the green and urban environments had sound effects incorporated to increase the vividness and realism of the virtual environments. As *Grahn* (2010) explains in his *eight perceived dimensions* and his first dimension, *peacefulness*, nature sounds are soothing and that they are an important part of a persons experience in a green environment. The green environment (the tropical island) had typical jungle sounds such as, bird song, the sound of the wind, crickets and other insect sounds. The urban environments sound effects were mostly based on car sounds and general traffic noise that could be heard from a distance.

4.2 Results

This summary portrait the preferences of twelve users (10 women and 2 men of age 55-65) in the context of a second environment that was tested at the same time. We have taken into account the observation template we created:

1. Chose active urban environment before the passive urban environment. The participant preferred to have an activity before not having one. The trip in the passive urban environment felt like going on a bus where the driver did not know where he was driving. The quiz was perceived as a fun and interesting activity.
2. Passive green environment was chosen before the passive urban environment since the participant appreciated the green environment before the urban environment.
3. The passive urban environment was chosen before the active green environment since the participant perceived herself as a “city person” and did not appreciate nature. She did not appreciate the passive urban environment much either, since the route was way to swingy and she was frequently commenting on driving on the wrong side of the road.
4. Active urban environment was chosen before the passive urban environment since the participant appreciated the activity more than the passive environment. The time went by much faster when there was an activity in the virtual environment.
5. The participant chose active urban environment before the passive green environment since she appreciated the quiz, which was more stimulating. Although she appreciated the passive green environment and thought it would be better for a recreation purposes.
6. The participant preferred the active green environment before the urban active environment. She enjoyed the tasks performed in the virtual environments, but considered the urban environment as boring. The participant has always preferred the nature before urban environments in her life.
7. The participant preferred the active green environment before passive urban environment, since it was more entertaining to do a quiz than not doing one. The participant also preferred the tropical environment before the urban environment, which he perceived as extremely boring.
8. The participant preferred the active urban environment before the passive urban environment - since she liked the quiz but perceived the green environment as most pleasant. She wished she could have done the quiz in a green environment instead.
9. The participant preferred the active green environment before the passive green environment. It was more entertaining to do a quiz that “just runs along the track”. The nature experience more prominent in the passive environment thought.
10. The participant chose the green active environment before the passive urban environment. She did not mention the activity, but preferred the green environment.
11. The participant had to interrupt the test halfway, since she experienced motion sickness and did not want to continue.
12. The participant preferred the passive green environment before the active green environment since she preferred to look around and contemplate the nature and the quiz was in the way of that experience.

Table 3. **A:** Urban passive environment, **B:** Urban active environment, **C:** Passive green environment, **D:** Active green environment. The preferred environments are marked in bold type, e.g. active urban environment (B) was preferred by Participant 1 before passive urban environment (A).

Participant 1: A+B	Participant 2: A+C
Participant 3: A+D	Participant 4: B+A
Participant 5: B+C	Participant 6: B +D
Participant 7: C+A	Participant 8: C+B
Participant 9: C+D	Participant 10: D+A
Participant 11: X	Participant 12: D+C

Figure 7 summarizes the participant's preferences of virtual environments in relation to the other tested environment.

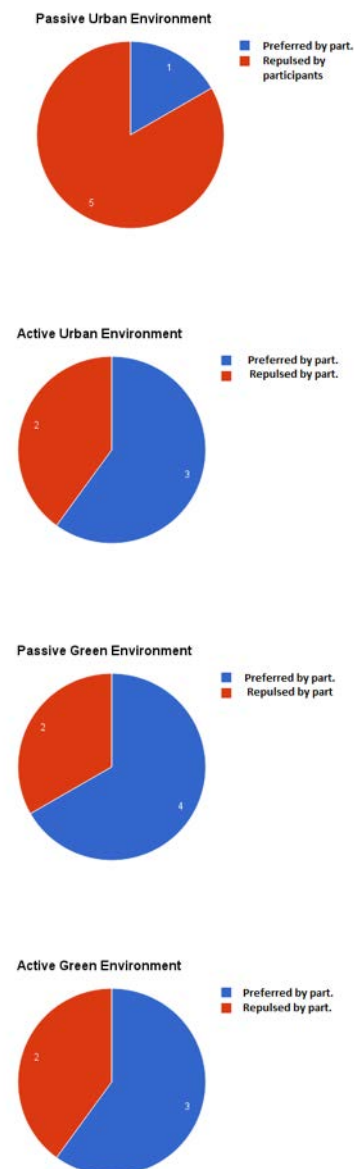


Figure 7. The Pie-charts above summarize the participant's choices of environments

Participants preferred the passive green environment least, one of totally six chose the passive urban environment before another environment. The other three environments: active urban environment, passive green environment and active green environment received the same “score”.

4.2.1 Diagram results

Below the diagrams for the dependent variable are presented. **Serie1** = Passive Urban Environment, **Serie2** = Active Urban Environment, **Serie3** = Passive Green Environment and **Serie4** = Active Green Environment.

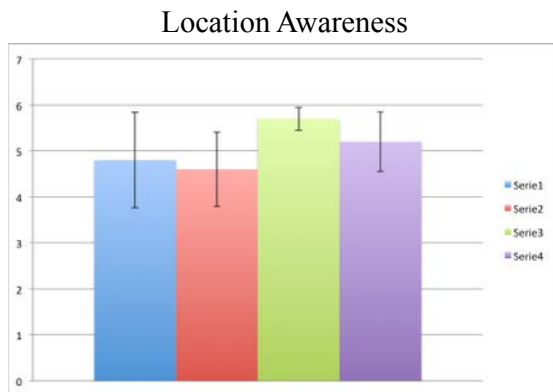


Figure 8

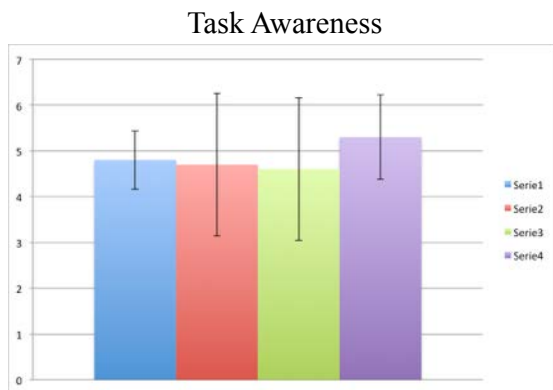


Figure 9

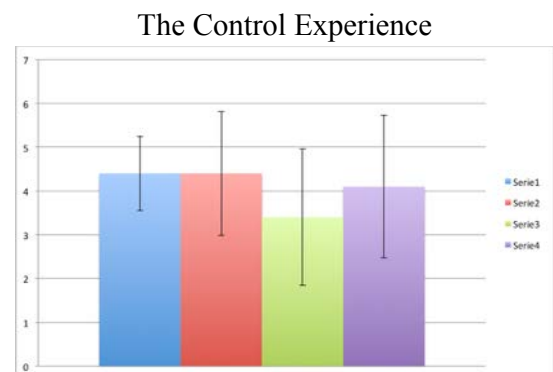


Figure 10

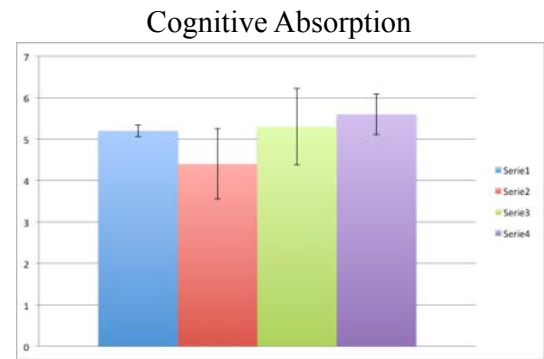


Figure 11

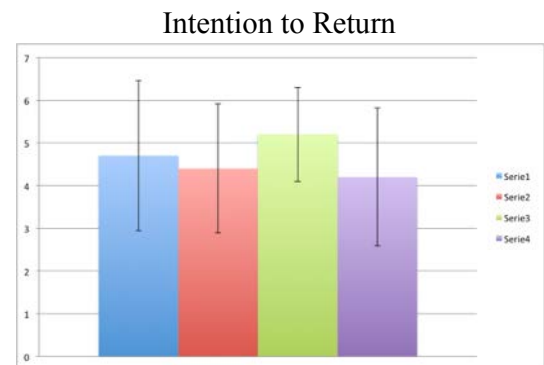


Figure 12

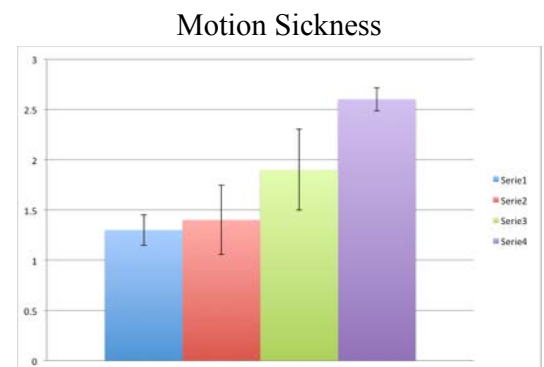


Figure 13

4.2.1 Passive Urban Environment

The study showed that the five indicators of *awareness* in passive urban environment did not distinguish themselves substantially from indicators in other environments. The experience of motion sickness was least substantial in passive urban environment. The participants preferred this environment the least, only one out of six participants chose this environment before another. Participants did not choose this environment mentioned following reasons: the trip felt like going on a buss where the driver did not know where he was going, active environment was more appealing and it felt extremely boring comparing to the environment with an activity.

4.2.2 Active Urban Environment

The diagrams show that *Awareness* levels for the different environments was not very distinguishable. The levels *Cognitive Absorption* did though distinguish itself compared to the others with a lower mark. Amongst the participants this environment was chosen three out of five times. The times it got chosen was against passive urban environment two times and at one occasion passive green environment. The reason the participants chose this environment was because they preferred to have an activity before not having one. They also felt like the time went by faster when having an activity than not having one.

The active urban environment was compared with passive green environment on two occasions, the first time the active urban environment was chosen before the passive green environment since the participant enjoyed the activity, but wished she could do the quiz in a green environment instead. At one occasion passive green environment was turned down due the participants felt the quiz did the active urban environment more stimulating.

In the cases where the active urban environment was turned down it was compared to active green environment the participants preferred the green environment since she preferred nature before the urban environment.

4.2.3 Passive Green Environment

It received higher levels of *Awareness* in the tests for passive green environment in form of *location awareness*. Motion sickness was higher in the case of passive green environment compared to the two urban environments. *Intention to return* also received the biggest score, but surprisingly received the lowest score on control experience.

Passive green environment was chosen four out of six occasions. The two times it was deselected were against the active green environment - the participant enjoyed the activity and at one occasion against the active urban environment, since the quiz was more appealing than passive participation.

This environment was preferred before passive urban environment at two occasions, since nature was more appealing. One of the participants preferred passive green environment before active green environment since she preferred to look around and contemplate the nature and the quiz was in the way of that experience.

At one occasion it was chosen before active green environment - the nature experience was more appealing than quiz.

4.2.4 Active Green Environment

The levels of *Awareness* were significant higher in the case of *task awareness*. *Cognitive absorption* was the highest score in test by a little compared to the others. Motion sickness for the active green environment was the highest score in the test, almost double the score of passive green who had the lowest score.

The active green environment was chosen in three out of five occasions.

The times it did not get chosen was against passive urban environment and passive green environment this due to, in the former case, the participant saw herself as a city person and did not appreciate nature. In the later case the person

thought it was more enjoyable to sit back and contemplate nature then do an activity in it.

Two out of three times it was chosen due to its activity and at one occasion the participant only mentioned the green environment (and omitted the activity) when it was compared to the passive urban environment.

4.3 Analysis

Active environments have been set against passive environments at eight occasions during the study. At five of these occasions active environments were chosen before passive environments. Urban environments were set against green environments on seven occasions. The green environments were chosen before the urban environments at six out of seven occasions.

On two occasions when active environments were set against each other (D+B) - the green active and urban active environments were chosen one time each. On two occasions when passive environments were compared to one another (A+C) - the passive urban and the passive green environments were chosen one time each.

We had four cases of shortfall in our study, the participant could not proceed at one occasion because of motion sickness.

At other three occasions the participant's glasses could not fit in the *Oculus Rift*.

The results of this study make it hard to draw any substantial conclusions. None of the virtual environments showed any significant differences in all *awareness subcategories*.

It is not possible to draw any clear conclusions from the collected data, the passive urban environment distinguished from the other environments and was chosen last by participants. The active green environment was favoured before urban environment at five out of eight occasions and green environments were favoured before urban environments at five out of six occasions - giving a vague indicator that activity and green environments are more favoured.

Our first hypothesis that green environments impact more heavily in *awareness* confirmed partly (not due *awareness* indicators that were similar in other three environment), the users chose the urban passive environment one out of six occasions. An interaction effect was spotted in the active urban environment, activity in urban environment had a positive effect on user's preferences. Our second hypothesis that active environments would impact more heavily on *awareness* than passive environments could not be confirmed by the data.

One more aspect of the tests is that there might be an interaction effect in the active environments compared to the passive. This is due the task that is present in the former environments. This is something added to those environments that make the participants return and creates a value of it own. This is evident when comparing active urban environment to the passive urban environment. With activity the active urban environment gets the same number of selections as preferable environment as the two green environments and thereof is comparable to those two.

The green environments compared to the urban environments had more terrain, which was to its disadvantage. We figure this to contribute quite greatly to motion sickness and this is also confirmed by our collected data. There are no

other evident factors that contributed to the higher motion sickness values in the green environments as we see it. For future studies this would be something to consider since anything that can contribute to a lower expectation of motion sickness is favourable compared not to.

5 Work process in the development phase

Working and developing the virtual environments in the developing program *Unity* for the *Oculus Rift* was not always pleasurable. For a person without any previous experience with *Unity* the most difficult part was to learn and understand what all the different tools and functions did and what possibilities and or restrictions the program had. Since we used virtual environments that were already “made” we had to learn and understand those environments before modifying and altering them to our own specifications. We had a few problems with especially the urban environment since it was very large and the fact that it had so many functionalities built in. We could not simply remove elements from it as we pleased, first we had to assess that the elements we wanted to remove did not have crucial functional connections to the parts of the environment we wanted to keep. This process took a bit of time and was a bit frustrating when the camera view in the environment did not always cooperate in the way we wanted.

The work process involved a whole lot of searching on the web for what you were looking for. Furthermore since time was scarce, there was always a feeling of stress and pressure in the back of our heads, questioning if we were going to make the deadline since we were such novices working with *Unity*. After a few days of hard work and late hours, we started to understand more and more and things got clearer. A key factor was the way we learned to search the web, to know which keywords to use for maximum search results. *Lars Thern*, an acquaintance of ours and someone with a lot of *Unity* knowledge helped us when we needed it, he was a great asset for us to have, especially in the beginning of the development process.

After we had developed the environments it was a bit of a hassle to make the environments sync with the *Oculus Rift*. E.g. initially we used gui windows for the questions in our quiz walk when developing in *Unity* without encountering any problems at all. However, we were quite surprised when we plugged in the *Oculus Rift* that they did not work. This was very frustrating for us, how were we supposed to know that you can develop things in the development environment but later when using the *Oculus Rift* it would stop working. Luckily, we were a bit ahead of our decided deadline so we solved the problem rather quickly. There were always these little problems appearing during the development process, which seemed so unnecessary and they made us quite frustrated from time to time since we felt we were doing everything right but it still would not work.

6 Discussions

There were some problems with our tests regarding the number of participants. It was quite low. The study got the minimum number of participants to test every number of combinations but not any more than that. This might be a problem for the study to really say something. The result we

got might have been completely different if we had more participants. Trends that we did not notice in this study might have become evident if you did more tests or the results we got might have been misleading. So more testing should therefore be performed in future studies.

Another problem with the study is that none of the participants had tried *Oculus Rift* before. This might mislead the results in a way that the first environments got chosen more often than the second one due to the news value. One way to take away this effect is to make the participants start the test with a short introduction to *Oculus Rift* with a different environment than those later to be tested.

The experience of *Local Awareness*, *Task Awareness* and *Cognitive Absorption* got high results among participants, while the *experience of control* and *the intention to return* got lower result. The holistic experience of *awareness* and *Cognitive absorption* was high in all environments. Possibly, the outcome would have been different if the participants got to spend longer periods in each environment e.g. 10 instead of 5 minutes. The difference between the active and passive environment would possibly be more apparent.

Urban environments were very flat with asphalt surface, while the natural environment was much more hilly and provided more discomfort and motion sickness in our participants. Users experienced more motion sickness and less control in the green environments. The uneven surface may have had a major impact on the experience. Two more similar environments (flat both) could possibly have had a different effect on the participant's experience.

Test sessions for each environment was relatively short, 5 minutes each.

In the active environments, the participants could not take in the surroundings in the same way as in the passive environments, one of the participants commented that the quiz was in focus and the environment ended up in the periphery. We believe that an extension of a test time to 10 minutes, where the distance was twice as far between questions would give users a different perspective on the environments.

One conclusion that can be drawn from our study in relation to motion sickness is that *Oculus Rift* still has not reached its peak in terms of development and that the designers should invest in flat environments with minor ups and downs to create as comfortable experience for elderly users as possible.

7 Conclusions

There were no significant differences between the virtual environments in our study in terms of *Awareness* and *Cognitive absorption*. Our first hypothesis that green environment would be preferred before the urban environments was confirmed by the study. Our second hypothesis about activity would be more appreciated in relation to passivity was not confirmed. An interaction effect was discovered in the active urban environment between activity and environment. Our third hypothesis about interaction effect should occur in the active green environment did not come true. More studies have to be conducted with greater numbers of participants and carefully designed tracks that minimizing the effect of *Oculus Rift* and *Motion sickness*. The time spent within each environment should be increased slightly for users to be able

to appreciate both the nature and activity, consequently this will expand differentiation between passive and active environments.

Acknowledgements

We would like to thank Mattias Wallergård and Susanne Frennert for their guidance and support. Lars Thern for his help with the developing environment. The personnel at “Försäkringskassan” made a great contribution to our study through the user testings.

References

- Agarwal, R. and Karahanna, E. 2000 *Time Flies When You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage*, MIS Quarterly (24:4), pp. 665-694
- Borg, E., & Westerlund, J. (2006). Statistik för beteendevetare. Stockholm: Liber
- Grahn P, Ottosson Å, 2010 *Alnarpsmetoden Trädgårdsterapi, Bokförlagaet Bonnier Existens*.
- Gaën Plancher, Sylvain Hauptert, Valérie Gyselinck, Hubert Tardieu, Serge Nicolas and Pascale Piolino, *no date Virtual reality for the study of true and false memory*
- Goel, L., Johnson N. A., Junglas, I., Ives B. *From Space to Place: Predicting Users' Intentions to return to Virtual Worlds* MIS Quarterly Vol. 35 No. 3 pp 749-771/ September 2011
- Richard Yao et al. 2014 *Oculus Rift Best Practice Guide*.
- Paasche-Orlow, M., Silliman, R., Winter, M., Cheng, D., Henault, L., Bickmore, T. 2012
- Efficacy of a Computer-based Intervention to Promote Walking in Older Adults Annual Meeting of the American Geriatrics Society (abstract)*
- Shaughnessy J. J., Zechmeister, E. B. & Zechmeister, J. S. 2010
- Research methods in Psychology, Eighth Edition*
- Tanja-Dijkstra K, Pahl S, White P. M, Andrade J, Qian C, Bruce M, May J, Moles R. D, 2014
- Improving Dental Experiences by Using Virtual Reality Distraction: A Simulation Study*
- Wallergård M, Eriksson J, Alce G, 2013
- Virtual Reality i teori och praktik (MAM101)*
- <http://www.eat.lth.se/fileadmin/eat/MAM101/VR-Kompendium2013.pdf>

Appendix

Frågeformulär

Fråga 1, Jag var medveten om omgivningen och dess element när jag färdades genom den virtuella världen

- 1, Instämmer inte alls
- 2, Instämmer inte
- 3, Instämmer tveksamt
- 4, Instämmer i liten utsträckning
- 5, Instämmer till viss del
- 6, Instämmer helt

Fråga 2, Jag såg element i omgivningen som jag ville ta en närmare titt på

- 1, Instämmer inte alls
- 2, Instämmer inte
- 3, Instämmer tveksamt
- 4, Instämmer i liten utsträckning
- 5, Instämmer till viss del
- 6, Instämmer helt

Fråga 3, Jag var nyfiken på objekt i min omgivning

- 1, Instämmer inte alls
- 2, Instämmer inte
- 3, Instämmer tveksamt
- 4, Instämmer i liten utsträckning
- 5, Instämmer till viss del
- 6, Instämmer helt

Fråga 4, Jag visste vad jag skulle göra i den virtuella världen?

- 1, Instämmer inte alls
- 2, Instämmer inte
- 3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 5, Gav objekt i den virtuella världen mig ledtrådar som indikerade hur jag skulle genomföra uppgiften?

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 6, Jag kände upplevelse av kontroll när jag befann mig i den virtuella världen

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 7, Jag avklarade uppgiften på egen hand och behövde ingen hjälp från testledarna

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 8, Jag tappade bort tidsuppfattning medan jag befann mig i den virtuella världen

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 9, Tiden gick väldigt snabbt när jag befann mig i den virtuella världen

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 10, Jag tyckte det var kul att vistas i den virtuella världen

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 11, Jag skulle gärna vilja ha Oculus Rift med virtuella miljöer hemma och använda den i min vardag i underhållningssyften

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 12, Jag var engagerad i aktiviteten i den virtuella världen

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 13, Jag upplevde åksjuka under mitt besök i den virtuella världen

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 14, Jag upplevde yrsel när jag befann mig i den virtuella världen

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Fråga 15, Jag upplevde att det var ansträngande för ögonen när jag befann mig i den virtuella världen

1, Instämmer inte alls

2, Instämmer inte

3, Instämmer tveksamt

4, Instämmer i liten utsträckning

5, Instämmer till viss del

6, Instämmer helt

Jämförelse av virtuella världar

Fråga 1, Vilken virtuella miljö föredrar du utifrån de miljöer som du har testat?

Fråga 2, Vilken miljö upplevde du var roligast och mest intressant?

Fråga 3, Föredrar du att ha en aktivitet i den virtuella miljön kontra att inte ha det?

Fråga 4, Hade du velat komma tillbaka till någon av de virtuella miljöerna?

Fråga 5, Viken av miljöerna hade du valt att komma tillbaka till utav de två som du har testat?

Digitala läromedel: Feedback för 2000-talet

Axel Duvebäck, Line Flank, Fredrik Lindholm, Christoffer Toft

ras10adu@student.lu.se, adi09lfl@student.lu.se, adi10fli@student.lu.se, dic11cto@student.lu.se

Huvudsyftet med detta arbete var att ta fram en prototyp till ett digitalt läromedel med relativt sett mer välutvecklad feedback och undersöka om barn då lättare tar till sig ett material. Ett ytterligare syfte var att prototypen skulle utnyttja det digitala mediet i högre grad än vad som generellt görs i digitala läromedel. Qnoddarnas värld, från Natur och kultur, var det digitala läromedel vi utgick ifrån, och efter en analys valdes två uppgifter för årskurs två ut att fungera som grund för prototypen och som jämförelsematerial i användartestet. Användartestet gjordes på en skola i Lund med tjugo elever i årskurs ett, som delades upp i två gånger fyra grupper och fick testa varsin uppgift (Qnoddarnas värld Hallon 1.2, Hallon 5.2 eller prototypmotsvarigheten på uppgiften) för att därefter byta plats med gruppen som gjort motsvarande typ av uppgift. Avslutningsvis följde en semistrukturerad intervju. Följande skillnader uppträdde. Eleverna var mer positivt inställda till prototypen än Qnoddarnas värld. De upplevde också svårighetsgraden i olika uppgifter som mer likartad i prototypen medan de upplevde motsvarande uppgifter i Qnoddarnas värld som antingen för lätta eller för svåra. I båda fallen var det så att när eleverna förstått en uppgift, inte nödvändigtvis matematiskt men interaktionsmässigt, gick det snabbare och det krävdes mindre hjälp från testledarna. Samtliga elever behövde muntliga instruktioner och var, enligt testledarnas bedömning, lätt distraherade. Detta kan delvis bero på att uppgifterna eleverna – från årskurs 1 – fick är avsedda för årskurs två, men vi drar också slutsatsen att digitala läromedel för yngre barn inte bör byggas med tanken att barn bara ska sättas framför det och lyssna på instruktionerna och börja lära sig. En del av den informativa feedbacken som gavs i vår prototyp togs inte in av eleverna utan de frågade istället testledaren. Inslagen av konsekvensfeedback i prototypen fungerade dock bättre. Eleverna lyckades lösa problem med hjälp av denna i större utsträckning än med den informativa feedbacken.

1 Syfte

Syftet med projektet är att finna alternativ till den feedback som idag ges i digitala läromedel. De läromedel som i dag är i majoritet på den digitala marknaden innehåller generellt undermålig feedback, vanligen inget annat än enkel rätt/fel-feedback, vilket är helt på sin plats i ett testverktyg men inte i ett lärverktyg. Begreppet "bok på burk" används ofta för att beskriva dessa läromedel då de i stor utsträckning bara fungerar som en digital version av en lärobok. I detta projekt

har vi sökt ta fram alternativa exempel där det digitala mediet utnyttjas till sin helhet och där feedbacken anpassas efter användarens handlingar, speciellt vilka slags fel som görs och hur många.

2 Introduktion

Projektet *feedback för 2000-talet* har utgått från ett digitalt läromedel som heter Qnoddarnas Värld och som avses täcka svenska och matematik i årskurs ett till tre. Läromedlet beskrivs enligt hemsidan som heltäckande och fyllt med "specialdesignat, riktigt bra och pedagogiskt innehåll!". Qnoddarnas Värld är utvecklat för att användas på iPad.[10] Vi har i vårt projekt utgått från uppgifter som idag finns i Qnoddarnas Värld, analyserat dem med avseende på feedback samt tagit fram förbättrade alternativ. Avslutningsvis genomfördes ett användartest med ursprungsuppgifterna samt våra alternativa versioner, implementerade i en prototyp. Uppgifterna är anpassade för årskurs två, men vi valde att genomföra studien med elever i årskurs ett så att inte förkunskaperna i matematik skulle bli avgörande för resultatet. Med andra ord ville vi undvika en takeffekt, där för många barn skulle tycka att uppgifterna var för enkla och klara dem oavsett vilket stöd som gavs eller inte gavs.

Bakgrund

De senaste åren har det blivit allt vanligare att använda iPads i skolundervisningen. Till dessa har det utvecklats en stor mängd digitala läromedel. Applikationerna utvecklas antingen som ett komplement eller som en ersättning för vanliga läromedel. Många av dem är dock bristfälliga ur ett pedagogiskt perspektiv. I en studie av Sjöden och Tärning (under bearbetning) testades 42 applikationer på den svenska marknaden. 62 % av applikationerna hamnade i trial-and-errorkategorin, där den enda feedback som ges är huruvida svaret är korrekt eller inte och utövaren får ange ett nytt svar. 21 % visar det korrekta svaret på uppgiften direkt eller en kort stund efter att ett svar angetts oberoende av om personen svarat rätt eller ej. 17 % av applikationerna hamnade i kategorin informativ, där det ges ledtrådar eller förklaringar om man svarat fel. Dock var informationen ifråga oftast mycket begränsad. Det fanns inga applikationer som gav feedback i de fall då uppgiften löstes korrekt. Det enda som gavs var en bekräftelse att personen hade utfört uppgiften rätt.[14] Studi-

er pekar på att även om svaret är korrekt kan informativ feedback ge elever möjlighet lära och utvecklas. Detta eftersom eleven då får en förståelse för varför svaret var rätt och kan se möjligheter att förbättra sin problemlösnings-strategi. [6]

Avgränsningar

Då tiden för projektet är begränsad och projektet i princip kan bli hur stort som helst har vissa avgränsningar fått göras. I beskrivningen av projektet nämns att fokus ska vara på hur man utnyttjar feedback i ett lärandesyfte. Det står även att det ska genomföras en pilotstudie för att testa resultatet. Den första avgränsning vi beslutade oss för var att fokusera på uppgifter inom matematik för årskurs två.

Den slutgiltiga prototypen skulle dessutom vara tillräckligt avancerad för att användas som underlag för utvärdering för de uppgifter som gruppen tagit fram och ställa dem i kontrast till de uppgifter som redan finns i originalprodukten. Fokus för utvecklingen bestämdes därför läggas på feedback och inte på teknisk innovation eller design. Under projektets gång har gruppen fört en kontinuerlig dialog där kvantitet har vägts emot kvalitet. En viktig fråga var hur många nya uppgifter av olika sorter vi skulle konstruera. Diskussionen ledde till att två olika typer av uppgifter kom att utvecklas i ett flertal steg. Målet var att erhålla mer träffsäkra resultat från vår studie. Under processen beslutade vi till exempel att lägga vikt vid att minimera rent utseendemässiga skillnader mellan ursprungsuppgifterna och alternativuppgifterna. Användartester med slutanvändare genomfördes inte på Lo-Fi prototyperna som utvecklades under delkurs 1, då det fanns en risk att eleverna skulle ha svårt att se förbi den enkla designen. Vi beslutade i stället att enbart göra användartester på Hi-Fi prototyperna.

3 Frågeställningar

Inför och under projektets gång formulerades ett antal frågeställningar som är centrala för projektet och som vi önskade besvara genom projektet. Några är inspirerade av artikeln ”Cognitive Guidelines for the Design and Evaluation of Early Mathematics Software: The Example of MathemAntics” [3].

- Vad för feedback finns i dagsläget i produkten (Qnoddarnas Värld), och hur hjälper den eleverna?
- Vad för interaktionsmöjligheter finns och hur stödjer dessa inläring?
- Hur väl utnyttjar produkten det digitala mediets möjligheter?
- Vilka förbättringar kan göras i termer av feedback?
- Hur väl uppmuntrar upplägget till korrekta och effektiva problemlösningsstrategier?

Under rubriken diskussion går vi in djupare på dessa frågeställningar samt söker svar.

4 Metod

Processen inleddes med en studie av läromedel för digitalt lärande. Det finns idag ett stort utbud av läromedel för olika typ av läsplattor och vi ville bilda oss en bred uppfattning av vad som fanns på marknaden. Det finns idag 42 olika applikationer med fokus på inläring enbart i App Store vilket gör att valmöjligheterna för vilken utgångsprodukt vi skulle använda oss av var många [14]. Dessa 42 applikationer hittar man om man söker på ”Education”, age = ”6-11”, ”matematik”, ”läsa svenska” och ”skriva svenska”. Tillsammans med handledare valde vi Qnoddarnas Värld.

Studier inleddes för att få en uppfattning om vilken typ av feedback applikationen använde sig av. Vi spelade igenom olika typer av uppgifter för att få en uppfattning om hur de hanterades av läromedlet behandlade olika typer av uppgifter. Både uppgifter inom svenska och matematik gick igenom och därefter beslutade vi oss att gå vidare med matematik. Främsta motivet var att vi ansåg oss kunna bidra med kreativa lösningar på ett bättre sätt inom matematik än svenska eftersom vi själva har en gedigen utbildning inom området matematik.

Vi gjorde även studier av lärarklienten för att få en helhetsbild av Qnoddarnas Värld. Med hjälp av tidigare publicerat material inom pedagogik, kognitionsvetenskap och interaktionsdesign samt diskussion i samband med genomgång av det digitala läromedlet valde vi ut fem uppgifter att ha som utgångspunkt för projektet. Uppgifterna dokumenterades i ett användbarhetslabb där allt filmades. Under inspelningen testade vi hur läromedlet reagerade när både rätt och fel gjordes för att se vilken form av feedback som erbjöds. I ett senare skede av projektet valde vi att fokusera på två av uppgifterna, för att kunna prioritera kvalitet över kvantitet samt hålla tidsramarna.

Två Lo-Fi prototyper togs fram utifrån en tidig brainstorming följt av ett flertal vidareutvecklingssteg på basis av diskussioner med handledare samt tester på andra studenter. Valen av de olika typer av feedback som används i prototyperna är grundade i litteratur om feedback. Många olika designalternativ vad gäller Lo-Fi prototyperna diskuterades, men slutligen kom gruppen gemensamt fram till att prototypernas design skulle utgå från originaluppgifternas motsvarigheter i Qnoddarnas Värld. Skälet var att undvika att resultatet från användarstudien skulle påverkas för mycket av designskillnader mellan Qnoddarnas Värld och prototyperna, då det är skillnaderna i feedback som är i fokus för användarstudien.

Implementering

Merparten av arbetet under den andra delkursen las på att implementera det vi hade kommit fram till under delkurs 1. Då ingen i gruppen hade tidigare erfarenhet att jobba med en grafikmotor börjades processen med att studera olika typer av verktyg för implementeringen. Ett flertal verktyg undersöktes och beslutet föll på att utföra arbetet med hjälp av programmet Unity. Unity är en spelmotor gjord för att göra enklare spel i 2D eller 3D. Nästa beslut gällde val av programspråk. Unity stödjer C# (C sharp), JavaScript och Unitys egna språk med namnet boo. Vi beslutade att använda C# då det är mest likt programspråk gruppen tidigare hade använt sig av. C# skiljer sig även emot de andra då det är ett

programspråk och inte ett script.

Efter dessa beslut om utvecklingsmiljö gjordes en uppdelning i tre olika delar där en person var ansvarig för varje del. Den fjärde personen i gruppen fortsatte med att bygga på teoridelen och att ta fram en testplan för testerna. Trots att olika personer ansvarade för olika delar var kommunikationen omfattande inom gruppen. Om liknande problem skulle lösas i olika uppgifter gjordes bara en lösning. Arbetet gick inte helt smärtfritt och tog betydligt längre tid än vad som förutspåts. Ett viktigt skäl var att ingen i gruppen tidigare gjort något liknande. Under implementeringen fördes omfattande diskussioner inom gruppen om hur olika problem skulle lösas utan att feedbackaspekterna skulle försvagas.

B. Testpersonernas förkunskaper

Materialet var gjort för årskurs två medan eleverna som deltog i användartestet gick i årskurs ett. Eleverna hade tidigare kunskap av digitala läromedel och hade använt iPads förut men ingen hade testat Qnoddarnas Värld. De tidigare digitala läromedel de hade testat var bland annat Math ohoy och Af-fären.

Användartestet

Eleverna (N=20) delades in i två gånger fyra grupper, och hälften av eleverna testades åt gången. Varje testledare fick två eller tre elever åt gången som testade den digitala läromedelsuppgift som testledaren skötte. Efter ca 10 minuter bytte grupperna station och fick testa den uppgift som motsvarades av den tidigare gjorda uppgiften fast av den andre utvecklaren. Planen var att eleverna skulle hinna göra tre uppgifter på varje station, men på grund av skillnader i svårighetsgrad samt elevernas matematiska kunskap blev det skillnader i antalet utförda uppgifter. När eleverna testat båda varianterna av uppgiften genomfördes en semistrukturerad intervju, där vi ställde frågor om svårighetsgrad, i vilken grad de tyckte det var roligt att arbeta med de två varianterna på digitala läromedel, om någon variant var roligare än den andra och hur de stod sig i förhållande till matematikboken. Efter experimentet diskuterade testledarna igenom experimentet och resultatet för att få en så korrekt bild som möjligt av den insamlade datan.

5 Teori

Feedback

En aktör utför en handling och får sedan information om utförandet; denna information kallas feedback. Informationen kan komma ifrån en lärare, kompis, bok med mera. Man kan också beskriva feedback som ett instrument för att minska diskrepansen mellan det man redan kan och det man vill lära sig [6]. Exempel på detta kan vara en elev som får veta om hon/han gjort rätt eller fel på en matematisk uppgift eller en ståuppkomiker som ändrar sitt material när hon/han upptäcker att stora delar av publiken inte skrattar.

För att feedback ska fungera på bästa sätt måste den komma snabbt efter att någonting utförts och vara tydlig för att personen ska kunna koppla den till rätt utförd handling. Feedback bör ges oavsett om eleven gör rätt eller fel. [2] Designen av feedback i digitala läromedel bör ta hänsyn till individuell progression. En del elever behöver mer hjälp och andra mindre. Eleverna som utvecklas i en snabbare takt hin-

dras om det finns för mycket scaffolding (erbjudande av stöd för ny inlärning, exempelvis hjälpanimationer och ledtrådar) och de som utvecklas långsammare behöver mer av detta. Därför bör uppgifter vara designade så att de känner av vilken typ av hjälp eleven behöver och svarar utifrån det och om eleven gör rätt ska den också få veta det. Uppgifterna bör även vara uppbyggda så de går från enklare till svårare men också från ett mer konkret plan till ett abstraktare [2]. Man bör också noga överväga vilken slags feedback som används. Nedan följer en genomgång av olika typer av feedback.

1) Rätt och fel-feedback

Detta handlar om feedback som enbart indikerar rätt eller fel. Många digitala läromedel använder denna typ av feedback [14], bland annat Qnoddarnas Värld. Men rätt-och-fel-feedback är i första hand feedback för en test-situation och inte för en lärandesituation. För en del uppgifter där "enkel fakta" ska memoreras [6], som till exempel att lära sig multiplikationstabellen, är denna typ av feedback motiverad. Men i de flesta situationer är informativ feedback betydligt mer lämplig. [2]

2) Informativ feedback

Detta innebär att feedbacken ändras beroende på vilka signaler från aktören som tas emot av den person eller system som ger feedbacken. Utifrån dessa signaler ger personen/systemet ledtrådar till aktören om hur denne kan utvecklas[6]. Fisch [2] diskuterar vad som bör ske om aktören inte lyckas få rätt svar efter tre eller fyra försök och rekommenderar att svaret då bör avslöjas, men är öppen med att det är situationsbaserat. Allra mest centrala är de tips som ges var gång användaren gör någonting fel. En risk med informativ feedback är att vissa kan utnyttja situationen och tar sig framåt utan att lära sig innehållet. Begreppet kallas för "The Illusion of Learning" och innebär att personen lär sig en strategi för att ta sig igenom uppgiften men lär sig inte det som ska läras[7]. Det finns många problem med informativ feedback i digitala läromedel. Ett är att utvecklarna måste förutspå vilka fel eleverna kommer göra och vilken hjälp de då behöver. Därefter måste dessa implementeras vilket är betydligt mer resurskrävande än att bara visa om svaret blev rätt eller fel. Därtill kommer möjligheten/risken att en del inte använder produkten såsom utvecklaren tänkt sig och avsett. Att använda hjälpfunktionen för att komma framåt men utan att lära sig innehållet, som nämndes ovan, är ett exempel på detta. Andra exempel är om användaren tolkar frågan på ett sätt som utvecklaren inte förutsåg, trycker på någonting icke relevant av misstag eller trycker där i syfte av att upptäcka vad som sker.

3) Implication feedback

Detta innebär att aktören direkt får se konsekvensen av sitt handlande [1]. Säg till exempel att en elev ska lära sig hur olika mått relaterar till varandra och har till uppgift att fylla ett litermått med vatten med hjälp av ett decilitermått. Om eleven väljer att fylla på med för många fyllda decilitermått blir konsekvensen att vattnet rinner över. Det blir konsekvensen i den fysiska världen, och en virtuell motsvarighet kan implementeras digitalt. En tumregel är att konsekvens-

feedback (implication feedback) ska stämma överens med verkligheten i den utsträckning som passar bäst för syftet med uppgiften. Givetvis är det svårt om inte omöjligt att uttala sig med säkerhet om vad som är bäst för inläring, men det går att göra kvalificerade gissningar. Ett exempel på hur en alltför verklig implication feedback i vårt exempel kan ställa till är, om när vattnet hålls över och användaren gör det för snabbt och det skvätter ut vatten. Det leder till att när tio decilitermått hållts i är litermåttet fortfarande inte helt fullt, för en del av vattnet skvätt ut. Det går att koppla implication feedback till Linderoths [7] diskussion om *affordances* i olika situationer. Hans uppfattning grundar sig i den ekologiska psykologins framställning där varje situation innebär olika *affordances* beroende på vem som är i situationen och situationen i sig. *Affordances* skiljer sig både mellan och inom arter. En *affordance* gör att organismen "vet" vad som kan göras i situationen, exempelvis gör fiskens gälar att den kan andas under vatten. Gälarna ger *affordance* till fisken som gör att den kan vara i vattnet. En vuxen människa kan se ett glas med vatten och uppfatta en *affordance* att ta upp glaset samt en *affordance* att dricka ur det, medan en bebis inte kan lyckas med det. En koppling till implication feedback och digitala läromedel är att man vill lära eleven *affordances* som är överförbara till andra situationer. Följer *affordancen* i det digitala läromedlet verkligheten öppnar det upp för lärande genom att eleven då kan använda sig av gammal kunskap för att klara uppgiften men också använda uppgiftens kunskap för ett framtida problem. Men som visat i exemplet tidigare, så ska *affordancena* i första hand vara mappade mot det som ska läras och inte bara mot verkligheten. Det bör tilläggas att detta är möjligt även om det finns invändningar mot ekologiskt psykologi som metodologisk grund.

Synlighet

I designen ska de relevanta delarna vara tydliga.[11] Det digitala mediet har möjligheten använda denna princip till fullo men i många digitala läromedel faller det ändå bort. I gungrädsuppgiften i Qnoddarnas värld, som vi haft som en utgångsuppgift, finns all information på skärmen direkt, istället för bara den nödvändiga. I vår version byggde vi istället upp uppgiften flera steg, där information inte finns med på skärmen innan den ska vara där eller efter den är använd.

Variationsteori

Variationsteori är en modell med en empirisk bas, som försöker förklara lärande. Viktigt i teorin är "the object of learning", lärandeobjektet.[13] Lärandeobjektet kan ses utifrån tre perspektiv, det som ska läras, det som kan läras exempelvis kan en observatör sitta med på en lektion och se vad som kan läras på lektionen, samt det eleven lär sig. [8] Variationsteorin förklarar lärande som aspekter av det som ska läras. Om läraren sedan kan variera mellan dessa aspekter, ökar möjligheterna till lärande. En av de viktigare aspekterna är att ta upp vad det som ska läras *inte* är för någonting. Runesson [12] tar upp exemplet med två lärare och en kvadrat. Den förste läraren beskriver hur många sidor den har, vilka vinkelstorlekar och avslutar med en matematisk definition av begreppet. Den andre läraren använder sig utav

samma aspekter men varierar dessa. Antalet sidor kontrasteras genom jämförelser med andra geometriska figurer såsom en triangel. Vinklarna kan varieras och kontrasteras mot till exempel en romb, samt hur kvadraten förhåller sig till en kub.

Lärare kan variera sin undervisning baserat på vad eleverna förstår. I digitala läromedel måste detta vara implementerat och innebär en avvägning mellan för kraftig scaffolding vilket leder till att en del elever stannas upp av uppgifterna, samt för lite scaffolding vilket kan leda till att en del elever aldrig når fram – och idealt sett ska scaffoldingen varieras automatiskt för att passa de enskilda eleverna. Termen scaffolding kommer ifrån det engelska ordet som betyder byggställning och som i detta sammanhang står för något som kan fungera som en byggnadsställning för lärande. En problematik med scaffolding är att hjälpmedlen som är nödvändiga för vissa, kan utnyttjas av andra på ett ur pedagogiskt perspektiv oönskat sätt. Anta till exempel att svaret på uppgiften ges efter ett antal försök. För en del är detta ett stöd för att lära och komma framåt, medan andra kan utnyttja funktionen för att slippa tänka [2]. Feedbacksystemet eller avsaknaden av ett utvecklat sådant gör att eleverna inte får in variationen av aspekter som skulle behövas för att öka lärandegraden. I vår prototyp finns "gömd" information som eleven kan fråga Professor Qnodd om. Och om användaren trycker på olika ställen ges varierande feedback, som ska fungera som variation av aspekter. En av poängerna med variationsteorin är att den kan kopplas till begreppet *transfer*, som handlar om möjligheter och svårigheter med att använda kunskap som lärts i ett visst sammanhang och överföra den till ett annat.[4] Om eleverna istället lär sig aspekterna, samt hur dessa förhåller sig till varandra ökar man möjligheten att de även ska kunna använda sig av dessa aspekter i andra, nya sammanhang.[9]

Gärdenfors sex kriterier

I boken Lusten att förstå tar Gärdenfors [5] upp sex kriterier för tekniskt lärande. Det första kriteriet är att tekniken ska stödja interaktiviteten, i bemärkelsen att användaren ska ha möjlighet att påverka programmet i varierande utsträckning. Beroende på vad det är för digitalt program finns det olika lösningar på detta och är en av fördelarna framför en bok där det inte finns någon interaktivitet. I ett Worddokument går det exempelvis att söka efter ett ord, i en Tedtalk går det att pausa om man behöver smälta informationen som sagts. Kriterium nr 2 säger att tekniken ska ge återkoppling till eleverna, dvs. feedback till eleven. För det tredje ska tekniken använda narrativa former i presentationen av materialet. Om materialet presenteras i form av en berättelse eller liknande kommer det att göra användaren mer engagerad i läromedlet. I det digitala mediet går det att använda allt ifrån text och tal till bilder och film vilket bör utnyttjas. Detta har inte varit i fokus i vårt arbete då uppgiften inte var att skapa ett nytt läromedel utan fortsätta utveckla ett existerande. Kriterium nr 4 säger att tekniken ska anpassas efter individens lärstilar. Detta baserar sig på att människor har olika sätt som de lär sig bäst på, om det bara funnits böcker hade alla fått lära sig på det sättet. Men i och med tekniken går det att presentera materialet på flera sätt och på så sätt kan användaren välja det som är fördelaktigt för sig själv. För det femte ska

tekniken erbjuda former för samarbete. Här går det att utnyttja det faktum att det med hjälp av datorn går att kommunicera utan att behöva vara på samma plats och det är möjligt att göra digitala läromedel som uppmuntrar till diskussioner och andra former av samarbeten.

Enligt kriterium 6, slutligen ska tekniken stödja metakognition. Till exempel kan man låta eleven få gå tillbaka för att uppleva en lärsituation, något som är särskilt fördelaktigt om läraren är med. Detta används redan inom sportvärlden då tränaren filmar atleten som sedan kan titta på filmen och se vad som gick rätt och fel, vad som kan ändras och förbättras.

6 Resultat

De uppgifter som testades var Hallon 1.2 och Hallon 5.2 från Qnoddarnas Värld samt gruppens prototyper som motsvarade dessa uppgifter. Tiden varierade mycket mellan de fyra uppgifterna som eleverna gjorde och om de hade gjort en likadan uppgift tidigare. Uppgiften Hallon 5.2 tog på första försöket 7 minuter och 12 sekunder för en grupp medan tredje försöket på uppgiften Hallon 1.2 gick på 5.4 sekunder. Generellt tog det längst tid på hallon 5.2 ($M=3.4$ minuter) och snabbast var Hallon 1.2 ($M=17.62$ sekunder). För prototypen finns ej exakt data men prototypuppgiften som motsvarade Hallon 1.2 tog ca 4-5 minuter. Dock fanns några mycket snabba tider (<30 sekunder) som drar ned medelvärdet. På prototypens bana motsvarande Hallon 5.2 låg tiderna mellan 3-4 minuter men även här fanns det några snabba tider (<30 sekunder).

Eleverna rankade de digitala uppgifterna som roligare än matematikboken. Endast tre av tjugo elever sade att de föredrog matematikboken eller tyckte att båda var lika roliga. Av de två programmen föredrog de flesta prototypuppgifterna och ansåg att uppgifterna från Qnoddarnas Värld var för enkla eller för svåra. Suget att fortsätta spela var dock stort, och de flesta ville spela prototypuppgifterna igen. Den uppgift som var minst omtyckt var Hallon 5.2, som ansågs vara tråkig och obegriplig.

Alla elever behövde någon form av hjälp i initialt skede, men för de flesta grupper gick det bra efter första uppgiften, med undantaget Hallon 5.2 där många grupper även behövde hjälp i senare skede. I prototypuppgiften som motsvarade Hallon 1.2 klarade två personer uppgifterna helt utan hjälp.

A. Prototypens motsvarighet till Hallon 1.2

I del två (se Fig 2) av gungbräde-uppgiften var det bara 25 % som inte kopplade ihop klossarna med talstrecken. Hur grupperna använde klossarna respektive siffrorna varierade kraftigt mellan alla grupper. En grupp drog direkt upp rätt antal genom att dra upp rätt siffra medan andra började med klossarna och sedan fortsatte med siffror.



Fig 1. Författarens version av Hallon 1.2

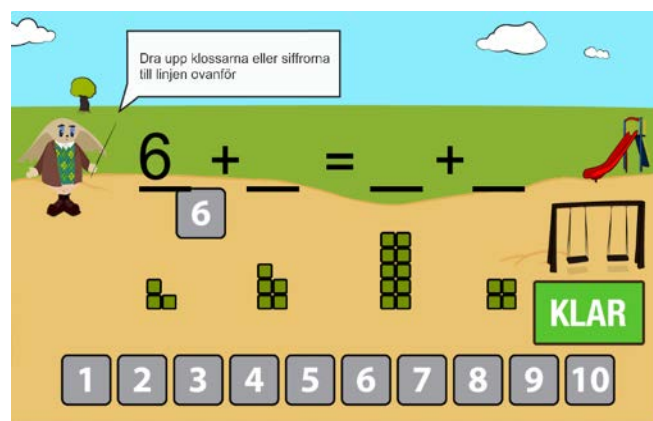


Fig 2. Delsteg 2 i uppgiften från Fig. 1

B. Hallon 1.2

I början försökte majoriteten av eleverna att dra klossar från fel hög (den högra högen på vänster sida av gungbrädan), se Fig 3. Sannolikt var det svårt att förstå vilken hög som menades med "högra tornet". När inget hände då de försökte blev de väldigt osäkra och vågade inte fortsätta. Därför fick testledaren förklara vilken hög med klossar de skulle flytta ifrån. I steg två, se Fig 4, uppkom ett nytt problem. Ca 80 % av eleverna förstod inte vad som menades i uppgiftens förklaring. Det underliggande problemet var att ingen av de elever som inte redan hade sett någon utföra uppgiften innan förstod hur de skulle skriva. När de väl förstod hur de skulle få fram tangentbordet fick ca 40 % av eleverna problem, de kunde inte komma ihåg hur många klossar som var över och eftersom tangentbordet täckte hela gungbrädan kunde de inte räkna om klossarna. Då de inte heller förstod hur de skulle få bort tangentbordet igen fick testledaren hjälpa till även med detta.



Fig 3. Hallon 1.2 steg 1

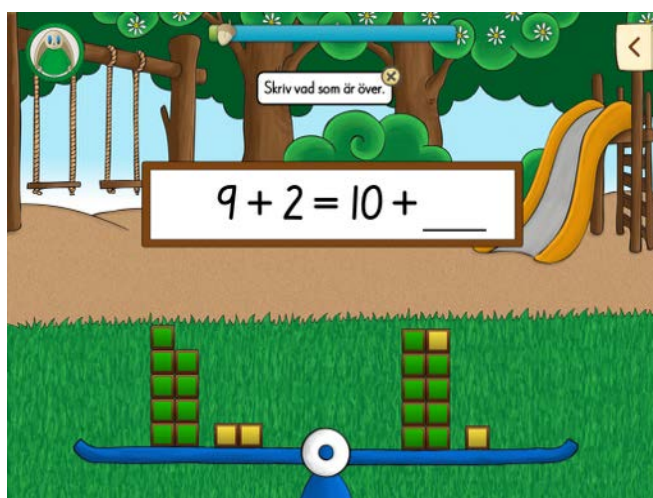


Fig 4. Hallon 1.2 steg 2

C. Hallon 5.2

En majoritet av eleverna hade inte lärt sig vad ett tiotal är vilket ledde till stor förvirring på grund av frågan (se Fig 5). När de väl förstod vad de skulle flytta Qnodden var det många som inte förstod hur den skulle flyttas. Nästa steg förklaras inte i frågan och det var ingen som klarade av att förstå vart man skulle hoppa härnäst. Även när testledaren hjälpte till var det många som inte förstod vart de skulle hoppa. I de fall där de lyckades med detta var det ingen som kopplade hur talet skulle skrivas och två grupper trodde de skulle skriva talet som Qnodden stod på. Totalt fanns 3 individer i 3 olika grupper som löste uppgiften på egen hand men de andra gruppmedlemmarna förstod ej. Huruvida dessa elever förstod matematiken är oklart, det kan även vara så att de lärde sig en strategi för att komma vidare.

D. Prototypens motsvarighet till Hallon 5.2

Ingen av testdeltagarna klarade av uppgiften utan någon som helst initial vägledning men fyra av fallen var av den enkla sorten medan två krävde mycket vägledning. Musen var ett stort problem även i denna uppgift, och en surfplatta hade varit att föredra. Detta påverkade resultatet i tid.

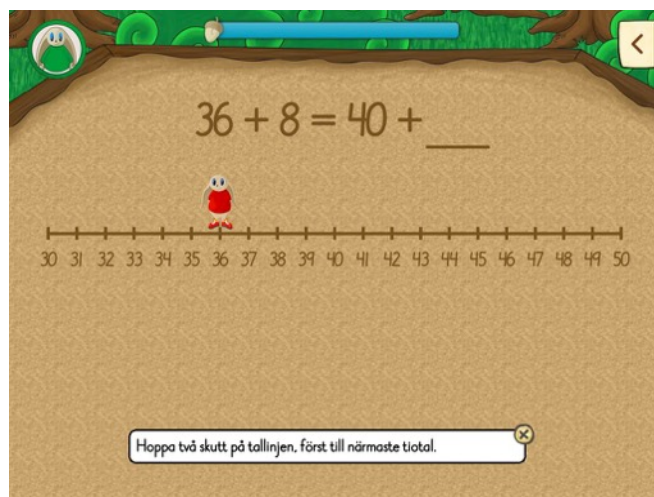


Fig 5. Hallon 5.2

Uppgiften tog allt från 15 sekunder att klara till över 2 minuter (se Fig 5). Trots att uppgiften genererar nya tal mellan 21 och 79 att hoppa till efter varje försök (vilket är stora tal för elever som inte har utmärkt koll på tiotal och talövergångar) verkade testdeltagarna efter hand klara uppgiften på kortare tid. Som uppgiften är utformad erbjuder den ingen strömlinjeformad lösningsmetod som skulle kunna ge en Illusion of learning. Viktigt är att barnet först ska hoppa till tiotalet och sedan till ett nytt tal mellan 1-9 för att bilda svaret. Utfallet – att eleverna klarade uppgifterna på allt kortare tid – kan tolkas som att prototypen ger stöd för lärande.

En av grupperna tyckte att Qnoddarnas Världs uppgift var lättare men att författarnas var roligare. Detta sågs som positiv feedback även om detta högst kvalitativa resultat kan ha färgats av eventuella diskussioner som den ansvariga läraren haft före ankomsten till skolan. En funktion som författarna lagt mycket tid på var implementationen av röster och texter för att förklara uppgiften. Denna funktionalitet verkade dessvärre gå över testdeltagarnas huvud. De hoppade direkt in i interaktionen istället och knappt lyssnade på vad professor Qnodd hade att säga. Kanske behövs mindre heltäckande förklaringar och mer korta och koncisa inledningar. Detta visar på att man bör göra reguljära tester med fokusgrupper efter varje iteration av mjukvaran för att få till en så riklig mängd feedback som möjligt. Det kan bli för mycket, speciellt för en så ung målgrupp.



Fig 5. Författarnas version av Hallon 5.2

VII. Socialiseringsprocess

Projektgruppen kom bra överens från början. Redan innan Mathias Osvaths föreläsning hade vi diskuterat lite om målen och planerna med FoU-kurserna. Efter föreläsningen höll vi ett mer djupgående möte och diskuterade problem som kan uppstå. Vi har också hållit "socialiseringsmöten" för att medlemmarna skulle lära känna varandra bättre. Bland annat diskuterade vi vilken ambitionsnivå vi hade och samtliga var överens om att ett högre betyg i kurserna är att önska. Uppdelning av arbetet i delkurs 1 var ganska fritt och alla jobbade med det mesta och försökte lösa uppgifterna i grupp. Vi höll brainstormingmöten för att tillsammans komma fram till bra Lo-Fi prototyper och förbättringar i metaforerna för inläring av olika matematiska begrepp.

I delkurs 2 delade vi upp arbetet mer för att hinna med implementeringen utav Hi-Fi prototyperna men trots detta har samarbetet fortsatt att vara nära och gruppmedlemmarna har hjälpts åt vid behov

8 Diskussion

Den feedback som i dagsläget finns i Qnoddarnas Värld är som tidigare nämnts framförallt rätt eller fel-feedback. Denna typ av feedback minskar inte gapet mellan vad eleven förstår och dit den behöver nå, något som definitivt är önskvärt, framförallt i en så förståelseorienterad och komplex lära som matematik. Ett ämne som dessutom lätt blir väldigt tråkigt för eleverna om de inte förstår den underliggande logiken och inte har de problemlösningsmetoder som krävs eller bara gissar sig fram. Den tekniska lösningen som tagits fram i denna rapport ämnar lösa detta problem genom en blandning av olika tekniker. Framförallt har uppgifterna delats upp i flera steg för att underlätta inläringen och tekniken, med feedback i varje steg (informativ feedback). En typ av scaffolding har införts, där problemet gradvis blir mer abstrakt och erfarenheterna från tidigare steg hjälper till att bygga upp förståelsen. Det har dessutom införts mer kvantitativ, implication-feedback i de olika stegen av uppgiften. Kan vi dessutom utveckla narrativet och utbytet med Professor Qnodd som en "expert", scaffoldas barnen till att nå sina läromål. Interaktionen i Qnoddarnas Värld är på många sätt onödig, den skapar inga kopplingar mellan abstrakt och fysisk representation och tillför inget till lösningar och förståelse. Faktum är att den ibland till och med är hinder för lösandet av uppgiften. Ett exempel på detta är uppgiften Hallon 1.2, där klossarna på brädan täcks av tangentbordet när man ska svara. Det gäller att man tänker på hur de olika elementen används när man implementerar interaktioner. I vissa uppgifter dyker bara "Klar"-knappen upp när man svarat rätt (och ibland finns det bara en sak att göra). I detta fall finns det risker med implementationen. Utan någon strategi och utan att räkna kan barnet svara rätt eftersom man vet att man är färdig när "Klar"-knappen dykt upp. Detta har flera konsekvenser, men framförallt kan eleven inte misslyckas och läromedlet har då misslyckats med att lära eleven vad som händer om det blir fel. Feedbacken i Qnoddarnas Värld går att förbättra på många sätt, exempelvis med implikationsfeedback som visar vad som händer när något går fel (eller rätt), constraints och affordances som guidar eleverna i sin interaktion och minskar de felaktiga handlingarna.

En risk med en begränsad värld är att felaktiga handlingar inte kan göras, vilket leder till att rätt handling görs men av fel anledning, det vill säga att den inte görs för eleven vet att den är rätt utan för att utveckla den gjort den alltför tydlig. Då uppstår det en illusion of learning.[7] I grundutförandet nyttjas inte dessa typer av hjälpmedel och signaler fullt ut. Den mest intressanta delen i det digitala mediets möjligheter är att kunna förbättra elevers inläring genom gradvis ökning av komplexitet och abstrakthet kopplat med tydlig, kortfattad och skraddarsydd "intelligent" feedback. Man kan skriva applikationer som känner av de fel som eleven gör och som kommer med adekvata råd och tips på förbättring bättre än en bok kan.

Upplägget i Qnoddarnas Värld handlar inte om relevanta problemlösningsstrategier för matematik, uppgifterna är repetitiva och erbjuder inget nytt sätt att lära ut matematik. Detta är något som hade kunnat lösas genom ovan nämnda förbättringar. Ett konkret exempel är gungbrädans originalutförande, Fig. 3, där momenten för att klara uppgiften är att flytta en kloss från ena högen till en annan och sedan skriva uttrycket. Första momentet skall repeteras i varje ny instans av uppgiften och erbjuder efter första gången inget mer än en repetitiv, separerad handling, fri från någon som helst inlärningspotential. I författarnas exempel förstärks istället likheten mellan två uttryck genom att gungbrädan är manipulativ och förhoppningsvis en metafor som eleverna kan förstå, se Fig. 1. När gungbrädan är jämn får eleverna möjligheten att gå vidare, varpå en kontrolluppgift ställs. Detta får eleven att tänka efter; gjorde jag rätt nu? och ger en möjlighet att korrigera eventuella fel. I senare steg kopplas klossarna till aritmetiken. I detta steg kan man räkna på flera olika sätt, ett sätt att tillåta elever att lösa uppgiften trots deras eventuella olika förkunskapsnivåer, se Fig 2. Designfaktorer som relaterar till synlighet i Qnoddarnas Värld såsom placering av relevant information och gruppering av element, se Kap. 5.1.4., förekommer sparsamt. Eftersom vår kulturella standard är att läsa uppifrån höger och till vänster verkar det märkligt att ha frågan längst nere i vyn, se Fig. 5, och vår lösning är istället att alltid ha frågan längst uppe eftersom det är det första man ska läsa.

Narrativ verkar saknas i många delar i Qnoddarnas Värld. Det är t.ex. inte osannolikt att barnen skulle vilja veta vilka Qnoddarna är, vad de gör där eller varför man ska hjälpa dem. Det är visat att narrativ kan stärka barnens intresse och inlärningsförmåga [3], så det vore naturligt att utveckla ett historieberättande för att fånga barnens uppmärksamhet. En förbättring som hade kunnat göras är att införa en elevqnodd som man kan klä upp istället för att inreda sin stubbe, vilket är en betydelselös funktion som idag finns i Qnoddarnas Värld. Denna avatar hade man kunnat hjälpa igenom Qnoddlandet genom att hjälpa honom lösa de problem som varje stig innehåller. Norman diskuterar att produkter med ökat antal funktioner också ofta får en ökad komplexitet. Det är möjligt att detta även sker med digitala läromedel. Om det funnits allt för stora interaktionsmöjligheter hade det troligen lett till mer förvirring än inläring. I författarnas version av Hallon 1.2 lades det till kedjor för att visa att interaktionsmöjligheterna är begränsade. Tyvärr uppfattade barnen inte detta och en del försökte lägga klossar på fel sida. Den manipulativa egenskapen inför också vissa begränsningar och även om implementationen blivit perfekt, hade eleven i så fall kunnat fylla på så högarna blivit likadana på båda sidor

istället för bara samma totala antal. Detta är samma problem som uppstår om det digitala läromedlet innehåller för få eller för många affordances eller constraints. Då riskerar uppgiften att bli blir en slinga som eleven går längs och då lär barnet sig den slingan men inte substansen och upplever en så kallade Illusion of Learning. Även om det finns goda läromöjligheter längs slingan utnyttjas inte mediet till fullo, då en av möjligheterna är att man kan testa flera olika sätt att lösa uppgiften på.

En av poängerna med feedback är att den ska vara tydlig och lätt att förstå och koppla. Därför går det inte bara att lägga till ett feedbacksystem baserat på teorier, utan de måste korrigeras och anpassas till det digitala läromedlet och även ner på uppgiftsnivå vid behov. Om en del uppgifter har som syfte att eleven ska memorera fakta, är den bästa typen rätt-och-fel feedback medan informativ feedback blivit hindrande. Med digitala läromedel är det också lättare att göra korrigerande åtgärder i senare stadier, alltså att patcha sin produkt. Är däremot läromedlet i tryckt format måste en ny bok tryckas. Alltså går det att utvärdera och implementera förbättring i större utsträckning. Ur ett variationsteoretiskt perspektiv måste det finnas flera kontrollfrågor, så eleven faktiskt förstått området som ska läras och dess aspekter. Det hade även varit bra med ytterligare frågor utanför det scenariot som lärts för att se om det eleven lärt sig fått en transfer-effekt. Gärdenfors sex kriterier är utarbetade efter ett komplett digitalt läromedel medan prototypen varit en del av ett fullständigt läromedel. Fokus för projektet har varit feedback för 2000-talet och därför har kriteriet att ge återkoppling till användaren varit i fokus. I den största möjliga utsträckning har även det första kriteriet uppfyllts även om det finns fler förbättringsmöjligheter. De sista fyra kriterierna har det inte funnits tid eller möjlighet att använda sig av, med undantaget att författarna skapade professor Qnodd och elevqnodden som kan rymmas inom narrativkriteriet.

A. Analys

Mycket av den insamlade data är kvalitativ vilket gör det svårt att dra säkra slutsatser mellan hur bra prototypen är i jämförelse med Qnoddarnas värld. Dessutom var eleverna yngre än målgruppen, på grund av att de inte skulle ha några tidigare kunskaper som eventuellt kunnat skapa brus i undersökningen. Vi önskade testa applikationen på fokusgrupper för att i lugn och ro kunna samla in data, men på grund av tidspress och kravet på rättvis fördelning (hela klassen skulle testa) var gruppen underbemannad och klarade inte hantera insamlingen av data på ett önskvärt sätt. Detta gjorde det svårt att urskilja hur eleven skulle löst uppgiften om eleven varit ensam. Hallon 1.2 bedömdes som väldigt enkel av många grupper och det var ingen som hade några större problem med den, medan alla hade problem med Hallon 5.2 trots att båda var riktade till elever i årskurs två. Båda uppgifterna i prototypen verkade mer jämna i svårighetsgrad. Eleverna som fick testa på prototypen som motsvarade Hallon 5.2 hade inte lika stora problem vilket tyder på att den har egenskaper som är fördelaktig i inlärnings syften. Några elever som tyckte matematikboken var roligare än något av de digitala läromedlen. Detta kan bero att eleverna är vana vid digitala läromedel och ser det som en vardagssak till skillnad från när författarna gick i skolan. Testledarna som skötte prototypen rapporterade att en del elever hade motoriska problem med att använda datormusen vilket kan ha

speglat deras omdöme om den. I Hallon 5.2 märktes det tydligt att barnen inte hade tillräckligt med kunskap då flera rapporterade att de inte visste vad tiotal var för något (första delen av uppgiften är att hoppa till närmsta tiotal). Andra delen av uppgiften när uttrycket skulle beräknas var första uttrycket $36+8=40+?$ och eleverna hade inte hanterat tal över tjugo och många sa att de bara lärt sig räkna till tio. Ett problem som framkom var att eleverna frågade testledarna om det var rätt eller fel istället för att testa det och det var ytterst sällan som de tryckte mer på iPaden än det som behövdes. Alla barnen var också i behov av en muntlig instruktion i början och de var mer benägna att komma till testledaren snarare än läromedlet för information. Detta gör att det är svårt att kolla hur vårt feedbacksystem är jämfört med Qnoddarnas Värld. För att komma runt detta skulle eleverna fått lära sig läromedlen men sedan fått använda dem själva med nya uppgifter medan testledarna observerat om dem lyckas med hjälp av den ökade feedbacken. Ett annat problem var att många elever var tystlåtna och inte så nyfikna på att testa vad som sker om de trycker på olika ställen. Därför fick vi inte så mycket underlag för att se om deras strategier ändrades eller förbättrades.

Under testningen varierade resultatet en del. Den största orsaken till det var att förkunskapskraven varierade väldigt mycket mellan de olika testpersonerna. Mycket av fokus hade lagts på att göra prototypuppgifterna lätta att förstå. Både tal och flera instruktioner för att förklara uppgiften hade lagts till. Barnen hade dock väldigt svårt att ta sig till den informationen. Koncentrationen var en bidragande faktor till det. Det ledde till att alla uppgifter var tvungna att förklaras ordentligt första gången. Vi noterade att testpersonerna hade svårt att förstå vad de skulle göra. Exempelvis var det inte självklart för dem att dra klossarna och placera dem på gungbrädan trots att det tydligt stod på skärmen att de skulle göra det. De kunde helt enkelt inte knyta informationen som gavs till att lösa uppgiften. De testade inte heller vad de kunde göra utan var bara frågande tills de fick mer information.

B. Prototypens motsvarighet till Hallon 1.2

När de väl förstod prototypens uppgift var det en övervägande majoritet, nämligen tre grupper (75 % av grupperna) som ville lägga en kloss på den högen som hade nio klossar på sig. Den klossgruppen hade ett hänglås och kedjor runt sig men barnen verkade inte förstå budskapet med den metaforen. Det intressanta var att de fortsatte att försöka fast ingenting hände. Något som noterades under uppgiften var också att "jämvikt" var svårt att förstå. Alla testpersoner klarade på tredje försöket att lösa uppgiften helt på egen hand. Verifikationssteget i uppgiften måste förstärkas då de förstod hur uppgiften fungerande kunde ta sig vidare utan att förstå vad de hade gjort. Detta testades muntligt på testet genom att be de räkna ihop hur många klossar som låg på varje sida. Deluppgift två av gungbrädan där de skulle placera klossarna på respektive talstreck och bilda tal fungerade bra. Det var en av fyra grupper som inte kopplade ihop klossarna med talstrecken se Fig 1. Det kan nämnas att det även var den gruppen som tog längst tid på sig. Hur grupperna använde klossarna respektive siffrorna varierade kraftigt mellan alla grupper. En grupp drog direkt upp rätt antal genom att dra upp rätt siffra medan andra började med klossarna och sedan fort-

satte med siffror. En slutsats vi drar är att de lade ihop så stora tal de kunde räkna i huvudet. Det ska även noteras att de räknade först innan de började dra upp klossarna, vilket betyder att metoden "trial and error" inte används i det här steget.

Under framtagning av uppgiften pratade vi även om att ta fram en demo på hur man löser uppgiften. Testerna gav tydliga svar på att de skulle vara en bra lösning då de inte förstod första gången hur uppgiften skulle lösas. Även starkare verifikationssteg skulle behövas så de skulle tvingas att verkligen förstå vad som händer.

C. Prototypens motsvarighet till Hallon 5.2

Slumpfunktionen som finns i prototypen används för att i en slutprodukt skulle det vara mer pedagogiskt med en uppgift som slumpas fram då det finns fler uppgifter samt att eleven inte bara lär sig vad dem uppgifterna blir utan hur man kommer fram till svaret. I ett digitalt medium finns denna möjlighet men i likhet med en bok har Qnoddarnas värld valt att bara använda sig utav nio uppgifter.

D. Framtida undersökningar

I denna undersökning var urvalet litet och homogent och för att kunna generalisera skulle det behövs fler testpersoner men också elever ifrån fler skolor. I studien fick eleverna bara testa tre uppgifter ifrån de olika digitala läromedlen men det skulle behövs att de hade fått jobba med läromedlen över en längre period så de förstått produkten, till exempel att det går att klicka på professor Qnodd för mer information. Efter en sådan lite längre period hade man kunnat gett eleverna nya uppgifter och undersöka hur pass väl och hur pass självständigt de hade kunnat lösa dessa.

Eleverna testades i par eller tre stycken tillsammans vilket påverkar mycket. Det sista som ska has i åtanke är att det är en Hi-Fi prototyp och inget fullt utvecklat läromedel vilket gör att det kan finnas buggar, samt att alla idéer inte fullt kunnat implementeras. Exempelvis hann inte verifieringskraven implementeras fullt ut. Det fanns tyvärr ingen tid att implementera dataloggar för att logga all data.

Referenser

- [1] Blair, K. P. (2013, June). Learning in Critter Corral: Evaluating three kinds of feedback in a preschool math app. In Proceedings of the 12th International Conference on Interaction Design and Children (pp. 372-375). ACM.
- [2] Fisch, S. (2005). Making Educational Computer Games "Educational". MediaKidz Research & Consulting
- [3] Ginsburg, H., Jamalain, A. och Creighan S. (2013). Cognitive Guidelines for the Design and Evaluation of Early Mathematics Software: The Example of MathemAntics
- [4] Gulz, A. (in press) kognition och lärande.
- [5] Gärdenfors, P. (2010). Lusten att förstå *Om lärande på människans villkor*. Natur och Kultur.
- [6] Hattie, J., & Timperley, H. (2007) The power of feedback. Review of educational research, 77(1), 81-112.
- [7] Linderöth, J. (2012). Why gamers don't learn more: An ecological approach to games as learning environments. Journal of Gaming & Virtual Worlds, 4(1), 45-62.
- [8] Marton, F., Runesson, U. och Tsui, A. (2004). The space of learning. Marton, F och Tsui, A. (red.), Classroom Dis-
- course and the Space of Learning. (pp. 3-40). Mahwah, New Jersey: Lawrence Erlbaum.
- [9] Moxley, S. (1979). Schema: The variability of practice hypothesis. Journal of Motion Behavior, 2, 65-70.
- [10] Natur och Kultur. Qnoddarnas värld – för alla som går i första, andra och tredje klass. (Hämtad: 2014-10-11) <http://www.nok.se/Laromedel/F-9/Grundskola-1-3/Digitala-laromedel/Qnoddarnas-varld/>
- [11] Norman, D. (2002) The Design of Everyday Things. New York: Basic.
- [12] Runesson, U. (1999). Variationens pedagogik: skilda sätt att behandla ett matematiskt innehåll. Göteborg: Acta Universitatis Gothoburgensis.
- [13] Runesson, U. (2005). Beyond discourse and interaction. Variation: A critical aspect for teaching and learning mathematics. Cambridge Journal of Education, 35(1), 69-87.
- [14] Sjöden, B. och Tärning, B. (2014). Appsolute Padagogy - Reviewing digital functionality of Swedish iPad apps for early math and literacy skills. Poster presentation.

APPENDIX 1

Produkt

Detta dokument beskriver upplägget av en användbarhetstestutvärdering där testpersonerna både får testa Qnoddarnas värld och vår version av Qnoddarnas värld där feedbacken har förändrats i förhållande till vad kognitiva modeller säger är att föredra.

Syfte

Syftet med studien är att se hur barn tar till sig olika typer av feedback vid inläring. Fokus kommer att ligga på att se hur man kan utnyttja möjligheterna i ett digitalt media. Resultatet som fås ut av studien kommer senare att sammanställas i en rapport där resultatet även kommer att diskuteras och kopplas till teori.

Frågeställningar

1. Hur mycket hjälper feedbacken i Qnoddarnas värld eleverna att lära sig innehållet?
2. Hur väl uppmuntrar upplägget till korrekta och effektiva problemlösningsstrategier?
3. Vad för interaktionsmöjligheter finns och hur stödjer den inläring?
4. Vad blir skillnaden i inlärningsnivå mellan eleverna som provar Qnoddarnas värld och den versionen med utbyggt feedbacksystem.
5. Finns det skillnad i inlärningsförmåga kopplat till interaktionsmöjligheterna

Testuppgifter

1. Usefulness
 1. Hur lätt är det att använda författarnas version av Qnoddarnas värld och är det någon skillnad mellan produkterna. (Hur många frågor behöver eleverna ställa relaterade till att bara använda produkten.) Hur användbar är produkten?
2. Effectiveness
 1. Hur bra är läromedlet på att lära ut materialet till elever.
3. Learnability
 1. Hur lång inlärningskurva har produkten.
4. Satisfaction
 1. Vad tyckte eleverna om produkterna.
5. Accessibility
 1. Finns det någonting som hindrar en grupp elever från att använda läromedlet.

Test

Hälften av eleverna testas en av två uppgifter på Qnoddarnas värld på en iPad och sedan motsvarande uppgift av vår prototyp på datorer. Resten av eleverna testas den andra uppgiften i de två olika utförandena. Varannan testperson börjar med Qnoddarnas värld och varannan börjar med prototypen, detta gäller båda uppgifterna.

Upplägg för genomförande

Användbarhetsstudien kommer att rikta in sig på barn i årskurs 1. Studien kommer att genomföras i en skola som är en trygg miljö för eleverna för att inte påverka resultatet allt för mycket. Eleverna kommer att delas upp i två grupper där den ena gruppen får göra uppgiften Hallon 1.2 i Qnoddarnas värld och i prototypen. Den andra gruppen kommer att få göra uppgiften Hallon 5.2 på motsvarande sätt.

Kommunikation med testledarna ska försöka att vara så begränsad som möjligt då försökspersonerna ska i så stor utsträckning lösa problem själva. Dock med tanke på åldern hos försökspersonerna kommer en del kommunikation att ske, till exempel om eleven inte kan läsa.

Efter testsessionerna kommer det ske en semistrukturerad intervju för att få reda på vad barnen tyckte. Det är även ett bra tillfälle att samla in kvalitativ subjektiv data. Då försökspersonerna är unga och tänker på ett annat sätt är det ett utmärkt tillfälle att få feedback på vår prototyp.

Översikt och tidsåtgång för sessioner

20 stycken elever kommer att delta i testet. Dessa elever kommer att delas upp i två olika testsessioner. Varje session kommer att vara ca 30 minuter långa. Ca 5 minuter kommer att användas till en efter-test intervju där försökspersonerna ska svara på några frågor.

Introduktion (1 min)

En snabb introduktion av oss samt utav testet.

Test av första digitala läromedel (10 min)

Försökspersonerna utför uppgifter ifrån ett av de två digitala läromedel. En fjärdedel testar gungbrädan i Qnoddarnas värld, en fjärdedel testar gungbrädan i prototypen, en fjärdedel testar tallinjen i Qnoddarnas värld och övriga testar tallinjen i prototypen.

Test av andra digitala läromedel (10 min)

Försökspersonerna utför uppgiften motsvarande den de precis testade. Alltså de som började med Qnoddarnas värld får nu testa prototypen med motsvarande uppgift.

Efter-test intervju (5 min)

Kort intervju med testpersonerna i grupper om 2-3 personer. Detta ger kvalitativ data.

Dat typer samlade med de olika testuppgifterna

1) *Objektiv/Kvantitativ*

- Antalet uppgifter som kan genomföras utan hjälp samt med hjälp, tre nivåer av hjälp analyseras:

- Ingen hjälp - Försökspersonen genomför en uppgift utan hjälp
- Vägledning - Försöksperson genomför en uppgift med enkel vägledning från testledaren
- Mycket hjälp - Försökspersonen genomför uppgift med mycket hjälp

- Tiden det tar att utföra uppgifterna.

- Tiden det tar att utföra provet.

2) *Subjektiv/Kvantitativ*

- Hur väl eleverna gillade de olika digitala läromedlen (motivation att använda igen)
- Hur lätt dem tyckte det var att använda det (hur intuitivt det var)
- Hur hjälpsamt det var

Urval av försökspersoner

Uppgifterna som har tagits fram inom kursen är gjorda för årskurs två elever. Men diskussioner med våra handledare har vi kommit fram till att vi vill testa uppgifterna på personer som inte har några förkunskaper om denna typ av uppgifter. Det gör att testen kommer att ske på elever i årskurs ett. Kontakter har skett med lärare på Lerbäckskolan i Lund och därför kommer försökspersonerna komma därifrån. Vilket gör att gruppen som testerna kommer att utföras på är väldigt homogen. För att kunna dra vetenskapliga slutsatser skulle en större urvalsgrupp behövas och det geografiska området vara större.

Testmiljö/utrustning

Platsen som valts för studien är Lerbäckskolan i Lund. Valet är gjort för att försökspersonerna ska vara i en så trygg miljö som möjligt. Testerna kommer att dokumenteras med hjälp av anteckningar och tidtagarur. Det räcker för att få ut den information som behövs. Vilket leder till att kameror inte kommer att användas, då det inte skulle kännas naturligt för försökspersonerna.

Testledaren och övrigas rollfördelning

Testledaren kommer att gå igenom instruktionerna och finnas tillgänglig för att svara på frågor som försökspersonerna har. Då varje testledare ansvarar för varsin uppgift kommer denna även att observera och anteckna resultatet i testerna. Testledaren kommer att leda den öppna intervjun efteråt.

Hur resultaten ska rapporteras

Rapporten kommer ha en inledande del där syfte, bakgrund och mål med testen kommer att beskrivas. Det är för att läsaren på ett enkelt sätt ska komma in i samma tankebana inom problem och frågeställningar som författarna.

Frågor efter-test intervju

Vilket av spelen tyckte ni var roligast?

Vilket hade ni velat testa igen?

Vilket av spelen var lättast att använda?

Tyckte ni någon av uppgifterna var svår att förstå?

Vad hade gjort dem bättre?

Hur tycker ni detta var mot att jobba med matteboken?

1st Grade Students Learn Number Sense with Math in Space

Per Grahn, Charlotte Leifland, Kristoffer Leo

The purpose of this paper is to present a pedagogical mathematical game concept, accompanied with arguments for the design decisions taken. The overall goal is that the game shall eventually be used in school environments for learning and teaching purposes. The game embraces a space theme and targets 1st grade students. As an extension of the Magical Garden play-&-learn game project, the present game, as well, is based on inclusive pedagogy with the intention to support each individual student in developing his or her mathematical thinking, and to support teachers at their work as well. The focus of our study has been to create a novel concept and find out whether it motivates the target group: girls and boys alike of an approximate age of 7 years, and whether it is suitable for developing the number sense for children of this age. In the first phase a lo-fi prototype was created and evaluated by two math teachers. The feedback from the teachers was of considerable help for us to achieve our goals. In the second phase the game was implemented and is ready for testing and evaluation.

1 Introduction

Learning can be seen as an evolving form of membership while participating in a community of practice. The participation is tied to motivation (Lave and Wenger, 1991). With help of collaboration, the learner can learn to accomplish things that are in her zone of proximal development and reach an understanding which otherwise would be out of her reach (Vygotski, 1978). In a collaborative setting the *teacher* may provide hints or cues to help the student reach an even higher level of thinking (Lave, Wenger, 1991). The teacher role can also be offered to a student. This is systematically done within the pedagogy *learning by teaching*. In the game concept that will be presented in this paper – *Math in Space* – hints and cues are provided from the game as such, and the students are acting as teachers to teach digital tutees in the form of game characters, in technical terms *Teachable Agents, TAs*. The first grade students can be said to participate in the game together with their TAs and with them share the social membership in the game world community – something that hopefully will increase their motivation of learning mathematics.

Magical Garden is a game that has been developed with the idea of the zone of proximal development in mind. Both the notion of scaffolding and the notion of inclusion guide the design of the game in order to provide the learner with support to be able to reach a higher mental development than she would on her own. In addition, children get a chance to practice learning by teaching by helping their chosen TA to learn the mathematical challenge. In this setting they are also

allowed to develop their skills phase by phase (Gulz, 2014). Malmberg (2014) emphasizes the importance of motivating all children to learn math with the help of educational games. However, motivational structures (Haake et al., 2014; Robertson et al., 2004) and distractibility (Axelsson et al., 2013) seem to differ between girls and boys in younger ages. To keep the motivation up an educational method called *adaptive learning* (Malmberg, 2014) is implemented. Adaptive learning ensures that each learner gets the right scaffolding depending on her individual development level and learning curve.

As an extended concept of Magical Garden, the play-&-learn game concept “*Math in Space*” will be presented in this paper. It has been designed to motivate first graders, boys and girls alike, into learning mathematics with the help of Teachable Agents. A new concept, novel TAs and novel graphics, compared to Magical Garden, have been created to meet the needs of the target group, since adjustments of the mathematical level and the game concept are needed when aiming to motivate older children than pre-school age children (Anderberg et al., 2013). In addition our goal was to cater for gender related differences with respect to motivation, such as the motivational effect of interacting with a TA (Haake et al., 2014; Robertson et al., 2004).

2 Background

This paper is based on the Magical Garden project. Magical Garden is a software solution developed to be used by teachers and children in preschool. One focus is to identify children who are weak in early math, specifically concerning number sense development, while at the same time aiming at improving *all* childrens’ number sense skills – without pointing out the weak (Gulz, 2014). Educational Technology Group at the Universities of Lund and Linköping in Sweden has developed the teachable-agent based game with its four modes, recurring for each novel kind of task: Mode 1) The child plays on her own and receives feedback depending on her performance. This is the mode where the child first gets to work on the tasks on her own in order to be able to take on the teaching tasks that follows. Mode 2) The child is offered to take the role of a teacher, and show a digital creature, their teachable agent, how to solve the task (which implicitly means to show how the mathematical concept works).. Mode 3) the TA takes more initiative and tries to solve the assignments more independently, while the child still gets to guide and tell whether the TA makes the right choices or not in solving the tasks. Mode 4) the TA solves tasks on her or his own and is as successful as the teaching provided by the child has been (Gulz & Haake, 2014). One benefit with having these four modes is that the high degree of repetition that

is often needed for early math skills get naturally motivated and infused with some variability too: do oneself, show to TA, guide TA, look at TA solving on her/his own. Magical Garden contains a set of sub-games, at present four, all created by master students, and 60 so called pedagogical scenarios, that target different early math concepts and skills, where most of the scenarios can be run in all four sub-games to obtain variation. The Magical Garden is a long-term, ongoing, R&D undertaking. Our project seeks to extend Magical Garden with a game targeting somewhat older children.

The use of TAs has many benefits. When allowed to be teachers children feel responsibility and seem more motivated to work with the material. At the same time they get to externalize their thoughts. They may discover something that completes and reflects their teaching. The role as a teacher is generally experienced as motivating and in contrast to the regular classroom also the weaker students or students with a weak belief in their own ability are able to – and invited to – take a teacher role in an educational game like Magical Garden, and thereby gain more confidence (Gulz & Haake, 2014). In addition, while using digital educational games of this kind, weak students are not exposed to the same extent as in the classroom. The concept of learning by teaching encourages the student to explain mathematical problems for the TA. In order to explain something you need to understand it yourself, however in many cases it is not until you get to the process of explaining to someone that you really develop your own understanding. There are many positive learning results from contexts where students are tutors, and so seems excellent to structure a game for pedagogical aims according to the idea of learning by teaching (Murphy Paul, 2011)

To sum up, two of the leading design notions of Magical Garden are i) *provide adequate scaffolding in relation to the childrens' zones of proximal development* and ii) *cater for an inclusive pedagogy*. We will present these concepts below and get back to them when we elaborate on the *Math in Space* concept.

Vygotskys theory about *zones of proximal development* holds that if the learner gets a bit of additional support – *scaffolding* – she will be able to reach a new level of knowledge. *Scaffolding* refers to the help provided to the students in order to support them in acquiring more and more advanced skills (Sawyer, 2006). Support and help is primarily needed for development of a skill, and later the supporting ingredients can be removed as the learners understanding grows. Some features of scaffolding are predictability, playfulness, focus on meaning, role reversal and modeling (Lave, Wenger, 1991). Playfulness, predictability and the role reversal in relation to Teachable Agents (TA) can support the learner to reach - little by little - a higher understanding of mathematics and number sense in a way that is based on their individual level. The level can be adjusted according to the abilities of the player. Teachable Agents are virtual characters that invite the learners to interact. The TAs role is to force the student to really understand the problem since it needs to be guided through the mathematical problems. While taking the role of a teacher the learners process the information and learn themselves.

Inclusive pedagogy according to Florian & Black-Hawkins (2011) targets both inclusion and individual development. In the average classes students have different skills

and abilities. This is very common in the age of 6-8 where each individual child thinks and acts differently (Nielsen, 2010). In an inclusive pedagogy the goal is to create an environment where all students feel they can achieve something, and where students with different abilities are not separated from each other (Malmberg, 2014). One of the teacher's most important assignments is to make sure that all students achieve the learning goals for their age. This is a hard challenge, while a teacher that has many students in a class can have a variety of students with very differing levels of mathematical understanding and mental development age. This could limited the learning process and students with less advanced skills might get invisible or the opposite, get a lot of attention pointing out their weaknesses in front of the other students.

Children who start school without appropriate skills in mathematics will often fall behind in school, they rarely catch up. It is therefore important to find these children at an early stage. With the help of pedagogy games like Magical Garden the goal is to possibly identify potential learning problems. In this sense, Magical Garden can be considered a pedagogical instrument (Gulz & Haake, 2014). A digital play-and-learn –game can have many benefits. It can support the teacher with material to follow-up the development of the children's skills in many ways. In Magical Garden and in our extension each student will be registered and her results is stored. The statistics is supposed to send alerts to the teacher to point out the weak students. In this way the teacher can get an overview of how the students understand and perform with number sense. This should help the teacher to give special support to the students who need it. If a child has difficulties in learning in a specific area, a teacher can identify this challenge at an early phase and provide additional support to the child during the learning process. Such formative evaluation of the development allows interfering at an early phase without pointing out the weaknesses or strengths. Inclusion and scaffolding are also some of the benefits that the games like Magical Garden offer. The children get a chance to take new roles and get more space within the existing cultural or gender frames (Gulz & Haake, 2014).

In order to test roles in a game, a child must be motivated to play the game. From the gender inclusion point of view we considered that it is very important that a game reaches and supports the development of girls and boys equally. From the point of view of including and motivating boys and girls in the play-and-learn activity, the design group started exploration of the design space in order to create the concept *Math in Space*. The aim of the space theme is to show both genders that everything is possible and nothing is specific developed just for one gender.

The principles of inclusion, scaffolding, play and learn that are a part of Magical Garden game, were included in the *Math in Space* game concept to provide and support the underlying theory for creating a great pedagogical tool.

3 Presentation of Concept

This section briefly describes the final results of the *Math in Space* standalone platform, which is an extension of Magical Garden. The concept is made with the presumption that the students have played the original Magical Garden in

preschool and when beginning first-grade they switch to the *Math in Space* platform. This platform is designed to lead to closure from the previous game and keeping a sense of familiarity with the new platform and game environment.

3.1 Moonbase

The moonbase is the home state of *Math in Space* and is built on the same foundation as the garden state in the original *Magical Garden*. Instead of watering and growing the garden the student will use stardust collected in subgames to replace parts and adorn their spaceship. This spaceship is also used to transport the student and the TA to explore the subgames. This gives the students a relevant link between exploring the subgames and improving their spaceship.

When replacing parts the student is given the option of *where* to replace a part, but not with *what* to replace it. This is done using the same line of thinking as in the original *Magical Garden*; it makes it harder to compare your abilities with your fellow students. *Math in Space* also allows a student to replace a part an infinite number of times. This is done because if a student is unhappy with a piece they might lose motivation to continue playing and gathering stardust unless they are able to replace it.

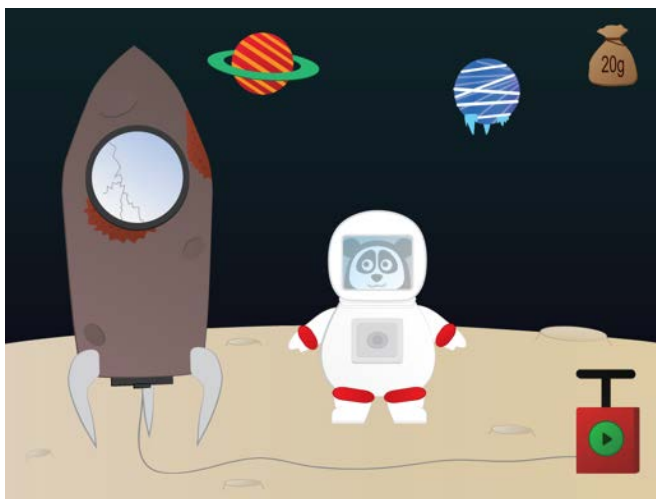


Figure 1: Graphics for moon state

3.2 Teachable Agents

All of the TAs from the original game will be available in *Math in Space* but in a new form more suitable for the new theme. This will tie the game together with the original and give a sort of familiar constant which we believe will be helpful for the students, especially those with lower self efficacy. It is also good for students who formed a good relationship with their old TA.

There will also be new TAs (an alien and a robot to fit with the theme) available for students who might be bored of the original ones or for some other reason did not like them.

3.3 Friends of ten - “At the bus stop in the world of popplers”

The one subgame implemented during this project is based around the math concept “friends of ten”. The subgame takes place in the world of popplers where ten popplers are waiting

at a bus stop on a hill. (A *poppler* is a small alien character that we have designed. They live in harmony with their planet, using environmentally friendly fuel and living in cozy caves underground. They are also very fond of jellybeans). Above the bus stop there is a busy space-road. Suddenly a moving truck jerkily flies by and accidentally drops two empty boxes over the now very confused popplers. X popplers get caught under one of the boxes and Y under the other. The popplers run chaotically under the boxes down the hill. Then a gust of winds grabs one of the boxes and it blows off, revealing Y popplers. The student is now prompted to determine how many popplers are under the other box. Input should be made using a scroll wheel, and when the remaining popplers have been correctly determined all the popplers happily return to the bus stop. Just in time a bus comes and picks them up thus ending a round of play.

3.3.1 Feedback in “At the bus stop in the world of popplers”

Feedback is given both visually and vocally. It is especially important to give feedback when the student is wrong. The amount and kind of feedback is designed in order to let the student benefit and learn from a mistake. When a student proposes a too small number, this number of popplers will come out of the box and join the other popplers. A voice will tell the student that more popplers remain under the box and the student is given a new problem based on the error from the first guess. This gives the student visual feedback on the number they guessed and an opportunity to learn from the new problem. When the student answer correctly we hope our feedback solution will show the student how much she was wrong before.

When the student answers with a number that is too big all the popplers will come out of the box. A voice will tell the students that there are no more popplers under the box. The popplers that were under the box will get curious and will go under the box to check, thus reversing the problem. The student now has to answer again. All vocal feedback can be found in Appendix B, where the feedback is explained together with the corresponding scenario.

3.3.2 Scaffolding in “At the bus stop in the world of popplers”

Three different kinds of visual scaffolding are currently available for this subgame. One is based on relations between X and Y, and two are based on the size of the hidden variable X.

- Moving as a scale. Popplers move slower if they are more under a box because they are pulling in all directions. This leads to that boxes with fewer popplers underneath will reach further down the hill, and gives students the clue that the gap between the boxes correlates to the difference in number between the popplers underneath. The levels are mathematically divided to give a correct weigh feeling.
- Bumps on the box. The popplers underneath a box will occasionally bump into the walls. The numbers of bumps on the box are equal to the number of popplers underneath, and more frequent bumps correlates with more popplers. This gives a clue about the size num-

ber.

- Different width on the box. The more popplers the wider box. The box size will be proportional to the amount of popplers. When five popplers in each, the boxes will have the same width.

4 Design process and design decisions

4.1 Design goals and structure of the process

At an early phase the design group did set the goal to keep the principles and structures of the Magical Garden in order to obtain a consistent user experience and because of Magical Garden's sound foundation on research and well-known cognitive theories. Using this foundation the design group aimed to introduce a stand-alone extension that would appeal equally to boys and girls at 1st grade age. In Sweden, this means children in the age of 7-8.

Managing the design strategy process and implementation of the prototypes followed the iterative design process that Kathryn Best (2006, p. 112) describes with preparation, incubation phase, insight, evaluation and elaboration. Characteristics of the design process are that it evolves practically from tried and tested ways of problem solving and is continuously refined by a design team. The design team

1. Defines the problem
2. Develop better understanding of it
3. Conceptualize the problem
4. Detail the design solution and
5. Test or implement a solution.

The process is not sequential or linear. The process of the actual project was divided in iterations since these are needed for adapting the changing conditions of the design space, making mistakes and backtracking important and necessary. This is the reason this design group utilized possibilities to evaluate and motivate the design decisions continuously. When developing software it is important to be able to adjust and improve the product during the whole project. When implementing and performing tests the problems not found in the lo-fi prototyping process are shown.

1. Defining the problem

To define the problem, the current Magical Garden subgames were played and analyzed. The literature, getting familiar with other mathematical games, supervisor/apprenticeship meetings with Haake and Gulz at Lund University and speaking with teachers of first grade students offered good ingredients for the brainstorming sessions. The focus was on presenting ideas freely and not limiting them with any resource issues. Benefits of the brainstorming method for problem-solving are that the design group does not need to be experienced in this technique, the inhibition and defeatism is not blocking ideas so easily, the enthusiasm in group is contagious and the group members tend to be eager to top other's ideas (Adams, 2001).

2. Develop better understanding of the problem

Having visualized and presenting ideas in different forms is a prerequisite of evaluation and helps to share the idea and create understanding of the problem. Sketching helps with refining the ideas. Ideas from head to hand and to paper can be presented in different ways: for example sketches, moodboards and storyboards (Krisztian & Nesrin, 2006). Sketching was used for the mind mapping, and lo-fi prototyping was a combination of sketches and moodboarding, that express more the atmospheric qualities of the design. The first sketches on paper were represented early in the process. Supervisor meetings were the source of feedback and helped the design group to evaluate the sketches.

3. Conceptualize the problem

At the conceptualization phase the pieces start to fall into place (Best, 2006). Formulating the ideas further included: refining and detailing the sketches and adding mood boarding. With the mood boarding ready describing the items that the game players can collect, and conceptualizing the mathematical task in form of a sketched storyboard were made here. Storyboarding was not done in a sketch form, instead of sketching three films were edited in order to present the idea more convincingly and consequently, getting more and better feedback from the evaluators, present on a mid-presentation.

4. Detail the design solution

While designing the game the design group tried to keep some design principles in mind. Norman presents seven principles for transforming difficult tasks into simple ones. Norman advises to (2002, p. 188).

1. Use knowledge in the world and in the head.
2. Simplify the structure of tasks.
3. Make things visible.
4. Get the mappings right.
5. Exploit natural and artificial constraints
6. Design for errors.
7. Use established standards

These principles have been helpful while building up our foundation, creating a good concept where the feedback and scaffolding solutions helps instead of knocking over. They were also helpful to us when brainstorming up solutions for each scenario.

5. Test or implement a solution

During the iterations, the feedback was collected from the supervisor meetings, meetings with the other design groups and finally from an interview with two teachers concerning the concept. After that, a more detailed prototype version of the concept was created on basis of the evaluation of the films. Here the new graphics were designed and used.

4.2 Brainstorming

The first step was an exploration of existing games in Magical Garden in order to find ways to adjust the Magical Garden concept to older target group of 1st grade students and

something that attracts both genders. Before the first brainstorming session we talked to primary school teachers about their teaching experience and how they use technology in their teaching today. We asked what possibilities they have in their schools and how much they integrate technology in their teaching. Since we went to elementary school a lot of the teachers teaching processes have changes, so our own experiences are not up to date.

Almost directly after we got assigned our project we all agreed on making something new and creative where we would set the standard. The theme was already born before the brainstorming – space – and little effort was made to find alternatives. The idea is that after learning with Magical Garden during the pre-school years something completely new will inspire the children to keep learning math. We proposed a new theme for two reasons. First, from our own experience, starting 1st grade is a big step in life and we want this game to follow that step and provide the students with the feeling that they reach a new level of learning. This is also supported by Anderberg et. al (2013). Second, our goal is to motivate both genders. We hope this new theme and graphics will motivate the 1st grade students to keep learning with TAs.

We decided quickly that we did not just want to do a sub-game, but develop a novel concept including at least one sub-game. Our goal should be to create a product ready to be used. For the product to be complete we needed a home state where the student can use its collected stardust and a game where the student can learn math and earn more stardust to its collection.

Our first idea was to create a sub-game that would teach "the number neighbors". This changed after some more research, when we learned that this was too easy for the average 1st grade student. We wanted a challenging and fun game, in an environment where everything is possible. The concept aimed to support the knowledge in the head of a level of 1st grade student and apply this knowledge gradually to the tasks that equal their impression of a world. That is when the concept of "friends of ten" was chosen.

The design group aimed to keep the task structure simple, available, and visible. An example of this design considerations is the game's new start button: During testing of the present Magical Garden sub-games, the design group considered parts (like the start and exit) of the user interface difficult. For example, an anchor as a start button did not show clearly enough that the game actually starts when being pressed. Also the water pot was visually confusing, because the part where the water would be dropping was pointing out of the garden. In addition, the graphics and the environment in the game did not meet the requirements of 1st grade students.

4.3 Mindmap

During the brainstorming we created a mindmap with our best ideas. We connected all the different parts required for each sub goal. Quickly the functions and things we wanted to evaluate escalated. A timespan was made to structure our goals for each week. Minigoals and deadlines were decided to easily give us an overview of reasonable goals and limitations to succeed. Step by step we broke down the project in parts to focus on in Interaction 1 and decided what to be the

main focus in Interaction 2. We decided to save some functions to implement if there would be time over. We wanted to keep a high quality and focus on getting the game finished, ready to be played.

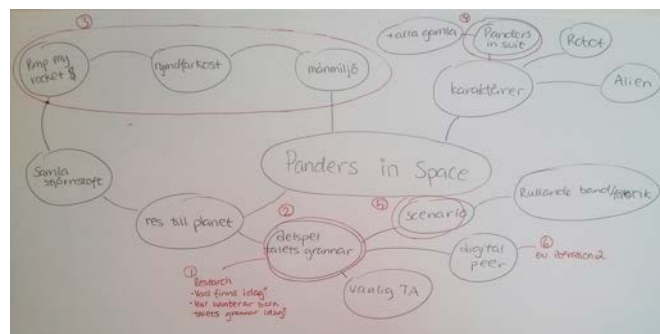


Figure 2: Mind map over project scope

4.4 Low-fidelity prototyping

Creating a low-fidelity prototype is the perfect step when developing a new project. It is cheap, takes little time and helps the whole project group to visualize all thoughts about the desired result. In our low-fidelity prototype we simply used paper, pen and scissors to draw all parts and create the environment we wanted. This was to more easily make our vision of the concept clear to everyone in the group. This was really helpful when deciding how the game should be played and how and where the scaffolding should take place.

4.4.1 TA:s

We decided to include the already existing characters of Magical Garden, and embody them in our space theme. The existing TA:s would be dressed in space suits. This would not only give them the right equipment to easily survive in space, but also help the students to recognize the characters they may have grown to befriend. We also wanted the space concept represent a next step in their mathematical development to be reflected by the TA:s.

In addition to the old TA:s we also wanted to develop new TA:s. One of the reasons for this was that the panda bear Panders had received poor response from boys and we wanted to leave room for experimenting if the character Panders was responsible for this or if there was a more general problem with respect to gender in the implementation of TA:s in Magical Garden. The other reason was that we wanted to provide fresh alternatives for students that might be bored or for other reason didn't like their old TA.

A benefit of using a space theme concept is that it gave us opportunities to experiment and try out a wide variance of different, crazier character types. As a proof of concept we developed several different alien types to later possibly be tested on students to see if girls versus boys are drawn to any specific character traits or characters. Afterwards we decided to go forward with the old characters and two additional agents: a friendly robot and a cute alien. The new characters plus Panders in a spacesuit have a slightly new graphic profile. We hope this design will be more appropriate for 1st grade students.



Figure 3: New space-themed TA:s

4.4.2 Moonbase

The moonbase is the home state of *Math in Space*. This is the place where the student can repair or adorn her or his spaceship. The only way to get to do this is by playing subgames and collect stardust. The spaceship is used to transport the student and the TA to the mathematical subgames. This gives the students a relevant link between exploring the subgames and improving their spaceship. The spaceship will be visualized in the background for each subgame in some context and when the student is finished with the subgame the rocket will travel back home to the moonbase. The aim of the graphic design choices is to inspire the student to play subgames on the colorful planets rather than just be inspired to build her or his rocket. That is why also only a few changes can be made before the student is out of stardust.



Figure 4: Lo-Fi over moonbase

4.4.2.1 Motivational role of stardust

Stardust is a main part of the logical cycle to get the students to repeat the rather monotonous mathematical subgames. The students collect stardust from completing a subgame, travel back to the moon base to build or change new parts on their rocket, and quickly get back into the subgames to gather more dust. Stardust is magical, creative and can easily be stored in an infinite amount in a bag. It was important to use something that belongs in space, but still is believable enough to be used to switch parts on a spaceship. It will hopefully be a magical visualization and give the feeling that everything can happen in space.

The stardust will be collected and presented in a bag in the upper right corner and will be displayed throughout the whole game. The more dust you collect, the bigger the bag will grow. To give the students passive learning about numbers and about larger-smaller number relations, the amount of stardust in the bag will be printed on it using its numerical representation.

4.4.2.2 Motivational role of rocket parts

The rocket will serve the same role as the garden in *Magical Garden*, but in a more motivating way. When replacing parts the student is given the option of where to replace a part, but not what to replace it with. This gives each student an obvious personal game progress, yet makes it hard to compare progress with fellow students.

It will be possible to build up the rocket in a theme. Cheese theme, spaceship theme, animal theme are the designed themes for now. The goal is to give the students the possibility to have a randomly or symmetric rocket.

When changing a part the student has to press a part of the rocket and the chosen part blinks to give the student feedback on which part to adorn. The student can now chose to buy a part with her or his stardust. Depending on the result the TA will give a vocally relevant comment. If the part is changing the stardust will randomly be thrown over the part and the new part will emerge. At the same time the bag in the upper right corner will become thinner and the amount will decrease with minus nine. Nine was decided because the learner will passively be exposed to all numbers. When the bag contains less than nine units of stardust the TA will suggest to press play. More about the vocals can be found in Appendix B.



Figure 5: Lo-Fi over rocket parts

4.4.2.3 Motivational role of planets

The moonbase will be a purposely rather bare and cold place with limited activities to take on. The purpose behind this is simply to not keep the students in this state rather than playing the mathematical subgames. In contrast to the gray moon, the students will see colorful and fun planets representing the mathematical subgames, hopefully pushing the students to quickly want to keep playing rather than putting too much time in rocket design. One is not able to interact with the planets, nor to choose which planet one will travel to. The only interaction outside the rocket will be to simply press play.

4.4.2.4 Play button

To give the play button a stronger meaning of linking the build-state and subgames together we chose to sketch it as a detonator. The students will see their TA take a seat in the rocket they have built and with the rocket the TA and the student together will fly to the subgames out in space.

4.4.3 Friends of ten - "At the bus stop in the world of popplers"

1st grade students in Sweden should automatically know the friends of ten, nine, eight and so on. Our subgame will only have the capacity to play out the 10 friends, thus with the possibility to implement further alternatives. This is an important feature while one of our goal is to create a game at a higher level of difficulty than Magical Garden and also to contribute to students learning something central to the Swedish grade 1 mathematics curriculum.

Before the subgame is ready to be played a short scenario will introduce the story of it. The scenario is designed to be self explanatory to the students so they quickly understand what to do. It is also designed to motivate students to want to help the popplers to get to school. As described in section 3.3 a space-road will be located up in the space air. This gives the student some time to see all the ten popplers before the moving truck accidentally drops down two boxes over them to get a sense of how many popplers were waiting at the bus stop. The scenario will only be played in the introduction and our goal is to keep it short and not draw too much attention from the actual math tasks.

The first time the student is playing the game a lot of scaffolding and feedback will be present and while the student improves her or his understanding of the math concept clues and help will gradually be removed.

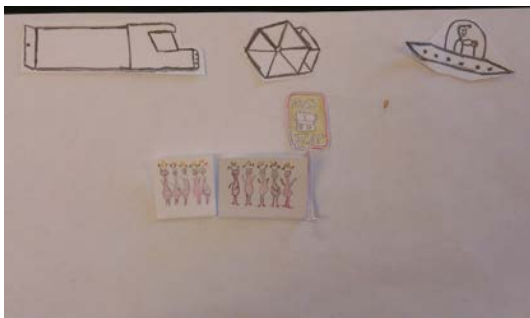


Figure 6: Lo-Fi of subgame

4.4.3.1 Play Scenarios

Similarly to Magical Garden we have chosen to play the game scenario in three steps. At first the student plays by herself to get a feeling on how the game is played and to get an understanding of the math concept. Depending on how well developed skills the student has, corresponding feedback will be provided.

In the second step, when the student has figured out what to do, the TA will get out from the spaceship and vocally describe its curiosity for how the student can know the correct amount of popplers hidden under the box. The TA is still insecure and prefers to watch the student play one more time.

After this comes the third step where the TA is excited to learn and wants to try to play. The aim is that the TA answers approximately compare to the students earlier results. The student will then need to approve the TAs answer before the two find out if the TA was correct. The goal is for the TA to get all answers right.

With respect to motivation we have decided that stardust will be collected for each correct answer. When the left box reveals the hidden popplers and blows away some stardust will be thrown and the money bag will increase with +1. We hope this will keep up the learners motivation during the whole game by them being continuously awarded. Depending on the learners mathematical level the amount of rounds will differ and by this solution there will be no comparison in collected stardust during one session between the students.

4.4.3.2 TA interaction

The TA will interact with the student at their level of communication, that is similar communication and language that 1st grade students' use between each other should be applied.

The underlying goal of the interaction with the TA is to force the student to take the "teachers" place and tell when the TA is right or wrong. The aim is that the TA and the student should create a bond between each other, and that the better friends they become and the more skills they achieve the more stardust and positive feelings will be given.

Depending on the different answers and if the TA is watching or playing, specific selected feedback will be given.

On the moon base the TA will speak to the student in a friendly way. The TA will explain the meaning of the rocket and give compliments when a part is changed.

4.4.3.3 Scaffolding using the boxes

When the student needs aid to understand the mathematical concept the boxes in the game will give clues on how many popplers are under the box.

The boxes will also indicate a feeling that relates to scale; they will differ in how far they can run down the hill. When there are more popplers under a box this will slow down the box's speed because all the popplers running in different directions. Also when all popplers are gathered together they will be jumping on the bus at the busstop. This gives the logical reason for one poppler to have the furthest way to the busstop.

The boxes will when running down the hill be wider if there are lots of popplers underneath and smaller with fewer. If

there are five popplers in each box the boxes will have the same width. When the student does not need all the scaffolding any longer, the boxes will have the same size and always be in a horizontal line.

4.4.3.6 Scroll Wheel

This was not implemented due to time constraints, but we consider it an important feature and have explored and discussed the feature quite extensively during the project.

The idea was to try out a scroll wheel as device for the learners to give their answer. The reason was to introduce something new compared to Magical Gardens answering opportunities. We believe that a scroll wheel is a feature that 7-8 year olds can handle. Nielsen (2010) holds kids in the age of 9 and older can manage such a feature but we believe that with the trend of usage of Apples products and Nielsen's article now being a few years old the age is already lower than Nielsen proposes.

The scroll wheel will be placed to the left of the box. On the box a question mark will pop up to give a direct link for the player what to do in the game.

4.5 Concept videos

The reason for making our concept movies was to be able to visualize all parts in our game and figure out if things were missing or overwhelming. The movie was also meant to bring up feedback from the mid-presentation to get ideas and criticism from people outside our project group.

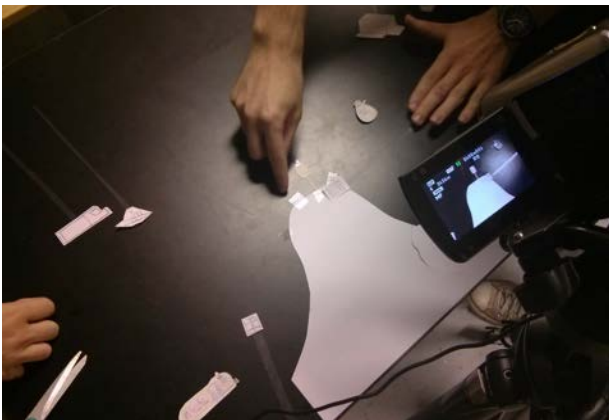


Figure 7: The group filming concept videos

4.6 Mid-fidelity prototyping

For our final presentation in Interaction 1 we animated our concept with our new graphic design. Here we could show what happens when buttons get pushed and show some of the feedback we were about to implement. This design also helped us with decision of colors and animations in order not to distract students attention from focusing on the math question and thinking about how to group the popplers in the best way.

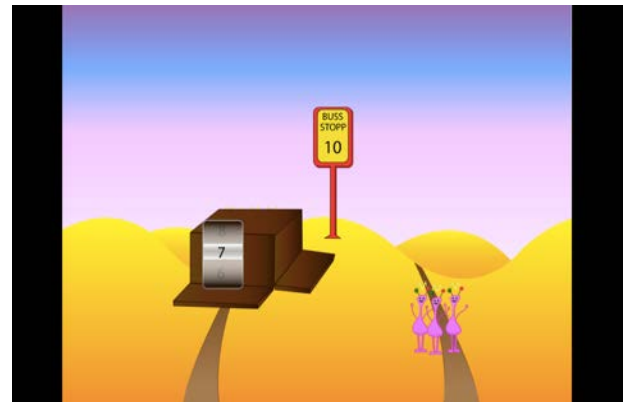


Figure 8: Mid-fi presentation

5 Prototype

5.1 Results

The following figures are screenshots taken from the final version of our prototype.

5.1.1 Moonbase



Figure 9: Character selection

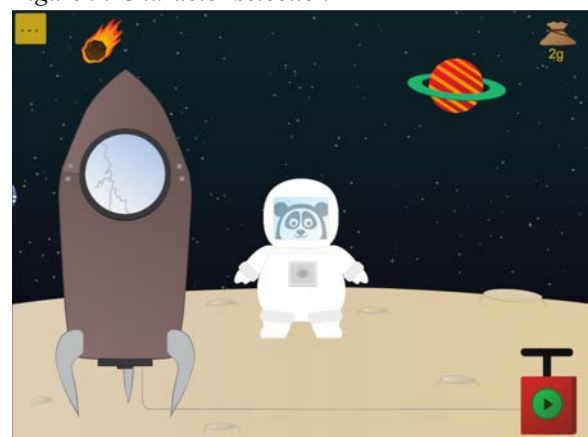


Figure 10: Default state of moonbase

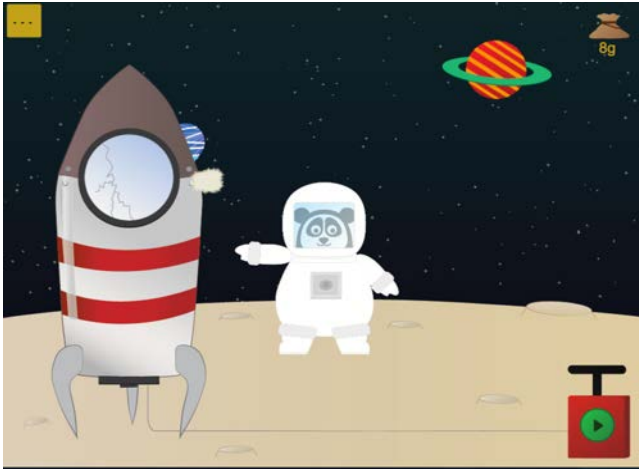


Figure 11: Panders throwing stardust to rebuild the window of his rocket

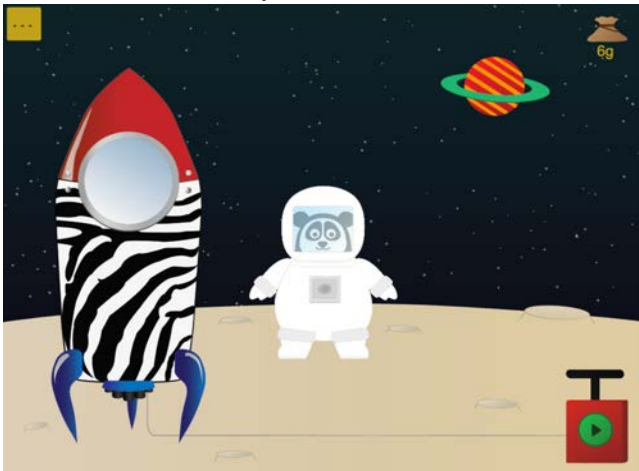


Figure 12: Example of how a repainted rocket could look like

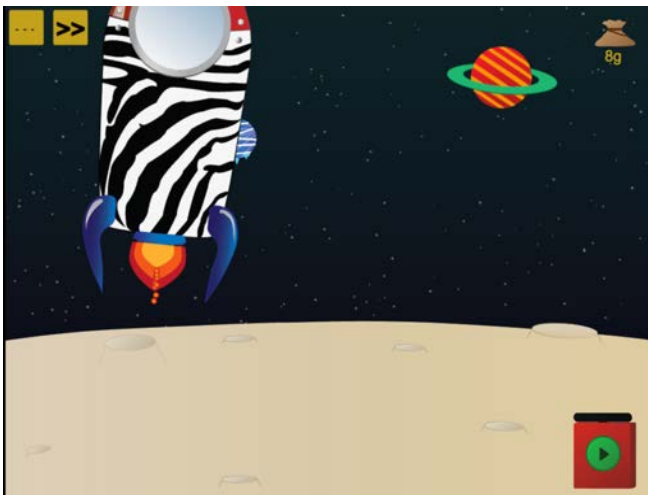


Figure 13: Rocket lifting off after pressing play button
5.1.2 Poppler World

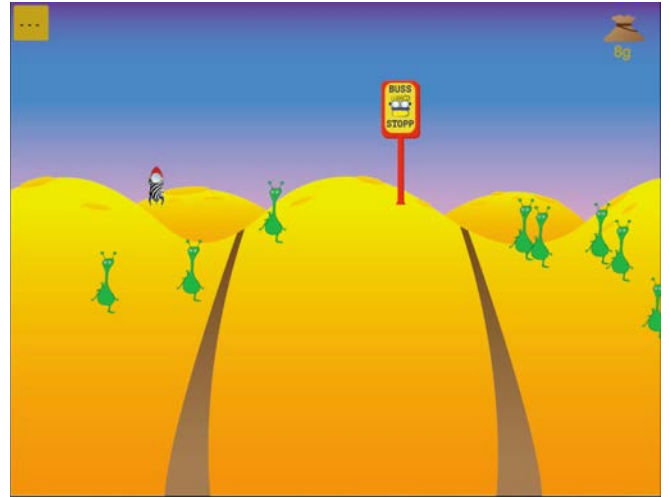


Figure 14: Poppler running towards the bus stop in the beginning of poppler world



Figure 15: Popplers running around in panic as a result of the truck dropping it's boxes

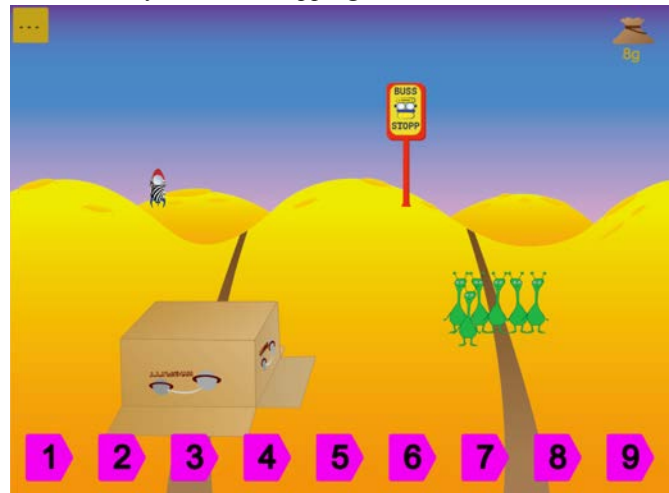


Figure 16: When answering how many popplers are in the other box

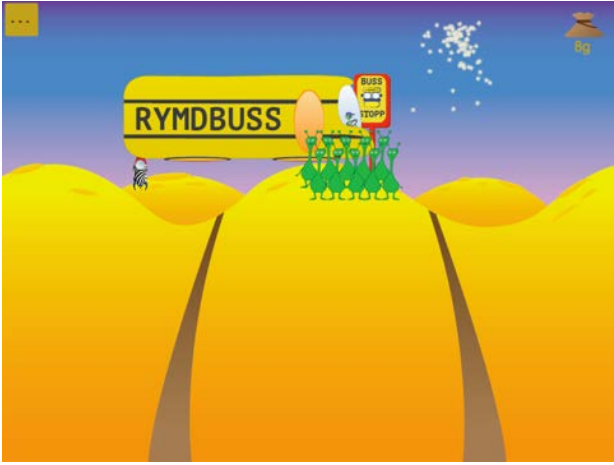


Figure 17: The bus arriving after the player has picked the correct amount. Stardust flies out from the buss and into the pouch.

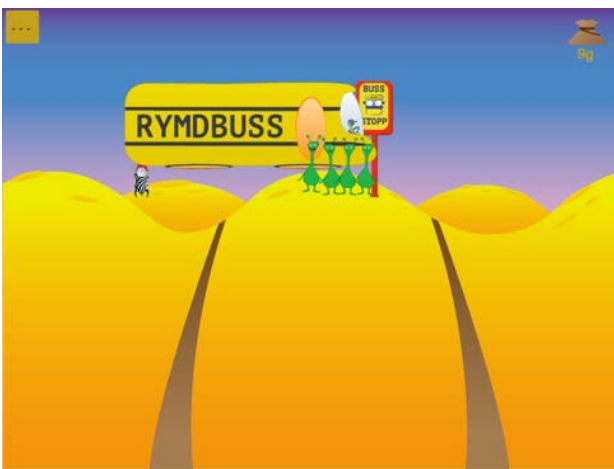


Figure 18: Popplers get on the bus.

5.2 Tools

We stored the code base in a Git repository through BitBucket. This allowed us to work in parallel and handle merge-conflicts and other configuration-management problems in a controlled environment. IntelliJ IDEA was used as a coding platform. We choose IntelliJ IDEA because it's a free and powerful tool for Javascript with good Git integration and error handling. Grunt is used for on the fly compiling and testing in the original Magical Garden project and we saw no reason to replace it. We used Trello as scrum board to easily keep track of which task every group member currently work on and which tasks that's next to be done. All illustrations are drawn and vectorized in Adobe Illustrator CS6 and can easily be scaled to bigger and smaller versions, in case of developing for platforms with higher resolution. Audio files have all been cut and edited using Final Cut Pro.

5.3 Implementation

The server side implementation, also known as the back-end, is kept the same as in the original magical garden implementation (Malmberg, 2013). The reason is that we want to have full compatibility between the data gathered playing Magical garden in preschool and the data gathered in our game. For this we have re-used Garden Data to store our spaceship. The Garden Data table stores four different variables x, y, content_type and level. the x variable is currently mapped to the

type of part (0 = top, 1 = mid, 2 =legs , 3=window ,4 =bot), the string name of the type is also stored in content_type for easier reading and debugging. The style of the part is stored in the level variable. the y variable is currently unused and always set to 0. When implementing add ons they will be stored at different y values. When starting a new Math in Space game the player's garden data currently need to be cleared from earlier magical garden data. To avoid this a new table have to be created in the back end for the space ship. Stardust is stored under a players water variable, no customization is needed for this to work.

To make it easier for developers to move between Magical Garden and Math in Space, the front-end of Math in Space is built on the same foundation. i.e. HTML 5 and Javascript using the Phaser game engine. Common elements were imported from Magical Garden into Math in Space.

5.4 Code base

The code base was imported from Magical garden. common objects have been left mostly intact and when needed to add and replace objects it has been done through the SpaceInterface file instead. The SpaceInterface also contains some functionality to replace name conventions with the backend so that for example stardust is used instead of water in the rest of the code base without having to change in the back end. Unused states and files from the original Magical Garden code base have been removed to avoid clogging up the code base. Some reformatting still needs to be done.

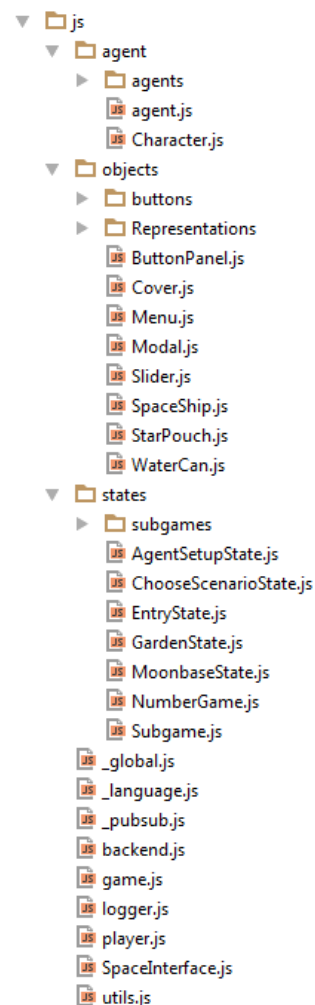


Figure 19: class structure

5.5 Animation

Animations were done using TweenMax and TimelineMax to chain them together. Due to time constraints only single sprites are used instead of sprite sheets. Therefore some animations may look a bit stiff in comparison to other games.

5.5.1 Moonbase

The animations on the moon base are predictable and monotonous. The planets move slowly, stars are twinkling slightly in the background and a comet is passing with a low frequency. When pressing the play button the detonation handle goes down and the fuse is lit. As the fuse is burning down the TA moves into the rocket. Then the engines start and the spaceship flies out in space.

5.5.2 Popplerworld

The main focus of the animations in Popplerworld has been to give a good flow in the game. The upper road is set to have traffic with a lower frequency than in our concept videos to not interrupt the learners' attention while still giving the world a more natural and alive atmosphere. The Popplers change their posture and expressions randomly to give them more personality. This will also make them easier to separate and count. The TA's animation is passive. It is an observer and will not disturb the gameflow.

5.5.3 Skippable

Some of the animations have a skippable feature to provide the students who already are already familiar with the game the ability to skip them. Currently only animations in the Moonbase are skippable. But in the future animations in Popplerworld will also be skippable. The skippable animations should be when the game round is being set up and when the answer is correct. Animations playing when the answer is wrong should not be skippable. This is because the animations provide valuable feedback and slows the game-play down, encouraging students to reflect on their answer and think before they act.

5.6 Audio

Sound design turned out to be one of the main difficulties producing this game. It's hard to tweak animations and graphics not knowing how the object sounds will interact with other factors in the game. Adding sound effects gives the motions a complete new dynamic, which makes it important to continuously add audio while developing new game features.

5.6.1 Sound effects

Every sound effect and voice in the current version of the game is either downloaded from free audio websites or created by our own voices using mobile phone. The sound quality might therefore vary during the different states of the game. We first thought it to be easy designing sound effects to a game concept taking place in space. Surprisingly it turned out to be a real challenge when it came to making every sound effect float smoothly together. The poppler voice is an example of sound effects we rather quickly decided on, but which later turned out to not fit in at all.

5.6.2 TA voice interaction

The TA will mostly interact with the player in much the same way as Magical Garden but with a completely new manuscript, as seen in Appendix B.

5.6.3 Narrator

Throughout the process we had problem with deciding how to introduce a narrator who gives feedback and introduces different aspects throughout the game. It isn't possible to have the TA do this since the TA isn't always available in the subgames and also because the TA is the student's tutor and can not all of a sudden know more than the student in order to provide feedback. The original Magical Garden solves this by introducing different narrative characters in all the subgames but we wanted the Narrator to be constant in all subgames. Our solution was to introduce the spaceship's navigator as the narrator. It seems natural when arriving in a subgame to have the navigator explain where we are and what to do. It is also expected of a navigator, or a computer in general, to be good at math which makes it reasonable for a navigator to give feedback.

5.7 Player Input

5.7.1 Moonbase

There is not much interaction in the moonbase state. The parts of the spaceship are clickable in order to have them change. The play button can be pressed in order to leave the state. There should also be some interaction with the Agent (currently unimplemented) in order to give the agent more personality.

5.7.2 Popplerworld

The interaction in Popplerworld is also sparse. The only interaction is when the game prompts for an answer. The game will react differently depending on the answer given as described in 3.3.1.

6 Test and evaluation

At an early stage in this project we interviewed and tested two teachers to receive feedback on our ideas and lo-fi prototype. The teachers encouraged us about the concept of learning by teaching and one of them used it already in school, but not with technology. The two teachers, with no previous experiences with the concept of TA in a digital mathematical game, were positive towards the idea. They also emphasized the fact that students in this age prefer different kinds of games. This was an important input to our knowledge. The teachers considered our idea of rewarding students with something important while it is likely to motivate the students to try harder and keep focus longer. Both our theme and game concept, design and structure were appreciated by the teachers.

One important feature we had not taken into account was that use of mathematical terms and vocabulary, which the two teachers had on their wish-list. We therefore thought about providing more feedback with the possibility to see the addition and subtraction in mathematical terms for the more ad-

vanced students. More detailed information about our results and analysis can be found in Appendix A..

7 User test evaluation material

7.1 Background

This section describes the evaluation test we have come up to, where everything is thought through.

The test is supposed to be performed in two phases. The goal is to use the first phase to collect information needed for the second phase where the final test is performed. The first phase includes a visit at the school where the test should be performed in order to get familiar with the students and the environment. Goals are to create a better understanding of the test environment and to get greater knowledge in how to communicate with students at the age of seven. It is important to be able to communicate on the same level as the test persons to receive better end results (Girard, 2011). The purpose of the first visit is to find out how to use the technology and figure out how to keep the students that are not doing the test busy. According to Girard it is important to entertain kids in this age and keep in mind that all kids are not the same. Our goal is to get the kids to understand that they are the experts and that we need their knowledge and input. To create a secure environment for the students the testing will be performed in pairs, decided by their teacher (Nielsen, 2010). Even if Girard holds that kids do not need anything in reward we have decided to create and give out a *Math in Space* certificate thanking them for their expertise. The aim is to create a bond with the students and hope they will go home and talk about what they have done in school.

7.2 Purpose

The test is structured to give data that will help us to answer if we have succeeded to fulfil our goals with the project. The purpose of the test is to see if the game is working without any bugs and the final test serves to find answers to the following questions:

- Do 1st grade students understand what they can do on the moon base, how to go to a subgame and how the implemented subgame is played?
- Is the space theme a good environment for both girls and boys?
- Is the subgame challenging enough?
- What are the responses on the new graphics compared to the Magical Garden graphics?
- Is the scroll function too complex or will the students master it?

7.3 Method

7.3.1 Test Environment

The test is supposed to be done at school in Lund. The aim is to let at least six girls and six boys play and evaluate the game. We have decided to split up the test in two visits.

7.3.2 First Visit

The purpose of the first visit is simply to get to know the students and find out how to perform the real test with the technology working. The first visit's goals is to understand the communication barrier between us and the students, come up with how to structure up the logistics with the teacher and where to execute the test. The presumption for playing *Math in Space* is that the students are familiar with the original game Magical Garden and with us. In this first visit all students involved in the test will play Magical Garden.

7.3.3 Second Visit

The second visit should be done close into the first visit so the students still will recognise us and remember the Magical Garden concept and graphics. The test will be performed with the same students and structured with the help of the teachers.

7.3.4 Data collection (observations, analysing, interviews with the kids)

The test will be mostly done with a short introduction to the students about their assignment. Observation methods will be used and an open conversation during the test might be needed if the students are shy to open up for discussions. At the end the kids will individually get to answer a survey, see Appendix C. This is to be able to study possible differences between boys and girls.

7.4 The test

The students will test the game in pairs decided by their teacher. The aim is to test on two pairs of girls, two pairs of boys and two pairs with both genders. The test will take approximately 20 minutes and be followed by an informal conversation with the kids and with finally letting them answer a short survey. The survey will primarily use pictures and be written in Swedish.

7.5 Result

This section will not be done in this interaction course, while the time is too short. However the aim is that if a new group takes this game forward, this test should be the first thing performed, in order to illuminate whether both girls and boys like the space theme, whether the game is appropriate – attractive and sufficiently challenging – for 1st grade students.

8 Discussion

The *Math in Space* concept might be appealing to boys without losing the interest of the girls. This was held by the teachers that we interviewed, and we agree with their belief but it is yet to be tested with students themselves.

When analyzing the teacher feedback on our lo-fi concept we realized that mathematical terminology is way underused in the game and should preferably be expanded on. The only current mathematical languages used concern weight representation for the stardust pouch. The same, or other similar representations should be included elsewhere. It is especially important in our subgame in order to bring the knowledge

from the game over to other mathematical material used in education.

From the teacher feedback we also got our hands on their students current favorite math games and found the level of flow and game animation to be way above the ones in *Magical Garden*. With our version of the game we believe that game flow and realism is a great important to capture the student interest. We have therefore tried to push ourselves and go beyond the technical scope of *Magical Garden* in order to create more realistic animations. By doing this we have kept in mind that the students still should not be distracted by the actual purpose of the game; learning math. We believe it is possible to do this within the current existing game module, for example: this can be done in the Phaser game engine by working with more advanced spritesheets and implementing dragonBones skeleton animations for characters.

We hope that future project groups will expand on our version of the game. The possibilities of including new subgames are wide, both regarding scenarios and different mathematical aspects. Our subgame is based on friends of ten since this, according to one of the teachers, is one of the main problem areas in the first grade curriculum and is worked on during the whole year. The teachers requested for other numbers than ten as well, which is something we already have thought of in the design of our subgame. It will be very easy to expand to other friends than ten, both larger and smaller. When we chose friends of ten as the mathematical aspect for our subgame, one of the reasons was that we wanted something that was considered “harder”. The reason was that we felt *Magical Garden* was focusing primarily on students who could be considered weaker and we wanted to also appeal to students who are at a higher mathematical level. Because of this there is a large gap between *Magical Garden* and our *Math in Space* concept, which needs to be filled, from both sides. There should be games that are on the more difficult spectrum in *Magical Garden* and games that are on the easier spectrum in *Math in Space*. And there should be a large overlap between the games. All this is needed to catch all students regardless of skill level and create an individualized learning experience.

We hope that *Math in Space* could actually be something included in the lower elementary school schedule.

The goal of this project is still to be evaluated and analyzed and our wish is that a new group of students will continue with this project.

9 Future work

During the project we have come up with a lot of ideas that we were not given enough time to develop. Many of these ideas might bring about large conceptual differences from what the game is today but we are very open for future groups to further evaluate, develop and implement whatever they believe might make the platform better. All new changes and subgames created for *Math in Space* should aim to follow the same quality, graphics and underlying structures.

9.1 Mathematical terminology

When showing our version of *Poppler World* to first grade teachers they pointed out the lack of mathematical terminology during the mathematical tasks. They reasoned that the

students now are old enough to see equations in “normal math terms”, for example including an ‘X’ under the box, amount number under the popplers and a ‘+’ inbetween. Introducing mathematical terminology in the game, both visually and orally to support the mathematical operations will provide an interface for math in school books and learning in Math in space.

9.2 Subgames

To cover the gap between the mathematical level to *Magical Garden* more subgames need to be developed and included, for example implementing a game which supports the concept “number neighbors”. This is something that exists in the curriculum of first grade and that students should learn but in practice very little time is spent in class since students are expected to know this from before.

9.3 Improve Poppler world

In the current version of the subgame, *Poppler World* only includes friends of 10. In future versions the plan is to extend and provide other number friends to learn. The number of which number friends the student should learn might be determined by either the teacher, or how well the student performs in the subgame.

9.3.1. Scaffolding

As for futures scaffolding improvement we have had ideas about experimenting with different colors of different amounts of popplers. This is a scaffolding solution which could improve the subgame a lot, and for future project groups to further look into.

9.3.2 More variation on the popplers

Since The popplers are grouped together it is important to give them distinct and unique features so that the individual is easy to separate from the group. Popplers may have slight differences in size, tint and expressions.

9.3.3. Scroll wheel

We always intended to have player input in *Poppler World* in the form of a scroll wheel as we described in 4.4.3.6. Due to time constraints we decided that it was better to keep the current form of user input rather than create a badly implemented scroll wheel. In the long term plan we still want this to be implemented.

9.3.4. Role of TA

The TA is currently acting in the exact same way as in *Magical Garden*. We suspect that older children might find her or his behaviour predictable, boring and mostly non interactive. It is also probable that the current interaction is too slow for friends of ten, which requires some automatisation as it is considered a core knowledge for math in the first grade. Therefore we suggest that the role of the TA and how it acts in subgames to be evaluated and improved. We also suggest considering and evaluating how adding a fourth phase of a “hard” subgame where Panders has the role of a digital peer i.e. are on equal term as the student.

9.3.5 Voice and sound

The *Poppler World* audio are all last-minute made and could definitely be improved on. In the current version most sound effects don't go together and are unnecessarily annoying.

The current voice introducing the subgame are also last-minute made, based on an idea to have the rocket explaining how the subgame works. We had to use an external voice to introduce the rules since the TA doesn't enter until phase 2. The idea of having the rocket speak is just a fast fix solution which could be solved in other way. This is hopefully for future groups to further evaluate and develop.

9.4 Improve Moonbase

During our project we focused most time getting the moonbase as well developed as possible. There is little room for larger conceptual changes, but of course room for further extension of what already exists.

9.4.1 Rocket paint

There are now only 5 different painted parts of every rocket part. This is something we wish for future groups to extend.

9.4.1 Rocket addons

The base idea is to have two choices of changing your rocket, either paint or add. Paint is stated above and currently implemented in *Moonbase*. The idea with addons is to put smaller parts on the painter rocket, for example a tie or a hat. Some graphics to this exists in our rocket part atlas but currently not implemented and used. This is something we wish for future groups to implement and extend on.

9.5 TAs

A main benefit of developing in space environment is the great opportunities of introducing fun new TAs to implement and evaluate.

9.5.1 Porting Magical Garden TAs

The current version of *Math in Space* includes Panders in a space suit. We also intend to port the other TAs from *Magical Garden* to our space environment.

9.6 Testing evaluation

Included in the report is an testplan and evaluation material. We wish to perform more detail testing to found get more data. One test would be to investigate which character that is the most popular and figure out why. Some external randomized characters could be created to give a wider range to test on. One important thing to keep in mind is that it is a game supposed to only be used in school during mathematical sessions and is to be compared with a math book. The reason for mentioning this is that kids in this age plays a lot of games outside of school with a totally different design.

10 Future case

Projekt 1: Math in Space – Testa samt utveckla delspel för åk 1

Bakgrund: Spelet är implementerat enligt HTML5 och i dagsläget har vi ett ramverk till spelet i Math in Space samt ett delspel anpassade för 6-7-åringar. Spelet är utvecklat av en projektgrupp i Interaktion 1 & 2 år 2014 som hade stort fokus på utvecklingen. Spelet och dess teoretiska bakgrund är därför ännu inte testat.

Uppgift: Projektet handlar om att första testa den existerande versionen av Math in Space och där utefter utforma ett eget delspel (eller två). Konceptuell design med LoFi- eller MidFi-verktyg (pappersprototyper osv.) ingår i utvecklingsprocessen. Man kan även arbeta med HiFi-implementering i t.ex. HTML5.

Intressenter: Grupp 2b 2014, Educational Technology Group LU/LiU, Stanford School of Education, Elevdata, Natur & Kultur, Linnémiljön: Cognition, Communication & Learning.

Projekt 2: Math in Space – Teachable agents

Bakgrund: Learning by teaching är en väldigt kraftfull metod för lärande och en hörnsten i Magical Garden projekten. Den implementation som finns idag är linjär och i en trestegs-lösning. Det har också visat sig i pilottester att vissa karaktärer är ointressanta för speciellt killar och att de föredrar att spela på egen hand.

Uppgift: Utvärdera nuvarande implementation av Teachable agents i *Magical Garden* eller *Math in Space*. Se över vilka karaktärer som är intressanta och varför. Utforma förslag på förändringar i utformning av teachable agents och interaktionen med dem.

Intressenter: Grupp 2b 2014, Educational Technology Group LU/LiU, Stanford School of Education, Elevdata, Natur & Kultur, Linnémiljön: Cognition, Communication & Learning.

References

- Alseth, B., Kirkegaard H., Rösseland, M. (2007). Pixel Matematik, 1 A, Lärbok och 1 B Grundbok. Natur & Kultur
- Best K., (2006). Design Management, Managing Design strategy, Process and Implementation. AVA Publishing.
- Blair, K.P., Schwartz, D. L., Biswas, G. and Leelawong, K. (2007). Pedagogical agents for learning by teaching: Teachable Agents. *Educational Technology*, 47(1) 56-61.
- Brorsson, Å. (2013). Prima Matematik, 1 B. Gleerups utbildning AB.
- Chase, C., Chin, D.B., Oppezzo, M. & Schwartz, D. (2009) Teachable Agents and the Protege Effect: Increasing the Effort towards Learning. *Journal of Science Education and Technology*, 18(4), 334-352.
- Gulz, A., Haake M., (2014). Att stöja och utforska lärande med hjälp av digitala läromedel i Vetenskapliga perspektiv på lärande, undervisning och utbildning i olika institutionella sammanhang. *Media-Tryck, Lunds universitet, Sverige 2014. Internet: <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId>*

- =4437902&fileId=4437905
- Erik-Philip Sörensens Stiftelse (2013). Han bryr sig inte om hur många äpplen där är.
- Florian, L. & Black-Hawkins, K. (2011) Exploring Inclusive Pedagogy, *British Educational Research Journal*, 37, 5, 813-828.
- Girard, Sylvie (2011). Case studies illustrations of experimental settings issues in classrooms, University of Bath.
- Haake, M. (2009) Embodied Pedagogical Agents – From Visual Impact to Pedagogical Implications, 3-56, Dept. of Design Sciences, LU.
- Johnson, D., Gardner, J. & Wiles, J. (2004) Experience as a Moderator of the Media Equation: The Impact of Flattery and Praise. *International Journal of Human-Computer Studies* 61(3), 237-258.
- Lave, J., Wenger, E. (1991). *Legitimate Peripheral Participation*. Cambridge University Press.
- Malmberg, (2013). Combining Inclusion and Individually Adaptive Learning in an Educational Game for Preschool Children. Master Thesis in Computer Science / Interaction Design, Lund University, 2013.
- Norman, D. (2002). *The Design of Everyday Things*, Doubleday, New York.
- Nielsen, Jakob (2010). *Children's Websites: Usability Issues in Designing for Kids*. <http://www.nngroup.com/articles/childrens-websites-usability-issues/> (2014-11-03)
- Robertson, J., Cross, B., Macleod, H., & Wiemer-Hastings, P. (2004). Children's interactions with animated agents in an Intelligent Tutoring System. *International Journal of Artificial Intelligence in Education*, 14, 335-357.
- Sawyer R. K. (2006). *The Cambridge handbook of the learning sciences*, volume 2. Cambridge University Press New York.
- Yee, N., Bailenson, J.N., & Ducheneaut, N. (2009). The Proteus Effect: Implications of transformed digital self-representation on online and offline behaviour. *Communication Research*, 36 (2), 285-312.

Appendix A

Evaluation with Teachers on first Interaction

Two teachers were asked questions concerning digital tools and motivation when teaching 1st grade students and they got chance to give feedback on Panders Space concept. Their answers are presented below.

1 Digital resources in development of mathematical skills

Both of the teachers considered that it is important that the students learn mathematical vocabulary and learn to communicate in mathematical terms. The mathematical tasks should be represented in several different ways in the game. It is also important to adjust the level of the tasks so that they correspond the individual level of each student. There should be an attractive reward system in form of points, collecting or building something with help of the reward. The teachers had no experience of TAs in the games that their pupils have been playing at school, but it was considered to be positive feature. Also, one of the teachers mentioned that she uses learning by teaching in the classroom.

2 Important features of games in learning mathematics

Different students like different kinds of games. For example, some prefer competition or playing in groups and some prefer playing on their own. Different alternatives are available. A teacher mentioned two types of resources: one type that focus on more experimental mathematics and the other type focus on training the skills. GeoGebra and applications on web site of National Library of Virtual Manipulatives are good for experimental mathematics. Arcademic Skillbuilders or Ten friends of Digitalt Hjarta AB would fit better for training mathematical skills. King of Math Junior offers tasks with variation. The details that make a successful game are:

1. The game offers possibility to communicate with mathematical terminology in interaction with others.
2. The game offers different kinds of mathematical content in different kinds of representations.
3. The game takes error possibilities into consideration and supports avoiding the typical misunderstanding of a specific mathematical content.

3 Motivation of the 1st grade students

When using mathematical games the teachers had not experienced any difference between boys and girls. The games that are adjustable to the personal needs of the students are the most interesting ones. The rewarding is important and it is interesting if the student gets a reward in form of getting to build up something, develop something or transform oneself into some figure etc.

4 Panders in Space is a useful concept

The teacher's feedback on Panders in Space concept was positive. According to both teachers the space theme is fascinating and one of them mentioned that it is equally fascinating to boys and girls. The idea of hiding part of ten friends in the boxes was appreciated, a teacher mentioned that it is a good exercise also at school: part of matches will be hidden in fists, and then, the children are shown the matches on one hand and they will figure out how many matches are in the other hand. TA was considered as an excellent feature that the teachers use in the classroom, but has not been found in the games that the teachers have been using.

5 Improvements to be considered on the basis of the teacher's feedback

The fact that there are 10 friends concept in the game is good, but an improvement would be if the student would be able to practice also other amounts, for example 5 friends or choose any amount of friends in the same game. It would be useful to show the addition under the boxes and prepare the students with the other levels of development, for example subtraction or using more numbers.

On the basis of the feedback it is important to learn the mathematical terminology and use the mathematics in different ways. Scaffolding and learning by teaching are in use in classroom and would be beneficial in a game. When getting familiar with the other projects and observing during the seminars, the design group came to the conclusion that it is

very important to appeal as many senses of the students as possible. The seeing, hearing and appealing to emotional imagination in a positive way would be beneficial for different kind of learners. In addition, it would be interesting to offer the possibility to play in groups and individually.

6. Elaborated Examples the potential improvements

Ten friends under boxes: Add number ten and addition $x + x$ below the boxes. For example, 3 popplers could jump on one box and the number under that box would show 3. The student can then scroll the scroll bar below and select a correct number. As soon as the correct number is selected, the student gets the reward (stardust), and also audio feedback for the correct answer (1 fanfare): the TA makes a volt and says, “Jippii! More stardust! $3+7=10$ and $10-7=3$ ”. Also, $10-7=3$ will be shown in a thought bubble. This scaffolding prepares the student for subtraction and the mathematical terms can be spoken and shown in different ways. A new amount of popplers and new number will appear and the student can continue immediately and in that way adjust the speed of the game to the own skills.

If the student gives a wrong answer a sound reminding 3 funny farts comes out (pruut, pruut, pruut). If the answer given is below the correct answer, the TA gives a hint by showing its hands with 10 fingers of which 3 will disappear: “Three popplers are out. (the popplers are jumping in the rhythm of 3) Hmm, do I need to count again? Please help me to choose a bigger number”. If the student gives a wrong answer that is above the correct answer, the TA gives a hint by showing its hands with 10 fingers of which 3 will disappear: “Three popplers are out. Hmm, do I need to count again? Please help me to choose a smaller number”.

Feedback from final presentation

The final presentation in Interaction 1 gave us some great feedback and we came up to some parts to consider and take decisions about. The most important questions to bring further into development are:

- When and how will our TA be integrated and which role will it take?
- Is the scroll wheel quick and easy enough to understand?
- Will the boxes take more time than needed when each new math assignment is played?
- How many times will the student be able to guess, what and when will the negative feedback be?

These questions have been discussed in Interaction 2, thus the first two is still not solved and the third needs to be tested. The boxes speed are improved from the feedback. The last questions first part is not really solved while we have the solution that when the answer is wrong the student get a new problem. The feedback is discussed in the Appendix B.

Appendix B

Vocally Manus – Interaction and Feedback

Allmänt från Agenten:

Variabel: **ok1**
Mening: "Okej"

Variabel: **ok2**
Mening: "Okej"

Variabel: **ok3**
Mening: "Okej"

Variabel: **hej1**
Mening: "Hallå!"

Variabel: **hej2**
Mening: "Hej hej!"

Variabel: **hej3**
Mening: "Halli hallå!"

Variabel: **hej4**
Mening: "Morsning korsning!"

Variabel: **hmm**
Mening: "hmm"

Variabel: **laggAv**
Mening: "Du, lägg av!"

Månbasen

Variabel: **introMessage1**
Mening: "Det här är min raket. Den består av fem olika delar."
Händelse: *Besöker månbasen för första gången*

Variabel: **introTop**
Mening: "Ett tak"
Händelse: *Besöker månbasen för första gången, Top blinkar*

Variabel: **introWindow**
Mening: "Ett fönster"
Händelse: *Besöker månbasen för första gången, Window blinkar*

Variabel: **introBody**
Mening: "Ett Chassi"
Händelse: *Besöker månbasen för första gången, Mid blinkar*

Variabel: **introBot**
Mening: "En motor"

Händelse: *Besöker månbasen för första gången, Bot blinkar*

Variabel: **introLeg1**
Mening: "Ett par ben"
Händelse: *Besöker månbasen för första gången, Leg blinkar*

Variabel: **introLeg2**
Mening: "Och Ett par ben"
Händelse: *Besöker månbasen för första gången, Legs blinkar*

Variabel: **introMessage2**
Mening: "Den är inte så fin som jag vill ha den. Skulle du kunna Hjälpa mig att pynta upp den lite?"
Händelse: *Besöker månbasen för första gången*

Variabel: **introStardust**
Mening: "För att byta ut delar behöver vi stjärnstoft. Jag samlar mitt stjärnstoft i påsen där uppe"
Händelse: *Besöker månbasen för första gången, Agenten tittar på påsen och vinkar med höger hand.*

Variabel: **introLeaveMoonbase**
Mening: "Tryck på antändningsknappen, så tar vi raketerna och samlar lite stjärnstoft"
Händelse: *Besöker månbasen för första gången, Agenten tittar på startknappen och vinkar med höger hand.*

Variabel: **tooLow**
Mening: "Typiskt! Vi behöver mer stjärnstoft för att måla om raketerna. Vi ger oss ut i rymden för att hämta mer!"
Händelse: *'Klicka paint' utan stjärnstoft.*

(Fler varianter)

Variabel: **partPainted1**
Mening: "Nu blev det Mycket bättre."
Händelse: *En bit målas om*

Variabel: **partPainted2**
Mening: "Vad fint!"
Händelse: *En bit målas om*

Variabel: **partPainted3**
Mening: "Wow!"
Händelse: *En bit målas om*

Variabel: **letsGo**
Mening: "Nu kör vi!"
Händelse: *Klicka på startbutton*

Poppler World - TA speaking

Variabel: **popplerIntro1**

Mening: "Vad händer här?"
Händelse: Agenten entrar moment 2

Variabel: **popplerIntro2**
Mening: "Vad håller ni på med?"
Händelse: Agenten entrar moment 2

Variabel: **boxes**
Mening: "WOHA?!"
Händelse: när lådorna faller ner över popplarna

Variabel: **moment2**
Mening: "Vad är det som händer egentligen? Kan du visa mig?"
Händelse: När högra lådan har blåst bort

Variabel: **moment3**
Mening: "Jag tror jag förstår, nu vill jag försöka."
Händelse: När agenten tar över

Variabel: **correction1**
Mening: "Hade jag fel? Kan du visa mig hur det ska vara istället?"
Händelse: När man rättar agenten

Variabel: **wrong1**
Mening: "Ojdå, det blev inte rätt. vi försöker igen"
Händelse: När agenten gör fel

Poppler World - Instructions (rocket)

Variabel: **scanning**
Mening: "Skannar planeten"
Händelse: Innan man kört första gången

Variabel: **welcome**
Mening: "Välkommen till popplarnas värld, Det verkar som vi har TIO popplare som väntar på bussen. Avvaktar."
Händelse: Innan man kört första gången

Variabel: **howMany**
Mening: "Hur många popplare finns i den andra lådan?"
Händelse: När högra lådan har blåst bort

Variabel: **howManyLeft**
Mening: "Det finns fler popplare kvar i lådan. Hur många då?"
Händelse: När man svarat ett för lågt tal

Variabel: **notThatMany**
Mening: "Det fanns inte så många popplare i lådan"
Händelse: När man svarat ett för högt tal

Variabel: **howManyNow**

Mening: "Hur många finns det i lådan nu?"
Händelse: När man svarat ett för högt tal steg2

Appendix C

Evaluation form for students in Swedish*

*This is just a suggestion and should be developed depending the focus of the evaluation and testing that will occur

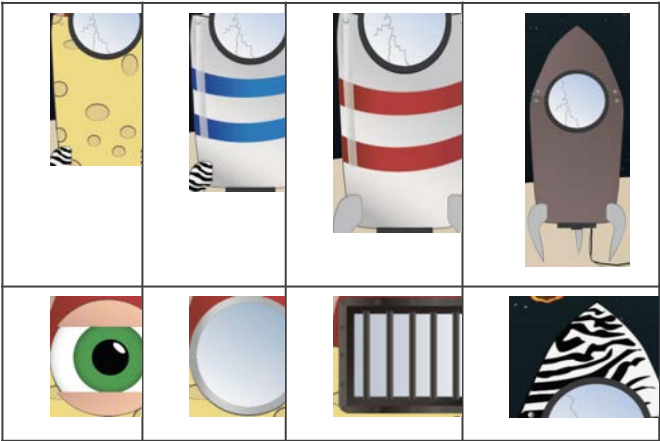
Ringa in om du är tjej eller kille

	
KILLE	TJEJ

Ringa in din favorit karaktär

		
add robot	add alien	

Vilken är din favorit Raket bit?



The effect of gesture-speech lag on comprehension

Sarah Alomari, Andrey Anikin, Joakim Bonnevier

sjawdat@hotmail.com, rty_anik@yahoo.com, joakim.bonnevier.213@student.lu.se

Background: *there is evidence that the time window for automatic integration of gestures and speech is narrow relative to the delays observed in natural speech. In particular, gestures often precede their lexical referents. We argue that the time is ripe to study the pragmatic effects of gesture-speech lag, since modern technology both makes such studies easier to conduct and increases the relevance of this question for human-machine interaction.*

Method: *60 participants reconstructed formations of geometric shapes based on six short instructional films, which could be in full sync or with a lag of ± 1.5 seconds between audio and video tracks.*

Results: *the total score on six tasks was significantly lower when the sound preceded gestures, but not vice versa. However, for participants with lower spatial abilities performance dropped even when gestures preceded sound. Those with better spatial abilities were able to compensate for this audio lag, but it appeared to take extra conscious effort, since there was a tendency for participants in the “delay” group to rate the task as somewhat harder and the instructor as less efficient.*

1 Introduction

It is widely accepted that gesturing serves a communicative function. Manual gestures facilitate speech production, which is evidenced by the fact that they persist when blind people speak among themselves (Iverson & Goldin-Meadow, 1998) or when the listener is not visible (Wagner, 2014). Furthermore, gestures may improve listening comprehension, especially when speech is ambiguous (Thompson & Massaro, 1986) or when there is a lot of background noise (Rogers, 1978). But how exactly are gestures temporally related to speech? How important is this temporal relation to normal communication? This is the point of departure for this study.

An influential view, first formulated by Kendon (1980) and then developed by McNeill (1985), is that “speech and gesture are two surface forms of a single underlying process” (Loehr, 2007, p. 182). In other words, gestures do not arise from speech or vice versa: the two simply share a common origin, namely Kendon’s (1980) “idea unit”. Their temporal relationship is then determined by the semantic synchrony rule and the pragmatic synchrony rule: if speech and gestures co-occur, they must either present the same semantic information or perform the same pragmatic function. Theoretically parsimonious, this theory does not necessarily explain all the details of speech and gesture use in natural communication. One controversy concerns the automaticity of gesturing. According to the “hand-in-hand” hypothesis, the production of gestures simply accompanies the production of speech; gestures reflect what is being said (Wagner, 2014). Alternatively, there may be a trade-off

between the two channels: speakers rely more on speech when gesturing is hard or ineffective, and when speaking gets harder, gestures are emphasized (de Ruiter, 2012). Compared to Kendon’s formulation, this view leaves more room for gestures and speech to be produced and processed relatively independently.

One robust finding is that gestures are generally synchronized with the accompanying speech or slightly precede it (e.g. Loehr, 2007; Treffner et al., 2008), albeit with a lot of variability in natural conversation (Morrel-Samuels & Krauss, 1992). Thus, according to a review by Nobe (2000), gestures and speech tend to be perfectly synchronized, but in 10% of cases the onset of gesture precedes speech by about 250 ms, i.e. the duration of one syllable. Furthermore, the preparation phase often anticipates speech. But a new question immediately arises: are they synchronized because this is necessary for successful comprehension or simply because speech and gesture, as claimed by Kendon and McNeill, stem from the same “idea unit”?

One way to answer this question is to see how comprehension is affected if we break down this natural synchronization. Obermeier and Gunter (in print) used event-related potential called “N400” to determine whether the integration of speech and gesture is automatic or effortful in different conditions, from perfect temporal synchronization to a considerable lag (up to 600 ms). Since speech and gesture exploit different modalities, this is a case of multisensory integration, which is known to happen automatically only when the two channels are fairly closely synchronized. That much said, the exact window within which a temporal mismatch is tolerated varies depending on the nature of stimuli (ibid). In this particular study the time window of automatic integration turned out to be unexpectedly wide: from approximately -200 ms (audio lag) to +120 ms (gesture lag). Notably, an audio lag was tolerated better than a gesture lag.

There is also some evidence from other studies which confirms that gestures are easier to integrate with speech if they are realized slightly before their verbal referent (Treffner et al., 2008). In another study N400 was greater to mismatched versus matched gesture-speech sequences only when speech lagged by 0 and 160, but not by 360 ms (Habets et al., 2011). The authors conclude that gesture and speech are integrated automatically when they fall within 160 ms of each other, so that a gesture which does not semantically match speech leads to effortful processing and evokes a larger N400. However, if gesture comes 360 ms before its verbal referent, this gesture is completely processed by the time speech is heard, so that effortful and presumably conscious processing is needed in any case, regardless of whether the gesture matches the verbal utterance or not.

The view which emerges from these studies is that

gesture and speech may be integrated either automatically or with some conscious effort, depending on how precisely they are synchronized. According to Habets et al. (2011), a delay of about 400 ms prevents speech and gestures from being automatically integrated. Presumably, at this delay the two channels are processed relatively independently and then may be fused into a single representation via high-level cognitive processing. This temporal delay, however, is well within the time frame reported for naturalistic conversations. For example, in a large study Morrel-Samuels and coauthors (1992) discovered that gestures were never preceded by their lexical affiliates in their data bank, but the onset of gesture usually preceded its lexical affiliate. In fact, the mean reported delay was 1 second, and for less familiar words it could be as long as 3.8 seconds! This result not only shows that gestures may be more readily accessible than lexical labels for complex concepts, but also emphasizes the simple fact that gesture-speech asynchrony is a normal feature of human conversation, for which humans must possess a compensatory mechanism.

An important question to ask pertains to the mechanisms of cross-modal integration, especially at longer delays. What exactly happens if gestures and speech are poorly matched temporally and fail to be integrated “automatically” back into a single “idea unit”? An influential position in psychology invokes the notion of “mental models” (Johnson-Laird, 1983) or “situation models” (Dijk & Kintsch, 1983), which accompany text-based representation. Situation models are the listener’s holistic representations of the described situation, which are integrated across sentences, modalities, sometimes even languages and multiple documents or conversations. Situation models also integrate the message with background knowledge and expertise (Zwaan & Radvansky, 1998). The temporal structure of messages is not always linear. Grammatical rules being what they are, the order of events in a narrative does not always correspond to the order in which they are mentioned in a sentence: for instance, we may say: “Before I opened the door, I had to search for my keys for a few minutes”.

Seen in this light, a gesture-speech lag of a few seconds leads to a special case of integrating information arriving from different modalities and at different times into a unified situation model. It may require a conscious effort, but it is a task that humans perform all the time, and there is no a priori reason to believe that it will present an insuperable barrier to comprehension. Historically, the practical implications of gesture-speech lag have remained relatively unexplored, partly due to methodological problems with generating naturalistic sequences with mismatched speech and gestures. In particular, lip movements quickly give the manipulation away, unless the face is hidden or computer animation is used to separate facial from bodily movements. Such technology has become available relatively recently. The same technological advances also make the practical implications of gesture-speech synchronization more relevant today, when digital agents are becoming increasingly common as chatterbots or virtual salesmen. Studying the effects of gesture-speech (de)synchronization appears to be an essential part of the effort to make computer-human communication smooth and effortless. If very precise synchronization improves comprehension and likability of the computer-generated character, temporal alignment should

be prioritized by developers. On the other hand, just because humans produce gestures slightly before or simultaneously with the corresponding speech units, it does not necessarily follow that this is always the optimal strategy from the listener’s point of view. In fact, perfect synchronization may even turn out to be suboptimal in certain conditions, in which the verbal and visual channels may best be exploited sequentially rather than in parallel. These questions remain entirely unexplored, but the answers could drastically change the way digital agents speak and move.

We suggest that the method of manipulating the temporal alignment of speech and gestures presents a promising new avenue for investigating the communicative function of gestures, and that a purely pragmatic approach to measuring the effects of gesture-speech asynchrony allows us to raise and solve an important question of both theoretical and practical significance: how does comprehension depend on the temporal lag between gestures and speech? Two alternative research hypotheses present themselves:

- a) If automatic integration of gesture and speech is the norm, their temporal misalignment can be seen as a special case of gesture-speech mismatch with a large negative effect on comprehension.*
- b) If gesture and speech are attended to as two independent channels, it should be possible to compensate for their temporal misalignment and tolerate a large delay without a noticeable effect on comprehension.*

A number of studies have already demonstrated that omitting gestures or presenting the listeners with semantically incongruous gestures negatively affects comprehension and memory. For instance, Cassell (1998) reports that mismatched gestures caused many more inaccuracies in retelling a story compared to matched gestures, and in a study by Goldin-Meadow et al. (1999) children were better at reiterating the teacher’s speech when it was accompanied by matching rather than non-matching gestures. We can take methodological inspiration from these studies but look at the effect of a temporal delay rather than semantic incongruity between gestures and speech.

We know that gestures may precede their lexical affiliates by 2000 to 0 ms (Morrel-Samuels & Krauss, 1992); we also know that a lag of over 200 ms is likely to prevent the two modalities from being integrated automatically (Habets et al., 2011). From these two findings it follows that humans routinely integrate gestures and speech with a varying degree of conscious effort, suggesting that hypothesis (b) may be correct. But is it possible to use gestural information at all if the temporal lag is in the atypical direction, i.e. if speech precedes gestures and not vice versa? Is the amount of information available after automatic integration of well-synchronized gestures and speech comparable to the amount of information available after more effortful conscious integration? How is performance on a task affected by the need to do this integration? A simple way to find out would be to design a practical task to which the auditory and visual channels contribute equally – a task which can only be solved if the participant attends to both. If we then manipulate the lag between the two channels, the change in performance will provide the most direct measure of how successfully speech was integrated with gestures at different time lags.

2 Methodology

Several tests were initially evaluated in a series of pilot trials. The one chosen for the experiment presented participants with the task of arranging geometric shapes based on a short instructional video, in which an instructor used gestures but did not show the objects. The gestures produced by the instructor in this task were relatively focused and short in duration, making their temporal synchronization with the accompanying speech clearly relevant to comprehension. In addition, this task could be performed incrementally, and missing a single step of the instructions did not preclude successful completion of later steps. Finally, it was relatively straightforward to assess performance by comparing each reconstructed array with a picture of the original.

Experimental tasks

Participants were either students recruited and tested at the University of Lund or personal acquaintances of the authors. Approximately every third participant was compensated with a cinema ticket.

At the beginning of the experiment subjects were asked to evaluate their ability to read maps (a skill judged to be functionally similar to the demands of the main test). As a practical pre-test, they also had to arrange small pieces of paper (“furniture”) in the drawing of a room based on verbal instructions (Fig. 1). The instructions were read by the experimenter slowly, but without repetitions, and the resulting arrangement was informally assessed on a scale of 1 (poor) to 3 (good). After this the subjects began the main experiment, which consisted of six tasks. In each task the participant was asked to reconstruct a formation of five geometric shapes after watching an instructional video. The shapes had to be selected from an array of 8 objects: two boxes, two sticks, a ball, a can, a tube, and a small cylinder (see Appendix). Participants were instructed to watch the instruction videos first and to build afterwards, so that they would not have to divide their attention between the videos and the objects. The reconstructed array was photographed for future coding, and the participant proceeded to the next task. Sometimes the participants could not recall all five objects; in such cases we did not press them to pick the last object at random but accepted incomplete arrays of four or, occasionally, three objects.

Three conditions were tested: unmanipulated film (“sync”), 1.5 s audio advance (i.e. the sound came first and the gestures 1.5 seconds later; “a15”), and 1.5 s audio delay (first gestures, then sound; “d15”). The value of 1.5 seconds was chosen based on the evaluation of the pilot study, where subjects sometimes commented that something was wrong with the video if the delay was 2 seconds or larger. At the same time, some subjects in the pilot study managed to perform very well even with a large delay, and shorter delays of <500 ms did not appear to have any effect, making it interesting to push this value as high as possible without revealing the manipulation. All experiments were double-blind: neither participants nor experimenters knew which condition was being tested.

After completing six tasks, participants filled out a short questionnaire, rating the difficulty of experimental tasks and the efficiency of the instructor. They were also encouraged to leave free-text feedback. Finally, each participant was

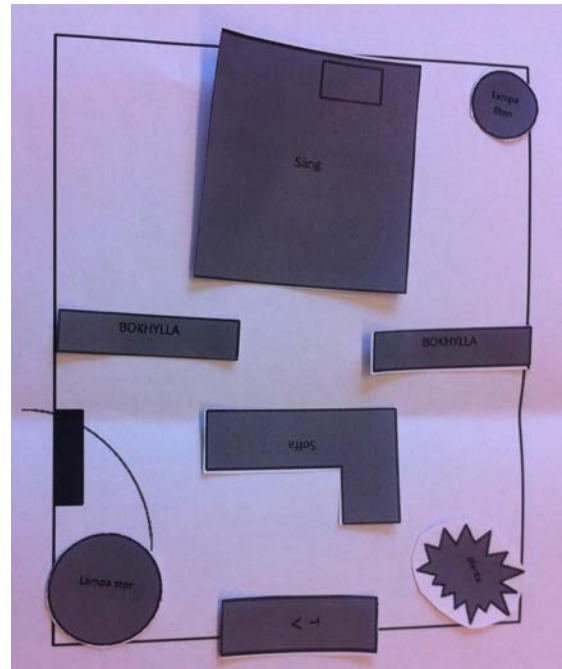


Fig. 1 A perfectly reconstructed “room” in the “furniture” pre-test.

debriefed and asked whether they had noticed anything strange about the video and sound. If they did not report noticing anything unusual, they were then asked directly whether the video- and soundtracks were in sync. The entire procedure took 15-20 minutes.

Instructional videos

Six short instructional videos were filmed by a hand-held camera, which was placed high over the shoulder of the demonstrator so as to provide an unobstructed view of his manual gestures but not the face. The demonstrator was a male student, who was told that the focus of this study was the effectiveness of communication but did not know that we were interested in the temporal relation between gestures and speech. He was not specifically asked to use gestures but simply encouraged to describe what needed to be done “as well as he could”. For each task, a picture of the target formation was shown on the screen of a smartphone, which the instructor kept in his lap (off camera) while explaining how to build it.

The recordings were then split into separate video- and soundtracks, which were presented to experimental subjects in PsychoPy (Peirce, 2007, 2009) in a randomly chosen condition (0, +1.5 or -1.5 sec). Each soundtrack was converted to a sample rate of 44100 Hz, filtered to remove background noise, and normalized. In case of mistakes or unwanted noise, such as the sound of a hand slapping the desk, a new take was filmed. The final videos contained no background noise and few cues that could allow the subject to detect an audio lag. The videos were between 42 and 82 seconds in length. All participants in pilot studies confirmed that they could hear the instructions clearly and were not bothered by the slightly unorthodox filming angle.

Coding

The “furniture” pre-test was coded informally by one of the

authors (SA) for all subjects. All six main tasks were coded independently by two of the authors (AA and JB) based on a carefully defined and pre-tested algorithm, which awarded points based on the correct choice of each object as well as its position in two dimensions, three dimensions, and in relation to the reference object mentioned in the instruction to indicate the location of the object in question (see Appendix). A maximum of 19 points could be awarded for each of the six tasks, amounting to a maximum total of 114 points per subject. The coders were, as far as possible, blind to the experimental condition, although it proved difficult to be consistent in this regard because of the need to monitor the numbers of participants in each condition and to debug the program. Any disagreements in coding were discussed by the two coders, and then either a compromise solution was reached or, when the decision appeared to be purely subjective, two different scores were entered in the database.

Statistics

All statistical analyses were performed in R (R Development Core Team, 2008). The tasks varied in their difficulty, and therefore it would not have been meaningful to do within-subject comparisons. Accordingly, each participant performed all six tasks in only one condition (between-subject design).

The *primary endpoint* (outcome measure) was implicit comprehension - the total score on all six experimental tasks, i.e. the ability to perform the tasks based on instructions. Each participant could receive maximum 114 points on all six tasks from each coder. Since some subjects did score close to this ceiling, it was judged inappropriate to apply a linear regression and the scores were treated as proportions. At the same time, simply adding the scores from both coders would double the number of observations and thus lead us to underestimate standard errors in binomial models. Accordingly, the scores from two coders were averaged and rounded to the nearest integer; the weights in the model were thus all equal to 114. Since on most occasions scores given by individual coders differed by 1 point, this rounding is conservative in that the higher of two numbers becomes the effective score used for analysis. This outcome variable was then modeled using generalized linear models (GLM) of the binomial family. Differences between the six experimental tasks were explored using a generalized linear mixed-effects model (GLMM) from the lme4 package (Pinheiro et al., 2014).

Secondary endpoints were explicit measures of comprehension and satisfaction with the instructor in the experimental videos. Participants were asked four questions pertaining to the difficulty of the tasks and starting with “Was it hard to understand...”:

- “...the language?”
- “...which shapes to use?”
- “...how the pieces were in relation to each other?”
- “...what to build?”

They were also asked to evaluate the instructor in three questions starting with “Was the instructor...”:

- “...clear?”
- “...self-assured?”
- “...professional?”

All secondary endpoints were measured on a Visual Analogue Scale (VAS) from 0 to 100 in PsychoPy. Note that

in this case binomial distribution would not be appropriate, since the score represents a freely chosen number, rather than the number of success on many separately graded sub-tasks. These variables were therefore assumed to be continuous and analyzed using linear regression.

Self-evaluated ability to read maps was measured on a VAS (0 to 100) and treated as a continuous predictor. The score on the “furniture” pre-test was coded manually on a scale of 1 to 3 and treated as nominal

3 Results

A total of 60 participants (29 females and 31 males) were enrolled in the experiment: 17 in the “advance 1.5 sec” [a15] condition, 23 in the “delay 1.5 sec” [d15] condition, and 20 in the “sync” condition. There were no significant differences between experimental groups in baseline characteristics. Chi-square tests demonstrated that all three groups were similar in terms of gender composition ($\chi^2(2, N = 60) = 0.84, p > .5$) and the score on the “furniture” pre-test ($\chi^2(4, N = 60) = 1.52, p > .5$). Linear regression of self-assessed ability to read maps also failed to discover any effect of condition ($F(2,57) = 0.09, p > .5$). Total scores awarded by both coders were very strongly correlated, demonstrating high inter-rater reliability (Spearman’s $r = 0.994$).

Primary endpoint: implicit comprehension

Implicit understanding of the instructional videos was assessed by comparing each reconstructed array with the original and adding up the scores on all six tasks. Individual variation of the total score per subject proved to be very considerable, but overall the level of success was high: mean absolute score ($\pm$ SD) was 92.2 ($\pm$ 11.2), or 80.8% ($\pm$ 9.8%). The distribution of raw scores in different groups is shown in Fig. 2. Linear regression failed to find a significant main effect of condition ($F(2,57) = 1.30, p = .28$). However, a logistic regression fit for the total score per subject and weighted by the maximum possible score (114) revealed a significant main effect of condition (likelihood ratio test: $L = 17.9$; $df = 2, p < .001$). More specifically, the total score was lower in the “a15” group compared to “sync” group (Wald test: $z = -3.86$; $p < .001$) and compared to “d15” group ($z = 3.57, p < .001$), while there was no difference between “sync” and “d15” groups ($z = -0.44$; $p > .5$).

It is quite likely that performance on the experimental tasks is strongly affected by the individual spatial abilities of

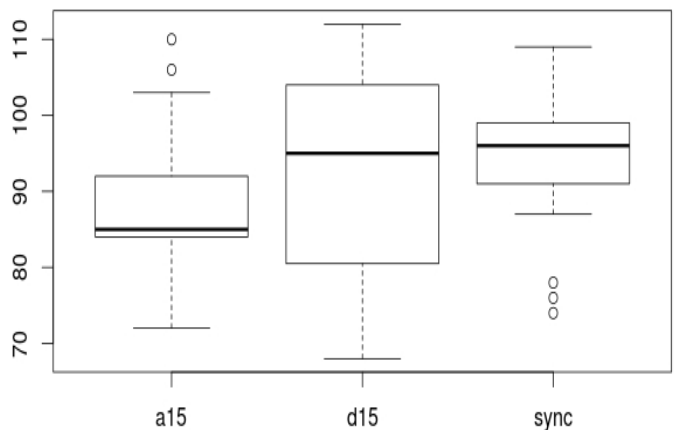


Fig. 2 Boxplots of raw total scores for each group

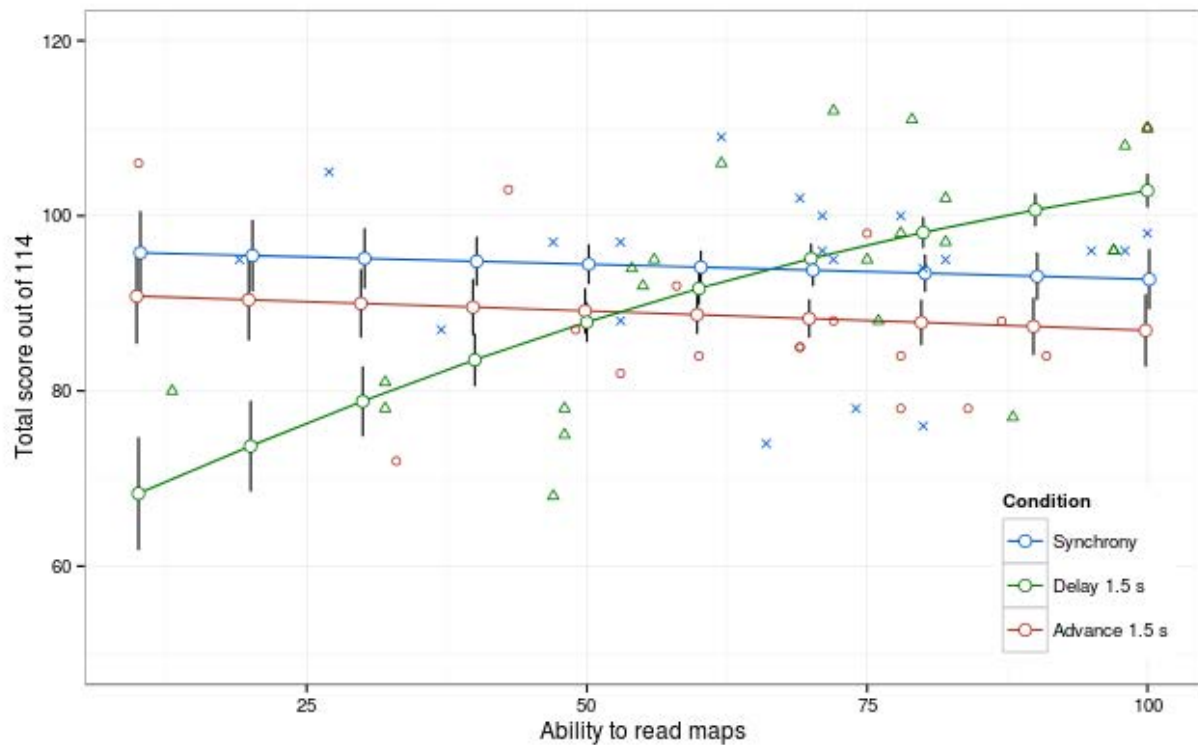


Fig. 3 Predicted total scores per subject from a logistic model with three terms (score, maps, and their interaction) over the range of reported self-assessed ability to read maps, with 95% confidence intervals. Raw data is overlaid as points marked by shape and color, using the same colors as for regression lines.

each subject, and the effect of condition may depend on these abilities. We therefore investigated possible interactions between condition and each of two measures of underlying spatial ability: a direct question (“How do you evaluate your ability to read maps?”) and the score on a pre-test, in which the participant had to arrange furniture based on verbal instructions.

Self-evaluated ability to read maps proved a strong and highly significant predictor of success. The inclusion of this variable in the model did not influence the effect of condition, i.e. the performance in the “a15” group was still lower, even controlling for possible differences in spatial ability. However, in a model with interaction between condition and the ability to read maps the interaction term was highly significant (likelihood ratio test: $L = 131.6$, $df = 2$, $p < .001$). For those with average self-evaluated spatial abilities, the effect of condition is predicted to be approximately the same as in the model without interaction (lower scores in the “a15” group and similar scores in the other two). On the other hand, scores in the “d15” condition are predicted to be high for people with above-average spatial abilities and much lower for people with poor spatial abilities. This model is presented visually in Fig. 3.

The left part of the graph should not be taken too literally, since very few subjects reported such low spatial abilities. What the model does suggest, however, is that the effect of delaying speech with respect to gestures leads to deteriorated

performance only in those subjects who are likely to find the task difficult, but not in those who are good at solving such problems. On the other hand, the effect of delaying gestures with respect to speech was equally deleterious for all participants, whatever their spatial abilities. This interaction effect with spatial abilities can also account for the extreme dispersal of results in the “d15” condition compared to the other two conditions (see Fig. 2).

To proceed to the “furniture” pre-test, 6 participants scored 1 (“poor”), 17 scored 2 (“moderate”) and 37 scored 3

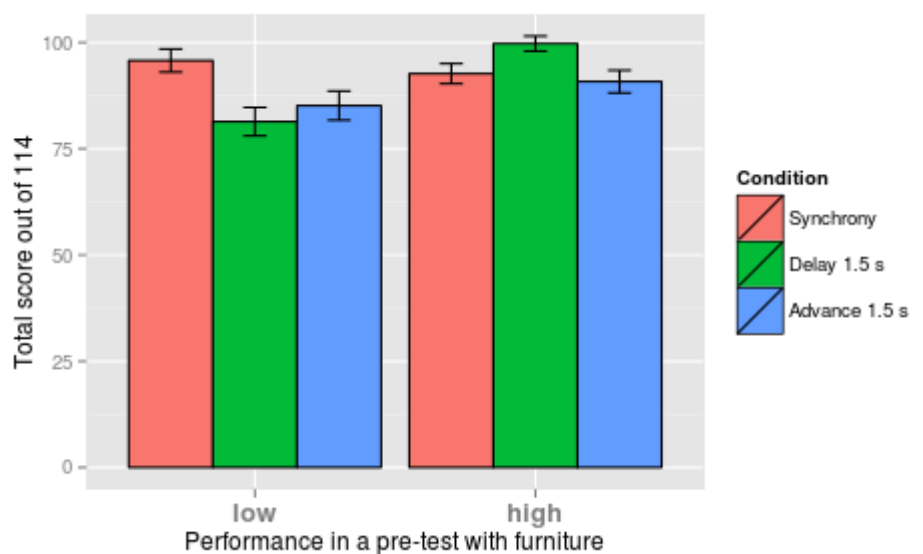


Fig. 4 Total score predicted for subjects in each experimental group with low vs high spatial ability as assessed by a pre-test with furniture. Error bars show 95% confidence intervals. Based on a logistic model for total score with three predictors: condition, score on the “furniture” task coded as “high” or “low”, and their interaction.

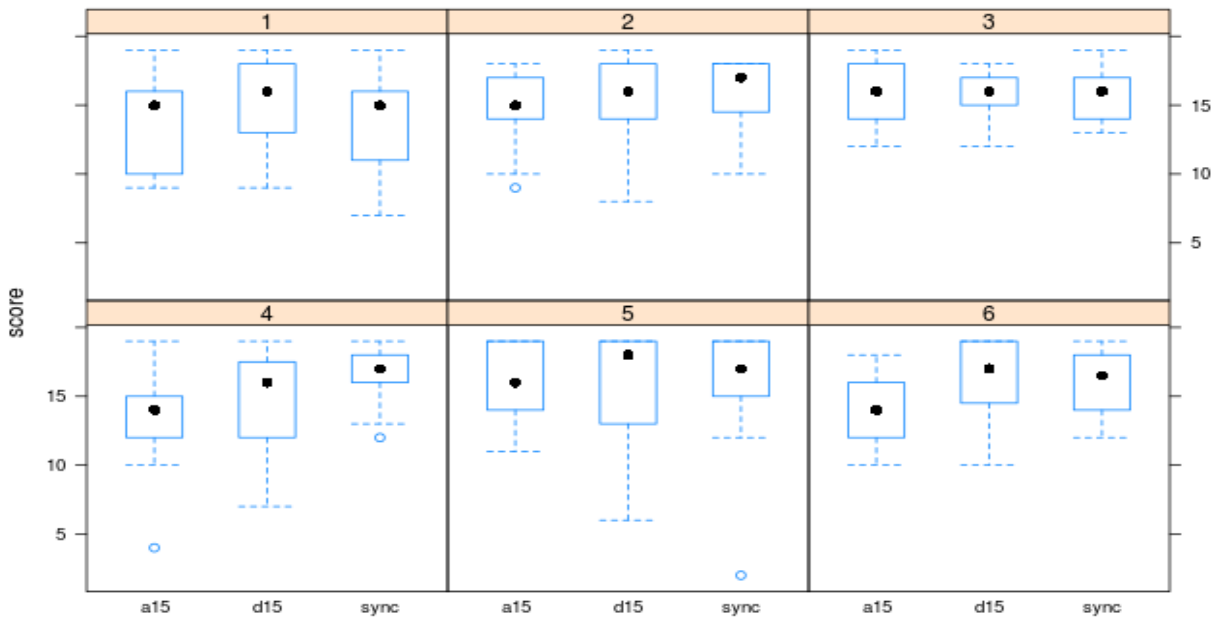


Fig. 5 Boxplots of raw total scores for each of the six experimental tasks, plotted by condition (n=60 for each task).

(“good”). Explorative analysis did not reveal any interesting differences between those who scored 1 and 2 in terms of the effect of experimental condition, and these two groups were pooled together. A multiple logistic regression was then fit to predict the total score based on condition, the score on “furniture” pretest (redefined as high/low), and their interaction. The predictions of this model are shown in Fig. 4. As we can see, among the group of participants who performed less well on the “furniture” pre-test, performance on six experimental tasks was significantly better in the “sync” condition compared to “d15” and “a15” (mean [95% CI] = 95.8 [93.0; 98.5], 81.4 [78.0; 84.7] and 85.1 [81.7; 88.6], respectively). On the other hand, among those who performed well on the “furniture” pre-test, scores were similar in the “sync” and “a15” condition and even somewhat higher in the “d15” condition (92.7 [90.3; 95.0], 99.7 [97.9; 101.5] and 90.8 [88.1; 93.5] for “sync”, “d15” and “a15”, respectively).

Another interesting aspect of the distribution of scores is that, if we fit separate models predicting scores on individual tasks rather than the total score per subject, the effect of condition is only found for three tasks out of six: 1, 4 and 6. For task 1, subjects in the “delay” condition scored higher than those in the other two groups (likelihood ratio test: $L = 8.5$, $df = 2$, $p = .014$). In task 4 and task 6 scores were lower in the “advance” condition ($L = 28.9$, $p < .001$ and $L = 11.9$, $p = .003$, respectively). Multiple comparisons of this sort may not be very reliable. However, in a GLMM fit for the score on each task as a function of condition, task and their interaction and nested for each participant, the interaction term was highly significant ($L = 36.3$, $df = 10$, $p < .001$), again strongly suggesting that the effect of condition was not the same for different experimental tasks. Tasks 4 and 6, in particular, stand out (Fig. 5).

Secondary endpoints: explicit comprehension

Exploratory linear regression models showed that there was a weak tendency to have higher scores on questions pertaining

to the difficulty of the tasks in group “d15”. All four questions presumably tapped into the same concept and showed the same tendency, so the scores were added up and modeled using linear regression (Fig. 6). No statistically significant differences were discovered between any of the groups, with or without controlling for individual spatial abilities, i.e. the score on “furniture” pre-test and the self-reported ability to read maps ($p > .5$ for all comparisons). The perceived difficulty of the experiment in the “sync”, “d15” and “a15” groups was (mean [95% CI]) 130.6 [100.1; 161.1], 152.8 [124.3; 181.2] and 137.3 [104.2; 170.4], respectively. On the other hand, as expected, those who reported having problems with maps evaluated the task as somewhat harder ($F(1,55) = 3.43$, $p = .07$).

As for the three questions that asked participants to evaluate their instructor (clarity, self-assuredness and professionalism), all three demonstrated a tendency to give lower scores in “d15” compared to the other two groups. Again, the scores on these three questions were added up and analyzed using linear regression. The effect of condition in this case was present but not very strong ($F(2,56) = 3.53$, $p = .04$), and the most marked difference was the lower rating that the instructor received in the “d15” group (Fig. 7). The instructor’s overall rating was (mean [95% CI]) 215.5 [181.8; 249.1] in the “sync” group, 155.5 [123.4; 187.6] in the “d15” group, and 197.5 [161.0; 234.0] in the “a15” group. The ability to read maps had no influence on how the instructor was evaluated.

Qualitative data: interviews, comments, observations

After each explicit comprehension evaluation the participants were given an option to comment on the tasks and latter the instructor through two separate free-text boxes. One test person can choose to comment in either one or both text boxes, hence, one comment or comment space is counted as a single entity. The comments were then compared to each participant’s condition, self-evaluation, pre-test, and final score to observe potential correlations. The comments were

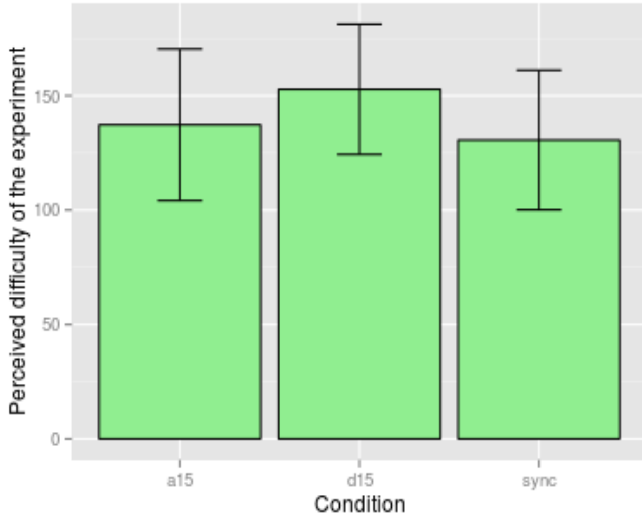


Fig. 6 The difficulty of experiment, as evaluated by subjects in three groups: the sum of four related ratings (each 0 to 100) with 95% CIs. The ability to read maps is assumed to be average (grand mean for 60 subjects).

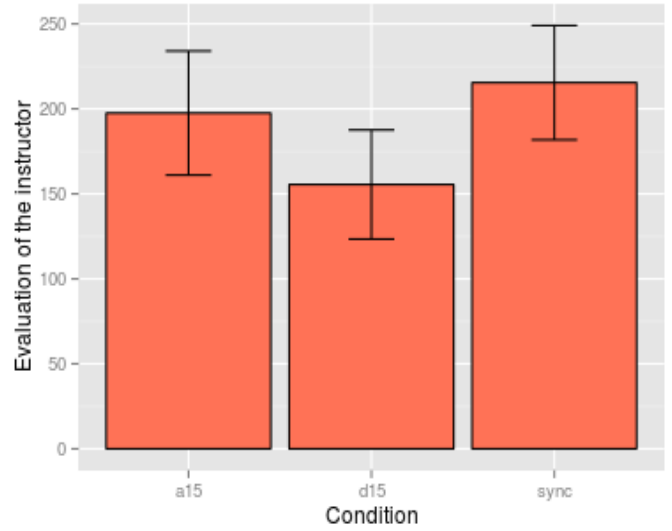


Fig. 7 An evaluation of the instructor by subjects in three groups: the sum of three related ratings (each 0 to 100) with 95% CIs. The ability to read maps is not a significant predictor and so is absent from the model.

organized into subgroups of tone nuance differences and comment direction (towards the self, the instructor, or the task) for pattern investigation. A single comment can include several references to both direction and keywords.

Condition	Comments	No comment
sync	7	31
d15	26	18
a15	16	20

Full synchronization (sync): 7 out of 38 comment possibilities were used in this particular condition. The existing comments show tendencies of self-criticism or the instructor's usage of different terminology for the same object. Clear references are made to the own memory as a potential causality factor in 4 comments though no direct comments of frustration or criticism were directed towards the instructor.

Audio delay 1.5 sec (d15): in this particular condition comments tend to explicitly pronounce difficulties regarding the given tasks and the unclarity and insecurity in what the instructor conveys. 26 out of 44 comment possibilities were used to explicitly express opinions about the test. Tendencies of experiencing difficulties, thus directing comments towards the instructor are very dominating.

Audio advance 1.5 sec (a15): 16 out of 36 comment possibilities were used in audio advance condition. The participants explicitly express insecurity regarding the tasks

and instructor and had difficulties in deciding whether the criticism should be directed toward themselves and their comprehension of the task and instructor, or towards the task and instructor itself, displayed but rather even results in direction, keeping a mild or moderate tone in all except for in 2 cases. Furthermore, tendencies of experiencing effort in completing the tasks are noticeable in the verbal comments.

Generally, no clear correlations to final score can be made in any condition though significant effort tendencies were explicitly most obvious in condition d15, as table 3 states. A15 was given a noticeable effort creating an experienced sense of insecurity regarding difficulty causality while the participants in synchronized condition did not tend to experience a need to explicitly expressing any opinion.

Conclusion

- Further evidence of d15 being "harder" effortwise from an explicit point of view
- No direct explicit references found to spatial abilities more than already proven through test results.

4 Discussion

Previous studies of speech-gesture synchronization suggest that their automatic integration is only possible within a narrow time window of no more than a few hundred milliseconds, but also that larger delays are relatively common in human conversation. Does this mean that it is

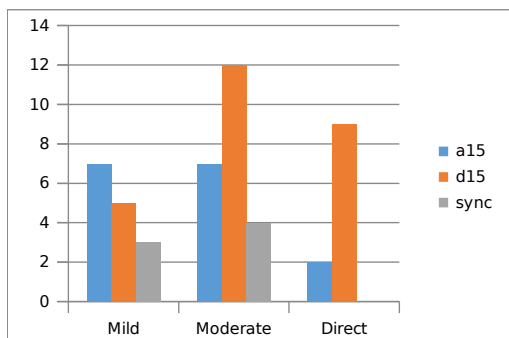


Table 1. Comment nuances

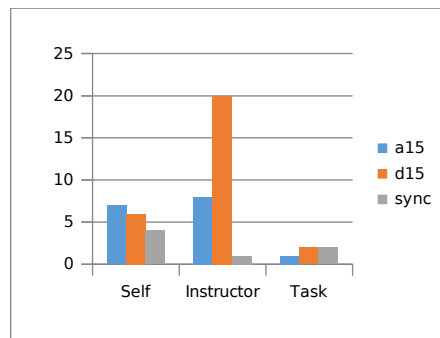


Table 2. Comment direction

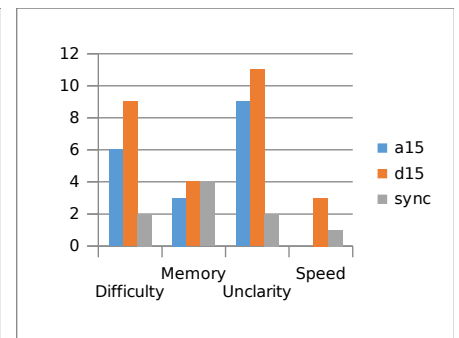


Table 3. Keyword references

subjectively more difficult to attend to gestures that are out of sync with the corresponding speech? Does comprehension suffer as a result? To address these questions, we performed an experiment, in which 60 participants performed 6 practical tasks after watching instructional videos either fully in sync or with sound coming 1.5 seconds before (“a15”) or after (“d15”) the video.

Since it is well established that gestures tend to slightly precede their verbal referents, we expected that the “a15” condition would prove more challenging for the subjects. However, we did not necessarily expect to see a strong effect of a lag in either direction on overall comprehension, believing that compensatory mechanisms might enable the subjects to overcome the problem. On the other hand, if extra conscious effort was required to integrate gestural and verbal information in case of a large delay, it appears plausible that participants in “a15” and “d15” conditions would rank the task as harder and the instructor as less competent.

To begin with subjective measures, no differences were discovered between experimental groups, except for a small but consistent tendency for more negative assessments in the “d15” group (Figures 6 and 7). On the contrary, an objective measure of implicit comprehension - the total score on all 6 main tasks - showed a clear effect of condition in the predicted direction: participants in group “a15” scored lower, but there was no difference in the overall score between “d15” and “sync” groups.

Clearly, individual variation in spatial abilities may strongly influence the results. Indeed, we discovered highly significant interaction between both measures of spatial skills (self-evaluated ability to read maps and performance in the “furniture” pre-task) and the effect of experimental condition. Participants with good spatial skills in “d15” group, it appears, were able to fully compensate for the temporal mismatch when gestures came before speech (just as so often happens in real-life conversations), while those with poor spatial skills were unable to compensate and performed worse compared to participants in the “sync” group. It is a tantalizing possibility (although the evidence is inconclusive) that participants in the “d15” group evaluated the task as harder and the instructor as less competent because they had to make a conscious effort to overcome the delay and to be able to use gestural information. On the contrary, the situation when speech comes before gestures (group “a15”) appears to be equally disruptive for all subjects, whatever their spatial abilities. In this group participants reported that the task was approximately as hard and the instructor as competent as did subjects in the “sync” group, possibly because participants in the “a15” group could make no sense of the instructor’s gestures, “gave up” and simply stopped trying to attend to the video, relying purely on the verbal part of instructions.

In this respect, it would be very interesting to see whether the performance of participants in the “a15” condition is still better than what would be expected without any gestures at all. It seems likely that the verbal part of instructions becomes the only source of information in this condition, in which case subjects presented with the same instructions without any accompanying video would perform at the same level. An extension study is currently under way to explore this issue, and we expect the results to become available early in 2015. Another interesting consideration is the nature of the tasks and the gestures used by the instructor. Since

there was such a clear difference between conditions for some of the six tasks, but not for others, it appears fruitful to investigate the instructional videos more closely and look for systematic differences between the tasks, e.g. the nature of gestures used by the instructor. Furthermore, given the high natural variability in gesture-speech temporal coordination, it is highly probable that the instructor’s gestures were not perfectly synchronized with their lexical referents in the original unmanipulated videos. This means that even in “sync” condition some gesture-speech pairs may have been less than perfectly synchronized, possibly further contributing to the differences between the six experimental tasks. These issues are also going to be explored fully as the study is being prepared for publication.

A possible limitation of the tasks used in this study is that performance is fairly high, making it harder to notice differences between conditions. Statistically speaking, it would be desirable to make the tasks harder and thus to lower the average performance, although it might have a devastating effect on the motivation of participants (who already complained that the instructions were hard to memorize and to follow). One final consideration: the tasks were constructed so as to invite short, precisely timed gestures, such as showing on which side of the box a stick should be placed. On the one hand, this is an advantage insofar as a lack of temporal synchronization in such contexts is likely to have measurable effects on performance. But on the other hand, this may limit the extent to which we can observe the full human power for integrating gestural and verbal channels: such gestural information simply becomes irrelevant after a delay of 1.5 seconds. With richer, more realistic and less structured tasks, such as demonstrating complex bodily movements involved in performing magic tricks or assembling pieces of furniture, it is possible that observers may still utilize gestural information even after extremely long delays, treating the two channels as relatively independent. The data from our pilot studies of magic tricks indicates that this may indeed be the case, since even very large delays (up to 4 seconds) did not prevent participants from attending to gestures.

The precise mechanisms of gesture-speech integration may thus vary depending on the informational requirements and the nature of activity being demonstrated. In the current study we report tolerance for a lag of 1.5 seconds in the audio channel, but not in the video channel, moderated by individual spatial abilities. It remains to be determined how far these results hold for other contexts and what the precise threshold values for a lag in either direction might be.

Conclusions:

1. Explicit measures assessed with a questionnaire were not significantly affected by delaying or advancing the soundtrack by 1.5 seconds. There are some indications that subjects found the tasks harder and the instructor less clear when the soundtrack was delayed, although the evidence for this is inconclusive.
2. Performance was worse when the audio came first and the gestures 1.5 seconds later.
3. When gestures came first and the audio 1.5 seconds later, performance dropped for those with poor self-reported spatial abilities, but those with good spatial abilities were able to compensate for this delay, perhaps with some conscious effort.

References

- Goldin-Meadow, S., Kim, S., & Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology*, 91(4), 720.
- Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience*, 23(8), 1845-1854.
- Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, 396(6708), 228-228.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.
- Kendon, Adam (1980). Gesticulation and speech: Two aspects of the process of utterance. In Mary Ritchie Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207-227). The Hague: Mouton.
- Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7(2), 179-214.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological review*, 92(3), 350.
- Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(3), 615.
- Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? *Language and gesture*, 2, 186.
- Obermeier, C., & Gunter, T. C. (in print). Multisensory Integration: The Case of a Time Window of Gesture-Speech Integration. *Journal of Cognitive Neuroscience* X:Y, pp. 1-16.
- Peirce JW (2009) Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* 2:10. doi:10.3389/neuro.11.010.2008
- Peirce, JW (2007) PsychoPy - Psychophysics software in Python. *J Neurosci Methods*, 162(1-2):8-13.
- Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2014). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117, <URL:http://CRAN.R-project.org/package=nlme>.
- R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rogers, W. T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. *Human Communication Research*, 5(1), 54-62.
- Thompson, L. A., & Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology*, 42(1), 144-168.
- Treffner, P., Peter, M., & Kleidon, M. (2008). Gestures and phases: The dynamics of speech-hand communication. *Ecological Psychology*, 20(1), 32-64.
- Van Dijk, T. A., Kintsch, W., & Van Dijk, T. A. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209-232.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2), 162.

Appendix 1: coding algorithm

For each of 5 objects, answer 4 questions:

1. Is the object part of the original array*? Max 1 point. (Exception: if 2 sticks are used instead of one, it's worth only 1 point.)
2. Is it correctly placed in 2D, i.e. more or less on the right spot on the desk (as if the desk was a coordinate space seen from above)? Max 1 point.
3. Is it correctly placed in relation to the surface of the desk, i.e. in 3D? For instance, is it standing vertically on its long/short side, lying, leaning / lying on another object, etc. Max 1 point.
4. [Except for the first object in each task] Is it correctly placed in relation to the reference object mentioned in the instruction (column E)? The exact nature of this relation is specified in column E based on the instruction. Max 1 point.

* If a wrong object is chosen, points can still be given for correct placement in relation to other objects, in 2D / 3D, etc. Ex: taking the rectangular instead of flat box does not lead to all zeros for this object.

Comments on tricky coding decisions

Task 1:

- the can should be noticeably lower than the box. Otherwise take off 1 point: 2D

Task 2:

- if the box touches the end of the stick with its flat end (rather than the sharp side), take off 1 point for that (relational)
- the can should touch the end of the stick. Otherwise take off a point for relational (if very far, also a point for 2D)

Task 3:

- the opening of the tube should face the observer. If not, take off 1 point (3D)
- the orientation of the box (long side left-to-right) is shown ambiguously through gestures and not mentioned orally. Don't take off points if it's rotated by 90°
- if the box is on the wrong side of the ball, take off 1 point (2D)

Task 4:

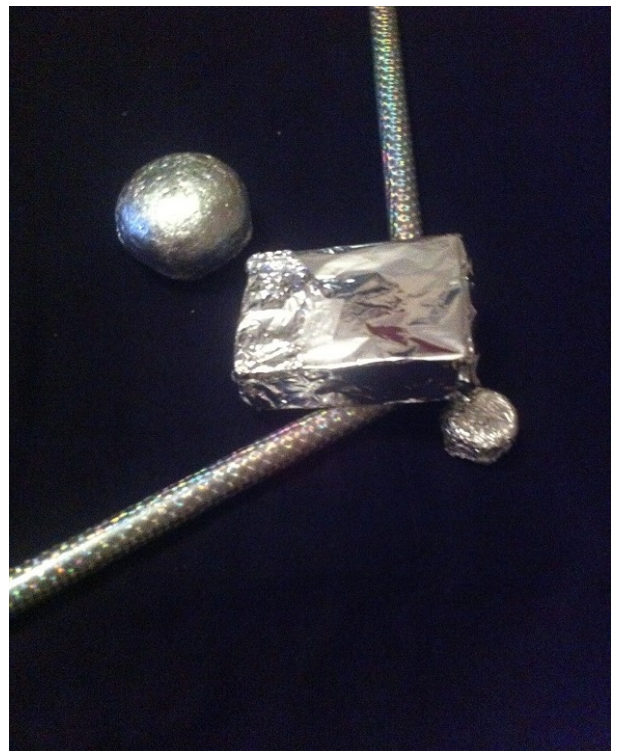
- stick_vertical may lie towards the right end of the box, don't take off points for that (not mentioned in the instructions)

Task 6:

- the orientation of the box is shown clearly (unlike in task 3). If it's rotated by 90°, take off 1 point (3D)
- if the tube is on the wrong side of the box, take off 1 point for 2D (but not for 3D or relational)



a



b

The original picture for task 4 (a) and a reconstruction by a participant in group “a15” (b). This reconstruction was given 10 out of 19 points.

Quieasy- ett digitalt testverktyg för grundskolan

Bernt Christensen, Kalle Palm, Niklas Sjöberg, Olga Fredriksson

bernt.e.christensen@gmail.com, kalle.palm@gmail.com, adi10nsj@student.lu.se, deograciass@me.com

I denna rapport presenterar vi vårt arbete med att utveckla en plattform för att göra digitala test. Plattformen är till för lärare i grundskolan, som vill genomföra kortare läxförhör på individnivå eller ge en test som underlag för klassdiskussioner. Målet med sådana test är att vara tidsbesparande för lärare, kunna användas i flera ämnen och innehålla möjlighet till formativ bedömning. Vi har genomfört intervjuer med både gymnasielärare och högstadielärare för att ringa in synpunkter kring sådana digitala test. Vi har vid två tillfällen testat våra LoFi-prototyper på ett flertal lärare på en högstadieskola samt utarbetat en HiFi-prototyp som vi också har testat under två dagar på lärare och elever. Den första dagen skapade fem lärare i olika ämnen var sitt test med tillhörande feedback. Två av dessa lärare genomförde samma test även i pappersformat. Följande dag kunde två elever från respektive klass genomföra dessa tester. Våra deltagare uttryckte tillfredsställelse med flera funktioner i vårt digitala test samt med möjlighet till återkoppling och, följaktligen, formativ bedömning.

1 Introduktion

Bakgrund

Utvecklingen av informationsteknologi och förbättrad digital infrastruktur har de senaste decennierna lett till att antalet datorer per elev i skolan ökat och att fler skolor får möjlighet att använda sig av digitala verktyg för bedömning (Skolverket, 2013). Utveckling och implementering av digitala arbetssätt pågår i många länder. Skolverket (2010) konstaterar att i Norge och Danmark har utvecklingen lett till nationella datorbaserade bedömningssystem. Parallellt med större bedömningsverktyg riktade på summativ bedömning utvecklas även mindre och mer lätthanterliga tester för formativ utvärdering. Diverse läroplattformar som används av skolor, såsom *Fronter*, *Moodle* och *It's learning* inkluderar verktyg för att göra test. Dessa tenderar att ha mycket omfattande funktionalitet, men användare förutsätts genomföra någon utbildning i dem för att kunna använda dem.

Som kontrast finns det även enkla tester på marknaden (Kahoot, Mentimeter m.fl.) som dock saknar möjligheter till individuell bedömning. Dessa ger endast statistik på grupp-nivå, ger enbart möjlighet till att ställa en fråga och är tänkta för klassrumsdiskussioner.

Anledningen till att lärare inte använder digitala test i särskilt hög utsträckning anges bland annat vara lärarnas låga digitala kunskapsnivå samt att det är tidskrävande att skapa ett test (Gulz, 2014). En annan förklaring som McLaren

(2012) föreslår är att lärare är skeptiska till digitala utvärderingar som sådana för att de anser att den digitala återkopplingen inte kan ersätta den personliga dialogen. Detta är dock inte syftet med ett digitalt testverktyg, som inte ska vara en ersättning utan snarare som ett komplement och stöd till den personliga dialogen.

Enligt Jenkins (2010) måste också digitala bedömningssystem vara så lätthanterliga och effektiva att varken lärare eller elever upplever dem som en börda. Här finns alltså en lucka i marknaden och det är där vårt projekt kommer in.

Syftet med projektet

Vårt uppdrag var att skapa en lätthanterlig plattform som kan användas av lärare på högstadium för läxförhör och andra liknande utvärderingsaktiviteter i klassrummet på regelbunden basis. Ett antal omedelbara fördelar kan ses med att byta format från analogt till digitalt. Den normala arbetsgången nu är att skriva ett test, skriva ut det, dela ut det och sedan rätta för hand. För det första ger ett digitalt test tidsbesparingar för lärarna vid producerandet av test. Om gränssnittet är lättfattligt borde det gå snabbare att skapa test då man inte behöver tänka på layout med mera. Vidare, och framför allt till nytta för lärarna, kan rättning ske automatiskt. Här finns enorma tidsbesparingar att göra samtidigt som rättning är en monoton syssla många lärare skulle slippa. Automatisk rättning ger också elever möjlighet att se de rätta svaren precis efter att de gjort testet och de behöver således inte vänta på svar. Som det fungerar nu kan det vara fallet att elever får tillbaka sina test så sent att de glömt bort dem. Dessutom kan statistik genereras snabbt och också automatiskt.

Vidare såg vi möjligheter att inkorporera formativ bedömning, bedömning av lärande, i vårt verktyg. Detta skulle generera ytterligare mervärde i vår produkt och dessutom vara en *unique selling point* mot övriga verktyg på marknaden.

Utifrån uppdragsformuleringen ringade vi alltså in syftet med det här projektet: att skapa ett digitalt testverktyg som är dels enkelt (tidsbesparingsaspekt) dels innehåller möjlighet till mångsidig och flexibel återkoppling (formativ bedömningsaspekt).

Terminologi

Vår marknadsundersökning tillsammans med forskningsgenomgång visade att det finns ett stort antal termer som används för att benämna digitala tester. Här är några exempel på svenska och engelska: digitala tester, quizzes, digitala

förhör, datorbaserade prov, formativ utvärdering. En av förklaringar kan vara att olika tester har olika syften, därmed olika namn. Ord som prov, förhör och quiz bär med sig förutfattade meningar om vilken typ av omfattning och svårighetsgrad det rör sig om. Vi vill, i kontrast till detta, poängtera mångsidigheten i vårt verktyg samt inte styra in användarna på någon typ av användning. Därför har vi valt att använda ordet test, vilket vi uppfattar som mest neutralt.

2 Förstudie

Varje programvara måste vara anpassad till sin användarmålgrupp för att vara framgångsrik. Inför skapandet av vårt test intervjuades således åtta lärare för att skapa en bild av vad som konstituerar en bra plattform för digitala test. Syftet var också att finna hur nuvarande plattformar används samt vilka funktioner som är önskvärda och vilka hinder som är frusterande.

Sex av lärarna undervisade på högstadiet och två var gymnasielärare. De undervisade i ämnen som täcker större delen av läroplanen, bland annat matematik, samhällskunskap, biologi, naturkunskap, historia, spanska, filosofi m.fl. Tre av dem var vana användare av någon befintlig plattform för digitala test och använde dessa regelbundet i undervisningen. Två av dessa var de båda gymnasielärarna. De intervjuade refererade även till hur kollegor använde digitala test.

Vad lärarna använde eller ville använda testen till varierade mycket. Somliga använde det för att examinera läxor eller för inlämning av prov med essäfrågor, medan andra använde det i instuderingssyfte och lät eleverna göra om testen flera gånger. För distansundervisning på Komvux användes det som närvarokontroll. Ytterligare ett användningsområde var att använda det som mentometrar i klassrummet för att skapa underlag för klassrumsdiskussion eller undersöka hur många som förstått något. Slutligen användes de också som förkunskapstest innan en kurs började.

Typen av frågor som ställdes var oftast på basal nivå. Även om en del ställde avancerade frågor, så påpekades att mer komplex typ av förståelse är svår att visa i ett kort test. Ett exempel på en avancerad fråga var en omvårdnadslärare på gymnasiet som lät eleverna se en film om ca 10-20 minuter med en patientinteraktion där elever fick ge längre svar på frågor om vad som gjordes rätt och vilka misstag som skedde.

Något som ofta nämndes som en av de digitala testens största fördel var den tidsbesparande funktionen att testet rättade sig själv. Det var ovanligt att läraren gav feedback muntligt eller skriftligt utöver det som de skrivit in i testet. Funktionen att ge det rätta svaret användes både av de som gav essäfrågor och de som gav flervalsfrågor eller liknande. Den automatiska rättningen av fritext litades inte på, utan rättning skedde manuellt av fritext. Dock gav läraren då bara poäng, eleven hade ju redan fått det rätta svaret i samband med inlämning. Lärarna tyckte att möjligheten till individualiserad feedback var viktigt. De ville kunna skriva färdiga kommentarer som var olika beroende på om man svarade rätt eller fel.

Den feedback som lärarna fick av testet användes på skiftande sätt. I de fallen som testen var att betrakta som övning tittade lärarna endast på huruvida elever gjort testen eller inte. I fallen då det användes som undersökning av den allmänna nivån var naturligtvis statistik på gruppnivå det

mest intressanta. Något som en lärare gjorde var att koppla frågorna till aktuella betygsmatriser och automatiskt låta den pedagogiska plattformen bocka i om någon förmåga uppfyllts. Detta användes huvudsakligen för att se om eleven uppnått E, det lägsta godkända betyget i nuvarande system. Data från testen användes dessutom för att individualisera prov, då endast förmågor som inte visats ännu behövde testas. Av de som inte använde test ännu önskades också denna typ av longitudinell funktionalitet.

Av dem som använde digitala test hade endast en låtit eleverna göra det i klassrummet. Övriga såg det som en fördel att det inte behövde användas på lektionstid. Ett hinder för flera av högstadielärarna var att skolorna de arbetade på saknade 1:1-policy (en-till-en), det vill säga att varje elev fick en egen dator av skolan. De påpekade dock att skolan tillåter lärarna att förvänta sig att alla elever har tillgång till en dator utanför skolan på något sätt. Flera av dem skulle dock få iPads inom kort och hoppades då att vårt verktyg skulle vara plattformsoberoende. Detta så att även eleverna skulle kunna göra det i sina telefoner.

Att elever gör test hemma skulle kunna öppna för fusk, men det fanns väldigt liten erfarenhet av detta. En lärare påpekade att denna risk finns med alla hemuppgifter, digitala eller analoga. Vissa ville hellre göra testen i klassrummet av denna anledning. Lösningar använda av en lärare var att ställa frågor som inte var lätta att googla upp ett svar på samt att ge begränsad tid från det att testet startade. Den senare lösningen i kombination med komplexa frågor borgade för begränsat fusk, hoppades han. Något som ingen diskuterade, men som kanske vore en möjlighet är att validera längre essäsvår genom en plagiatidentifieringstjänst såsom Urkund.

Flera av lärarna visade intresse för en bank med frågor som man kunde importera in i test. Flera lärare påpekade att de hellre ville spara test än frågor. De återanvände gamla test, men bytte ofta ut eller ändrade enskilda frågor då. Samtliga tyckte att det var viktigt att enkelt kunna organisera sina test. De ville även kunna dela med sig av test till sina kollegor. De ville också kunna importera test via hemsidor som kursnavet.se eller CFL (Centrum för Flexibelt Lärande). Att också kunna dela material via omodererade sajter som lektion.se nämndes också.

Nästan samtliga lärare visade intresse för att använda media såsom filmer, bilder och ljud i test. Matematikläraren ville gärna ha en Kikora-liknande funktionalitet eller i alla fall att kunna skriva in formler, snarare än att rita dem och bifoga som bilder. Några av lärarna önskade också att man skulle kunna svara med en bild, som man ritade direkt i iPaden. En språklärare ville kunna lägga en ljudfil för hör-förståelseövningar.

Samtliga lärare var intresserade av att kunna ställa många olika typer av frågor. De ville ha fritextsvar, flervalsfrågor samt matchningsuppgifter. De vill även att dessa skall vara flexibla så att man kan ställa in att det kan vara mer än ett rätt i flervalsuppgifter eller att ett svar kan kopplas till flera alternativ i en matchningsuppgift.

Kritik som framfördes mot nuvarande system var att bildöverföring fungerade dåligt och att bilder inte följde med om testen exporterades. Sorteringen av test var konstig och flera irriterade sig på att rubrikerna i testverktygen var ologiska och svåra att gissa var de ledde. För fritextsvar fanns begränsningar i antal tecken, vilket kunde ställa till problem i de fall test användes för att lämna in kortare uppsatser.

3 Teoretiskt underlag

Enkelhet

Fokus i projektet har ständigt varit på enkelhet. Det är till att börja med inte lätt att definiera enkelhet. Det som anses enkelt för vissa är inte enkelt för andra. Intervjuerna med lärare ledde oss till flera parametrar för enkelheten: test ska vara lätta att ta fram, de skall ha tydlig layout och ha tydliga knappdefinitioner. Eleven ska också förstå vad, var och hur. Med andra ord, designen ska vara anpassad till verktygets syfte.

Av allt som tas upp i lärarintervjuerna (Förstudie) är det mycket få funktioner som vi har för avsikt att implementera. Vi insåg att avancerade och kraftfulla funktioner i huvudsak efterfrågades av de två gymnasielärarna. Dessa var erfarna och ville ha ett dynamiskt system och var inte rädda för den komplexitet som oundvikligen skulle bli följden av detta. Att nå fram till grundskolelärarna handlade snarare om att göra ett så enkelt system att de överhuvudtaget skulle förstå det och använda det. Inom design säger man, enligt Lidwell, Holden och Butler (2003), att flexibilitet och användbarhet står i en ständig trade-off relation med varandra.

Ett sätt att göra så att lärarna kände att det gick snabbt var att minska antalet funktioner. Detta då Hicks lag (Lidwell, Holden & Butler, 2003) postulerar att tiden det tar att fatta ett beslut ökar logaritmiskt med antalet val. Med färre val skulle lärarna alltså snabbare kunna producera ett test.

Lidwell, Holden och Butler (2003) menar att en tumregel är att 80 % av användandet av ett system är med 20 % av funktionerna, samtidigt som 80 % av felen kommer från 20 % av funktionerna. Det känns rimligt att anta att det man använder sällan är det som avbuggats minst noggrant, som man är minst van vid som användare och som följaktligen borde stå för stor del av felen. Vi tänkte att om vi kompromissar bort 80 % av funktionerna kanske vi blir av med nästan alla användarhindren/felen.

En kompromiss vi gjort är att inte införa något inloggningssystem. Med ett sådant kan man skapa longitudinell funktiona-litet, fuskkontroll med mera. Anledningen att vi ändå valt bort det är att en inloggning är en typisk onödig försvårare. Det skapar trösklar och kan genast göra att en lärare avstår, för att man inte orkar skapa en användare eller ett konto.

Något vi däremot har valt och som vi försökt inkorporera i samtliga delar av vår design är Normans (2002) idéer om *affordances*, dvs. att en produkts konceptuella design ska kunna inbjuda till rätt handlingsalternativ och att den används för det den är avsedd för. Vi vill att knappar skall locka till att tryckas på och samtidigt utnyttja de normer som finns bland datoranvändare. Detta är sådant som att ett kryss betyder stänga av, stäng ner eller ta bort, medan ett minus betyder fäll ner eller dölj. Förutom denna typ av enkla *affordances* ville vi skapa *affordances* för formativ bedömning.

Ibland kan enkelhet också gå hand i hand med *affordances*. Något som kändes enkelt att införa var att man kunde klicka på rubriken till varje fråga för att ändra dem från "fråga 1" till något mer informativt. Vi valde dock bort detta inte bara för att det minskade antalet möjliga platser att klicka (fel) utan också för att siffran på frågan uppmanar lärarna till en kronologi eller progression i frågorna, vilket är bra för elevers prestation (Wise et al., 1989).

Ett annat exempel på enkelhet är att Kulhavy och Stock (1989) har visat att elever inte bryr sig om feedback när de haft rätt. Vi hade experimenterat med att lärarna skulle kunna ge olika typer av feedback om man hade rätt eller fel, men efter att ha läst denna artikel såg vi ingen mening med denna funktion. Återigen var det frågan om komplexitet som verkade nödvändig, men som kunde skalas bort eller avstås från att läggas till.

Återkoppling

En av ambitionerna för de digitala testen var förekomst av återkoppling som skulle baseras på formativ bedömning. Formativ bedömning är ett begrepp som lanserades i slutet av 1960-talet och utvecklades vidare av Black och Wiliam (Skolverket, 2010) i slutet av 1990-talet. Ett exempel på en definition av formativ bedömning är att göra en bedömning som identifierar elevers inlärningsbehov och svarar mot dessa (OECD, 2005).

En annan definition av formativ bedömning ges av Black och Wiliam (1998b) i form av tre faktorer. För det första, att lärare anpassar undervisning och bedömning till vad som finns evidens för att vara effektivt. För det andra, att elever får återkoppling om deras lärande och hur de blir bättre. Slutligen, för det tredje, att elever deltar i bedömningsprocessen genom självbedömning. Vi väljer här en bred tolkning av begreppet, då vi menar att de flesta definitioner är relativt lika.

Det skall dock påpekas att begreppet *formativ bedömning* är omtvistat och många konkurrerande (om än liknande) definitioner finns (Bennett, 2011). Formativ bedömning framförs både som motsats och som komplement till så kallad *summativ bedömning* som syftar till att fastställa elevens kunskap vid en bestämd tidpunkt genom att sammanfatta i ett betyg eller en poängsumma (Grönlund, 2010). Dock menar Torrance och Pryor (1998) att den pedagogiska debatten alltför ofta kommit att handla om distinktionen *summativ/formativ*, snarare än hur god formativ bedömning skall se ut för att vara den bästa.

Trots att begreppet formativ bedömning har funnits länge som utlärningsverktyg visar den internationella forskningen att lärare inte helt har förstått de formativa aspekterna av bedömningen (Heritage & Chang, 2012). Heritage och Chang (2012) presenterar vidare en omfattande lista av forskningsrapporter som pekar på att lärarnas förståelse för formativ bedömning inte stämmer överens med begreppets tänkta syfte. De kommer i sin studie fram till att de flesta lärare ser formativ bedömning som ett instrument för att lära om från början (*reteaching*) istället för att bestämma var eleven befinner sig just nu och hur man kan hjälpa eleven till nästa nivå. En viktig faktor angående lärare och feedback är att för att ge god återkoppling krävs djupa ämneskunskaper, så att man förstår de centrala problemtanken i sitt ämne (Hattie & Timperley, 2007).

Antoniou och James (2014) har vidare funnit att lärarna antingen ger alltför allmänna kommentarer som inte för eleverna framåt eller jämför elevernas senaste resultat med föregående och på så sätt fokuserar på produkten och inte på processen. Dessa feltänk kring formativ bedömning leder till att målet missas, menar Waters (2012) och påpekar att för att bedömningen ska anses som formativ krävs specifika strategier för förbättring. Det är även så att den feedback som syf-

tar till lärandemålen har större effektivitet än den som bara syftar till att öka elevers självförtroende (Atkins, Black & Coffey, 2001). Vidare fann Black och William (1998b) att även när lärare säger sig vilja fokusera på förståelse, handlar deras undervisning och test om utantillkunskaper.

Enligt Heritage (2010) kan det vara svårt i testsammanhang att dra gränsen mellan formativ och summativ bedömning. Det räcker inte med att bara definiera formativ bedömning som ett instrument som används för formativa processer genom ett antal mindre tester. Formativ bedömning, enligt Heritage (2010) ska inkludera även förståelse för Vygotskys modell om den proximala utvecklingszonen. Dessutom skall bedömningstillfällen kännetecknas av ett aktivt elevdeltagande i lärandeprocessen som ökar förmågan till självstyrning. Feedback måste vara av deskriptiv karaktär med syfte att fylla luckor (alter the gaps) och inte i form av poäng bara för att mäta skillnaden mellan elever.

Ett sätt att komma åt dessa luckor är enligt Hattie och Timperley (2007) att läraren i sin planering tar hänsyn till tre processer: Var är eleven nu? Vart skall eleven? och Hur kommer eleven dit? Dessa processer inkluderar eleven och läraren men även elevens kamrater. Redan vid planeringen av ett arbetsområde bör läraren tänka på återkoppling som påverkar frågornas utformning. En typiskt bra fråga är till exempel en som eleverna tror sig ha rätt på, men i själva verket har fel på. Anledningen till detta är att det är i just detta fall då effekten av feedback är som allra störst (Kulhavy & Stock, 1989). Det är också välkänt att endast information om att man gjort fel utan att man får reda på vad som krävs för att göra rätt nästa gång hjälper en elev föga i hennes eller hans lärande (Breakwell, 1983).

Fokus hos lärarens kommentarer bör alltså vara av sådan karaktär att de för lärandet framåt (Grönlund, 2010). Hattie och Timperley (2007) skiljer på fyra typer av formativ återkoppling: uppgiftsrelaterad, processrelaterad, metakognitiv och personlig. De mest verkningsfulla återkopplingarna menar de är de processrelaterade och metakognitiva kommentarerna. De tidigare avser processer som eleven använder för att lösa uppgifter, till exempel strategier som är generaliserbara. De senare refererar till elevens förmåga att styra sin självdisciplin (bland annat ansvar för inläring) och självvärdering (såsom att reflektera över sitt arbete).

Fisher och Frey (2014) hävdar vidare att för att lyckas med formativ bedömning som sådan behöver läraren först utarbeta ett bedömningssystem där de uppsatta lärandemålen kopplas till ett lämpligt verktyg för bedömning och återkoppling. I annat fall finns det risk att läraren samlar och analyserar meningslös information som inte heller leder elever till nästa nivå i processen. Här kan dock vårt verktyg inte hjälpa, utan vad läraren gör blir upp till denne.

Vidare fann McLaren och Stables (2007) att betyg och poäng blev mindre viktiga för eleverna då formativ bedömning användes. Istället blev deras eget lärande i fokus för eleverna. Att poäng blev mindre viktiga är också positivt i allmänhet då Black och William (1998a) visat i sin metastudie att kommentarer är mycket mer effektiva än poäng och betyg. De menar sig visat att bedömning enbart med poäng och betyg

knowledge which pupils soon forget.” (Black & William, 1998, s. 14)

Formativ bedömning är som synes ovan en komplex fråga, men avslutningsvis måste påpekas att Hattie och Timperley (2007) i sin meta-metastudie funnit att återkoppling har en större effekt på elevers lärande än många andra saker som ofta räknas som viktiga i utbildningssammanhang såsom socioekonomisk status, tidigare kognitiv förmåga och klasstorlek. Det är också så att formativ bedömning har den största effekten av alla pedagogiska interventioner (Black & William, 1998b). Det är alltså av stor vikt att lärares feedback till elever är genomtänkt och av vetenskapligt verifierat effektivt slag.

Teknologi och pedagogik

Enkelhet och återkoppling har varit de två ledorden för vårt arbete där enkelheten är mest kopplad till de tekniska lösningarna i projektet och återkoppling till de pedagogiska aspekterna. I detta projekt har vi försökt att integrera teknologi och pedagogik för bedömning, samtidigt som det var viktigt för oss att skilja på verktyg och bedömningsinstrument i digitalt arbete (Price, 2011). Gulz och Haake (2014) hävdar att man ska skilja på digitalt format och digital funktionalitet. Den senare innefattar både ett tidsbesparingsverktyg (automatiserad rättning) och ett pedagogiskt verktyg för bedömning. Det pedagogiska är sådant som vilka frågetyper som finns samt implementeringen av formativ feedback. Dock är det pedagogiken som styr användningen och utformningen av digitala verktyg. Price (2011) påpekar att själva klassrummet och de normer som finns i det påverkar hur dessa verktyg används. Dessutom menar Pellegrino och Quellmalz (2010) att integrering av teknologi och bedömning stödjer elevernas inläring genom feedback och ger ökad interaktion både mellan lärare-elev och elev-elev. Lärare kan med detta samla information mer effektivt om inlärningsprocessen, som kan leda lärandet framåt.

Men det finns flera andra pedagogiska fördelar med digital bedömning. Cassidy och Gridley (2005) hävdar att formativa digitala test minskar studenternas rädsla inför test. Tack vare testernas enkelhet kan man genomföra dem regelbundet och ofta vilket blir en utmärkt förberedelse inför summativa stora prov. Detta minskar studenternas oro. Även feedback som man får direkt efter testet bidrar till minskad oro. Enligt Stull et al. (2011) har digitala bedömningsverktyg fördelar för elever med lindriga till måttliga koncentrationssvårigheter. Oro bland lärare för att elever som har låg datorvana skall prestera sämre är dock obefogad, då Bocij och Greasley (1999) visat att de presterar lika bra på digitala test som datorvana elever.

Det är även värt att notera att effekten av att få feedback från en dator är jämförbar med att få feedback av en människa (Hattie & Timperley, 2007).

4 Arbetsgång

Instruktionen vi fick var att göra “ett sorts Google” för digitala test. Huvudfokus var på enkelhet och användbarhet för mycket oerfarna användare. Tanken var att närma sig en startskärm med endast en ruta och en knapp. När det bara

“typically encourages superficial and rote learning, concentrating on recall of isolated details, usually items of

finns en plats att skriva och en plats att trycka finns inga möjligheter att göra fel.

Med en ruta och en knapp, finns ändå två sätt att ge input. Ett sätt att ytterligare eliminera förvirring är att reducera ner till endast en ruta. Utgångspunkten för detta var att när läraren/användaren kommer in på hemsidan har man endast en tanke, man vill skapa ett digitalt test. Ett test består av frågor och inleds med den första frågan. Således är det enda man kan göra på startsidan att skriva den första frågan.

När man börjat skriva frågan tonar nya möjligheter in. Här inspirerades vi av den inlärning som finns i datorspel. Där kan det vara så att du börjar med att kunna gå omkring och hoppa. Efter ett tag hittar du en klubba som du kan slå på fiender med. Senare kan du hitta ett mer komplext och kraftfullare vapen som en pistol. Denna måste du hitta ammunition till och ladda om. I det sista steget har du mycket kraftfullare möjligheter att manipulera din värld, men på bekostnad av fler knapptryck. På samma sätt som i en sådan filosofi ville vi att användare möttes av ett väldigt enkelt gränssnitt med få möjligheter att göra fel. I takt med att de skapade sitt test kommer nya alternativ och redigeringsmöjligheter upp. Användaren lär sig alltså att använda verktyget medan denne använder det och komplexa funktioner är inte tillgängliga förrän enklare använts. I allmänhet, menar Lidwell, Holden och Butler (2003) att mängden kontroll en användare skall ha över ett system bör vara proportionell till hur mycket kunskap denne har.

Vi valde tidigt att ha mycket luft i designen. Inom konsten använder man begreppet *Horror Vacui* (rädslan för tomhet) när någon design eller något verk är helt täckt med detaljer. Ju mindre *Horror Vacui* något uppvisar, desto större upplevt värde har något (Lidwell, Holden & Butler, 2003). Vi upplevde också att hemsidan blev snyggare av detta. Det kan tyckas som ett trivialt faktum, med det är faktiskt så att vackra produkter upplevs som lättare att använda (Kurosu & Kashimura, 1995). Dessutom har användare en mer förlåtande attityd mot estetiskt tilltalande föremål/produkter (Asch, 1946).

I led med att skapa enkelhet har vi valt de två frågetyper vi identifierat som de vanligaste typerna, flervalsfrågor och matchningsfrågor. Flervalsfrågor ger möjlighet för eleven att träna på förmågan att göra val genom både en möjlighet att välja mellan rätt och fel och genom kunna välja flera svarsalternativ som kan vara rätt eller fel. Matchningsfrågor gör det också möjligt att göra val men även att utveckla förmågan att kategorisera och sortera.

Flervalsfrågor är vidare ett beprövat sätt att testa elevernas kunskaper och utveckling som baseras på ett samhälleligt behov av att kunna göra bra val (Chi et al., 2014). Chi et al. (2014) påpekar dock att digitala tester oftast erbjuder valet mellan rätt och fel utan möjlighet att resonera kring varför det ena valet är bättre eller sämre än det andra. Sådana rätt- och felsvar testar endast faktakunskap, men utvecklar inte förmågan att göra val.

Formativ bedömning är, som påpekats i teoridelen, en missförstådd och illa använd metod. Hur skapar man då affordances för något sådant? Vi tänkte att i en ruta för feedback skulle det cirkulera olika texter med inspiration. Vi har, i enlighet med teorin, försökt ge förslag på möjliga kommentarer som är antingen processrelaterade och/eller metakognitiva. Tanken är att läraren kommer på rätt spår

genom ett tips och utformar sina kommentarer därefter. Följande formativa återkoppling har vi föreslagit:

Processinriktade (strategier)

- Läs om frågan för att förstå den ordentligt.
- Läs om kapitlet/stycke för att förstå sammanhanget.
- Sätt alltid in nya ord i en mening eller ett sammanhang.
- Dela upp uppgiften i steg: först gör jag det, sedan gör jag det...

Metakognitiva (självstyrning)

- När du läser en text, anpassa lästempot efter uppgiften.
- Repetera alltid det vi gör på lektionen hemma.
- Innan du väljer ett alternativ, tänk på vad som är rimligt.
- Lägg gärna mer tid på att lösa uppgiften. Försök att inte ge inte upp direkt.

Dessutom har vi tänkt ha med texter av typen: "Tänk på att inte bara skriva 'bra gjort' utan att hjälpa eleven vidare i sitt lärande". Eventuellt skall också den typen av texter varnas för med en pratbubbla om man bara skriver "fel!" som kommentar.

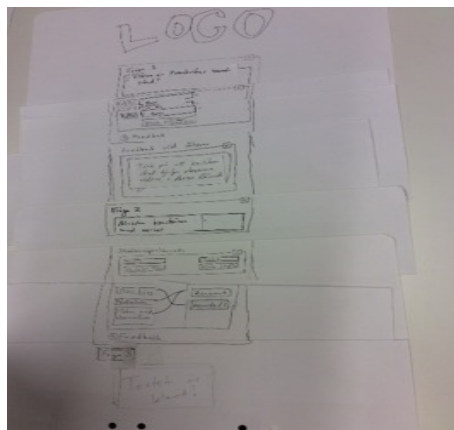
Slutligen har vi arbetat med defaults (standardinställningar). Vi vill att lärarna sällan eller aldrig skall vilja ändra standardinställningar och då måste de vara på ett sådant sätt att läraren får det den vill. Med andra defaults har vi valt en mer proaktiv position för att vi vill att läraren skall agera på ett särskilt sätt. Ett exempel på detta är att man får sina svar rättade efter att man fyllt i hela testet, då detta är bättre än att ge det efter varje fråga (Wise et al., 1986).

Prototyp 1

Vår första LoFi-prototyp (Figur 1) bestod av diverse pappersbitar, vilka vi tog till en grundskola för ett test. Till denna hade vi också ett scenario som löd:

Du kommer in på en hemsida och skall på den göra ett test för dina elever. Testet skall innehålla två frågor: en flervalsfråga och en matchningsfråga. På flervalsfrågan skall det finnas fyra alternativ varav två av svaren skall vara rätt.

Till prototypen hörde också en "skärm" vilken vi höll fram framför vår prototyp för att ge fokus på känslan att detta hände på en dator. Denna tog vi dock bort när testpersonerna började manipulera prototypen, då vi märkte att den var i vägen.



Figur 1. LoFi-prototyp som användes vid första testet

Första testet

I första testet deltog tre lärare, två yngre och en lite äldre. Till att börja med fanns en skillnad i tillvägagångssätt mellan dessa, den äldre hade mycket frågor och inte ville klicka på något hon inte visste vad det gjorde, medan de yngre provade sig fram. De kunde säga "undrar vad den här gör" och sedan klicka. De tre lärarna undervisade alla i olika ämnen.

En sak som noterades var att inte alla förstod hur man ändrade ett alternativ från rätt till fel och vice versa. Detta kan ha berott på prototypens råa, ofärdiga utseende som därmed gav viss tvetydighet. Vidare uttryckte samtliga att de var intresserade av någon typ av förhandsgranskning. De var osäkra på hur det skulle se ut för eleverna och i, med deras uttryck, "vad som var facit". Dessutom fanns en osäkerhet i vad som var sparad och inte.

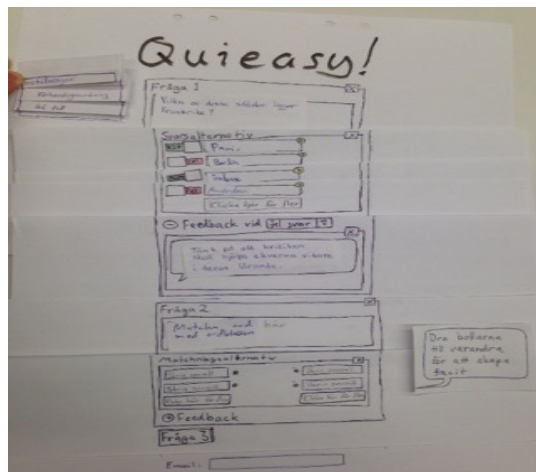
En effekt av intresset för förhandsgranskning var att det första de gjorde när de mötte skärmen med de två länkarna (en till eleverna och en till läraren) var att klicka på elevlänken. Detta fick effekten att de lämnade testet och kom in på elevtestet utan möjlighet att se länkarna igen. En potentiell källa till oerhörd frustration.

Slutligen framkom diverse småsaker, där det mest relevanta var att vi märkte att det fanns en dissonans mellan vad som innebar en radering och vad som uppfattades som en dölja-funktion.

Ändringar till prototyp 2

Den andra prototypen vi tillverkade (Figur 2) påminde väldigt mycket om den förra. För att undvika att misstag uppstod till följd av ett alltför rått och kladdigt utseende valde vi att göra den med större omsorg om detaljer och lade till färger på knappar med mera. Den kan således betraktas som någonting mellan en LoFi- och MidFi-prototyp. Vidare lade vi till en "sidebar" där man kunde se en miniatyr av testet, vilket vi hoppades skulle tillfredsställa användarnas önskemål på förhandsgranskning. Dessutom lade vi till ett obligatorium om att man skulle fylla i sin e-post innan man tryckte på knappen som indikerade att testet var klart.

Vidare ändrade vi symbol på feedback så att man inte skulle tro att man fällde ner när man tog bort. Dessutom ändrade vi så att plustecknet vid feedback ändrades till ett minus när man öppnade det, så att man ändå kunde fälla ner det. Slutligen lade vi till möjligheten att lägga till olika feedback vid rätt svar, fel svar och delvis rätt svar genom en dropdown-menyn.



Figur 2. LoFi-prototyp som användes vid andra testet

Andra testet

Vid nästa tillfälle på testskolan deltog fyra lärare varav en hade varit med vid det tidigare tillfället. En av dem var väldigt nervös och förstod inte alls uppgiften. Flera försök gjordes att lugna personen och förklara att det som testades var prototypens funktionalitet och design, inte försökspersonen på något vis. Detta misslyckades dock och vi lät personen genomföra testet och tackade sedan personen för input, trots att det snabbt blev uppenbart att personen inte hade förstått vad denne skulle göra.

En allmän reaktion från samtliga användare vid det här tillfället var att endast en person ens öppnade feedback-fliken. Den personen stängde den dessutom snabbt. Här finns utrymme för förbättringar som antingen lockar lärarna att klicka på den eller en förklaring som gör att de förstår dess vikt. En förklaring kan dock vara att det bara var ett försök på papper. I en skarp situation hade de förmodligen varit mer motiverade att skriva feedback till sina elever. Den lärare som ändå gick in på feedback påpekade att hon ville skriva specifik feedback beroende på vilket fel man gör. Detta fanns möjlighet att ändra i prototypen, men hon hann nog inte se detta eftersom det stängdes så snabbt.

Ytterligare två saker som var gemensamt för samtliga lärare var att de missade matchningsfunktionen och hade frågor kring förhandsgranskning. Att de missade möjligheten att skapa en matchningsfråga kan berott på att scenariot var otydligt, även om det sades explicit där att de skulle skapa en matchningsfråga. Vi spekulerade i om namnet är illa valt och att det kanske istället skulle heta "para ihop" eller något dylikt.

När det gäller förhandsgranskning hade vi uppenbarligen inte förstått den kritik vi hade fått på det första testet. Det lärarna ville se var allt från hur en specifik fråga såg ut för eleverna till hur testet i helhet skulle se ut. De ville dessutom se hur det såg ut när man svarade, hur facit/rättning såg ut för eleverna samt hur feedback presenterades. Denna typ av komplexa och dynamiska feedback kan vara svår att implementera och kanske är det rimligare att lärarna efter att de är färdiga med testet själva kan gå in på elevlänken. Om de är missnöjda kan redigering ske genom lärarlänken.

För övrigt noterades att de nya rätt/fel-knapparna inte orsakade samma huvudbry, utan direkt förstods. Detta kan dock vara en följd av att de var tydligare ritade och att de var färgkodade. Slutligen kan nämnas att syftet inte gick fram till

samtliga lärare, då en av dem ifrågasatte plattformen som sådan. Hon undrade hur man skulle skriva ut testet, men verkade få en aha-upplevelse när vi förklarade att det var självvrättande och skulle utföras digitalt. Detta är viktigt att lärarna förstår.

Design av HiFi-prototyp

När vi började arbeta med vår HiFi-prototyp var det på basis av tidigare diskussioner i gruppen tillsammans med handledare bestämt att denna skulle utformas i HTML 5 med ytterligare funktionalitet från JavaScript och biblioteket jQuery. Dessa språk och verktyg var helt nya för projektgruppen vilket har bidragit till en utmanande inlärningskurva. Efter en tids utveckling upptäckte vi också att det finns ramverk som bland annat innehåller HTML- och CSS-baserade mallar för knappar, textfält, typsnitt, navigering samt ytterligare JavaScript-funktioner. Ett av de ramverk som stämde bra överens med våra tidigare designval var Foundation 5 från Zurb, och vi har därför valt att använda detta för vår hemsida.

När vi har jobbat med HiFi-prototypen har vi utgått från den senaste LoFi-prototypen samt lagt fokus på den dynamik som inte riktigt går att uppnå med hjälp av pappersbitar och färger. Arbetet har fram till slutredovisningen fokuserats på att skapa ett fungerande formulär för läraren att kunna formulera en frågetext och lägga till svarsalternativ i en flersvarsfråga. Utöver detta har även energi lagts på att kunna skapa flera frågor och ta bort frågor samt att möjliggöra för läraren att lägga in flera svarsalternativ genom att hela tiden erbjuda ett nytt svarsfält nedanför det som läraren för närvarande skriver i. Hur detta ser ut kan man tydligt se i Figur 3 nedan.

Figur 3. Läraren skapar en fråga

Då tiden från när vi var klara med LoFi-prototypen (vid mitredovisningen) fram till slutinlämningen var ganska kort har vi endast fokuserat på delar av frontend-systemet i den första delkursen. Vi har därför i delkurs 2 fortsatt med designen av frontend-systemet samt utvecklat ett backend-system. Vi har bland annat lagt in mer dynamik i lärarens verktyg genom att erbjuda alternativ för att ta bort skapade frågor och

svarsalternativ. Förutom detta har vi även sett till att frågorna sparas undan i en databas (gjord i MySQL) och läraren får en bekräftelse på att frågorna är sparade. Bekräftelsen visas tillsammans med en länk som kan distribueras till eleverna, så att de kan göra testet. När eleven går in på länken genereras hemsidan genom att frågorna och svarsalternativen hämtas från databasen och visas med hjälp av PHP. När eleven sedan har gjort sitt test måste eleven fylla i sitt namn för att skicka in testet. Sedan rättas testet, resultatet från en sådan rättning kan ses i Figur 4 nedan. Där en grön ruta syns innebär det att svarsalternativet är korrekt. Där en röd ruta syns innebär det att svarsalternativet var fel. Alla rutor som eleven har valt markeras med en bock i. Om eleven svarade helt rätt på frågan visas texten "Du svarade rätt på frågan". I övriga fall visas texten "Du svarade fel på frågan" följt av texten som läraren valt att lägga in i feedback-fältet för frågan när testet skapades.

Figur 4. Eleven har svarat på frågorna och får sitt resultat

Tredje testet

Genomförande

Vi genomförde sedan ett användartest av vårt nästan färdiga testverktyg. Vi besökte återigen vår testskola vid två tillfällen. Den första dagen var en studiedag, då vi hade gott om tid med lärarna och kunde låta dem göra test med frågor och feedback. Medan de gjorde testen tog vi anteckningar och efteråt genomförde vi semistrukturerade kvalitativa intervjuer. Dagen efter lät vi två elever ur respektive lärares klass göra de faktiska testen. Till följd av ett fel raderades ett test efter att det skrivits in och därför, deltog inte den lärarens elever i utprovning av testet dagen därpå.

Lärarperspektiv

Observationer av och intervjuer med de deltagande lärarna visade att alla förstod vad de skulle göra och förstod hur hemsidan fungerade. Alla tyckte att det fanns ett naturligt flöde, det var enkelt (ingen överflödig information runt omkring själva testfältet), att layouten var snygg och det gick snabbt med hela förfarandet.

De önskemål som framfördes av lärarna var att de önskade ha tillgång till kursiv/fet stil samt möjlighet att kunna

ändra teckenstorlek. En lärare önskade att man kunde lägga in bilder och en lärare önskade att kunna spara testfrågor i en bank. Intressant att notera är att de två lärarna som deltog samtidigt hela tiden diskuterade frågeformuleringarna, men inte alls den tekniska sidan av testet. Detta pekar på att vi har uppnått ett av våra två mål, målet på den tekniska enkelheten. Ambitionen har aldrig varit att skapa en hemsida som får lärare att fascineras eller prata om. Snarare skall det vara ett tyst verktyg som "bara fungerar".

En av de centrala aspekterna för oss var tidsbesparingsfrågan. Vi har satsat på enkelhet i form av tekniska lösningar, men också möjlighet till snabb rättning, direkt återkoppling och statistik på grupp- och individnivå. Vi bad därför alla deltagande lärare att jämföra tidsåtgång för ett digitalt test och samma test i pappersformat för en hel klass på ca 25 elever. Förberedelser inför ett test, det vill säga frågeformulering och tid för att skriva in frågor är samma för båda tester. Man skulle kunna argumentera för att vårt verktyg är något snabbare, då man inte behöver tänka på layout eller utskrift. Tidsbesparingen skulle framförallt ligga dels i själva rättningsprocessen och registreringen av resultat dels i en automatiserad återkoppling, som genereras lika för alla elever. Två lärare delade ut samma test som de genomförde i pappersformat, samtidigt som två elever testade vår digitala version. En lärare uppskattade att ett test i pappersformat för en klass skulle ta ca 30 minuter att rätta, registrera och skriva in feedback. En annan lärare uppskattade samma arbete som en timmes arbete. Ett digitalt test som vårt minskar tidsåtgång för ovanstående arbete med 30–60 minuter. På en termin (och tanken är att man arbetar med den sortens tester regelbundet) genereras en väsentlig tidsbesparing.

En del diskussioner fördes kring feedback-knappen. Den föreslagna texten i feedback-knappen löd: Tänk på att kritiken ska hjälpa eleven att gå vidare i sin inläring. Den återkoppling som lärarna skrev in i sina tester var, framför allt "du svarade rätt/fel på frågan"; vid flera frågor rätt svar skrevs in; en lärare uppmanade elever att fundera över svaret och kolla i boken.

Två av fyra lärare märkte först inte feedback-knappen, men sedan kom snabbt igång med den, när de väl upptäckte den. Alla lärare undrade när feedbacken visas för eleverna och om den ges även på rätt-besvarade frågorna. När vi frågade dem om deras intuitioner svarade de att feedback nog bara skulle visas vid fel eller delvis fel, vilket också är vår intention. En av lärarna påpekade att när man skriver feedback manuellt är man mer noga och flerordig när man rättar de första testerna, sedan skriver man mindre och mindre för varje elev, eftersom man blir trött på att skriva ungefär samma sak. De sista eleverna får, i bästa fall, en stjärna eller ingenting alls. Detta är ett mycket starkt argument för att använda digitala test, då alla elever får likvärdig feedback.

Elevperspektiv

De sex elever som genomförde testerna tyckte att det var enkelt att förstå knapparna. Samtliga elever föredrog digitala test framför papperstest eftersom de upplevde att det var jobbigare att skriva med penna. En av eleverna tyckte att upplägget av digitala tester var ett bevis på att lärarna började ta eleverna på mer allvar genom att lägga ner tid på att utforma dessa enklare och roligare tester samt att det var bra att få se svaren direkt och att veta vad som gick fel och varför. Detta

ser vi som en bekräftelse på en lyckad integrering av teknologi och pedagogik.

5 Diskussion

I denna rapport har vi presenterat vårt arbete med utveckling av ett enkelt digitalt test för lärare på grundskolan. Enkelheten för användare i tidsbesparingssyfte och möjligheten till återkoppling genom formativ bedömning har varit två aspekter som präglat vårt arbete.

Allt gick inte smärtfritt. Från början hade vi missförstått uppgiften och satsade på utveckling av ett test i linje med de som redan finns på marknaden såsom Fronter eller It's learning, det vill säga tester med många funktioner, som kräver breda datorförkunskaper från användare. Lärdomen som vi drar från detta är hur viktigt det är att studera och förstå målgruppen innan man börjar utveckla produkten. Vi har gjort intervjuer med både gymnasielärare och högstadielärare men blev missledda av gymnasielärarnas kompetens. De kom förvisso med värdefulla synpunkter på detaljer de önskade skulle finnas i digitala tester, men vi förstod först inte att de inte var vår målgrupp. De knappa önskemålen från högstadielärare tog vi inte helt på allvar och fortsatte att utarbeta en lång lista på testfunktioner tills vi förstod att högstadielärarnas brist på förkunskaper krävde ett annat sätt att angripa uppdraget.

I pilotstudie 1 med LoFi-prototypen där lärare skulle göra en matchningsuppgift hade vi valt att de skulle para ihop konstnärer med konstverk som färdiga frågor. Det blev för oss en lärdom om hur man ska hantera testpersoner. Exemplet som vi använde oss av gjorde den första läraren nervös. Hon kände sig osäker och valde att inte matcha alternativen. Om detta berodde på att hon inte förstod att hon skulle skapa ett facit eller att hon var rädd för att blotta sin okunskap får vi aldrig veta. Lärare kan känna ett tryck på att alltid veta svaret och detta hade vi inte tänkt på. Istället för att koncentrera sig på utprovning av funktioner i testet gick energi åt till att inte göra bort sig. Till användare nummer två ändrade vi snabbt exempel till ord och enkla ordklasser.

Vidare märkte vi att vårt scenario kanske var dåligt. Ingen av användarna utforskade riktigt hur feedback fungerade i vårt andra test. Man kan anse att vi borde haft med i vårt scenario att de skulle lägga till feedback. Samtidigt ville vi att de skulle upptäcka detta själva, för att se om våra affordances fungerade.

Tredje omgången av testande, som genomfördes i en mycket liten skala, gav oss ändå vissa inblickar i hur vårt verktyg skulle kunna användas. Enkelheten i form av testdesign och återkopplingen genom formativ bedömning i digitalt format samt möjlighet till direkt statistik av resultatet ger uttalad tidsbesparing för lärarna och roligare inlärningsmöjligheter för eleverna.

Vi planerade i kurs 2 att utveckla alternativen för återkoppling. I kurs 1 hade vi föreslagit ett antal exempel på formuleringar av formativ feedback, baserade på processinriktade och metakognitiva strategier. Vi var inte helt nöjda med dem och planerade att fördjupa oss i detta och förbättra formuleringarna under kurs 2. Vi ville även utforska möjliga konkretioner av formativ bedömning i kurs 2. Vi fick låta bli denna satsning dels på grund av att vi endast hade ett svarsalternativ att erbjuda, flervalsfrågor, dels därför att vårt test inte skulle genomföras i en helklass. Det är rimligt att

anta att lärarna inte lade ner mycket tid på att fundera ut lämplig feedback när det bara gällde enstaka elever. Ännu en anledning till att låta bli fokusera på feedbackformuleringarna var insikten om att lärare kanske inte var helt förberedda på att fördjupa sig i denna problematik. Vi bestämde att den ursprungliga texten fick stå kvar. Våra farhågor besannades i viss mån: lärarna använde inte feedback i så hög utsträckning. Vissa gjorde det och skrev medioker återkoppling i den avsedda rutan, men inte tillräckligt god för att verkligen kallas formativ. Detta kan vara ett tecken på att lärarna inte helt förstod vad menades med feedback och att det behövs en teoretisk fördjupning i formativ bedömningsproblematik, helt i linje med vad Hattie och Timperley (2007) säger om formativ bedömning och lärares kompetens.

En viktig aspekt av ett projekt som detta är att få en grupp att fungera. Särskilt komplicerat kan det bli när en grupp är heterogen. I vårt fall var det en grupp med olika kompetenser, åldrar och utgångspunkter. För att ändå lyckas och ha fokus på själva uppgiften gäller att angripa skillnader och lösa konflikter, samtidigt som skillnader alltid innebär styrka man kan ta till vara på.

För att komma åt detta har vi använt oss av idéer från Osvath (2014) om hur man får en grupp att snabbt komma över sina sociala spänningar och nå en fas av högsta produktivitet. Detta har bestått i att diskutera våra inbördes förutsättningar och vara öppen med om något är frustrerande eller inte går ens väg. I vårt fall har det till stor del handlat om att ständigt ha så lång planering framåt som möjligt så att en av våra gruppmedlemmar har kunnat boka in möten och delta i grupparbete. När vi väl ordnat med detta i god tid, har det heller inte blivit några problem.

Sammanfattningsvis känner vi tillfredsställelse med vad vi har kunnat åstadkomma under utvecklingsarbetet med det digitala testet. Det hade varit intressant att genomföra ett antal av sådana tester på klassnivå med flera typer av frågor och se hur feedback och statistik fungerar i verkligheten, då med en genomtänkt återkoppling för elever under denna serie av tester. Man skulle kunna genomföra en jämförelse mellan summativa prov i en kontrollgrupp och en grupp som arbetade med våra tester och som fick regelbunden formativ återkoppling. Det är då man skulle kunna utvärdera fullt ut testets enkelhetsaspekt och se dess pedagogiska värde som ett verktyg för formativ bedömning på flera plan.

Referenser

- Antoniou, P. and James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. *Educ Asse Eval Acc*, 26:153-176.
- Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41(3), 258.
- Atkin, J., Black, P. & Coffey, J. (2001). *Classroom assessment and the National Science Education Standards*. Washington, DC: National Academy Press.
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5–25.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–75.
- Black, P. J., & Wiliam, D. (1998b). Inside the Black Box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80, 139-48.
- Bocij, P. & Greasley, A. (1999). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology*, 1(1).
- Breakwell, G. M. (1983). Formulations and searchers. In G. M. Breakwell (Ed.), *Threatened identities* (pp. 3–26). Chichester, UK: Wiley.
- Cassady, J.C. and Gridley, B.E. (2005). The effects of online formative and summative assessment on test anxiety and performance. *Journal of Technology, Learning and Assessment*, 4(1).
- Chi, M., Schwartz, D.L., Chin, D.B., Pilner Blair, K. (2014). Choice-based assessment: Can choices made in digital games predict 6th – grade students' math test scores? *Proceedings of the 7th International Conference on Education Data Mining*.
- Fisher, D. and Frey, N. (2014). Formative assessment: designing and implementing a viable system. *Reading Today*, July/August 2014.
- Gulz, A. et al. (2014a). Projektbeskrivningar MAMN10 & KOGP05, Projekt 4: digTests. *Studentinformation*. Lunds universitet.
- Gulz, A. & Haake, M. (2014b). SvD Opinion: Ett redigerat urval av läsarfrågor med kommentarer, 2014-01-02. Hämtad: 2014-09-25 från www.lucs.lu.se
- Grönlund, A. (2010). Formativ bedömning - en översikt. Skolverket. Hämtad: 2014-09-20 från http://www.skolverket.se/polopoly_fs/1.100658!/Menu/article/attachment/formativ_bedomning_fordjupad_lasning_ny.pdf.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-122.
- Heritage, M. and Chang, S. (2012). Teacher use of formative assessment data for English language learners. *The Annual meeting of the American educational research association*, april 2012, Vancouver, Canada.
- Heritage, M. (2010). Formative assessment and next-generation assessment systems: Are we losing an opportunity? *Council of Chief State School Officers*. September 2010.
- Jenkins, J.O. (2010). A multi-faceted formative assessment approach: better recognising the learning needs of students. *Assessment & Evaluation in Higher Education*, 35(5).
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1(4), 279–308.
- Kurosu, M., & Kashimura, K. (1995). Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In *Conference companion on Human factors in computing systems*, 292-293.
- Lidwell, W., Holden, K. & Butler, J. (2003). *Universal principles of design*. Gloucester, Mass: Rockport.
- McLaren, S.V. (2012). Assessment is for learning: supporting feedback. *Int J Technol Des Educ*, 22:227-245.
- McLaren, S., & Stables, K. (2007). Exploring key discriminators of progression: Relationships between attitude, meta-cognition and performance of novice designers at a time of transition. *Design Studies*, 29(2), 181–201.
- Norman, D. A. (2002). *The design of everyday things*. Basic book, New York.

- Pellegrino, J.W. and Quellmalz, E.S. (2010). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education*, 43(2).
- Organisation for Economic Co-operation and Development (OECD). (2005). *Formative assessment: Improving student learning in secondary classrooms*. Centre for Educational Research and Innovation.
- Osvath, M. (2014) Föreläsning om grupprocesser. Lund: 2014-09-17.
- Price, E. (2011). Complex interactions between formative assessment, technology, and classroom practices. *Physics Education Research Conference 2011*.
- Skolverket (2010). *Adaptiva och andra datorbaserade prov. En kunskapsöversikt*. Stockholm.
- Skolverket (2013). *It-användning och it-kompetens i skolan*. Stockholm.
- Stull, J.C., Varnum, S.J., Ducette, J., Schiller, J., Bernacki, M. (2011). The many faces of formative assessment. *International Journal of Teaching and Learning in Higher Education*, 23(1).
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment*. Buckingham, UK: Open University Press. s. 14.
- Waters, J.K. (2012). Resolving the formative assessment. Catch-22. *T.H.E. Journal*. September 2012.
- Wise, S. L., Plake, B. S., Eastman, L. A., Boettcher, L. L., & Luken, M. E. (1986). The effects of item feedback and examinee control on test performance and anxiety in a computer-administered test. *Computers in Human Behavior*, 2, 21–29.
- Wise, S. L., Plake, B. S., Pozehl, B. J., Barnes, L. B., & Luken, M. E. (1989). Providing item feedback in computer-based tests: Effects of initial success and failure. *Educational and Psychological Measurement*, 49, 479–486.

Projekt MonZ

Rebecka Alves-Martins, Frida Bredberg, Niklas Grahl, Felix Hedlund, Carl Larsson Danell

Kungliga Musikhögskolan, KMH, i Stockholm har eftersökt en lösning för att digitalisera sitt antagningsprov i musikteori efter att ha insett vinsterna i form av tids- och resursbesparingar med en lyckad digitalisering. Efter denna insikt la de en beställning hos LTH, vilket resulterade i ett projekt som kallas MonZ.

Projektet vars mål varit att ta fram denna lösning har bestått av iterativa designprocesser som krävs för att ta fram de idéer och koncept den slutliga framtagna prototypen bygger på. Systemet som ska fungera som det digitala antagningsprovet ska erbjuda en mängd funktioner, vilka är framtagna ur en iterativ process där beställaren hittat nya önskvärda funktioner och tjänster efter att implementering synliggjort dessa.

Den konceptuella modell som utgör kärnan i det digitaliserade provet har utformats efter en julkalender vilket betyder att provet består av en övergripande vy där hela provet utgör ett musikstycke. Varje fråga i provet utgör en takt i musikstycket och kan besvaras först efter att provtagaren tryckt på den specifika takten. Den implementerade lösning som utgör prototypen för projektet är baserad för moderna webbläsare och ska gå att köra på de datorer som finns på KMH. För att navigera i det digitala provet har ett gränssnitt utformats och för att fullfölja ett prov ska provtagaren ha tillgång till dator, datorskärm, tangentbord, mus och MIDI keyboard.

Lösningen utgör endast en prototyp av en färdig lösning och för att färdigställa den bör noggranna användartester och fortsatt implementering fortgå. Det finns svårigheter med att digitalisera prov i musikteori, inte minst för att musikteori är ett komplext ämne som kräver stora resurser för att kunna digitalisera allt som är möjligt med vanlig penna och papper

1 Introduktion

Kungliga Musikhögskolan är Sveriges mest prestigefyllda utbildningscenter för musikstudier och varje år söker ett stort antal personer in med förhoppning om att deras musikkunskaper ska räcka för att bli antagna. De 500-600 personer som ansöker om de 20 platser som finns på jazzutbildningen får först göra ett prov, som fungerar som en gallringsprocess, där ca 100 sökande väljs ut för att göra ytterligare prövningar lokalt vid KMH i Stockholm. Detta inkluderar ett musikteoriprov där de sökande får visa sina kunskaper inom musklära, satslära och gehör. De som går vidare efter det steget kan tillbringa upp till sex dagar i Stockholm, medan de som sällas ut spenderar en till två dagar på KMH i samband med antagningen. För lärarna på skolan innebär rättning och bedömning av alla sökande mycket administration och jobb, då proven än idag görs analogt, med penna och

papper som enda verktyg - bortsett från ett keyboard. Uppgiften, som vi fick av beställare Torbjörn Gulz som ansvarar för Jazz-, Musik- och Medienriktningarna på KMH, var att digitalisera teoridelen av antagningsprovet och därmed föra in Kungliga Musikhögskolan i 2000-talet.

1.1 Bakgrund

Teoridelen av antagningsprovet till Kungliga Musikhögskolan består av tre delar: musklära, satslära och gehör, där vardera del tar en timme i anspråk och kräver olika hjälpmedel.

Musikläran är den del som består av fakta- och kunskapsfrågor, transponering och andra frågor som kan besvaras antingen rätt eller fel. Satsläran är däremot mer kreativt provande och kräver djupare förståelse i harmoni- och melodilära. Den innefattar även en kompositionsuppgift som måste bedömas. Slutligen har vi gehörsdelen som prövar de sökandes gehör och musikalitet. Uppgifterna på denna del går ut på att lyssna på ett musikstycke och transkribera detta, även andra uppgifter som testar sökandes gehör kan förekomma.

1.1.1 Mål med digitaliseringen

Ett mål med digitaliseringen är att sammanfoga dessa tre delar (musklära, satslära och gehör) av teoridelen till ett enhetligt prov, där provtagarna har tillgång till samma hjälpmedel genom hela provet. De direkta fördelarna med en väl genomförd digitalisering är att provtiden för de sökande minskar och därmed blir genomförandet av provet mindre uttröttande. Belastningen på provvakter kan minska betydligt då utdelande av material och administration av musikuppspelning kan upphöra helt. Då ett digitalt prov inte kräver lika sträng administration som ett analogt öppnar det upp för möjlighet till att låta de sökande själva boka in vilken tid de vill utföra provet. Även rättning har varit en omfattande process för personalen på KMH, vilket önskas minska tack vare digitaliseringen.

1.1.2 Tidigare försök till digitalisering

På KMH har man gjort ett tidigare (internt) försök med digitalisering, som resulterade i ett prov med enbart flervalsfrågor. Det gjordes i Moodle och riktade sig endast till lärarstuderenter. Det faktum att provet bestod av enkätfrågor gjorde att dimensioner av kunskap och kreativitet försvann, vilket därmed sänkte korrelationen mellan vad KMH ville att de sökande skulle visa att de kunde och vad de faktiskt fick visa att de behärskade. I detta projekt har nya vägar prövats och funktioner har satts ihop i nya konstellationer, vilket gör Projekt MonZ unikt.

1.2 Beställaren

Digitaliseringen medför större möjligheter i digital funktionellitet och en av utmaningarna har varit att undersöka och utforska dessa möjligheter, med tanke på att detta är ny mark. Gulz har under hela projektets gång också varit uttryckligen mån om hög användbarhet för att göra övergången från ett analogt till ett digitalt system smärtfri och smidig. Fördelen med en innovativ lösning som spränger tidigare gränser är att "bäst före"-datumet skjuts längre fram i tiden vilket betyder att lösningen ska vara hållbar en lång tid framöver och leder till sparad tid och kostnad för Kungliga Musikhögskolans IT-avdelning.

Målet på KMH är att använda vår lösning skarpt redan våren 2015, vilket sätter stor press på oss att leverera ett resultat som är funktionellt, relevant och framför allt rättvist för alla sökande, eftersom antagningsprovet kommer påverka de ansökandes framtid.

1.3 Intressenter

Den största intressenten inom detta projekt är KMH som har mycket att tjäna på en förenkling av den långa och resurskrävande process som utgör antagningsprovet i dagsläget. Andra intressenter är musikskolor som även de kan vara intresserade av en sådan lösning samt skolor inom helt andra områden men med samma typ av antagningsprocess.

1.4 Målgrupp

Målgruppen är av varierande åldrar då det inte finns en övre åldersgräns på att ansöka till utbildningen. Ansökande ska ha avklarat gymnasiet och lämpligen ett par år på folkhögskola vilket betyder att ansökande kan vara av endera kön och minst 20 år gammal med få undantag.

2 Generering av idéer

2.1 Genomförande av prov

2.1.1 Brainstorming

Under denna första och inledande brainstormingen diskuterades hur antagningsprovet skulle gå till för att provtagaren instinktivt skulle kunna lösa uppgifterna på ett prov. Gruppen var snabbt överens om att en webbsida var enklast och smidigast för att implementera det digitaliserade provet, då dagens webbläsare kan hantera avancerade funktioner och dessutom är plattformsoberoende. Ett program skapar dessutom problem när det kommer till uppdateringar och nedladdningar. Det enda som skulle krävas för att köra MonZ är en modern webbläsare. Däremot har valet av webbläsare som plattform genererat problem som att bibliotek kan behövas uppdateras och underhållas då webbläsaren uppdateras. Därav kom vi överens om att ett dokument för att sätta upp inställningar för datorerna som ska användas måste skapas där det beskrivs optimerade versioner av webbläsare och inställningar för dessa webbläsare, även vilken typ av webbläsare bör specificeras då vanliga webbläsare som Internet Explorer kan vara otillräckliga för denna typ av applikation.

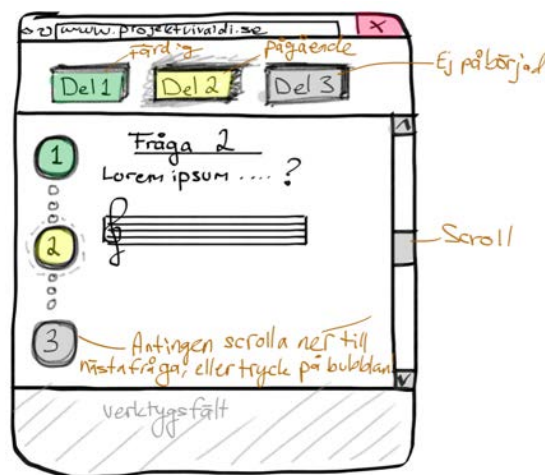
I samband med "brainstormingen" togs en linjär modell fram där man klickar sig vidare till nästa fråga sekventiellt.

Ett annat förslag handlade om en vertikal vy där provtagaren scollar vidare till nästa fråga vilket ger ett naturligt flöde. Samtidigt var beställaren av projektet var ute efter idéer som kunde bryta det traditionella konceptet av ett antagningsprov och kom därför själv med egna idéer som vi tog till oss och använde i vår slutliga lösning. De olika idéerna och koncepten som föreslogs under den här fasen baserades till största del på information från kurserna "Interaktionsdesign, grundkurs - MAMA15" och "Kognition - TEK210" givna vid Lunds Tekniska Högskola under läsåret 2012.

En stor del av projektet handlar om inmatning av noter och möjligheten att redigera dessa noter. Något som skulle utforskas vidare efter brainstormingen var vilka bibliotek som skulle kunna tillåta denna typ av inmatning och redigering. Det skulle komma att bli en kritisk punkt i projektet om inmatning av noter inte kunde accepteras i en webbläsare. Därmed blev detta något av det första som undersöktes efter brainstorming.

2.1.2 Förkastad idé

Efter den inledande brainstormingen tog vi fram en första Lo-Fi-prototyp enligt en iterativ process som präglades av gruppens egna åsikter samt beställarens krav och idéer. De första skisserna till Lo-Fi-prototypen gjorde vi efter ett relativt traditionellt provkoncept. Fokus låg på användbarhet och design i avseendet att webbsidan skulle vara så enkel att förstå och navigera i som möjligt (se *figur 1*). Provet skulle i det skedet bestå av tre delar - satslära, musiklära och gehör - varav vardera del bestod av ett antal frågor. I denna fas var orientering i fokus, vilket vi löste genom genvägsknappar längs vänstra vertikalen, för att navigera sig till respektive fråga.



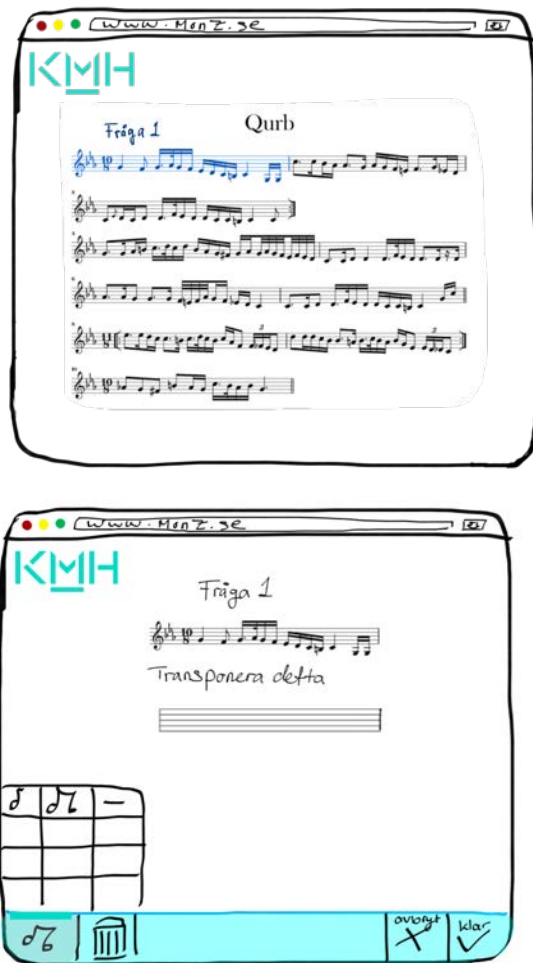
Figur 1: Lo-Fi-prototyp version 1

Avklarad/pågående/kvarstående fråga och delprov är färgkodade med de vedertagna färgerna grönt och gult, samt grå för att användaren enkelt ska kunna orientera sig och få en tydlig överblick av var i provet hen befinner sig. På så sätt skapas naturlig mappning. I och med att användaren besvarar en fråga och går vidare, ändrar den färg till grön parallellt med att den valda frågan blir gul, vilket ger tydlig feedback till användaren. Det var viktigt att få alla navigeringsknappar att se klickbara ut, för en tydlig synlighet.

2.1.3 Nytt koncept

Efter överläggning med beställaren som hade mycket idéer kring upplägget tog vi fram ett nytt koncept och en ny prototyp därtill ("LoFi-prototyp version 2"; se figur 2). Den traditionella modellen släpptes och ersattes av ett koncept med en julkalender som konceptuell modell [2], läs mer om denna modell i avsnitt 4. Huvudsidan består av ett notblad där frågorna vävs in i takterna, vilka är färgkodade, för att ge en tydlig överblick av var frågorna är i gränssnittet. Frågorna öppnas genom att klicka på dem och efter att en fråga besvarats kommer man alltså tillbaka till notbladet.

2.1.4 Lo-fi prototyp



Figur 2: Lo-fi-prototyp version 2

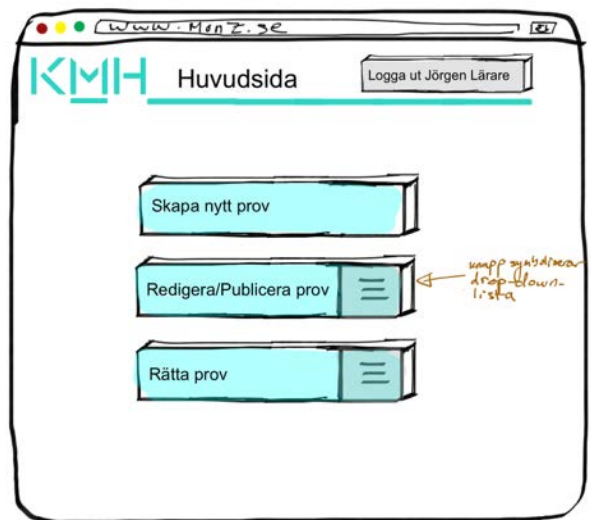
Den vänstra bilden i figur 2 ovan visar en möjlig översiktssvy över provet där takten som utgör den första frågan är markerad i blått. Tanken var att flera frågor skulle markeras, möjligtvis i olika färger, en idé som sedan avsågs vid implementering. Den högra bilden illustrerar vyn efter att frågan valts. Verktygsfältet består av en roll-up innehållande noter för notinmatning, radera-knapp samt dubbelkodade avbryt- och klarknappar. När användaren klickar på klarknappen ändras frågans färg på översiktssidan för att indikera att den är besvarad. Vid klick på avbryt-knappen sparas eventuella ändringar och frågans färg förblir oförändrad för att indikera att den ännu ej är besvarad så att användaren inte ska missa att fullborda någon fråga.

2.2 Konstruktion av prov

2.2.1 Brainstorming

Att systemet ska vara enkelt att använda har varit ett klart önskemål från beställaren. En annan faktor som påverkade tankegångarna till den provskapande delen av systemet var att provet ska fungera som en julkalender som tidigare beskrivits. Med denna utgångspunkt och att provet ska genereras från en bild skapat i ett externt musikprogram var den mest självklara och enkla idén att efter provbilden laddats upp ska läraren som skapar provet rita ut på bilden var frågorna ska hamna. När en fråga ritats ut ska läraren därefter ange vilken typ av fråga, svarsalternativ och eventuell mall för rättning.

2.2.2 Lo-fi prototyp



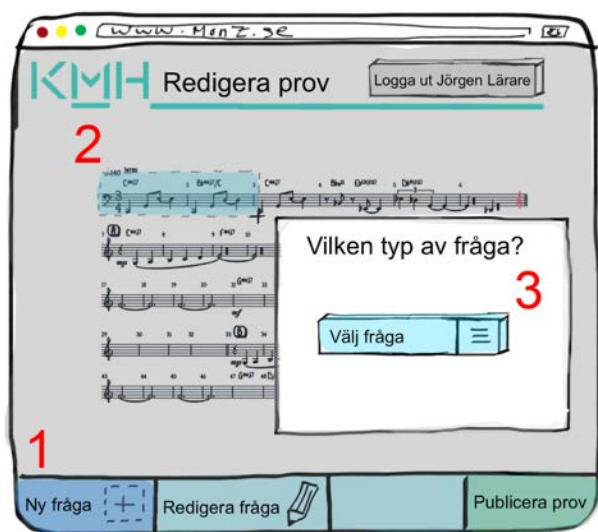
Figur 3: Lärarens huvudsida

När läraren har loggat in ska en huvudsida visas med alternativ för att skapa ett nytt prov enligt figur 3, redigera eller publicera prov samt för att rätta prov. De alternativ som kräver att läraren ska välja ett specifikt prov att redigera eller rätta ska tydligt indikera att ett knapptryck genererar en lista med möjliga prov där läraren sedan väljer vilket prov som ska hanteras.



Figur 4: Skapa nytt prov

När en lärare vill skapa ett nytt prov går man från huvudsidan till provskaparsidan genom att trycka på knappen för att skapa nytt prov. Provskaparsidan ser ut som figur 4 och låter läraren ladda upp en provbild som ska vara en horisontell bild skapad i Sibelius. Läraren ska även namnge provet och trycka på knappen för att skapa prov.



Figur 5: Redigering av prov

När läraren precis skapat nytt prov eller valt att redigera ett befintligt ska en sida visas med funktioner för att lägga till och redigera frågor samt en funktion för att publicera provet och därmed göra det tillgängligt för provtagarna. Enligt figur 5 går det att lägga till en fråga genom att:

- Tryck på knappen för ny fråga.
- Muspekaren är nu ett kors och läraren ska markera vilket område som ska utgöra en fråga.
- En pop-up ruta lägger sig ovanpå provet och bakgrunden får en grå ton. I denna ruta ska läraren ange vilken typ av fråga som skapas och därefter ange extra information som behövs i frågan innan pop-up-rutan stängs ned med ett knapptryck.

På liknande sätt ska en redigering av fråga vara möjlig. Då ska dock steg 1 bytas ut mot att trycka på knappen för att redigera fråga.

2.3 Rättning av prov

2.3.1 Brainstorming

Den sista stora delen som utgör systemet och som ännu inte berörts är hur rättning ska gå till. Med utgångspunkt i att systemet ska vara lätt att använda kunde en brainstorming genomföras för att ta fram en enkel idé. En matris över provtagarnas resultat blev en självklar och enkel lösning då läraren snabbt kan få en överblick över resultaten. Om rutorna i matrisen är klickbara och visar provtagarens svar på den specifika frågan blir det även väldigt lätt för läraren att hitta och rätta de frågor som kräver manuell rättning.

2.3.2 Lo-fi prototyp

Provtagare	Fråga 1 (Svarvalfråga)	Fråga 2 Gehör fast rytm	Fråga 3 Ackord gehör	Fråga 4 Ackordmål Ackord från notbild	Fråga 5 Notbild från ackordsymbol	Fråga 6 Inmålning av taktart	Totalt
LisaSten	2/4	6/8	2/3	2/3	7/4	4/4	7/26
CarSch	4/4	5/8	3/3	3/3	7/4	3/4	7/26
SarWoo	3/4	8/8	3/3	3/3	2/4	2/4	21/26
ChiLam	4/4	8/8	3/3	3/3	7/4	3/4	7/26
ManElm	4/4	7/8	3/3	3/3	4/4	4/4	25/26
SteRoo	0/4	0/8	0/3	0/3	7/4	0/4	7/26
CarLin	0/4	5/8	2/3	2/3	7/4	1/4	7/26
RasPer	0/4	4/8	1/3	1/3	7/4	0/4	7/26

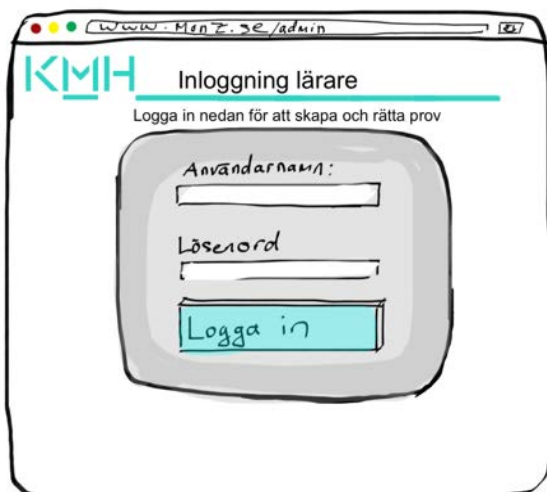
Provtagare	Fråga 1 (Svarvalfråga)	Fråga 2 Gehör fast rytm	Fråga 3 Ackord gehör	Fråga 4 Ackordmål Ackord från notbild	Fråga 5 Notbild från ackordsymbol	Fråga 6 Inmålning av taktart	Totalt
LisaSten	2/4	6/8	2/3	2/3	7/4	4/4	7/26
Lisa Stenström					7/4	3/4	7/26
					7/4	2/4	21/26
					7/4	3/4	7/26
					4/4	4/4	25/26
					7/4	0/4	7/26
					7/4	1/4	7/26
RasPer	0/4	4/8	1/3	1/3	7/4	0/4	7/26

Figur 6: Rättning av prov

Rättning av ett prov ska bestå av en matris där besvarade frågor står poängsatta för varje provtagare enligt figur 6. Varje ruta i matrisen ska vara klickbar. När en ruta klickas ska en pop-up visa hur provtagaren har svarat i gällande fråga. Om frågan rättas automatiskt blir den grönmarkerad och poängen visas enligt figur 6. Om frågan kräver manuell rättning blir den gulmarkerad och läraren kan trycka på frågan för att se provtagarens svar och därmed själv kunna skriva in hur många poäng provtagaren ska få och sedan spara rättningen, rutan blir då grönmarkerad.

2.4 Inloggning

2.4.1 Standard inloggningssida



Figur 7: Lo-fi-prototyp inloggning

Gällande provtagarens inloggning har det tidigare bestämts att provtagaren ska få inloggningsuppgifter av institutionen. Därmed är det rimligt att provtagaren loggar in på en traditionell inloggningssida enligt figur 7 och det blir lätt för systemet att kontrollera vilken provtagare som hör till vilken session. Detsamma gäller lärarens inloggning där tanken är att även lärarens inloggningsuppgifter ska komma från institutionen.

2.5 Iterativa genomgångar av systemet

En mängd idéer har tagits fram genom iterativa genomgångar av det befintliga systemet. Då funktionalitet testas och utvärderas i dessa genomgångar och man får en bild av specifika funktioner i en större kontext - en samlad bild över hela systemet - är det en bra metod för att generera idéer och tankar kring hur systemet bör se ut och fungera. Exempel på detta är hur vi testat att dra rutor för att skapa frågor i provskaparens vy, då insåg vi att det blir svårt att placera rutorna jämt vilket ger ett rörigt och inte särskilt överskådlig bild över provet. Därmed bestämdes att ett antal pixlar ska hoppas över för att lätt kunna placera frågor i raka led.

3 Funktioner och tjänster

Viktiga funktioner och tjänster som ska ingå i systemet har delats in i tre övergripande delar nedan.

3.1 Genomförande av prov

- i. Notinmatning - En provtagare ska kunna mata in noter med muspekare/styrplatta och/eller MIDI keyboard då denne besvarar frågor.
- ii. Olika typer av frågor - Det ska förekomma frågor på sju olika inmatningsalternativ:
 1. Enkätfrågor.
 2. Inmatning av melodi på gehör.
 3. Inmatning av ackord på gehör.
 4. Inmatning av ackordsymbol utifrån notbild.
 5. Inmatning av ackord i notbild utifrån ackordsymbol.
 6. Inmatning av taktart.
 7. Inmatning av rytm.
- iii. Besvara frågor - Frågor ska kunna besvaras delvis eller helt, det ska gå att återuppta en fråga och fortsätta besvara den.
- iv. Hjälpmedel för att besvara frågor - Musikspelare samt verktygsfält för att radera, göra om, ångra och återställa frågan.
- v. Frågor ska markeras som ej påbörjade, delvis färdiga eller färdiga med lämpliga markörer.

3.2 Konstruktion av prov

- i. En lärare ska kunna ladda upp en provbild skapad i det externa musikprogrammet Sibelius.
- ii. En lärare ska kunna skapa ett flertal frågor av olika typer.
- iii. En lärare ska kunna publicera ett prov för att möjliggöra provtagning.

3.3 Rättning av prov

- i. Om möjligt ska frågor rättas automatiskt.
- ii. Frågor ska visa vilka poäng provtagaren fått.
- iii. Läraren ska kunna rätta de frågor som kräver manuell rättning.

4 Konceptuell modell

En konceptuell modell är helt enkelt en vanligtvis mycket förenklad förklaring till hur någonting fungerar[3]. När vi refererar till den konceptuella modellen menar vi på både systemets konceptuella modell som beskriver hur det övergripande flödet av information ska synliggöras i systemet och även provtagarens konceptuella modell. Provtagarens konceptuella modell är däremot den mentala modell som ska fungera som en förklaring till hur systemet ska styras.

4.1 Konceptuell modell

Vår konceptuella modell är en julkalender, där provet - notbladet - utgör själva julkalendern och frågorna motsvarar luckorna. Genom att klicka på en markerad takt, som sym-

boliserar en fråga och därmed också en lucka, öppnas frågan i ett popup-fönster så att användaren fortfarande ser notbladet bakom frågan, precis likt en julkalender. Således kan man öppna och stänga en lucka utan att tappa bort provet. Systemets och provtagarens konceptuella design är densamma då systemet ska fungera som en julkalender och användaren ska bilda en mental modell av en julkalender då det är på precis samma sätt som en julkalender information visas i systemet.

4.2 Konceptuell design

Vi valde att ha en horisontell design, snarare än vertikal. Detta för att ge användaren en tydligare översikt över arkitekturen. Tack vare att frågorna öppnas i ett popup-fönster, ser användaren enkelt var hen befinner sig, och var hen kom ifrån. Att vara konsekvent med placering av knappar, underlättar navigeringen för användaren. Därför valde vi att utforma alla popup-fönster, oavsett typ av fråga, exakt likadant, med ett kryss i övre högra hörnet och sparaknappen i nedre högra hörnet. Detta För att förtydliga ytterligare för användaren.

5 Interaktion med systemet

5.1 Gränssnitt - visuell interaktion

Ett gränssnitt är utformat för att visa information för användare av systemet. En betydelsefull interaktion som däremot omöjliggör användning för en människa med synskador.

5.2 Musikspelare - interaktion med ljud

Musikspelaren utgör en kompletterande interaktion till frågor som kräver att provtagaren lyssnar på musikstycken. En interaktion som utesluter användare med hörselskador.

5.3 Interaktion via tangentbord

Ett tangentbord bör vara anslutet till användarens dator då en del frågor kräver inmatning från ett tangentbord. Även vid provskapande och rättning återkommer denna interaktion.

5.4 Interaktion via datormus eller styrplatta

Denna interaktion är avgörande för att kunna navigera i systemet.

5.5 Interaktion via MIDI keyboard

Ett av grundkraven var att inmatning via ett MIDI keyboard ska vara möjlig då klaviatur är ett vanligt förekommande hjälpmedel vid antagningsprovet på KMH. Denna interaktion möjliggörs genom att koppla in ett MIDI keyboard i den dator provet ska genomföras.

6 Implementerad lösning

Designprocessen har under hela projektets gång varit iterativ, vilket innebär att den slutliga Hi-fi-prototypen kom att skilja sig från den ursprungliga Lo-fi-prototyperna. Vi implementerade Hi-fi-prototyper redan från en tidig fas i projektet vilket har lett till att flera idéer har förkastats och nya designbeslut har fattats även på Hi-fi-nivå.

6.1 Genomförande av prov

Designen för provsidan är den vi lagt absolut mest tid på under projektets gång eftersom det är denna sida som kommer synas utåt för provdeltagarna. Att provsidan ska vara så avskalad och användarvänlig som möjligt är något vi lagt stor vikt vid.

6.1.1 Förkastade idéer

I och med en iterativ design process har flera idéer efter diskussion förkastats och ersatts med en nya, bättre lösningar.

Verktygsfältet

Verktygsfältet har designats om flera gånger på grund av ändrade krav från beställaren och nya lösningsförslag. I början av projektet implementerade vi verktygsfältet så att det skulle fungera för fri rytmik, det innebar att vi hade en knapp som dolde en roll-up-menü med olika musikaliska tecken, se figur 8. Efter diskussion med beställaren valdes istället att endast ha fast rytmik varvid roll-up-menyn ersattes till att endast innehålla förtecken och notsymbolen ändrades till en mer beskrivande symbol, dvs. ett förtecken.



Figur 8: Roll-up-menü

Även redigera-knappen (pennan i figur 8) och lägg till-knappen (plustecknet i figur 8) raderades från verktygsfältet då vi senare insåg att det gick att implementera så att det gick att lägga till eller redigera en not utan några knapptryck, dvs. endast m.h.a. muspekaren. I detta skede av projektet betydde papperskorgen att provtagaren kunde radera en inmatad not genom att trycka på ikonen och därefter markera den noten som skulle tas bort.

Ganska sent i designprocessen beslöt vi oss för att radera Avbryt- och Klar-knapparna till höger i verktygsfältet och istället ersätta dessa med en enda knapp i verktygsfältet, Spara, och med ett kryss längst uppe i högra hörnet. Detta designbeslut togs då ordet "klar" inte indikerade att lösningen sparades och för att avbryt-knappen inte indikerade att den stängde ner frågan. Eftersom vi eftersträvade en avskalad

design valde vi att bara ha en knapp till höger i verktygsfältet och istället flytta upp en knapp i högra hörnet för att förtydliga det och undvika en eventuell feltryckning.

6.1.2 Hi-Fi-prototyp

Nedan visas de prototyper över provtagande som blivit den slutliga implementeringen. Ändringar från tidigare diskuterade alternativ har skett.



Figur 9: Övergripande vy över provet

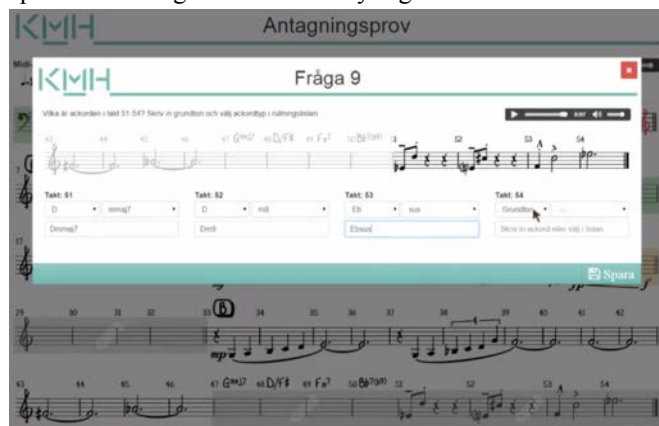
Ett stort genombrott i projektet var när det implementerades en ny färg för att indikera en icke fullt besvarad men påbörjad fråga. Frågan markeras alltså i en gulaktig färg och därmed ser provtagaren direkt vilka frågor som är helt besvarade enligt figur 9, vilka som inte har påbörjats och vilka frågor som ännu inte besvarats fullständigt. Därmed kan provtagaren sedan besluta sig för att lämna in provet fullt medveten om vilka frågor som saknar fullständig lösning. Musikspelaren har även fått en mer betydande roll i denna slutliga implementering då hela provet går att spela upp då provtagaren befinner sig i översikts-vyn. Däremot när provtagaren besvarar en fråga ändras ljudfilen till den specifika takten som utgör gällande fråga.



Figur 10: Fråga med flera alternativ i verktygsfältet

Verktygsfältets utseende har ändrats en del. Då provtagaren svarar på en fråga har papperskorgen fått betydelsen att frågan återgår till initierat tillstånd innan provtagaren börjar svara på frågan. För att radera enskilda noter som matats in har ikonen radergummi införts som visualiseras i figur 10. Det har även införts ikoner och funktioner för att ångra och göra om senaste ändringen med de vedertagna pilarna för att ångra och göra om. Som även kan ses i denna figur har frå-

gorna ändrats till popup-rutor med en snygg animering, detta för att provtagaren lätt ska kunna orientera sig. Det blir en självklarhet att man återvänder till den översiktliga prov-vyn när frågan har besvarats. Den nya knappen för att spara en fråga syns även i figuren, där har även den traditionella spara-ikonen lagts till för ett förtydligande.

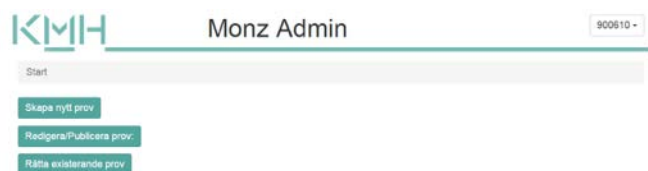


Figur 11: Fråga med begränsat verktygsfält

För att göra provtagandet dynamiskt och enkelt att använda begränsas verktygsfältet om frågan inte kräver extra funktionalitet enligt figur 11. I ovanstående uppgift ska ackord matcha beskrivna takter vilket endast kräver inmatning med rullistor och/eller inmatning i textfält med tangentbordet, därmed är det optimalt att enbart visa knappen för att spara då ingen annan funktionalitet behövs.

6.2 Konstruktion av prov

6.2.1 Hi-Fi-prototyp



Figur 12: Lärarens alternativ

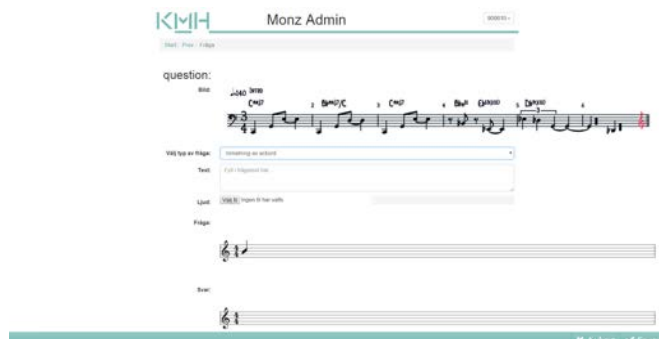
Enligt figur 12 visas alla lärarens navigation på en inledande sida efter inloggning. Delarna har betydande och mycket olika funktioner och bör därmed vara separerade på detta sätt.



Figur 13: Att skapa en fråga

När en lärare vill skapa ett nytt prov har denne navigerat till "Skapa nytt prov" där systemet ber läraren ladda upp en

provbild. Efter att provbilden laddats upp kan läraren gå vidare till vyn som visas i figur 13. Där går det att skapa frågor. Ny fråga väljs med plus-knappen nere i vänstra hörnet, därefter kan läraren dra med markören en rektangel som sedan blir den valda frågan enligt figur 13. Detta är även den vy som syns om läraren hade valt "Redigera/Publicera prov" i figur 12.



Figur 14: Formär vid skapandet av fråga

När en fråga har skapats med markeringsverktyget som användes i figur 13 kommer läraren till denna vy som syns i figur 14 där systemet ber om information till frågan. Det är sedan denna information som provtagaren ser när denne ska besvara den givna frågan.

6.3 Rättning av prov

6.3.1 Hi-Fi-prototyp

Provtagare	Fråga 1 text text text	Fråga 2 text text text	Fråga 3 text text text	Totalt
Niklas	4 / 4	0 / 4	7 / 4	4 / 12
Calle	4 / 4	4 / 4	4 / 4	12 / 12
Frida	4 / 4	4 / 4	4 / 4	12 / 12
Rebecca	4 / 4	4 / 4	4 / 4	12 / 12
Felix	4 / 4	4 / 4	4 / 4	12 / 12

Figur 15: Rättning av prov

Om läraren väljer "Rätta existerande prov" i vyn som visas i figur 12 navigerar systemet vidare till denna vy som syns i figur 15. Där visualiseras alla provtagares resultat med siffror som anger poäng. En fråga som kräver manuell rättning är gulmarkerad och läraren ska därmed kunna trycka på denna besvarade fråga och manuellt skriva in poäng på uppgiften. Därefter ska den manuellt rättade frågan grönmarkeras och poängen läggas till i provtagarens totala poäng.

7 Grafisk design

7.1 Färgval

För att bestämma färgpalett för provet besökte vi Kungliga Musikhögskolans hemsida [1], för att se vilka färger som redan används. Rött, svart och vitt dominerar, varpå de olika institutionerna (jazz, klassiskt, musik- och media osv.) representeras av vardera egen färg. Utifrån denna information fattade vi beslutet att anamma färgerna svart och vitt, för att

följa KMH:s färgtema och få provet att kännas som en del av musikhögskolan, och den dominanta färgen röd bestämde vi oss för att ersätta med en färg som inte redan existerar i sammanhanget, och som är neutral men ändå utmärkande med avsikt att ha en färg som förknippas med MonZ. Resultatet blev en nyans av turkos. Färgen används till logotypen, verktygsfältet, samt som markör för takterna. Nyansen av turkos ändrades en aning inför hi-fi-prototypen, då den i sammanhanget blev för utstickande och tog för mycket visuell uppmärksamhet.

Takterna är färgkodade med grått som indikerar en obesvarad fråga, gul för halvt besvarad fråga, och grön för besvarad fråga. När man håller musen över takten, vilken det ligger en fråga i, blir den turkos för att förtydliga var musmarkören befinner sig samt att klargöra att den går att klicka på. Efter att en fråga fullständigt besvarats blir tillhörande takt grön för att indikera att frågan har besvarats. Om användaren väljer att svara trots att något saknas, blir den gul. När man hovrar över en besvarad fråga förstärks den gröna färgen, återigen för att förtydliga att den går att klicka på. På samma sätt förstärks den gula färgen.

7.2 Typsnitt

Eftersom det är en webbapplikation är texten sans-serif, vilket gör att texten blir lättare att läsa på grund av att ögat har lättare att följa text som inte har fötter på en skärm. Vi har valt typsnittet Arial för att det är ett sans-serif-typsnitt och som dessutom är ett typsnitt som är anpassat efter en skärm. Arial skapades med Monotypes Grotesque som mall, men har modifierats för att passa ännu bättre till datorskrmar[11], med en förbättrad ordbild. Dessutom finns det installerat på de flesta datorer.

7.3 Layout

Målet var att ta fram en design som är ren och minimalistisk, eftersom fokus ska ligga på själva provet och dess uppgifter. På grund av omständigheterna kring antagningsprovet - pressen på provdeltagarna och tidsbristen - är det viktigt att de ansökande inte blir distraherade av gränssnittet. Detta tänk resulterade i en diskret bård längs nedre kanten som utgör verktygsfält och består av de få verktyg provtagarna har tillgång till. Notbladet som utgör provet ligger centrerat i mitten av sidan för att fänga provtagarens uppmärksamhet.

7.3.1 Verktygsfältet

Vid utformandet av verktygsfältet tog vi hänsyn till lagen om närhet. Knapparna är grupperade efter deras respektive funktionskategori. Till vänster ligger knappar som är relaterade till redigering och inmatning, och till höger de som hör till hela fönstret. Ikonerna vi valt är enkelt utformade och helt vita för att de ska smälta in i designen, men även sticka ut i sammanhanget. När man för musen över en ikon markeras den med en mörkare nyans av turkos för att tydligare visa var musmarkören befinner sig men även att ikonerna är klickbara.

8 Teknisk implementering och funktioner

I detta avsnitt beskrivs hur implementeringen har gått till på ett övergripande plan. Här behandlas kodstruktur, val av bibliotek samt implementering av specifika funktioner och tjänster.

8.1 Struktur

Koden är uppdelad i 7 mappar, se *figur 16*. Klientkoden ligger i mapparna “admin”, “main” och “shared”. Mappen “admin” innehåller filer för administrationssidan och “main” innehåller filer som endast används på huvudsidan (där studenterna genomför provet). Filer som delas mellan “main”- och “admin” ligger i mappen “shared”. Jämte dessa ligger mapparna “sass” (.sass filer) och “data” (original till bildfiler). Mappen “tests” innehåller filer för enhets- och integrationstestning.

För att strukturera upp klientkoden så är den uppdelad i komponenter. Komponenterna har varsin javascriptfil samt en HTML-mall i “partials”-mappen. Eftersom webbläsaren inte förstår sass direkt så genereras en css-fil utifrån våra sass-filer. Denna fil hamnar i “shared/css”-mappen. Mapparna “img” innehåller bildfiler som visas på hemsidan. Mapparna “js” innehåller de javascript-filer som används. I “api”-mapparna ligger filer för kommunikation mellan klient och server. Filer från externa bibliotek, bland annat AngularJS och Bootstrap, kopieras in i en ny “vendor” mapp.

```
~admin/  
|+api/  
|+js/  
|+partials/  
|-index.php  
|-login.php  
|-logout.php  
+data/  
~main/  
|+api/  
|+audio/  
|+img/  
|+js/  
|+partials/  
|-index.php  
|-login.php  
|-logout.php  
+sass/  
+server/  
~shared/  
|+css/  
|+img/  
|+js/  
|+partials/  
|+uploads/  
+vendor-static/  
+tests/
```

Figur 16: Kodstruktur

8.2 Bibliotek

De bibliotek som använts för front-end är: AngularJS[4], Lodash[5], VexFlow[6], Bootstrap[7], Web MIDI[8], HTML5 Audio[9] och SASS[10].

- AngularJS är ett ramverk som används för single-page hemsidor. Det ger också en mer fast struktur och översikt över koden.
- Lodash är ett bibliotek för vanligt förekommande JavaScript-funktioner.
- För att rita noter har vi använt VexFlow, som är ett JavaScript-bibliotek under utveckling, använts.
- För designen av gränssnittet har “front-end”-ramverket Bootstrap utnyttjas, varifrån olika designkomponenter kan hämtas till sin egen design.
- För inmatning av noter från klaviatur har Web MIDI använts. För att spela upp ljudklipp används HTML5 Web Audio API.
- SASS är ett språk som bygger ut CSS med extra funktionalitet och används i vårt projekt då det ger oss större renare och mer strukturerade stilmallar.
- De programmeringsspråk som används för back-end är PHP och MySQL eftersom det är dessa programmeringsspråk KMH använder på sina servrar och uppgiften är att interagera detta system med deras.

8.3 Funktionalitet

I den första delkursen låg störst fokus på notinmatning och design. De implementerade funktionerna var därmed främst till för provdeltagarna då uppgiften till hands till största del handlat om hur ett prov utförs och inte hur en lärare tillverkar provet. Detta var dock också en mycket viktig del som vi la störst fokus på under den andra delkursen. Här nedan presenteras några utvalda funktioner vi har implementerat.

8.3.1 Notinmatning

Notinmatning är den viktigaste funktionen eftersom det är denna funktion som skiljer detta projekt från den tidigare digitaliseringen.

Noter kan matas in med hjälp av muspekaren och verktygsfältet, eller keyboard, enligt fast rytmik. Fast rytmik innebär att det de ansökande ska göra är att mata in rätt not, utan att behöva ta hänsyn till rytm. I roll-up-menyn (se figur 5) kan man välja tillfälliga förtecken. Genom att markera en redan ifylld not, kan noten flyttas vertikalt. Suddgummit raderar den valda noten. Papperskorgen återställer hela frågan. Den vänstra cirkelpilen ångrar och den högra cirkelpilen gör om.



Figur 17: Roll-up-menyn

8.3.2 Musikspelare

I gehörsdelen av teoriprovet behöver provtagaren tillgång till en musikspelare för att spela upp musikstycket men även spola, pausa och ändra volym efter eget behov. Som musikspelare används HTML5 audio[8] elementet, som i utseende och funktionalitet följer en numera etablerad standard.

8.3.3 Ackordväljare

Ett ackord består av en grundton, ters och kvint, och inom jazz är det dessutom vanligt med så kallade färgningar. Detta ger utrymme för närmast oändligt många ackordkombinationer. En del provfrågor går ut på att mata in själva ackordnamnet och vår lösning till detta är två listor, där den ena innehåller alla grundtoner, och den andra alla varianter av ackord. Därtill finns en fritextruta där provtagaren manuellt kan skriva in önskat ackord via tangentbordet enligt figur 18. Då grundtonen valts, antingen i den vänstra listan eller via tangentbord, listas förslag på de vanligaste ackorden med den inmatade grundtonen. Vill man välja ett annat ackord är det bara att mata in resten av det önskade ackordet.

Figur 18: Ackordväljare

8.3.4 Taktart

Några provfrågor går ut på att skriva in taktart för takterna. En taktart beskriver antalet slag per takt och skrivs i bråkform, där nämnaren beskriver vilken typ av not (exempelvis åttondelsnot) och täljaren antalet sådana noter som finns i

takten. Vi valde att lösa denna typen av frågor med hjälp av två fritextrutor, en för täljare och en för nämnare, separerade med ett snedstreck. I rutorna kan taktarten antingen matas in via tangentbordet eller med uppåt-/nedåtpilarna längst till höger i rutorna. En annan idé var att istället ha två dropdown-listor med taktartsförslag. Efter diskussion valdes dock detta lösningsförslag bort eftersom det hade blivit för långa listor med alla möjliga taktarter som kan existera.

8.3.5 Markera som påbörjad/avslutad

För en provtagare är det viktigt att se huruvida en fråga är besvarad eller ej. Detta lade vi stort fokus på och gick från att bara ha två olika status (ej påbörjad och avslutad) till att istället ha tre olika (ej påbörjad, påbörjad, avslutad). Detta beslut togs då vi diskuterade oss fram till att det är viktigt för provtagaren att kunna se att hen påbörjat en uppgift, men ännu inte besvarat den fullständigt. En fråga markeras som påbörjad (blir gul) om en fråga inte besvaras fullständigt, exempelvis om alla noter inte har fyllts i. En fråga markeras som avslutad (blir grön) om all förväntad input har fyllts i. Vi tittar här alltså inte på huruvida svaret är korrekt eller ej, utan bara om provtagaren fyllt i det hen ska eller ej.

8.3.6 Rättning av prov

En viktig del av digitaliseringen var att implementera automatisk rättning för att underlätta för lärarna. Under den administrativa sidan finns möjligheten att se olika provtagares resultat på prov och även möjligheten att rätta frågor som inte kan rättas automatiskt.

Rättningsalgoritmen är konstruerad så att den jämför det svaret som provskaparen har lagt in, när frågan skapades, med det svaret som eleven angav. Algoritmen fungerar bra på de enklare frågorna där det finns ett tydligt rätt och fel men på de lite mer avancerade frågorna är den inte lika effektiv. På dessa frågor blir det istället upp till rättaren att göra en manuell bedömning av det inlämnade svaret.

Sammanfattningsvis så kommer systemet att underlätta rättningen genom att rätta de enklare uppgifterna automatiskt och på så sätt frigöra mer tid åt att rätta de mer avancerade uppgifterna.

8.3.7 Konstruktion av prov

Att underlätta för provkonstruktörerna vid provskapande var en stor och viktig del vid digitaliseringen av antagningssprovet. Vi tog därför beslutet att det var enklast för provkonstruktörerna att använda sig av ett notprogram som de redan är vana vid, i detta fall Sibelius, för att skapa notstycket och sedan importera detta till hemsidan. När provkonstruktören laddat upp notstycket och ska skapa frågor markerar hen takterna frågan skall omfatta. Markeringsvertyget erbjuder i dagsläget ingen hjälp med att markera takterna på ett enhetligt vis utan det är upp till provskaparen att placera ut dem på snyggt över notstycket.

Efter att takterna blivit markerade på ett önskvärt sätt trycker provskaparen på klar för att komma vidare. Provskaparen kan sedan välja vilken typ av fråga som takterna ska innehålla, ladda upp eventuellt ljud och ange frågans rätta svar. När detta är färdigt sparas frågan och provskaparen möts återigen av notbladet, fast denna gången med en mark-

ering för den nyligen skapade frågan. Det är också möjligt att avbryta och då kommer den påbörjade frågan inte att sparas.

Slutligen när provskaparen har konstruerat tillräckligt med frågor kan hela provet sparas för att sedan göras tillgängligt vid önskvärt provtillfälle.

9 För framtiden

I samband med att systemets funktionalitet allt mer motsvarat de krav som ställts har vi sammanställt en grund för användartester för att i framtiden kunna testa viktiga delar av systemet. Vid kontinuerliga användartester säkerställs därmed att systemet behåller den funktionalitet som behövs för att kunna utföra ett prov. Testscenarion för dessa användartester återfinns i bilaga A.

Projektet som avslutats med en stark prototyp, full av funktionalitet har den blivit en bra grund att bygga vidare på. Ett förslag är att antingen återuppta projektet vid nästa iteration av kurserna MAMN10 och MAMN15 alternativt göra ett exjobb av det för att fortsätta förbättra, implementera och färdigställa lösningen.

10 Diskussion

Vi insåg tidigt att den största osäkerheten i projektet låg i den faktiska implementeringen (speciellt notinmatningen), varpå vi bestämde oss att arbeta vidare med en HiFi-implementering. Detta betyder inte att vi har låst den konceptuella respektive grafiska nivån -- utan vi forsatte att iterera detaljer på alla designprocessens nivåer. Vårt mål var att använda input från användartester för att bekräfta de designval som är bra och eventuellt ändra de som inte fungerar som vi trott. Ett exempel är hur vi markerar frågorna på provet, obesvarade frågor är gråmarkerade med en penna som ikon och besvarade frågor är markerade i grön färg med en bock som ikon. Ett problem kan visa sig vara att provtagaren tror att frågan är rättad och markerad som "rätt svar" istället för enbart "besvarad", vilket är den underliggande tanken. Det är även intressant att se närmare på om gulmarkerade frågor anses vara påbörjade eller om en provtagare har en helt annan bild av denna aspekt.

Vi valde att utforma strikta användartest för att testa så att all implementerad funktionalitet fungerade på ett önskvärt sätt. Detta medför dock att testpersonerna kan bli väldigt låsta och eventuella designfel eller andra logiska problem kanske inte upptäcks. Vi är medvetna om detta men valde trots allt, i de första testerna, att lägga fokus på att interaktionen och funktionaliteten i systemet skulle fungera så bra som möjligt. Ifall mer tid hade funnits hade en logisk uppföljning varit att vid nästa testtillfälle använda friare användtester för att upptäcka mer av de eventuella problemen.

Vid skapandet av ett prov läggs stort ansvar på den lärare som skapar provet. Det finns inga begränsningar kring var på musikstycken läraren placerar ut frågorna eftersom musikstycken består av en vanlig bild. Detta medför att läraren själv ansvarar för att provet får en bra helhetsbild. Det är däremot möjligt att låta markeringsverktyget låsa sig till pixlar och därmed underlätta för provskaparen att skapa frågor som ligger i bra placering i förhållande till varandra.

10.2 Problem och motgångar

En svårighet i projektet har varit att vi endast har haft kontakt via e-post med vår beställare då han befinner sig på annan ort (Stockholm). Skriftlig kommunikation är lätt att missförstå och ger inte heller den direkta återkoppling som man ibland hade önskat. Vi tror att arbetet hade flutit på smidigare och att färre missförstånd hade alstrats om vi istället haft möjlighet att kommunicera ansikte-mot-ansikte med vår beställare.

Användartesterna har även de varit svåra att genomföra av den anledning att testpersonerna, dvs. elever och lärare på KMH, befinner sig på annan ort. Detta har inneburit att vi har fått formulera testfall som vi skickat via e-post till vår beställare och bitt honom att ta rollen som testledare och sätta ihop en passande testgrupp som kan genomföra de olika testerna. Vi har därmed varit relativt handlingsförlamade vad det gäller användartester från hemmaplan vilket har lett till att användartester inte hunnit genomföras och vi har därmed inte heller fått den återkoppling som vi hade hoppats på från användartester. Men i och med våra testfall har vi har lagt en grund för användartester i framtida studier.

Det moment som tagit mest tid under projektets gång är implementeringen av notinmatningen och diskussioner kring hur denna ska gå till. Det finns många olika tänkbara sätt hur man kan implementera detta och för oss som inte har erfarenhet sen tidigare av notinmatning har detta varit en stor utmaning. Vi insåg redan tidigt att vi inte skulle ha möjlighet att skriva ett komplett system för notinmatning då det för det första skulle ta alldeles för lång tid att implementera och för det andra skulle göra gränssnittet väldigt komplext och svår-förståeligt. Till en början valde vi därför att implementera de viktigaste delarna för att senare under projektets gång bestämma vilken funktionalitet som bör läggas till.

Vi valde att först bygga notinmatning med fri rytmik och sedan skala bort den funktionalitet som inte behövs då fast rytmik används. Fri rytmik innebär att provtagaren får friare tyglar och kunskapsmässigt blir det mer utmanande för provtagaren. Däremot uppstår problem med rättning, som vi vill ska skötas automatiskt. Fast rytmik underlättar med avseende på rättning, och dessutom innebär det att färre funktioner behöver implementeras i gränssnittet. Roll-up-menyn i sitt nuvarande utförande innehåller därför knappar som inte behövs i frågor med fast rytmik. I de tidigare iterationerna av notinmatningen delades inte koden för uppritning och koden för manipulation av noter upp i separata delar. Det blev dock ganska stökigt och vi insåg att koden behövde struktureras i olika delar.

Själva uppgifterna som ska lösas med vårt system är att antagningsprovets provdeltagare ska kunna lösa (mycket) avancerade musikteoretiska uppgifter. Konkret betyder det att det har varit svårt för oss som implementerare att veta exakt hur provet skall konstrueras och vad olika musikaliska termer betyder eftersom vår musikaliska kunskap är begränsad. Dessutom är både inmatning från keyboard och notritning ny mark för oss, och har varit väldigt tidskrävande att sätta sig in i. Dessa tekniker har ingen av oss behandlat innan och det finns inte heller någon redan existerande produkt att titta på eftersom alla intagningsprov till musikhögskolor är analoga i Sverige.

Sammanfattningsvis har balansen mellan implementeringsmöjligheter och önskemål från beställaren varit svår och vi har ständigt fått kompromissa och skalat av funktioner för att landa i en rimlig och realiserbar slutprodukt.

10.3 Arbetsprocess

Under kursens gång har projektgruppen jobbat iterativt och implementerat funktioner och design för att sedan efter samtal med beställaren ändra och komma fram till beslut som passar projektet. Olika projektroller som projektledare och utvecklingsansvarig delades ut i början av projektet. En överenskommelse var att vissa projektmedlemmar skulle byta projektroller för att skapa dynamik i gruppen. Med ny projektledare var det möjligt att komma med nya infallsvinklar då olika projektledare fokuserar på olika aspekter. Projektgruppen har även haft sessioner då projektmedlemmars mål och lojalitet till projektet synliggjorts, därmed har alla i projektgruppen varit medvetna om varandras engagemang och förståelsen för hur mycket tid och energi olika projektmedlemmar kunnat lägga på projektet har ökat.

11 Sammanfattning

Detta projekt har varit mycket omfattande och har utmanat och utvecklat oss såväl som programmerare och som designers. Beställaren har under hela projektets gång kommit med önskemål om ny funktionalitet, vilket tillsammans med implementeringssvårigheter har gjort projektet mer omfattande än vad vi till en början kunnat ana. Detta resulterade bland annat i att vi inte hunnit genomföra användartester, vilket vi från början hade räknat med.

Vi är överlag mycket nöjda med projektet och med vad vi har lyckats åstadkomma. Även om våra ursprungliga mål inte uppfylldes tillfullo så har vi har kommit en god bit på väg och har lagt en bra och stark grund för vidare utveckling av antagningsprovet.

Referenser

- [1] <http://www.kmh.se> - Hämtad 2014-10-16 13:26
- [2] http://www.eat.lth.se/fileadmin/eat/MAMN25/Konceptuell_design_MAMN25_HT13.pdf - Föreläsning i kursen MAMN25 Interaktionsdesign. Hämtad: 2014-10-17 09:22
- [3] Normann, Donald. Konceptuell modell, The Design of Everyday Things - Revised and expanded edition 2013, s. 25-37
- [4] <https://angularjs.org/> - Hämtad 2014-10-17 09:23
- [5] <https://lodash.com/> - Hämtad 2014-10-17 09:23
- [6] <http://www.vexflow.com/> - Hämtad 2014-10-17 09:22
- [7] <http://getbootstrap.com/> - Hämtad 2014-10-17 09:22
- [8] <http://www.w3.org/TR/webmidi/> - Hämtad 2014-10-17 09:23
- [9] <http://www.w3.org/TR/html5/embedded-content-0.html#the-audio-element> - Hämtad 2014-10-17 09:23
- [10] <http://sass-lang.com/> - Hämtad 2014-10-17 09:23
- [11] <http://www.fonts.com/content/learning/fyti/typefaces/arial-vs-helvetica> - Hämtad 2014-11-17 09:00

13 Bilagor

Bilaga A: Testscenarion för användartester

Scenario 1 - Öppna och besvara fråga 1

	Steg	Förväntat resultat	Ok (O) Fel (X)	Kommentar
1	Klicka på den första markerade takten	Tillhörande fråga (fråga 1) öppnas i en popup		
2	Klicka på krysset i hörnet	Popupen stängs och takten förblir grå		
3	Klicka på takten igen	Popupen öppnas		
4	Välj ett alternativ genom att klicka i en av cirkelarna	Cirkeln markeras		
5	Klicka på Spara	Popupen stängs och takten blir grön		
6	Klicka på takten igen	Popupen öppnas och alternativet som valts är sparat		
7	Välj ett annat alternativ genom att klicka i en av cirkelarna	Den första cirkeln blir omärkerad och den nya cirkeln markeras		
8	Klicka på Spara	Popupen stängs och takten är grön		

Scenario 2 - Musikspelaren och notinmatning

	Steg	Förväntat resultat	Ok (O) Fel (X)	Kommentar
1	Klicka på den femte markerade takten.	Tillhörande fråga (fråga 5) öppnas i en popup.		
2	Klicka på play-knappen i musikspelaren.	Ett musikstycke spelas upp och reglaget rör sig åt höger samtidigt som sekunderna tickar. Play-knappen har bytts ut till en paus-knapp.		
3	Klicka på volymslidern och dra det åt höger/vänster.	Volymen höjs/sänks och ljudindikatorerna förändras		

4	Klicka på reglaget och dra det åt vänster	Musikstycket spolas tillbaka och spelar därifrån reglaget släpptes		
5	Klicka på pausknappen	Musiken slutar spela och reglaget stannar där det var då knappen klickades på. Paus-knappen har bytts ut till en play-knapp.		
6	Klicka på play-knappen	Musiken fortsätter spela från där den var pausad. Reglaget rör sig åt höger. Play-knappen har bytts ut mot en paus-knapp.		
7	För muspekaren över slash-noterna i inmatningsnotraden	Slashnoterna blir till noter som går att bestämma tonhöjd på		
8	Klicka på noten på önskad plats	Noten fastnar		
9	Klicka på knappen längst till vänster i verktygsfältet (åttondelsnoterna)	En roll-up-menü dyker upp		
10	Välj ett förtecken genom att klicka på det	Förtecknet läggs till vid den nyligen inmatade noten och roll-up-menyn försvinner		
11	Klicka på redigeringsknappen i verktygsfältet (pennan)	Knappen markeras		
12	Klicka på den inmatade noten och dra den i vertikalt led	Noten och förtecknet rör sig i vertikalt led		
13	Släpp musen	Noten har fått en ny tonhöjd		
14	Klicka på knappen längst till vänster i verktygsfältet (åttondelsnoterna)	En roll-up-menü dyker upp		
15	Klicka på återställningstecknet	Korsförtecknet vid noten försvinner		
16	Klicka på raderaknappen i verktygsfältet (suddgummit)	Noten raderas		

17	Klicka på ångraknappen i verktygsfältet (cirkelpilen som går moturs)	Noten kommer tillbaka		
18	Klicka på "gör om"-knappen i verktygsfältet (cirkelpilen som går medurs)	Noten raderas igen		
19	Mata in en not genom att klicka på en slashnot i notraden	Slashnoten byts ut mot en fjärdedelsnot		
20	Håll muspekaren över noten	Ett nothuvud dyker upp på noten		
21	För musen i vertikalt led	Nothuvudet följer muspekaren		
22	Klicka på nothuvudet	Nothuvudet läggs till på den redan existerande noten och ett nytt nothuvud dyker upp		
23	Flytta muspekaren från takraden	En not med två toner är inmatad		
24	Mata in en not på nästa slashnot genom att klicka på den	Noten har inmatats där slashnoten var		
25	Klicka på töm-knappen i verktygsfältet (papperskorgen)	Båda de inmatade noterna raderas		
26	Klicka på ångra-knappen i verktygsfältet	Noterna återskapas		
27	Klicka på Spara	Pop-upen stängs och takten markeras gul		
28	Klicka på takten igen	Frågan öppnas i en popup och de två noterna är kvar		
29	Fyll i resten av noterna genom att klicka på slashnoterna	Slashnoterna ersätts av fjärdedelsnoter		
30	Klicka på Spara	Popupen stängs och takten markeras som grön		

Scenario 3 - Ackordinmatning

	Steg	Förväntat resultat	Ok (O) Fel (X)	Kommentar
1	Klicka på den sista markerade takten	Tillhörande fråga (fråga 9) öppnas i en popup		
2	Välj grundton i takt 51 i den vänstra drop down-listan	Den valda tonen visas i listan och i textfältet		
3	Välj färgning i takt 51 i den högra drop down-listan	Ackordets namn skrivs ut i textfältet nedanför drop down-listorna		
4	Ange grundton för takt 52 genom att skriva in tonens namn i textfältet.	Förslag på olika ackord kommer visas i en drop down-lista		
5	Välj färgning i drop down-listan som dyker upp vid textfältet	Ackordets namn skrivs ut i textfältet		
6	Ange ackord för takt 53 genom att skriva in ackordets namn i textfältet	En lista med möjligt/-a ackord dyker upp vid textfältet		
7	Välj ackord i drop down-listan	Ackordnamnet kommer att visas i textfältet		
8	Välj på valfritt sätt ackord för takt 54	Ackordets namn visas i textfältet		
9	Klicka på klar	Popupen stängs och takten är grön		
10	Lämna in provet genom att trycka på "Lämna in prov" längst ner till höger i verktygsfältet	En varningsruta med texten "Är du säker på att du vill lämna in provet? (Lämna in prov/ Avbryt)"		
11	Tryck på knappen "Avbryt" i pop up-rutan	Pop up-rutan försvinner och översiktssidan visas igen.		
12	Lämna in provet genom att trycka på "Lämna in prov" längst ner till höger i verktygsfältet	En varningsruta med texten "Är du säker på att du vill lämna in provet? (Lämna in prov/ Avbryt)"		
13	Tryck på knappen "Lämna in prov" i pop up-rutan	Provet lämnas in och utloggningssidan visas		

Behavior Emulating Non-humanoid Decision Enacting Robot

– an implemented cognitive model of social attention

Thobias Bergström, Erik Jansson, Tim Nilsson

kog13tbe@student.lu.se, atf08eja@student.lu.se, dic11tni@student.lu.se

The aim of the project is to first study, and gather data on, individuals' visual attention and resulting behavior in a group setting, and to use that data to create a cognitive model ready for export into the Ikaros open-source environment for controlling a robot arm (the agent). The ultimate goal of the project is to have the agent represent the attention of a human in a group setting by moving its head towards salient features of its surroundings, in the same natural way that a human would. Given the focus on social cognition, the most important features of its environment are the (heads of the) people in the agent's view and the direction of their gaze. The cognitive model is based on the affordance competition hypothesis and features action plans that compete for execution at any given moment. The weight given to each action plan depends on the current environment (what others are doing) and the data from our study. In this report, the study is presented and the cognitive model, its implementation and the underlying theories from cognitive science are described.

1 Introduction

Our original goal was to study the workings of human attention in the presence of other people. That is, where do we look when in the company of others? What draws our attention and what is the relation between our attention and that of the other group members? A substantial part of the project was to also demonstrate our findings by programming a robot arm (henceforth referred to as *the agent* or *the robot*) to move in accordance with the attention of a human in a group situation. To do this, we created a cognitive framework as a basis for a tailor-made AI.

We were given the freedom to allot our efforts and time between these two parts of the project, and as it turned out we put a majority of our efforts into the latter part. Of course, without data to support our choices in the implementation, producing a natural and convincing behavior on the part of the agent would be a challenge indeed. Attempts to find supplemental data in the literature yielded little in terms of human behavior data of the sort that would fit our aims. Our approach early on was therefore to get started with the software framework for the agent's movement and cognition

while also conducting a simple study, the details of which are described later in this report.

One of the main goals of present-day robotics is to bring a social dimension to our fellow humanoids (Balkenius, 2014). For robots to be a natural part of our everyday life – be it at home, in the workplace or in other human-centered environments such as health care – they must to some extent behave naturally, i.e. human-like. An early example of a social robot is the Kismet robot developed by Dr. Cynthia Breazeal at MIT in the late 1990s. The robot supports visual, audial and proprioceptive sensory input and transforms this into an output in form of a facial expression, a sound or a movement (MIT, 1998), together these features make it possible for Kismet to impersonate human behavior. Even though this robot was created in 1998 the area of research is still of importance. Newer projects such as Ikaros, which this project is a part of, use a similar way of thinking.

It is important to understand the intentions of the robot and to be able to predict its behavior in the same way that we can predict the behavior of humans (Leite, et.al., 2012). In our project we work under the assumption that being able to see and follow the attention of the robot is of considerable importance to achieve this aim. It is also important to make humans feel acknowledged by the agent before any real interaction can be established. There needs to be a simple way for a human to get the agent's attention. Humans usually do this by simply looking at someone. This prompts the other person to initiate eye contact. We will return to the implementation of these features, and data supporting these assumptions, in the *Attention* section of the paper.

Although human-like behavior is the result of myriads of such aspects, creating a module with a stand-alone cognitive model for attention is a contribution to the overall robotics project since it can readily be hooked up to a larger framework with rather simple means. So, what do we mean by "module" and "framework" in this context? This is where the second motivation behind our efforts comes in.

Our work is part of the Ikaros project, which was established in 2001 by Lund University Cognitive Science with the purpose to develop a platform for modeling the human brain and its functions (Ikaros-project, 2001). The

project is open sourced and any contributions that provide new or better functionality is added to the standard library.

The system-level approach means that a complete "brain" consisting of these modules would not be a replica of a human brain, but rather an emulation of its functions (Balkenius, 2010). Since work on implementations of this kind requires extensive knowledge of human brain functions on the one hand, and yields new insights about human cognition and relations between brain functions on the other, the interplay between robotics and cognitive science more generally is one of positive feedback (Brooks et.al., 1999).

The main sections of this paper cover distinct cognitive aspects of the agent: *perception*, *attention* and *action/cognitive control*. Each of these cognitive aspects has their own section where they are defined in a theoretical part and a practical part. The theoretical part consists of cognitive terms that were applicable for that specific cognitive aspect and why they are of relevance. Then in the practical part we describe how we actually implemented them, and what kind of generalizations that have been made.

2 Prerequisites

In the next sections of this report we will explain our cognitive model and its implementation. This will generally be done on a functional level. To fully understand our reasoning it is pertinent to know a little bit about the prerequisites and limitations of the hardware and software used.

Our robot consists of three servos (Dynamixel AX-12+) mounted atop each other to create a robot arm with three degrees of freedom. The bottommost servo is installed so as to give horizontal movement in 360 degrees while the other two are mounted to give vertical movement. However, it is not possible to allow for full 360-degree motion as this would cause the cables to wind around the arm and potentially disengage the connections. This limits all the servos to about 330 degrees of movement, thereby essentially designating a "front" of the agent where the "back" is inaccessible (the servos cannot move here).

To enable sensory input we employ a Microsoft Kinect (model 1414). The Kinect provides ordinary 2D visual input as well as a 3D representation of what it sees, as explained in the next chapter.

On the software side we've used a UNIX-system (Mac OS X) to code for, and run, the Ikaros program. All the code is written in C++ or C and can be found here: <https://github.com/MechaCorp/ikaros/tree/refactor/poid>

3 Perception

Since our agent is limited to visual input, its cognitive model must make the most of the available visual information and extrapolate from it whenever possible (e.g. draw conclusions

about who's talking based on information about what others are looking at). For our purposes the agent needs to know where people are and where they are looking.

Our implementation makes use of the 3D-input from the Kinect and puts it through a set of modules which interprets the information and delivers it to modules that determine which set of actions are possible based on the visual input.

Going into the setup and the actual hardware of the Kinect we can scale it down to two main features that are important to our project, the infrared-projector and the infrared-camera. The retrieval of the depth data starts with the infrared projector which projects "speckle patterns". The infrared camera perceives the speckled area and estimates where all the speckles are located. Since all speckle patterns are unique, like fingerprints, the two components can identify and pair their measures. Then it is just a matter of triangulating in order to estimate the distance from the Kinect to a certain speckle pattern area.

To find out where people are looking we use a head tracking algorithm which uses decision trees to find heads. These decision trees can take in some intricate data, in this case depth data from a Kinect-camera and then split it up to more manageable data.

The process starts with using a few reference images that works as a foundation for training the algorithm in its decision making (Fanelli, et.al., 2012). In each reference-image the goal position is marked, in this case the nose. The next step consists of choosing random sub-areas or so called patches in each individual image. Since the nose-position is known and by using enough training data the decision tree "learns" how to interpret the patches and the depth data which they consist of and most important where every patch is located in relation to the nose.

Based on this information, the algorithm creates binary tests that are used in the trees to come to as accurate conclusions about future images as possible. The structure of a tree is built around two concepts; an element is either a non-leaf node or a leaf node. The difference is that the non-leaf nodes contain the previously mentioned binary test and are responsible for guiding the collected data to a good estimation. The leaf nodes are end nodes of every branch. When these preparatory assignments are completed the algorithm is ready to be used.

This algorithm is a typical example of an artificial neural network using random forests (Fanelli, et.al., 2012) which essentially entails that it learns what to do by some sort of training-process. Generally during cognitive development a person constantly encounter faces which correlates to the training part in the algorithm. We learn how to interpret and distinguish the characteristics of a face and by this a type of expertise in facial-recognition is obtained (Gauthier, et.al., 1999).

Once the algorithm is applied to the images the information gained concerns both how many faces that are located in the image but also where they are located which was essential to achieve a solid foundation. These two types of functionality can be mapped to two main streams in an actual human brain.

The first one, known as the ventral stream, is in charge of identifying the type of object (the “what?”-stream) within visual vicinity, which the implementation somewhat imitates. A specific part of the ventral stream, which was of particular interest in this project, was the fusiform face area (the FFA). This area in the brain provides the functionality necessary for facial recognition, however it has been subject to a vast amount of discussions concerning if this area solely is responsible for facial-recognition or if it also can provide a functionality of recognizing objects of a type which a specific person is quite acquainted with, e.g. cars (Gauthier, et.al.,1999). Perhaps it is an “expert” area, and its fondness for faces is the result of social beings such as humans being “experts” at faces. No direct conclusions have been made concerning this area, but we can assume that this area is, at least, responsible for the facial-recognition in the human brain. So in conclusion, parts of the functionality, which the random forest algorithm contributes with, can be compared to the FFA in a biological brain.

The second stream is the dorsal stream, which estimates a position of an object (the “where?”-stream) and this is also information which the Random Forest-algorithm indirectly can retrieve. In reality, the dorsal stream transforms the image-input to a spatial map, which organizes the visual environment, and thereby creating positions for all relevant objects in relation to one another. According to some researchers, the dorsal stream might also be a “how?”-stream in the sense that its ultimate functionality is to guide visual motor control (Goodale & Milner, 1992).

To sum up, our agent is capable of recognizing heads; track their position in three dimensions; and to determine their pose. Its world, then, mainly consists of vectors terminating at the position of each visible head on one end and at the edge of its visual field on the other (or in some cases – e.g. when a person is looking at the robot – at the Kinect). It can of course “see” any other object, but for our purposes only heads (and objects close enough to “frighten” the agent) are salient.

4 Attention

At the heart of our cognitive model is the agent's ability to shift its attention towards the features of its environment in a way similar to that of a human. In accordance with our goal, such features consist mainly of the people (and their behavior) in the agent's view. What or whom are they looking at? It does, however, also take into account the

bottom-up feature *proximity*, which is used to avoid objects that are too close. Bottom-up means attention-pulling (as opposed to top-down, pushing) features (Henderson, 2007). There are many such features we could have taken into account (contrast, bright colors, geometric shapes etc.) had we had the time, but we settled for sudden, large movements in the agent's proximity.

Given the relative simplicity of our cognitive model, the interplay between input, our modules and the environment is straightforward. It is perhaps even misleading to talk about the agent's attention since our goal is to model human attention and so the agent is in a sense *a representation of attention*. The agent doesn't pay attention, it embodies it. The agent's attention is the result of its awareness of a series of predetermined types of affordances in its environment (Gibson, 1979; Norman, 1999) That is, potential, competing action plans which are categorized and prioritized by our cognitive model and its constituent modules. In each moment one of these plans will be the preferred one and as such will be carried out until completed or aborted in favor of another, “stronger” plan or the onset of idle behavior. This also reflects what Paul Cisek (2007) calls *the affordance competition hypothesis*. According to Cisek, neurophysiological data suggests that this kind of decision making is parallel and not serial as traditionally believed. The agent does not need to construct representations and plan ahead, but rather has all the action plans already available and merely compete for execution.

In our agent, the carrying out of a model is a representation of the agent's overt attention. Human attention and its tie to action plans and ultimate action are of course far more complex, but there are examples even less complex than our agent in nature. Consider Uexküll's tick:

“...this eyeless animal finds the way to her watchpoint [at the top of a tall blade of grass] with the help of only its skin's general sensitivity to light. The approach of her prey becomes apparent to this blind and deaf bandit only through her sense of smell. The odor of butyric acid, which emanates from the sebaceous follicles of all mammals, works on the tick as a signal that causes her to abandon her post (on top of the blade of grass/bush) and fall blindly downward toward her prey.” (Uexküll, 1957)

A digital version such a simple attention system can be found in Agre & Chapman's (1992) computer game playing program *Pengi*. It uses indexical-functional aspects of its local environment to guide its behavior from moment to moment. That is, instead of representing an object symbolically and labeling it, the program only notes its relevant features, e.g. “the-block-next-to-me”. In the same way, our agent attends to “the-head-at-the-intersection-of-two-or-more-gaze-vectors”. Of course, once this identification is made, our system can label that head to keep

track of it until its no longer visible, but this is not a necessity for a functioning attention system. Instead of representing its entire *Umwelt*, it keeps track of what is of importance in the here and now creating the agent's *Merkwelt* (Uexküll, 1957). It has only a few rules to follow, and they are only adhered to when the situation allows or requires it.

As it turns out, a lot can be accomplished in terms of function with the simplest of mechanics. The trick is, for our purposes, to strike a balance between complexity and (apparent) function. Efforts to emulate human behavior are subject to diminishing returns and a good approximation can go along way. This is especially true when the agent looks nothing like a human and as such stays clear of the uncanny valley effect (Masahiro, 1970).

5 The study

Before going into the details of our solutions, it is useful to present the study we conducted. Since the behavior of the robot arm is to be modeled on human behavior and attention, a study of this kind was needed for head movement and observer behavior data. One of the participants (henceforth referred to as 'the observer') unknowingly had the role of the agent whereas the others represented people with whom the agent interacts. The behavior of the observer would be the basis for the behavior of our agent.

The goal of the study

The overall goal of the study, then, was to gain a better understanding of how a person encountering a group of people behaves. Questions to be answered include

- What are the factors that contribute to the shifting attention of the observer and how often do these factors cause a response?
- What are the standard durations of observer focus, e.g. how long until eye contact is broken?
- To what extent does the observer exhibit idle behavior, “detached” from the situation?

To answer these questions, a simple experiment was carried out where one participant was the observer (the same role as the robot will have) and three others were the observed.

Participants

Twelve participants (nine women) were recruited for the experiment. All were students from the LUX center for humanities and theology. Age data was not collected. Participants did not receive payment or reward for their participation.

Material

One laptop and one tablet was used to record the group from different angles. The laptop was placed on the table in front of the observer with its screen blacked out during recording so as to not create distraction. The tablet was placed in the corner of the room to catch an overview of the room and the participants.

Procedure

The experiment was carried out with one group of four at a time. One of the participants in each group was asked to come into the room first and have a seat. This was done under the pretense that we had not brought enough consent forms for all of them to read before entering. Once the first participant (the observer) was seated (recording was started at this time) and they had all read the consent forms, the other participants were asked to enter the room and sit down. They were then given a simple instruction: to discuss the LUX building, which was a new addition to the campus at the time, and to agree on the best and the worst thing about it. The experimenter then left the room and the group was left to their discussion. After three minutes, the experimenter entered the room again, announcing the end of the discussion and revealing the actual purpose of the experiment.

Analysis

The analysis of the recorded data was done using the ELAN video annotation software with tiers for several types of behavior. This was done in two stages: first the videos were studied with the goal of identifying the observer's behaviors relevant to us (see the compiled list below). With this list of behaviors as tiers in the ELAN annotation system, the videos were watched again and all occurrences and durations of those behaviors were noted. Statistics was collected for number (per minute), minimum-, maximum-, and average duration (in seconds) of:

- The observer looking at the one currently speaking.
- The observer looking at someone other than the one speaking (for no apparent reason).
- The observer looking at someone other than the one speaking (reacting to conversational support such as a nod or an “mhm”).
- The observer exhibiting idle behavior.
- The observer reacting to a sound or event (e.g. the door opening).
- The observer looking at what someone else is pointing at.
- The observer looking at someone entering the room (first glance).

Taking into account the fact that the agent will not be doing any talking—whereas our observer was not asked not to talk—meant that we did not record behavior data for the sections where the observer was talking. We also collected data for durations of behavior leading up to idle behavior. An example of this would be the observer attending someone who was talking and then averting attention (for no apparent reason) even though the person was still talking. Lastly, we attempted to establish to what extent stimuli such as 5, 6 and 7 in the list would shift the observer's attention. More on this in the discussion at the end of this section.

Results

Statistical analysis of the ELAN data showed that, as suspected, the majority of the observer's attention was focused towards the person talking (52.5%). See table 1 for the complete results. The average duration of attention before onset of idle behavior was found to be 5.43 s ($SD = 1.76$ s) on average.

Table 1. Result of the analysis

Annotation type	No. of occurrences per minute	Min. duration (s)	Max. duration (s)	Avg. duration (s)
1	12	0.6	11.1	3.5
2	4.7	0.4	2.3	1.0
4	5.1	0.4	4.5	1.9

Discussion

Given the overwhelming majority of the observer's attention being put towards the person speaking, it would make sense to make this the default behavior of the robot (whenever someone is, in fact, speaking). The other behaviors observed in the study could then be seen as exceptions, some with concrete causes, such as another person performing a non-utterance as conversational support (“mhm”, “yeah...”); some without any apparent cause, such as idle behavior or the apparently random switching of attention towards another person who is not talking. If no one is currently speaking, idle behavior or looking at a random person are likely candidates for default behavior. Details on how we chose to use this data in our cognitive model to follow.

In the case of 5, 6 and 7 in the list, it isn't very useful to talk in terms of occurrences per minute, since those behaviors are very much contingent on something triggering the behavior, that is bottom-up saliency. What is interesting is to see how often such events do in fact trigger the behavior in question. Collecting this data proved very difficult, however, since two cameras were not enough to cover the

entire room and all the participants from every required angle. Our best estimate from examining the videos is that every time someone new entered the room, the observer turned to look at them. We also found that every “event”, like the beep from the door's keypad, drew the attention of the observer. These last data points being somewhat unreliable, we decided to program the agent to attend such events every time they happen. For our purposes this seems a reasonable approximation.

Conclusions

The decisions we made based on this data are as follows (numbered according to the list of behaviors from the analysis):

1. If someone is talking (i.e. if someone is the focus of the others in the group), then the default behavior of the agent is to look at the person who is talking. Given the data, this is an obvious choice.

2. The observer's decision to look at someone other than the person who's currently talking appears to either be caused by conversational support expressed by someone else (e.g. nodding or agreeing verbally) or some seemingly random internal trigger. For our purposes, and with our limited hardware (e.g. no audio recognition), implementing this as a function for randomizing whom to look at makes sense. That is, if there are no other attractors of higher salience, such as someone talking, an event or a new person entering the room, randomly scanning the group members will be the default behavior.

3. Reacting to conversational support from others than the one currently talking proves nearly impossible since there is no audio support in our system. This behavior was not implemented.

4. Idle behavior. There are several ways this could be implemented, e.g. completely random, with increasing probability of activation. We settled on a “boredom” solution that is reflected in our data. If the same action plan has been executed for a given amount of time (this interval may be randomized within the variance of the data) then the agent will become bored and start looking elsewhere.

5. We concluded that reacting to a sound or an event should be of such high priority that it is triggered every time an event occurs. Since there is no audio recognition in our agent, joint attention based on the head poses of other persons will be used. The higher the number of people who look at something, the faster the motivation to do the same (look at the same point in space) will increase. It will appear as if the agent is curious about what others find so interesting.

6. Someone pointing only occurred once in our study, so there is no way to draw any conclusions from the data. Furthermore, we did not prioritize the implementation of a

pointing detector (partly because we did not expect pointing to be a common occurrence) and ended up not including it.

7. As stated above, to the best of our knowledge, every time a new person entered the room the observer at least glanced at that person. We decided to have our agent always turn to a new person. For our purposes approximations such as this produce quite natural behavior with simple means.

Our study and subsequent evaluation resulted in several modules or behavior action plans. These will be covered next.

6 Implementation (attention)

Meet gaze / eye contact

The *meet gaze* module uses the fact that the angles for head pose tracking provided by our algorithm originates from the center point of the Kinect, meaning that yaw, pitch and roll are all 0.0 when staring directly into the device. The closer to looking at the agent (i.e. straight into the Kinect), the higher the action plan strength gets. Our current implementation lacks a temporal dimension (except for some smoothing applied to all strength values). One obvious way to handle this would be to let the strength be low until a time threshold is reached whereupon the strength would rise until reaching its maximum value of 1.0. After this, some form of habituation could be implemented whence the strength would start to decline again even though the same person is still staring at the robot.

Joint attention

Our agent does not have a sense of the room it is in, it has no object permanence or even object recognition. It only knows about the people (or heads, rather) that are in front of it. This makes it rather hard for it to attend objects or surfaces in the room, let alone things that are outside the limits of its visual perception.

Since we have limited our scenario to just three people in front of the robot there are never that many objects to attend to, so we decided to focus on situations where the joint attention of the group is explicit. Currently this module responds to people looking upwards (presumably at the ceiling) or downwards (at the table in front of them). By checking how many people are looking upwards (a pitch higher than 20 degrees from looking straight at the robot) or downwards we calculate the action plan strength using an algorithm that simply multiplies the amount of people looking a certain direction with a factor that differs between the ceiling and table. This means that the strength increases faster the more people that are looking the same way. We originally created the algorithm used to accomplish this to be a general purpose joint-attention algorithm, that is, capable of making the agent look at *whatever* others are looking at, but as it turns out this only made things too complicated

given the fact that its field of vision is restricted. Apart from the people in front of it and the space above and below them, there was little for the agent to attend to. If our perceptual system (including its hardware) could be improved to include the whole room and a greater depth, the algorithm could in principle again be generalized. Alternatively, a coordinate-based representation of the room could be used.

Person of Interest

Without being able to track sound location it is very hard to know which person is talking. By using head tracking data we can approximate this by assuming that the other people will turn towards this person while listening. However, most of this movement is accounted for by moving ones eyes instead of the entire head. To compensate for this we can, again, take advantage of our limited use case of three persons in front of the robot by simply assuming that if two people are looking the same way they are looking at the third person (if they are not looking away that is). This allows us to use rather small head movements to detect when someone is talking.

7 Action / Cognitive control

The path from the agent's forming a goal and its performing the appropriate action very much reflects the motor pathway in the human brain. In a given situation, the visual input presents the agent with the affordances of its environment and activates the corresponding action plans. The choice of which of them to act upon is then made in our attention modules and an instruction (goal position in degrees of separation from the current position) is sent to the servos. In a human brain, this is obviously very complex, but the principle is the same. Action goals are generated in the association cortex and signals propagate to the supplementary motor cortex and the basal ganglia where movement selection and switching between actions, respectively, is carried out. The signals also reach the lateral cerebellum where movement patterns are prepared. The path goes through the intermediate cerebellum where the movement patterns are corrected and implemented and finally to the motor cortex which initiates the actual movement of the muscles (Gazziniga, 2014; Ward 2006).

The servos have different parameters that can be manipulated to create the desirable movement. Throughout the project, two parameters have been of interest. These are the speed and torque parameters. Tuning these two, and combining them with an algorithm that imitates a critically damped spring, achieved a satisfyingly smooth and organic movement.

The combination of the servos and the modeling of the brain in the actual code are comparable to the concept of proprioception in an actual human body. In the human body

some nerves cells are assigned to observe and notify the brain regarding the whereabouts of a certain limb. The direct link to this project is that the servos send back information about their current location and the code (our cognitive model) takes in that information and can by this adjust the position of a limb in a satisfying manner. In a normal human being this is only half the truth, in the human mind the proprioception is combined with the location of an object based on the perceptual input. The perceptual and the proprioception are then weighed together and combine to an estimation of a limbs position, this is known as the sensory-motor transformation (Ward, 154). For our purposes, retrieving the information from the servos has sufficed, so there is no need for visual feedback of the robot arm's position.

Normally as a human being moves their head this changes the field of view, nonetheless since the robot's field of view in this project is predefined to a constant area this gives a more concrete explanation to why the Kinect was not placed on top of the robot. In this context it is more accurate to say that the head movement is restricted by the field of view as opposed to reality where the head restricts the field of view. So to summarize this from a functionality perspective, the system is constructed around external stimuli, in this case from the Kinect, which is then converted to a possible action that may or may not be perform by the system.

8 Implementation (action)

Our cognitive model has several "action plan" modules. These modules run in parallel and respond to external stimuli (in most cases head movement data). Every module is sensitive to one type of action (as previously listed in the attention implementation chapter) and outputs a plan (coordinates for the servos to move to) and a plan strength (a float between 0.0 and 1.0) representing the strength (sort of like saliency leading to motivation) of the action plan. This is analogous to how the human brain reacts to affordances: *"The brain uses the constant stream of sensory information arriving from the environment through sensorimotor feedback loops to continuously specify and update potential actions and how to carry them out. That's the affordance part"* (Gazzaniga, 2014, p. 347).

The action plans are then sent to a main module (simply called "Brain" in our code) which chooses which plan to execute based on their strengths and the order which they are specified in the .ikc-file. The action plan strengths are also smoothed using a simple boxcar-smoothing algorithm in order to make the system less jittery. The brain module also handles idle-behavior by running a scripted action plan when another action plan has been executed for a long time (based on data from the study).

While an action plan is executed the coordinates are continuously sent through the movement smoothing module and finally to the servos. This creates a smooth movement in the agent as well as easing the transition between action plans when the agent changes behavior. This means that we never have to send "end" coordinates (to where the servos will be positioned at the end of a movement) to the servos, instead we continually update the current position of the servos throughout the entire movement. This allows the agent to dynamically switch between action plans without stuttering or having to finish movements.

Our agent has no hierarchical representation of these sequences; all actions are represented on the same basic level of movement instructions. This means that it neither uses world coordinates, nor (as previously mentioned) the object location to navigate. Although we use afferent signals from the servos to find out how they are currently positioned, if one were to disable this feedback, the agent would be unable to complete the action. This is in contrast with how humans do it (Gazzaniga, 2014) and it would probably be preferable to use some kind of coordinate system instead to avoid these potential issues.

9 Discussion

Ultimately, the goal of this project was to make the agent appear engaged and attentive. The question is: did we succeed? Well, the agent looks nothing like a human, in fact even its human features (head, eyes, mouth) are based on a fictional robot (Bender from *Futurama*). But even a fictional, cartoon character can appear human-like, so to a large extent it comes down to the agent's movement rather than looks. So, the question becomes, is the movement convincing? We will get back to this in the concluding remarks below, but first we will suggest a few potential improvements that we have considered but not had time to implement.

Beyond our project: suggestions for next year's project

One limitation that we encountered was the Kinect and its flaws and weaknesses. Its narrow field of view restricted the area in the room where people could be seen and ultimately limited group size to three people. Furthermore, there was not enough room for them to move around so we ended up having them sit at the table, facing the Kinect. A possible improvement would be to use multiple Kinects and correlate their information to create a greater *Merkwelt*. Of course, this alternate solution would present problems too since the synchronization-process would be quite demanding, but it would make the system more flexible in terms of the amount of people who could join the conversation. Another solution, as previously mentioned, might be to place the Kinect on top of the servo constellation. This would make all world coordinates relative to where the agent is looking at and thus

create a more “embodied” agent by merging perception to motor action.

Our implementation of the *PersonOfInterest*-module does not work properly when only two persons are detected. Since they probably will be looking at each other it is impossible to tell who is doing the talking.

The current solution uses the known location of the servos in terms of degrees and if a change is desired new goal-destinations expressed in angles will be sent. However it is probably more desirable to use coordinates, meaning that a coordinate system is created around the robot and by this you would send coordinates instead of angles to adjust the robots position. This is actually something that is being worked on by another Ikaros-project and by the end of this project a satisfactory solution was achieved in that project but not incorporated into our model due to time constraints.

The *Joint-attention*-module’s implementation could perhaps be improved using a statistical approach, possibly using the standard deviation of all the observed individuals’ visual focus. If the focuses of the observed individuals have a strong correlation, the standard deviation would be low and this would increase the strength parameter in the Joint-attention module.

Another candidate for implementation is what is called *displacement activities*. A displacement activity can arise when the brain encounters multiple potential actions to perform where it is incapable of selecting the accurate action at that time (Sevenster, 1961). When this occurs it is not uncommon that a random behavior emerges and is performed instead. A typical example of how this is displayed is when a person scratches their head when facing a dilemma where decision-making is required. As mentioned earlier, our agent works with the strength of different action plans; some rise and decay much faster than others. Ideally to create authentic behavior, including on the theory of displacement activities, the agent should perform a random lower prioritized behavior if it encounters multiple concrete actions with similar action plan strength. Given our time constraints however, it was decided that the randomized behaviors already implemented (idle behavior and scanning the group of people) lended enough realism.

Currently, the Kinect’s field of view is divided into circle sectors where the agent detects one face per sector. If a person would leave its designated sector and a new person enters that particular sector, our model interprets them both as the same person. In that sense, the agent cannot discern different faces. It would be more useful if the implementation could remember the characteristics of a unique face to achieve more human-like behavior. The idea of sectors is a quite rigid solution, which requires little movement of the observed individuals. So identifying every

unique individual and being able to separate them from one another would be preferable for further development.

The 3D-tracker for head movement is a good start in terms of interpreting an individual’s focus, but one thing that should also be taken into account is the eye-movement. An ideal solution would be to combine the head tracking with eye-direction. A small movement of the eyes may not seem like that big of a deal but it can in fact change the conditions considerably. Another improvement would be to train the Random Forest neural network on a longer distance than one meter (which is the current distance) and to also train it on tougher angles i.e. when an individual is directed ± 90 degrees from the Kinect’s zero-angle.

As we neared the completion of our cognitive model, the behavior data from the study became more important. We had also realized a few of the study’s shortcomings. Our finished agent does not have the ability to react to movement in its field of view. This behavior was not taken into account in our study, and so the only bottom-up attention feature is the proximity response module.

Concluding remarks

In all honesty, spending a minute interacting with our agent will reveal flaws that break the illusion rather abruptly. There are some bugs, or rather: quirks and eccentricities to work out. There will, however, also be moments where the illusion is striking. During trials, we have had some great moments where the agent appears to really have an inner life of its own. Sadly, we did not have time to conduct a study on the subjective experience of interacting with the agent. All we have to go on are our own assessments and those of our supervisors and of those who attended the presentation of our project. Most of the feedback has been positive.

Acknowledgements

We would like to thank Christian Balkenius and Birger Johansson for their valuable input and support throughout this project. Thank you especially for the 3D-printed Bender-head!

References

- Balkenius, C., Morén, J., Johansson, B., & Johnsson, M. (2010). Ikaros: Building cognitive models for robots. *Advanced Engineering Informatics*, 24(1), 40-48.
- Balkenius, C. (2014) Lectures on robotics, Ingvar Kamprad Designcentrum, Lund University.
- Brooks, R. A., Breazeal, C., Marjanović, M., Scassellati, B., & Williamson, M. M. (1999). The Cog project: Building a humanoid robot. In *Computation for metaphors, analogy, and agents* (pp. 52-87). Springer Berlin Heidelberg.

- Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485), 1585-1599.
- Fanelli, G., Dantone, M., Gall, J., Fossati, A., Van Gool, L. (2012). *Random forests for real time 3D face analysis*. Retrieved October 10, 2014, from <http://www.vision.ee.ethz.ch/~gfanelli/pubs/ijcv.pdf>
- Gazzaniga, M.S. Ivry, R. and Mangun, G.R. Cognitive Neuroscience: The Biology of Mind. IV edition. W.W. Norton.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature neuroscience*, 2(6), 568-573.
- Gibson, J. J. (1979). *The ecological approach to visual perception / James J. Gibson*. Boston, Mass. : Houghton Mifflin, cop.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1), 20-25.
- Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, 16(4), 219-222.
- Ikaros-project. (2001). *About*. Retrieved September 24, 2014, from <http://www.ikaros-project.org/about/>
- Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2012, March). Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 367-374). ACM.
- Maxim Integrated. (2014). *Daisy Chain*. Retrieved September 30, 2014, from http://www.maximintegrated.com/en/glossary/definitions.mvp/term/daisy_chain/gpk/71
- MIT. (1998). *Overview*. Retrieved November 19, 2014, from <http://www.ai.mit.edu/projects/sociable/overview.html>
- Mori, Masahiro. (1970). The Uncanny Valley. *Energy* 7(4), 33-35.
- Norman, D. A. (1999). Affordance, conventions, and design. *interactions*, 6(3), 38-43.
- Robotis. (2014). *Dynamixel*. Retrieved September 30, 2014, from http://www.robotis.com/xen/dynamixel_en
- Siciliano, B., Khatib, O. (2008). Springer Handbook of Robotics. Berlin Heidelberg: Springer-Verlag
- Sevenster, P. (1961). A causal analysis of a displacement activity (fanning in *Gasterosteus aculeatus* L.). *Behaviour. Supplement*, I-170.
- Uexküll, J. V. (1957). A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds. 1934. *Instinctive Behavior. The development of a modern concept*. London: Methuen and Co LTD.
- Ward, J. (2006). The student's guide to cognitive neuroscience. Hove and New York.

Animating natural robotic movements

Asas Innab, Bjarni Runar Birgisson, Trond Arild Tjøstheim

asas.innab@gmail.com, bjarni_runar.birgisson.608@student.lu.se, trondarild@gmail.com

Natural looking, goal directed movements are important for humans to interpret intentions and goals in robots. We studied the kinematics of four reaching type movements: push, pull, point and shove, and used recorded data to reproduce these movements in a robot arm. We developed a framework for recording and editing robot control parameters by the use of timelines with visual key points, connected by interpolation functions. This allows visual inspection and control of an arbitrary number of robot servos, as well as other time dependent parameters.

1 Introduction

As robots go from the factory floor to the living room, they also go from being inside cages to interacting closely with humans. For humans to be able to understand what a robot is doing, and to predict how it will move, the robot must move in human like ways. Human movement provides information about such variables as priority, mood and intentions (Pollick et al., 2001). Though we specifically aim for human-like motion in this work, the motion profiles we seek would also apply to any kind of motion involving joints and muscles.

According to Gielniak et al. (2013), it is easier for humans to collaborate with robots if the robots display human-like motion. One of the main effects is that predictions of the robots' goals becomes easier. Those predictions can then be used to assist the robot, or to simply get out of the way.

Examples of contexts where such robots might be useful are ones related to health, like hospitals or homes for the elderly, but also smaller factories, or the home. Associated tasks could be assistance during surgery; assisting patients with eating or using the toilet; and generally bringing, moving, or holding items. Using tools will also be desirable, though a robust implementation of this still appears a way off.

We have chosen to look at some movements that we judged both representative of natural movement in general and arm movement in particular. They are pointing, pushing, pulling and shoving. We also deemed these movements as useful for a robot to do in a human social context.

This report is structured into seven sections. In the first, background on human motion is presented, as well as a survey of previous work on producing human-like movement in robots. The second gives a short explanation of the employed hardware and software. The third is a section on an experiment we carried out, to gather data about characteristics of the particular movements we chose to consider. The fourth has details of the various control strategies we tested, and where this data is employed. In the fifth, we discuss problems and limitations of the strategies. In the sixth, we present suggestions for how our work can be extended. The seventh concludes the report.

2 Background

Characteristics of human motion

The study of natural movements are important to be able to replicate it, and production of general human like movement in robots is still unsolved. We concentrate here on the upper limbs and reaching motions, rather than the lower limbs and gait.

In humans, hand movements from one point to another are essentially straight, and the associated velocity displays a bell-like shape (Atkeson & Hollerbach, 1985). These linear trajectories imply movement planning at the hand or object level. The velocity curve can be decomposed into a coarse movement that brings the hand close to the target, and a finer movement at the end. Atkeson & Hollerbach also found that "peak velocity and acceleration varied linearly with movement distance, which implies a velocity profile invariance".

Human movements are also variant, with variance increasing with age (Cooke & Cunningham, 1989). In young people, variance is similar in both the acceleration and deceleration phase of a movement, while the deceleration phase has higher variance in the elderly. Additionally, there may be instances of movement decomposition, where a previously integrated movement is changed to consist of discrete sub-movements. Such changes in movement profiles may indicate cerebellar dysfunction.

Blake & Shiffrar (2007) investigated how humans perceive human motion using recordings of people wearing point lights on major joints. They found that visual form influence perception. The human form is hierarchical, with appendages connected to a central trunk. If this hierarchy is hidden or distorted, observers have greater difficulty understanding movement. This argues for a humanoid shape in robots that are meant to work in social environments.

In this work, we choose to focus on velocity characteristics at the expense of variability.

Reproducing human movement characteristics in machines

Flash & Hogan (1985) describe a mathematical model to produce natural velocity profiles, which they call the "minimum jerk model". Jerk is here the time derivative of the acceleration. They employ dynamic optimization with minimum jerk as the optimization criteria. Given initial and goal position coordinates, the model calculates a near straight trajectory, with a bell-like velocity profile.

Fligge et al. (2012) investigated human catching motion in 3D, and found that the velocity profiles matched the minimum jerk model quite well. Their kinematic analysis suggested that humans use a predictive strategy modulated by corrective sub-movements just before a catch is made. The patterns they observed were similar to those of point-to-point reaching and grasping movements.

Pattacini et al. (2010) tried to implement a modified version of the minimum jerk model to enable point-to-point movements in a humanoid robot, with smooth, human-like motion characteristics. They found that the original model is too compute intensive to operate in real time (33Hz in their case) for their setup (a multi-core Intel (R) Xeon with 2.27 GHz clock frequency). But they approximated it by using a 3rd order controller, which they report being able to operate at the required frequency.

Berret et al. (2011) looked at different kinds of optimization criteria in human movement planning. Subjects were told to close their eyes and reach for a vertical bar from a sitting position, with markers placed on the reaching arm as well as the head. They found that a hybrid model of maximizing joint level smoothness and minimizing mechanical energy expenditure fitted the observed data well. This hybrid model was also superior to any single cost model.

Mohammad Khansari-Zadeh & Billard(2014) writes about an imitation learning approach with a set of autonomous dynamical systems (DS) called Control Lyapunov Function-based Dynamic Movements (CLF-DM).

A dynamical system is a concept in mathematics where a fixed rule describes how a point in a geometrical space depends on time. A Lyapunov function is a construct from the theory of ordinary differential equations, and can be used to prove that an equilibrium of a dynamical system is stable. Each Dynamical System (DS) codes a behavior (e.g. reaching for a cup). An estimate of these DS are obtained from a set of demonstrations of the desired task.

When a set of demonstrations has been obtained the new technique proceeds in three steps:

1. Learning a valid Lyapunov function from the demonstrations by solving a constrained optimization problem.
2. Using one of several regression techniques to model an (unstable) estimate of the motion from the demonstrations.
3. Using (1) to ensure stability of (2) during the task execution via solving a constrained convex optimization problem.

A set of theoretical and real environment experiments was done in order to evaluate the presented method. The conclusion from the evaluation is that CLF-DM is able to learn a larger set of motions, compared to existing approaches, because it learns the energy function instead of using a predefined one. Additionally, it allows users to choose the most appropriate regression techniques based on the requirements of the task at hand.

Zhao et al. (2014) sets up two new criteria for interpreting human arm movements. These are constructed to, "resolve the kinematic redundancy of human arm to generate a human-like arm posture". A model, based on a virtual torsion spring with variable stiffness, is used to represent the effect of muscles on motion. Specifically, the stored potential energy.

A distinction is made between the reaching and grasping motions and two differing criteria for generating these motions are used. For reaching motion the potential energy criterion is constructed and total energy potential and wrist discomfort for grasping motions.

The potential energy criterion is created to "resolve the kinematic redundancy of the human arm in the target position to generate the human-like arm posture in the

reaching movement". A simplification of the arm is made using only four out of seven of the available degrees of freedom in the human arm with only the shoulder and elbow accounted for. It is based on the principle that the human arm strives for the posture with the least amount of energy stored in the limb from gravitational pull as well as elastic energy stored in the muscles.

The total energy and wrist discomfort extends the before mentioned criterion to include hand orientation so that all seven degrees of freedom are accounted for. Here the effects of wrist discomfort are studied by having test subjects reach for an arbitrary point and see how the arm posture is adapted to lessen the discomfort.

These two new criteria are used to generate human-like reaching and grasping motions.

Gielniak et al. (2013) used a robot configured as a humanoid torso to investigate human-like movement generation. They used a three stage model, where stage one applied optimizations that increase spatial and temporal coordination, stage two added variability, and the third stage took into account constraints.

They also developed a similarity measure based on spatiotemporal isomap (Jenkins & Mataric, 2004), an algorithm taken from chaos theory. This enables objective comparison of time series containing variance. A common measure in these cases is to use qualitative judgements from a group of people, a time consuming and expensive task. Quantitative metrics thus has obvious benefits.

When modulating trajectories based on the optimization criteria, they start at some root joint (typically the pelvis in humans), and extend it out to the end effector. This is mapped to the available joints in the target robot. Their algorithm allows for substitution of alternative cost functions, such as optimal control (Lewis et al., 2012) or dynamic time warping (Itakura, 1975).

The same robotic system was used by Akgun et al. (2012) to create a framework for movement instruction based on key frames. Key framing is a well known technique for animating characters in films and video games (Burtnyk & Wein, 1971). Akgun et al. employed key frame clusters, and splining between them to let their robot learn skills such as scooping, pouring, and placing things.

Summary of previous work

Human-like motion can be produced in a robot by optimizing various dynamic properties of a trajectory. A reaching motion can thus be executed by having the system search a space of possibilities and from that generate a plan. Many of these approaches are computationally intensive and mathematically complex. Also, they may require a specific robot configuration. However, it is also possible to have humans show the robot how to move, by leading an appendage through a motion trajectory. Key positions can then be saved along the path. This is called key framing.

Given the time constraints on this work, we chose to avoid optimization-based approaches, and concentrate on computationally simpler methods. See the section *Test of control strategies* below for details.

Open Sound Control and the Duration application

The Open Sound Control protocol (OSC) (Wright, 1997) was developed as an alternative to the venerable MIDI protocol for interfacing digital musical instruments, aiming to



Figure 1: Robot arm with servo 1 at bottom, servo 2 in the middle, and servo 3 at the top

overcome some of MIDI's limitations (Moore, 1988). OSC allows communication using network protocols, and has support for descriptive strings to accompany control data.

OSC now has a wide variety of applications supporting it. Its ease of use and support for low latencies also makes it useful for robotic control.

Duration (2014) is an open source application which supports editing timelines of key frame data. It can transmit and receive interpolated data using the OSC protocol.

3 Material

We used a robot arm consisting of three Dynamixel AX-12+ (Robotis Dynamixel, 2014) servos (see fig. 1). The servos communicate with a host computer via a serial link. This allows an arbitrary number of servos to be chained together on the same link. The servos can receive parameter values to control goal position, speed, and torque. They can also send out feedback information for current position, current speed, and current torque. Position and speed are angular.

The Ikaros system (Balkenius et al., 2010) act as a scheduler which carry out the function of registered modules a certain number of times per second. Each computing slot is

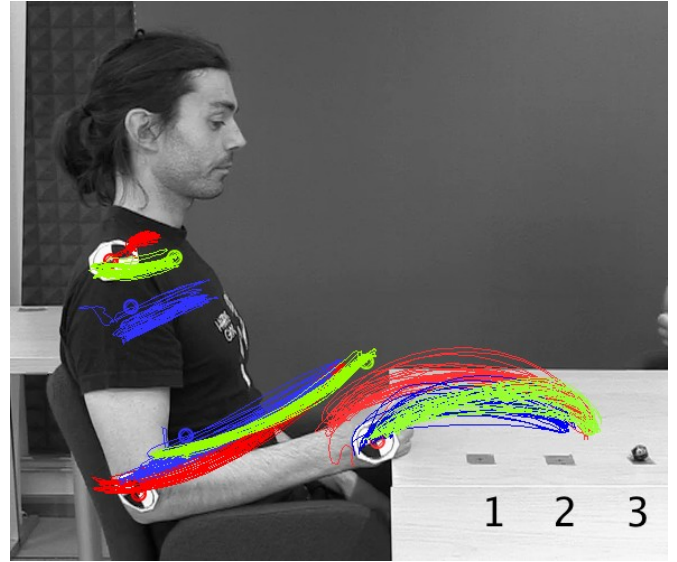


Figure 3: Experimental setup, showing position 1 at 11.6cm, position 2 at 22.6cm and position 3 at 32.7cm

called a tick. Thus, the servos can receive control information every tick, and feedback information can be read from them similarly. The minimum tick duration is limited by the speed of the host computer, but is typically between 10 and 20ms.

A graphical user interface can be connected to the Ikaros server, which allows inspection and real time control of parameters.

4 Experiments and data

We carried out two experiments: one to gather data about position and velocity changes in human reaching motion, and the other to get information about how to map such data correctly to robot servos. This was an explorative endeavor, and as such we did not have a hypothesis to test.

In this context "speed" or "velocity" refers to angular, or rotational speed. It is measured in degrees per second. In humans this corresponds to how quickly a limb rotates around a joint, such as the lower arm around the elbow joint. The robot servos rotate similarly. In reference to "torque", this is the rotational force necessary to produce movement.

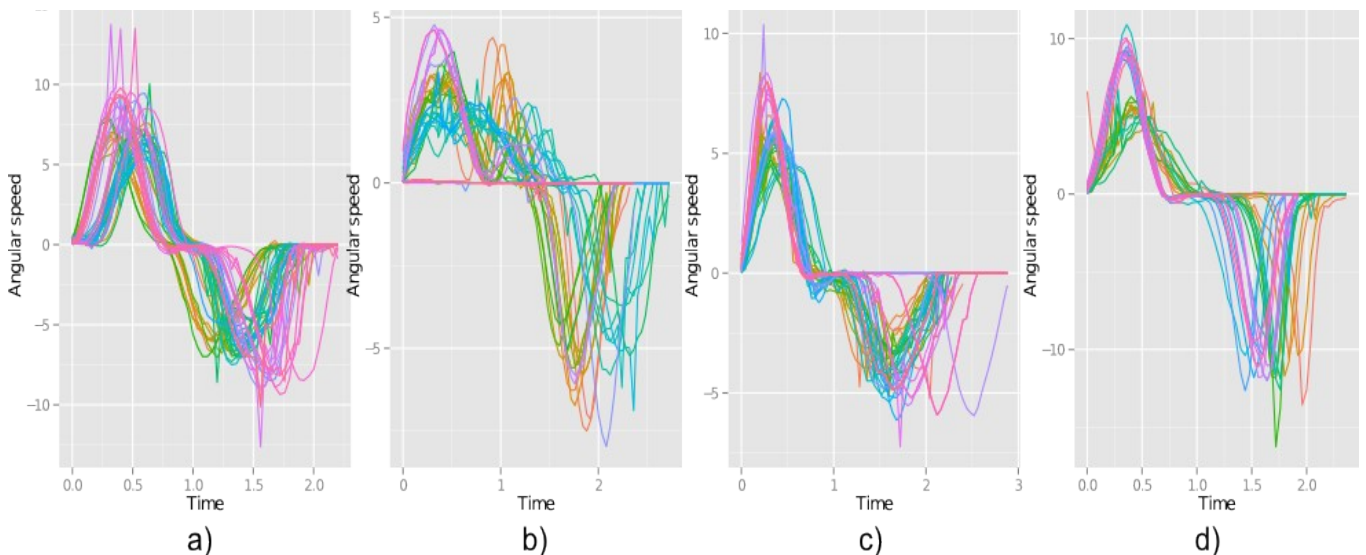


Figure 2: Speed profiles for a) pointing movement at pos. 3, b) pull motion from pos. 3 to pos. 1 c) push motion from pos. 1 to pos. 3 d) Shove movement from pos. 1

Both angular speed and torque can be used as control parameters on the servos.

The human hand moves in a Cartesian, or 3 dimensional space, where we only care about the sagittal plane. This is the one composed of the vector pointing out of the front of the body, and the inverse of the direction of gravity. Movement in this plane is mapped to angle between the body and the horizontal, the shoulder with the body, and the angle between the upper and lower arm.

As described in more detail below, we made measurements by connecting high-contrast markers to the shoulder, the elbow, and the hand, and filming the various movements. The film sequences were processed by software which can identify markers, and track their position across film frames. Each position is mapped to a coordinate system, and the resulting numbers can be exported. See fig. 2 for graphs showing angular speed data for all trials and all participants, for a selection of distances and movements.

Method

We recorded video footage of three male, healthy individuals performing types of reaching movements in the sagittal plane. These were pointing, pushing, pulling, and shoving. Each movement was repeated 17 times. The first and the last movement in the time series was discarded to avoid idiosyncrasies associated with beginning and ending a motion bout. Hence the plots for a movement for a given individual is the average and standard deviation based on 15 observations, while the same for all individuals is based on 45 observations.

This left us with essentially four types of reaching and retraction movements, but with differing kinematic profiles. A reaching motion in this context consists of shoulder flexion from about 0 degrees to approximately 60 degrees, and elbow extension from around 110 degrees to approximately

Table 1: Speed peaks (cm/s) for extension and retraction

movement	extension	retraction
point3	2.29	1.70
pull 2-1	1.42	0.83
pull 3-1	1.90	1.13
pull 3-2	1.74	0.77
push 1-2	1.15	1.06
push 1-3	1.68	2.05
push 2-3	1.50	2.04
shove 1	2.85	1.89
shove 2	2.34	2.16
shove 3	2.68	2.20

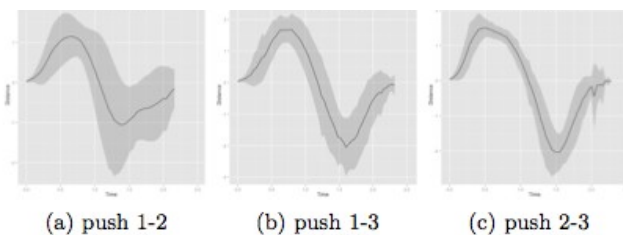


Figure 5: Average velocity of hand for push motion from various distances

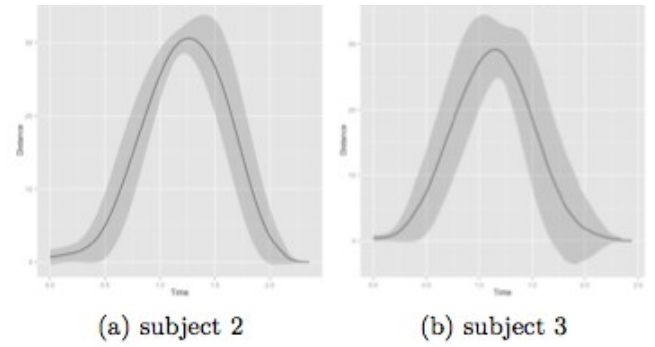


Figure 4: Comparison between two subjects for push motion 1-3, shoving movement distance of hand vs. time

30 degrees until the hand has reached the target object. After the motion has been completed, the arm is retracted to some approximate initial position. However, though the joint movements are very similar, the dynamics of the movements have individual characteristics.

We used a Canon video camera, recording at 25 frames per second at 1280 x 720 resolution, and saved in the MP4 format. The position data of the shoulder, elbow and hand was extracted using the Tracker video analysis and modeling tool for physics education (2014). The resulting time series were split into individual motion sequences using an algorithm we developed for this purpose, described in appendix C. These sequences were in turn imported into the RStudio (2014) software package, where average and standard deviation was calculated. Plots were generated from this data using the ggplot2 function.

The generated curves show the average distance (cm) and velocity (cm/s) over time (s) and is surrounded by the standard deviation marked in a darker background.

As mentioned earlier the movements were carried out from various distances.

When referring to numbers in the labeling column, the numbers represents a distance from the table edge (see also fig. 3)

1. 11.6 cm
2. 22.3 cm
3. 32.7 cm

See appendix A for a full complement of plots.

Results

In general, the velocity curves produced from the video footage replicate the findings of Atkeson & Hollerbach (1985), and show bell like shapes.

An interesting finding is the difference in the extension and retraction peak speed. For point, shove, and pull the extension it is higher than for retraction, while for push it is the opposite, see table 1.

We also observed some relation between distance to the target and standard deviation. It appears that as the target moves closer to the subject's body, the standard deviation increases (see fig. 5). It is possible that this could be explained by the discomfort zone of the arm when too close to the body.

As can be seen in the plots in appendix A, the standard deviation is somewhat large. We tried calculating the angular velocity of the elbow joint for the pointing movement to see whether this would yield a lower variance. This was done in

the Tracker software. The result did not, however, indicate any such effect.

Each subject has generally a characteristic standard deviation for a specific motion. Both in size and where it differs. For instance in subject 2's push movements the standard deviation tends to become larger after reaching the target whereas subject 3 tends to have less accuracy prior to reaching the turning point, see fig. 4.

Typically when reaching a turning point the standard deviation increases. We suggest that this is mainly due to our tendency to turn at different distances.

Mapping measured angular speed to servo torque and speed parameters

To measure the angular speed of servos, we apply time differentiation to position data. But in order to give accurate control instructions back to the servos, we need to be able to map back and forth between these. The servos take speed and torque values between 0 and 1. As noted above though, speed must not be equal to 0, as this triggers an internally set default value.

We set up a system that made each of the servos on our robot arm run through a range of speed and torque values, and measured the angular speed. The results indicate that the servos behave linearly and consistently when fed torque or speed values up to about 0.5. After this, the speed value levels off at about 7 degrees/tick. The torque behaves better, though the variability increases. The average mapping constant for the three servos 9.0 for torque, and 15.0 for speed. See table 2, and fig. 6. The full test plan is described in appendix B.

Summary of experimental results

We filmed three people carrying out four reaching type motions: push, pull, point, and shove, and extracted velocity profiles from this data. Each movement can be decomposed into a series of sigmoid curves representing angular velocity in time. To map this data to robot servos, we measured the servos' angular velocity for a series of speed and torque input values. The servos respond linearly for about half of the input range.

5 Tests of control strategies

Preliminary work

Before arriving at our final solution, we explored a range of alternatives. The simplest was recording and playing back position data from the servos. This works reasonably well on a fast computer, but may give staccato movements on less powerful machines. We tried various ways of setting goal positions and varying speed and torque in ways consistent with experimental data. But deducing the correct dynamics algorithmically did not yield satisfactory results. We did however verify that the positional data from the experiment

Table 2: Mapping factor for angular speed and torque

Factor	Servo 1	Servo 2	Servo 3	Average
Speed factor	14.18	15.39	15.39	15.0
Torque factor	9.01	9.10	8.86	9.0

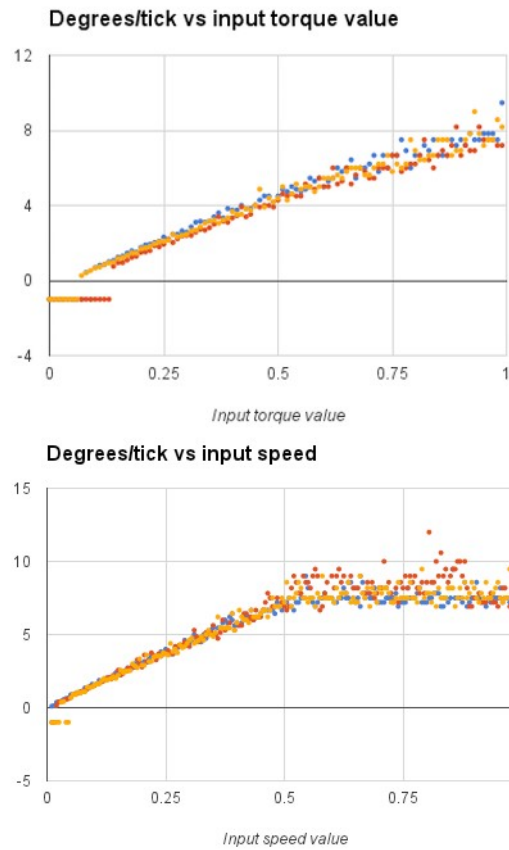


Figure 6: Degrees per tick for input speed and torque. Blue is servo 1 (bottom), red is servo 2 (middle), and yellow is servo 3 (top)

can successfully be mapped to the robot arm by streaming stored angular positions directly to the servos.

We tested using a PID controller (Bennett, 1993) to interpolate between positions. This was not ideal though, as there is no general way of determining correct parameters for such a system. Also, though the PID controller gives smooth movement at the end of a transition, the initial phase is abrupt.

A breakthrough was being able to extract position key frames from a recording. We operationalized a key frame to be the position at which the angular speed is zero. The challenge of playing back sequences of key frames with smooth dynamics, led us to the solution of mapping the key frames to visual timelines. This is described in more detail below.

Key frame editing and interpolation functions

One of the insights that came with practical experience was that leading the robot through motions does not necessarily mean that the reproduced movement will have high quality. This is mainly due to mechanical resistance in the servos. They are not very compliant, and quite a lot of force may be necessary to move them. This is the case even when torque is switched off. The recorded motion may have the correct key positions, but the resulting dynamics is not very natural looking. Also, several operators may be necessary to simultaneously move multiple appendages.

A solution to this is to use mechanisms from animation and audio software, such as time lines of editable key frames. Here, key frame points can be moved two-dimensionally, and various kinds of mathematical functions can be used to interpolate between them. An arbitrary number of channels

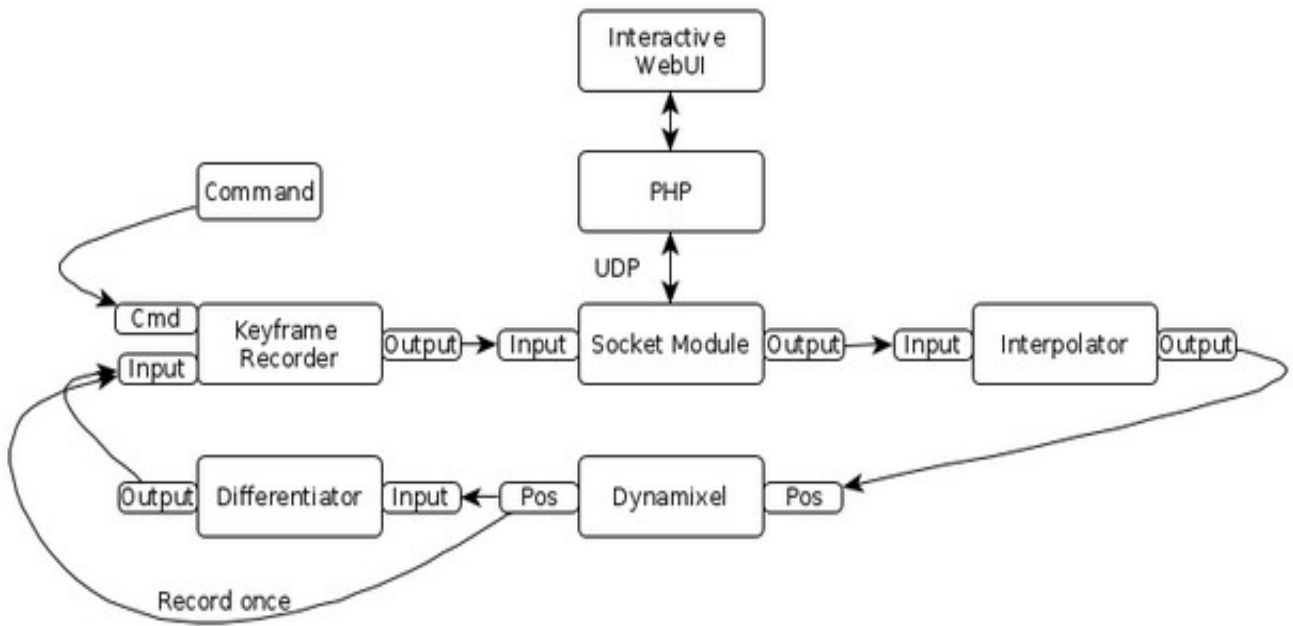


Figure 7: Block diagram of Ikaros modules and WebUI for key frame editing system

can then be set up to control different kinds of parameters. This maps very well to the servos in a robot, and the parameters that control them.

We set up a web interface (see fig. 8), written in Javascript, which enables the user to graphically define points in time, and have an interpolation function between them. This can be done for an arbitrary number of channels. This interface then acts as a server, and sends control data to a client module running in the Ikaros system. The communication happens via User Datagram Protocol (UDP), a communication protocol running on top of the Internet Protocol (IP).

The workflow then becomes to first record a coarse movement by manually moving robot effectors, then edit the generated key frames, and finally refine speed or torque data. The key frames are sent back to the Ikaros server, and an interpolation module assures that transitions are smooth. Sequences of key frames can be stored to, and read from file by using existing Ikaros modules.

We also implemented an alternative way of recording and editing key frames. This utilizes the OSC protocol to transmit data out of and into Ikaros, and using third party applications to do the actual editing and interpolation.

Information flow between Ikaros and the WebUI

The key frame recorder records all positions from the Dynamixel module and uses the Differentiator module for determining shifts in the motion (see fig. 7). The Command block are the buttons in the WebUI telling it if it should record or not. When recording is done and the command process is received, the recorder finds all the key frames and outputs them along with a sync signal. This signal tells the SocketModule to send the data to the WebUI. The SocketModule parses the data into a text string and sends it to the predefined destination address and port. The PHP server listens to this port, and hands the data off to the WebUI. Here, graphs are plotted and the motion data can be changed in various ways. Every time a change is made, all key frames are sent to the SocketModule. The data is parsed and output as two matrices with sizes. Values are continuously interpolated by the Interpolator module and subsequently output to the servos. Thus, the effects of a change in the WebUI is immediately observable.

The model could be simplified by removing the PHP server used to communicate with the WebUI. This can be

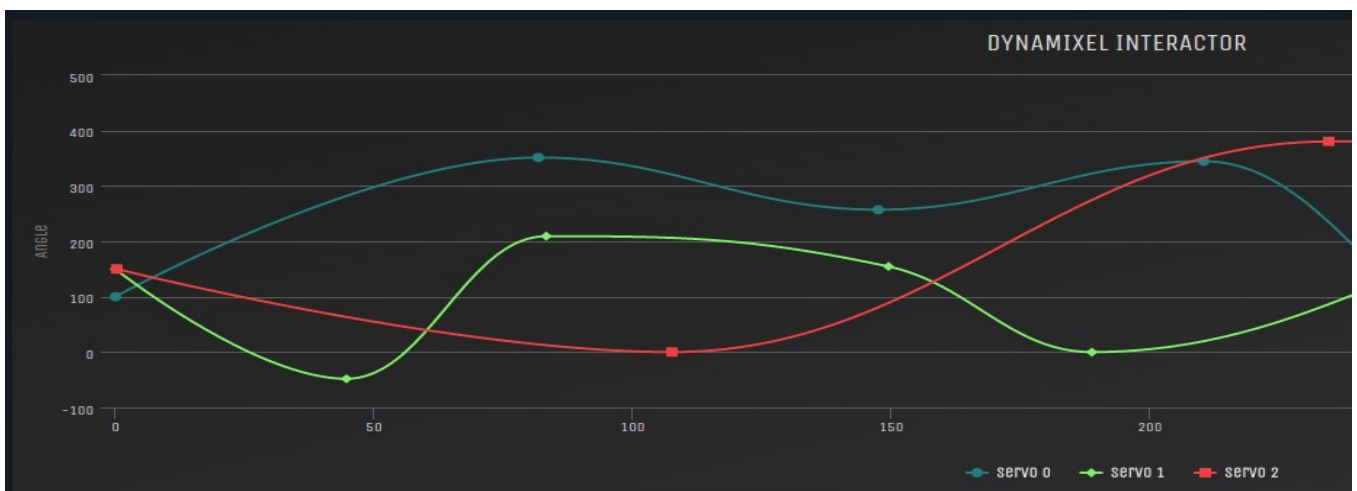


Figure 8: Web UI for editing key frames

achieved by letting the Javascript in the WebUI receive the packets itself via sockets.

Summary of control strategies

Recording and sending position data is a simple solution that works as a first approximation. In a fast computer, the resulting movement is quite smooth. The recording process has intrinsic problems though, such as the servos providing resistance to motion. The method could however be used to map data from filmed experiment to the robot.

Our final solution, using visual time lines with key frame points and interpolation functions gives a high level of control, and enables creation of high quality movement sequences.

6 Discussion

The problems we have had have been associated with correct coordination of movement, and with reproducing natural looking dynamics.

Much of the problems are due to noise in reading position data, and amplification of that noise when differentiating is performed on the data to get speed values. Even though we added an averaging stage, this was unable to remove all the high frequency components.

The smoothest results came from using torque to indirectly control speed. However, each joint has different load, and the load changes with the position of subsequent joints in the kinematic chain. This is hard to compensate for, and does not map easily to other robot types. Thus we initially preferred using speed, particularly as we discovered that using the Ikaros systems' realtime mode improves smoothness. However, the graphical system enables much more intuitive use of both speed and torque.

From the literature, the problems we encountered would appear to have been solved by other groups, with Gielniak et al. (2013) perhaps having the most generally applicable solution.

The limitations of our final solution are that manual editing work needs to be done for each type of movement. The two implementations also have their individual limitations.

For the time being, the solution using a web interface requires that a PHP server process is active to bounce messages between it and Ikaros. The work flow also requires a processing step to identify key frames in recorded data, which is not automatic but requires the pressing of buttons.

Using OSC and the Duration application requires that Duration is started in addition to Ikaros, and that recording is initiated in the Duration application. A recording currently involves a manual cleanup process, since a key frame is essentially created every tick. This can be laborious. Although Duration allows saving of data to its own file format, no module has been implemented in Ikaros to read such files. Making movement data available to Ikaros without using Duration would therefore involve streaming it to a file using Ikaros file modules, and then later loading up this file.

On the other hand, enabling bidirectional OSC communication in Ikaros opens up a host of possibilities, such as e.g. using external control surfaces such as TouchOSC (2014) or letting Ikaros control 3D models in the

Blender 3D modeling application (Blender, 2014). More speculatively, it might also afford multiple instances of Ikaros running on separate computers to be integrated into a single system, or to create something like a "cognitive synthesizer". Such a device might for instance be able to teach itself to produce speech-like sounds by controlling sound synthesis parameters in software like Pure Data (2014) or SuperCollider (2014).

Future work

The original task description for this project included implementation of goal directed movement. We did not have time to do this. Thus, work remains to enable the robot to activate stored motion sequences, and use them to direct its effectors to a target. This implies scaling the sequence in space, to reach targets that are closer or further away. As well it means doing some sort of rotation so targets can be placed to the left or the right of the robot. The best results are likely achieved by incorporating visual feedback into the control chain, so that motion can be finely controlled as the effector approaches the target.

7 Conclusion

We have tested strategies for algorithmically controlling a robot effect, but did not achieve a satisfactory result using this approach. As an alternative, we developed tools for recording and manually editing key frames for position and control data. This allows for efficiently creating very natural looking motion sequences, and affords control of an arbitrary number of simultaneous joints.

References

- Akgun, B., Cakmak, M., Jiang, K., & Thomaz, A. L. (2012). key frame-based learning from demonstration. *International Journal of Social Robotics*, 4 (4), 343–355.
- Atkeson, C. G., & Hollerbach, J. M. (1985). Kinematic features of unrestrained vertical arm movements. *The Journal of Neuroscience*, 5 (9), 2318–2330.
- Balkenius, C., Morén, J., Johansson, B., & Johnsson, M. (2010). Ikaros: Building cognitive models for robots. *Advanced Engineering Informatics*, 24(1), 40-48.
- Bennett, S. (1993). A history of control engineering. ,47, 1930-1955.
- Berret, B., Chiovetto, E., Nori, F., & Pozzo, T. (2011). Kinematic features of unrestrained vertical arm movements. *PLoS computational biology*, 7, e1002183.
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.*,58, 47–73.
- Blender (2014). Retrieved 2014-12-18, from <http://www.blender.org>
- Burtnyk, N., & Wein, M. (1971). Computer-generated key-frame animation. *Journal of the SMPTE*, 80 (3), 149–153.
- Cooke, J., Brown, S., & Cunningham, D. (1989). Kinematics of arm movements in elderly humans. *Neurobiology of aging*, 10 (2), 159–165.
- Duration (2014). Retrieved 2014-12-18 from <http://duration.cc>
- Flash, T., & Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *The journal of Neuroscience*, 5 (7), 1688–1703.

- Fligge, N., McIntyre, J., & van der Smagt, P. (2012). Minimum jerk for human catching movements in 3d. In *Biomedical robotics and biomechatronics (biorob), 2012 4th ieee ras & embs international conference on* (pp. 581–586).
- Gielniak, M. J., Liu, C. K., & Thomaz, A. L. (2013). Generating human-like motion for robots. *The International Journal of Robotics Research*, 32 (11), 1275–1301.
- Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 23 (1), 67–72.
- Jenkins, O. C., & Matarić, M. J. (2004). A spatiotemporal extension to isomap nonlinear dimension reduction. In *Proceedings of the twenty-first international conference on machine learning* (p. 56).
- Lewis, F. L., Vrabie, D., & Syrmos, V. L. (2012). *Optimal control*. John Wiley & Sons.
- Mohammad Khansari-Zadeh, S., & Billard, A. (2014). Learning control lyapunov function to ensure stability of dynamical system-based robot reaching motions. *Robotics and Autonomous Systems*, 62, 752–765.
- Moore, F. R. (1988). The dysfunctions of MIDI. *Computer music journal*, 19–28.
- Pattacini, U., Nori, F., Natale, L., Metta, G., & Sandini, G. (2010). An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots. In *Intelligent robots and systems (iros), 2010 ieee/rsj international conference on* (pp. 1668–1674).
- Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82 (2), B51–B61.
- Pure Data (2014). Retrieved 2014-12-18, from <http://puredata.info>
- Robotis dynamixel. (2014). Retrieved 2014-10-25, from http://www.robotis.com/xe/dynamixel_en
- Rstudio. (2014). Retrieved 2014-10-25, from <http://www.rstudio.com/>
- SuperCollider (2014). Retrieved 2014-12-18, from <http://supercollider.sourceforge.net/>
- TouchOSC (2014). Retrieved 2014-12-18, from <http://hexler.net/software/touchosc>
- Tracker video analysis and modeling tool for physics education. (2014). Retrieved 2014-10-25, from <https://www.cabrillo.edu/~dbrown/tracker/>
- Wright, M. (1997). Open Sound Control-A New Protocol for Communicationg with Sound Synthesizers. In *Proceedings of the 1997 International Computer Music Conference* (pp. 101–104).
- Zhao, J., Xie, B., & Song, C. (2014). Generating human-like movements for robotic arms. *Mechanism and Machine Theory*, 81, 107–128

Appendix A Statistics

Table 1: Point motion, hand - position

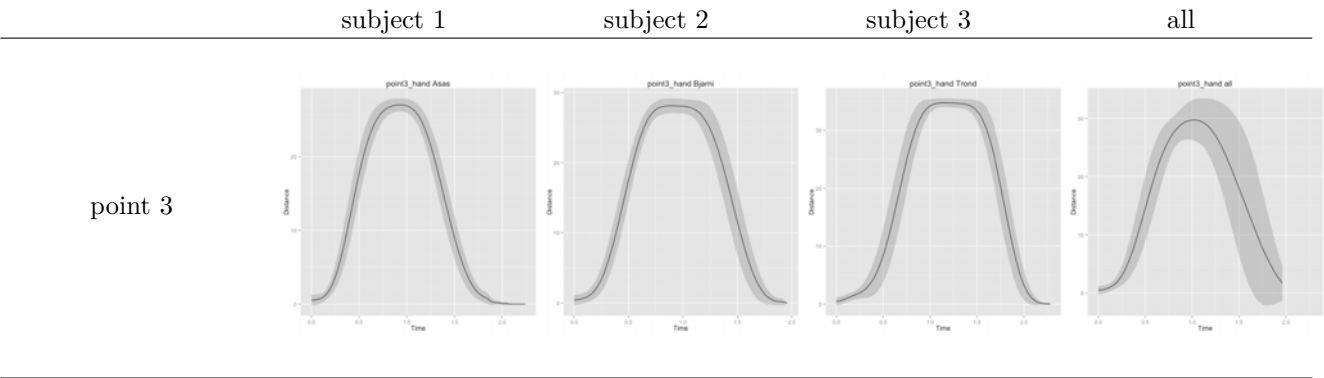


Table 2: Point motion, hand - velocity

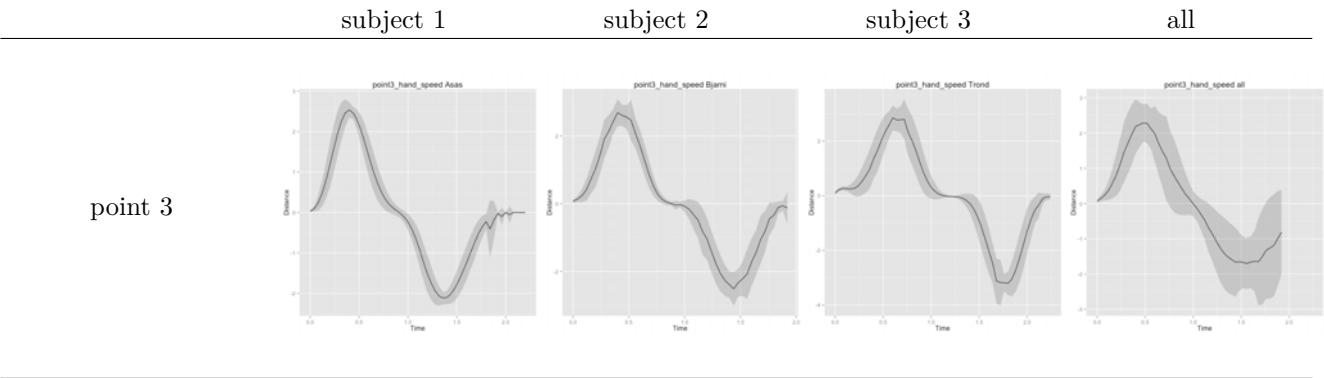


Table 3: Push motions, hand - position

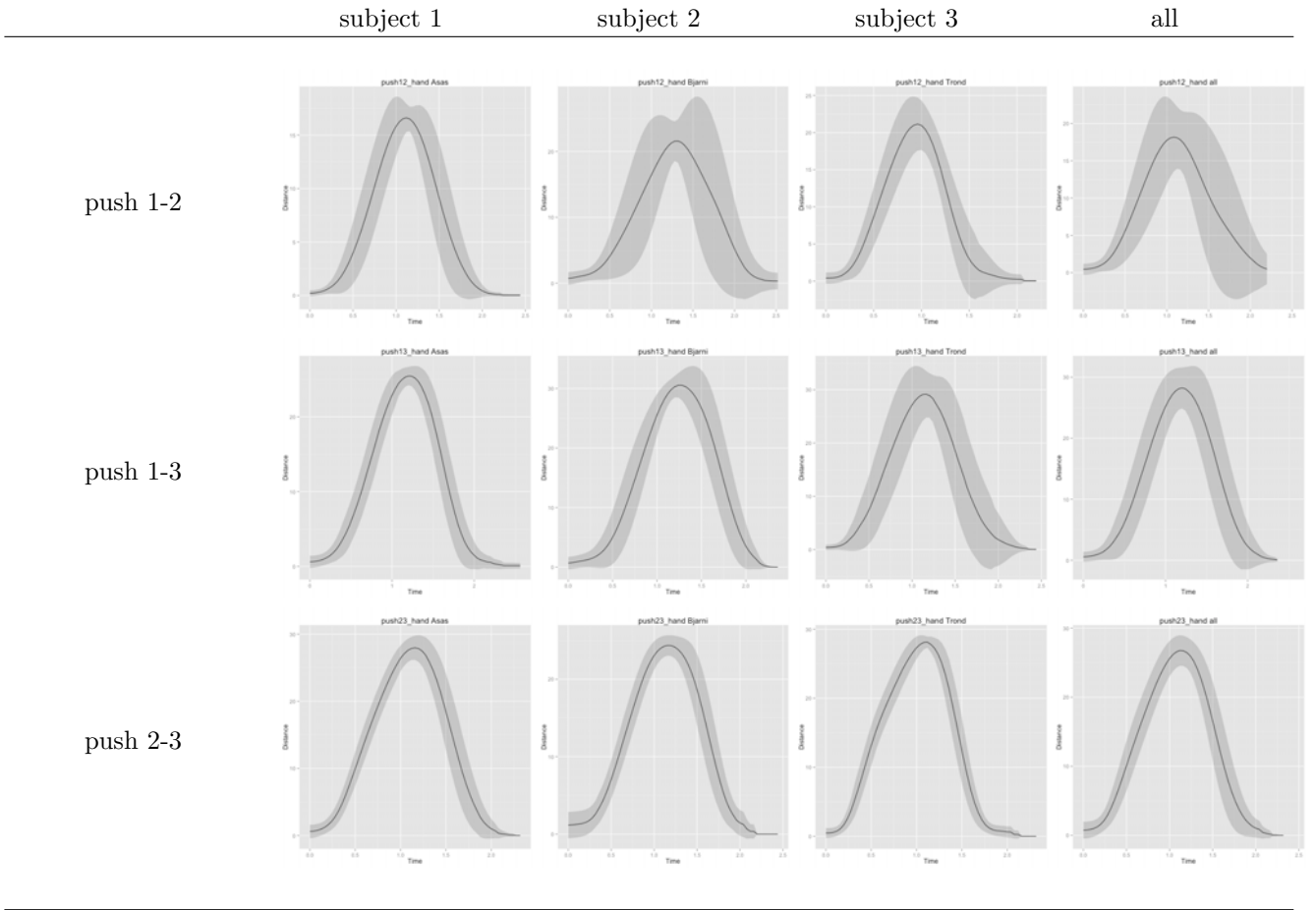


Table 4: Push motions, hand - velocity

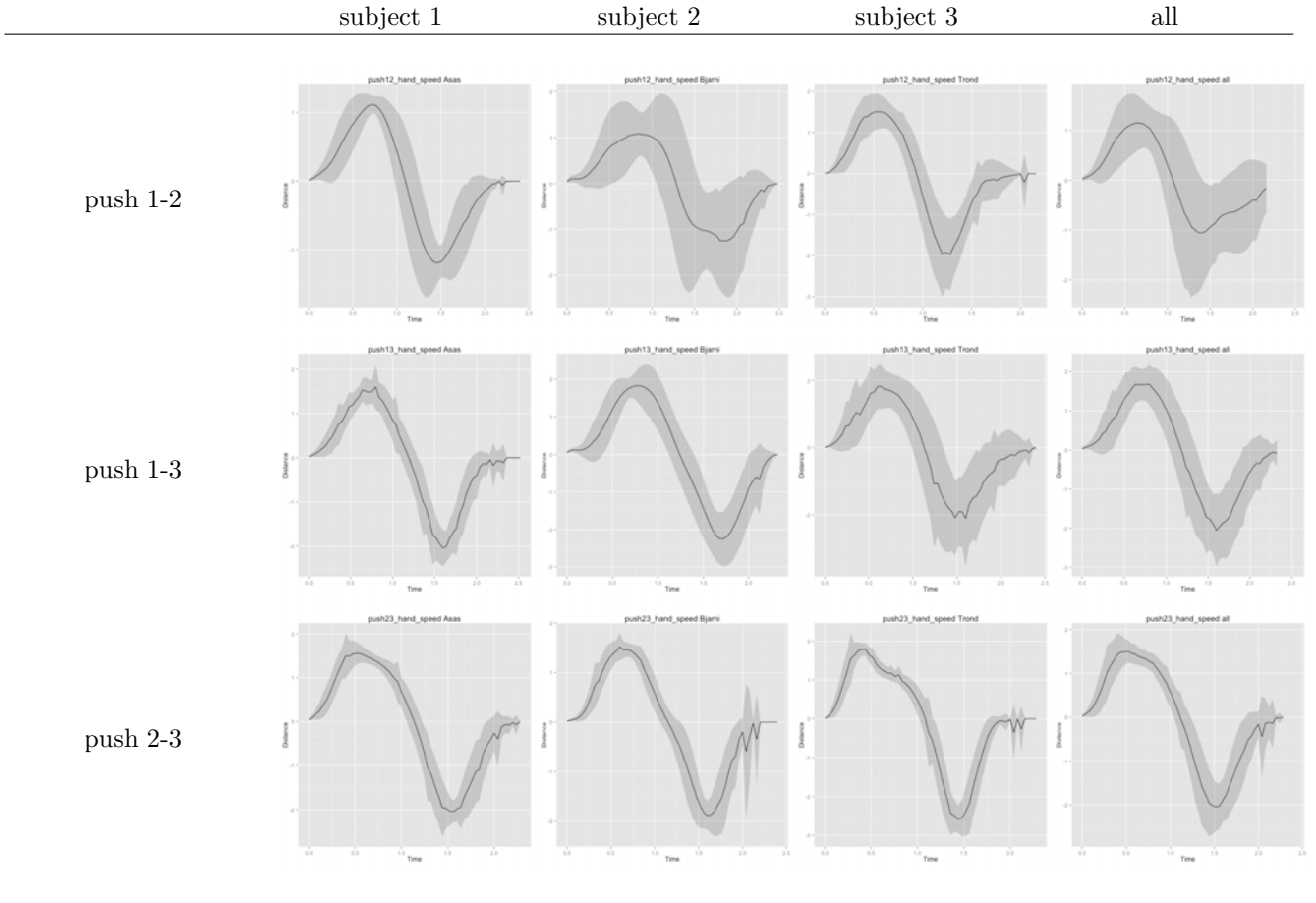


Table 5: pull motions, hand - position

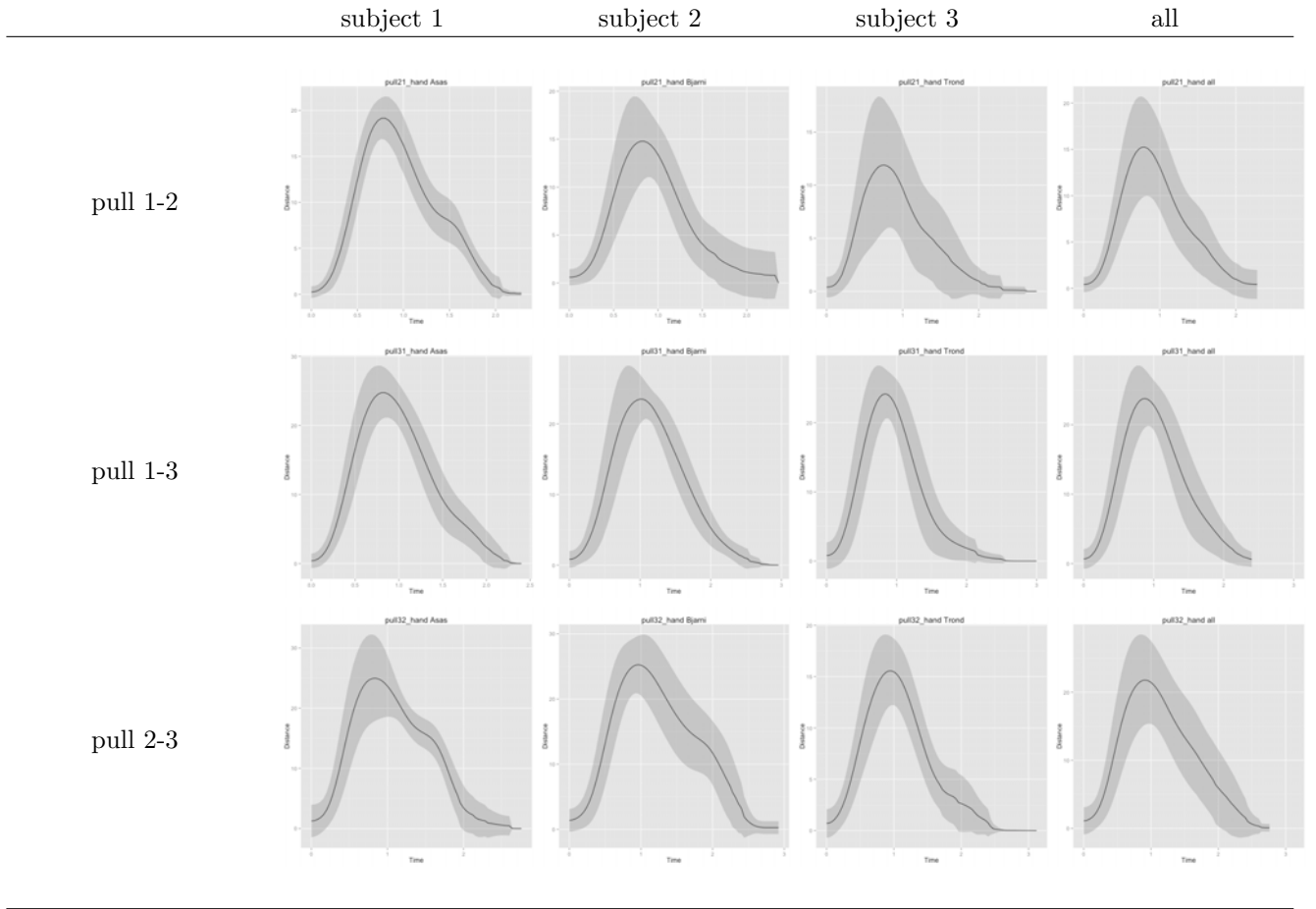


Table 6: pull motions, hand - velocity

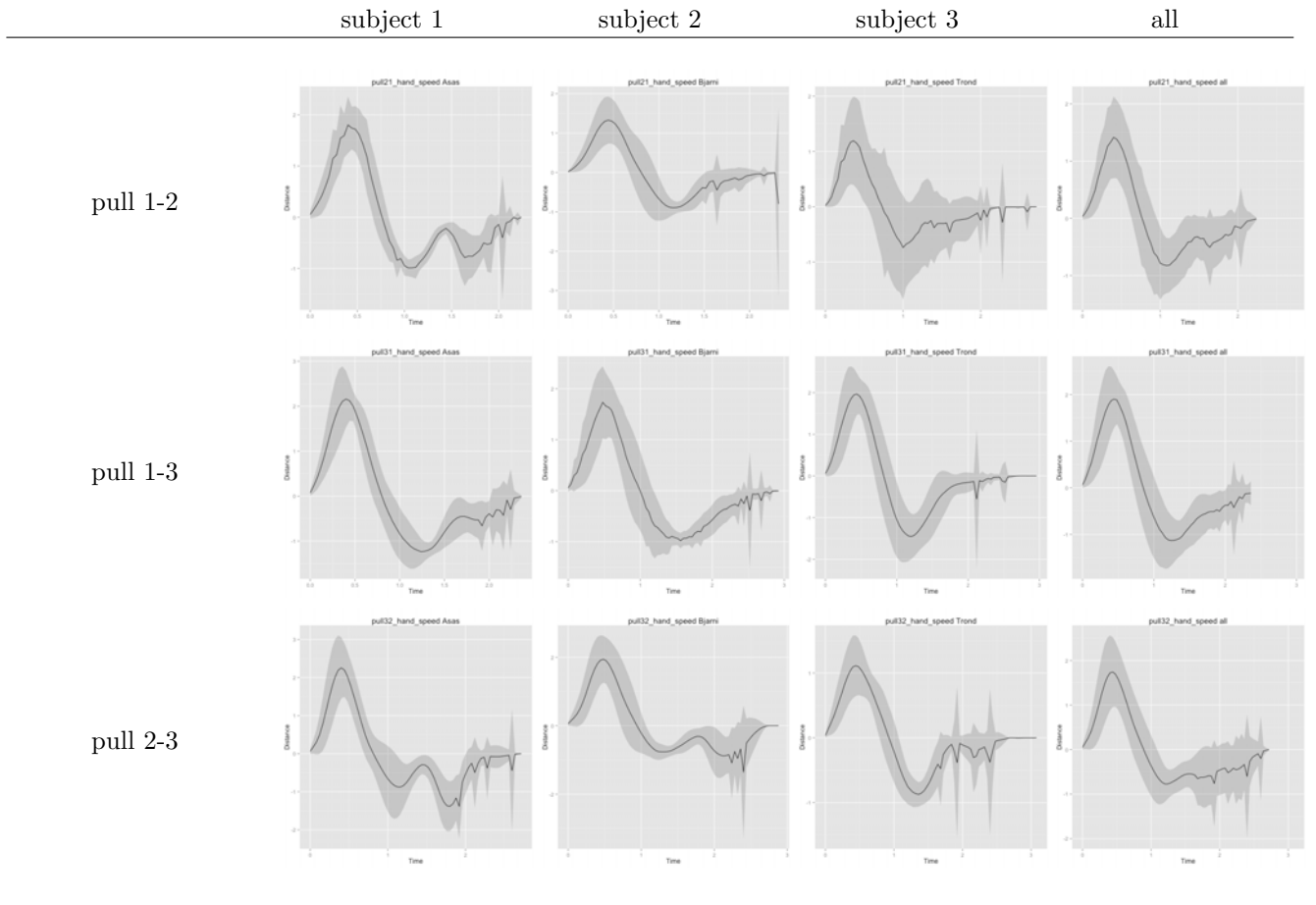


Table 7: shove motions, hand - position

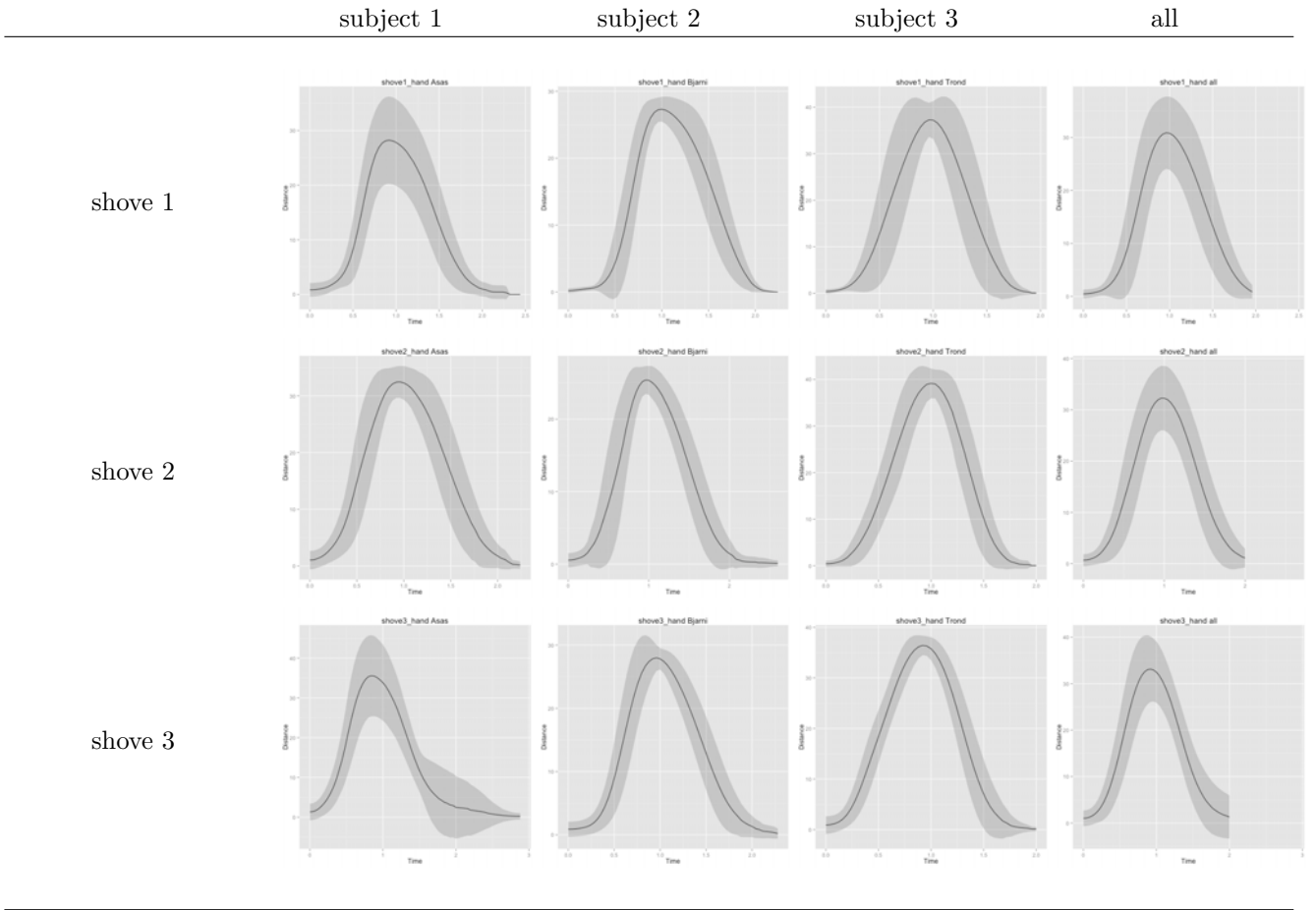
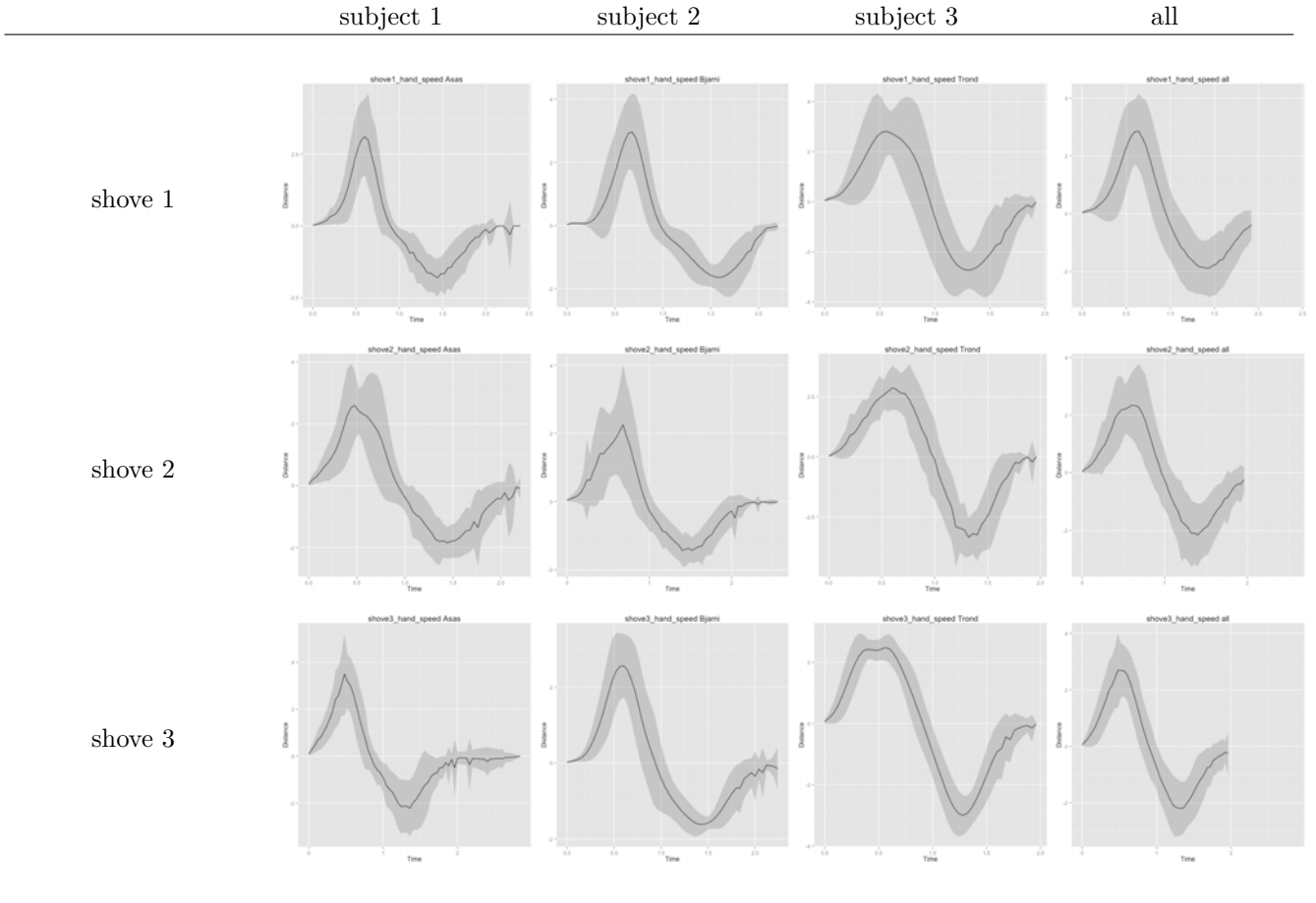


Table 8: shove motions, hand - velocity



Test 1

Goal

Find out the thresholds for lowest possible torque value for a given speed and load.

Why?

We want to find out how great impact the load differences during a motions has because the robot arm-extension and position (load) can vary during different movements. We want to find out whether this difference is negligible for our robot.

We also want to find out how the relation between the minimal torque-value and speed looks because we prefer to always use the lowest possible torque value. A smaller torque-value generates smoother motion, however a too low torque value would mean that we wouldn't arrive to our position in time.

Hypotes:

Our hypothesis is that different torque values is required for different loads at a given speed. We suspect that the difference is negligible for our robot, but that for a larger robot has a meaning that one needs to consider.

Task execution:

Before we begin:

- We need to check if the speed feedback from the servos are providing reliable results, or if we need to do our own calculations.
 - *Feedback speed of the servos was no real feedback in a servo current speed but rather so that it returns the speed input it receives.*

We are going to study the following cases:

Servo No. (0,1,2)

Load (heaviest / lightest)

Speed (0:25, 0:50, 0.75, 1.00)

Torque (0.00, 0.05, 0.10 1.00)

A total of $3 * 2 * 4 * 20 = 420$ different cases

We will be using the realtime value (-r 20) when running the Ikaros application.
The results for each separate case are to be written to a file.

What are we looking for?

- Minimum torque-value for a given speed and load
- How dependency between torque, speed and load looks like.
 - Is it linear? Is it possible to describe with a mathematical model?

Design for application:

```
for servoid in (0, 1, 2) {
  for load in (light, heavy) {
    set servo1_pos = load
    for speed in (0.25, 0.5, 0.75, 1.0){
      set servo2_speed = speed
      for torque in (0-1.0; 0.05){
        if servo2_pos == minpos
          servo2_target = maxpos
        else if servo2_pos == maxpos
          servo2_target = minpos
        else
          go to target

        // assume servo in startpos here
        set servo2_torque = torque
        for tick in (0..maxtick){
          speed_recording[tick] = current_speed
        }
        write speed_recording to file with name
        servo+load+speed+torque+date
      }
    }
  }
}
```

Appendix C Split data algorithm

The data collected from the experiments needed to be cleaned up and formatted so that it could be presented properly. So a short program was developed to automate this process. It was written in Java and below is a basic illustration of how it works in pseudo code.

Algorithm 1 Main application

```
1: procedure MAIN(input file)
2:   rowList  $\leftarrow$  read from input file
3:   rowList  $\leftarrow$  movingAverage(rowList)
4:   indexList  $\leftarrow$  find all curve tops(rowList)
5:   curvematrix  $\leftarrow$  parseCurves(rowList, indexList)
6:   curvematrix  $\leftarrow$  scaleCurves(curvematrix)
7:   writeToFile(output file)
8: end procedure
```

Algorithm 2 Low pass filter to smooth out the data

```
1: procedure MOVINGAVERAGE(rowList)
2:   for entry in rowList do
3:     entry  $\leftarrow$  entry + previousEntry/2
4:   end for
5: end procedure
```

Algorithm 3 Algorithm to find the 17 tops in the data, always looking for the largest values that are not to close together

```
1: procedure FIND ALL CURVE TOPS(rowList)
2:   indexList
3:   for entry in rowList do
4:     if indexList not full then
5:       add entry to indexList if not to close in
        time to previous largest value
6:     else
7:       replace lowest value of entry at
        index in list with current
8:     end if
9:   end for
10: end procedure
```

Algorithm 4 Algorithm for parsing whole curves from the *topIndex*

```
1: procedure PARSECURVES(rowList, indexList)
2:   cureveMatrix
3:   for entry in indexList do
4:     leftIndex  $\leftarrow$  look for bottom on left side
      of entry
5:     rightIndex  $\leftarrow$  look for bottom on right
      side of entry
6:     add to curveMatrix everything from
      leftIndex to rightIndex
7:   end for
8: end procedure
```

Algorithm 5 Scales the curves to that every curves lowest point is 0 (either start or end)

```
1: procedure SCALECURVES(curveMatrix)
2:   for rowIndex in curveMatrix do
3:     find smallest value in row
4:     for entry in row do
5:       entry  $\leftarrow$  entry - smallest
6:     end for
7:   end for
8: end procedure
```

Hur tittar människor?

Robert Bagge

bagge.robert@gmail.com

Olof Knape

dic11okn@student.lu.se

Henrik Siljebråt

henrik@siljebrat.se

Hur människor tittar i olika situationer blir alltmer intressant, framförallt beroende på det senaste årtiondets stora forskningsframsteg inom robotik som möjliggör att det inom en snar framtid kommer finnas robotar i hemmen som hjälper till med hushållsuppgifter såsom städning och tvättning. Det här projektet har fokuserat på att förstå hur människor tittar i en lärande situation, närmare bestämt i en situation när man ska imitera ett byggmönster med klossar. Ett experiment utfördes och resulterade i ett antal iakttagelser om människors tittbeteende, varav en av de viktigare är prediktion av lärarens handlingar. Med hjälp av en Microsoft Kinect och en robotarm implementerades sedan ett system baserat på dessa iakttagelser. Systemet kan sättas i en liknande situation som experimentdeltagarna och påvisa samma tittbeteende som dem. Detta väckte frågan om vad som behövs för att skapa en känsla hos läraren av att roboten förstår det som lärs ut, vilket diskuteras i kontexten av vardagsrobotar. Vårt nuvarande system behöver till viss grad generaliseras så att vi i framtida experiment kan utvärdera om roboten beter sig på ett vis som får läraren att uppleva interaktionen som att roboten förstått uppgiften.

1 Introduktion

Blickar kan döda. Uttrycket beskriver tydligt hur viktig blicken är för människors kommunikation. Om blicken inte stämmer överens med övrigt kroppsspråk eller den verbala delen av kommunikationen så uppfattar de flesta det beteendet som konstigt

Tittbeteende har tidigare främst undersökts i kontext av allmän social interaktion och inte på detaljnivå (Sidner et al., 2005). Men i takt med att forskningen inom robotik och artificiell intelligens blir mer avancerad finns idag ett stort intresse för att veta exakt hur människor tittar då detta gör att man kan replikera beteendet med robotar. Dessutom börjar man nu på allvar kunna se att det inom en inte alltför lång framtid kommer finnas robotar hemma hos gemene man som kan hjälpa till med enkla sysslor såsom att städa, diska och tvätta. Som tidigare nämnts kan kommunikationen kännas konstig om blicken inte används som den förväntas och detta gäller även för robotar, exempelvis när de får instruktioner. Litar du på att roboten förstår om den tittar konstigt? Att veta hur människor tittar blir därför en väldigt viktig komponent

när man ska förmänskliga robotar och på så sätt göra att människor känner sig bekväma med att ha robotar hemma.

Det är där vi står idag och det här projektet har syftat till att utforska just hur människor tittar, samt att göra en enkel implementation av detta i en robot med mjukvaran Ikaros. Först presenteras teoretisk bakgrund, sedan uppgiften vi valt att fokusera på och ett experiment med mänskliga försöksdeltagare där vi undersökt tittbeteendet för uppgiften. Därefter följer en beskrivning av roboten och hur experimentresultatet implementerats och slutligen en sammanfattning och diskussion av projektet.

2 Bakgrund/Teori

Förmågan att förstå vad andra tittar på uppstår redan under människans första levnadsår och brukar benämnas "joint attention" (Gazzaniga, 2013). I sin enklaste form används endast blickriktning men även huvudriktning och gester spelar in, varför begreppet innefattar inte bara blick utan även handlingar och aktiviteter.

Hos vuxna och äldre barn har denna förmåga utvecklats så att prediktiva ögonrörelser sker vid observation av andra människor. Webb et al. (2009) beskriver ett experiment där man undersökte försökspersoners ögonrörelser när de observerade en agent som greppade objekt på ett bord. Man fann ett tydligt sekventiellt mönster i att observatören rörde blicken från en förbestämd fixeringspunkt till agentens huvud och vidare till målobjektet. Det sistnämnda skedde innan agentens hand hunnit till objektet, alltså fanns en prediktiv förmåga hos observatörerna.

När människor observerar handlingar används till viss del samma neurala system som när man själv utför handlingar. Detta har visats på betendenivå i ögonrörelseexperiment där försökspersoner som beskriver en minnesbild rör ögonen som om de tittar på bilden framför sig (Johansson et al., 2013). På biologisk nivå har det påvisats genom upptäckten av spegelneuron - nervceller som aktiveras både när en handling utförs och samma handling observeras (Gazzaniga, 2013). Vad detta innebär är att man kan se hjärnan som ett system där vår kunskap om världen grundar sig i hur vi upplever och använder vår egen kropp - vi förstår världen genom oss själva - en teori som kallas "embodied cognition". (Wikipedia, 2014).

Denna utgångspunkt är en anledning till att kognitionsforskning använder robotar som verktyg för att förstå hur människor fungerar. Hur väl man än modellerar ett kognitivt system måste det sättas in i en kropp som efterliknar människans för att få en relevant kontext (Scholarpedia, 2014).

Robotar som interagerar med människor brukar benämnas som sociala robotar. Dessa kan enligt Fong et al. (2003) designas utifrån två huvudperspektiv - biologiskt och funktionellt. Det förstnämnda innebär att man i roboten försöker simulera social intelligens utifrån aktuell kognitionsforskning. Funktionell design innebär istället att roboten utåt uppvisar social intelligens, men att dess interna struktur inte behöver ha någon vetenskaplig grund. Den här inriktningen passar väl ihop med ett iterativt och utforskande arbetssätt som i vårt fall.

Joint attention hos robotar har bl.a. undersökts av Nagai (2005) som baserat på fynd hos människor om att rörelse har betydelse för blickförståelse implementerade ett inläringssystem i en robot där både huvudriktning samt optiskt flöde togs i beaktande och fann att detta förbättrade resultatet jämfört med andra metoder.

Liknande försök till blickförståelse har gjorts av Sakita et al. (2004), men i båda fallen är det robotens tolkning av människor man koncentrerat sig på, istället för hur människor tolkar roboten som är fokus i det här projektet.

Sidner et al. (2005) undersökte vad de kallar "engagement" (som för vårt syfte kan likställas med joint attention) mellan robot och människa. De genomförde först ett experiment på människors interaktion - där de tittade på verbalt innehåll och vissa gester - för att sedan implementera detta i en robot. Här har man alltså undersökt förståelse åt båda håll i interaktionen och sedan utvärderat robotens beteende i ytterligare ett experiment. En av huvudslutsatserna de kom fram till var att robotens gester var mer betydelsefulla än dess verbala yttranden för att försökspersoner skulle uppleva bättre "engagement".

Det verkar heller inte spela stor roll om verbala uttryck bygger på en förståelse av vad människoparten säger. I en studie där barn fick spela schack med en robot kom Leite et al. (2012) fram till att empatiska yttranden vid rätt tillfälle skapade ett starkare band mellan parterna. Roboten kunde känna igen ett antal olika känslor hos barnen och reagerade antingen inte alls, slumpmässigt eller "rimligt" på dessa känslor. Det slumpade fallet gjorde barnen mer nervösa och oroliga, vilket författarna menar ligger i linje med tidigare resultat om att fel beteende kan få en negativ upplevelse av roboten.

Samtidigt finns enligt Sharkey & Sharkey (2010) en djupt rotad benägenhet hos människor att tillskriva föremål med rätt egenskaper någon form av liv. Som exempel nämner de experiment där barn som fått interagera med robotar hittar på anledningar till varför robotarna inte betar sig som de "borde", t.ex. att roboten är döv om den inte svarar på tilltal.

Möjligheten att interagera med robotar på samma vis som med människor är som nämndes i introduktionen viktigt för att vem som helst ska kunna använda dem. Alla former av

gester, blickrörelser och verbala yttranden är en del av att skapa upplevelsen av en "naturlig" interaktion och det har visats att framförallt blicken används intuitivt av människor för att koordinera sin undervisning för att underlätta en robots inlärningsprocess (Brazeal 2009).

Vilket för oss fram till vår robot och frågeställningen vi arbetat fram under projektets gång: kan en funktionsdesignad robot uppvisa ett beteende som får en människa att uppleva förståelse hos roboten? Frågan är intressant inte bara för utveckling av sociala robotar men även för mer filosofiska diskussioner som vad det är som krävs för att något ska anses intelligent. Världen bryr sig trots allt inte om vad som händer på insidan, endast handlingarna som utförs.

3 Experiment

Scenario

För att uppgiften skulle vara av en storlek som var rimlig att genomföra under kursen var det nödvändigt att begränsa den till ett specifikt scenario. I Ikaros finns redan färdiga moduler för att identifiera kuber och därför togs ett scenario med fokus på kuber fram som en bra utgångspunkt. Vårt scenario har följande steg:

1. Två personer sitter på varsin sida av ett bord.
2. En av dem bygger med klossar (med en hand).
3. Den andra har i uppgift att imitera byggandet med klossar.
4. Den imiterande personen tittar på ett speciellt sätt för att lära sig byggandet.
5. Den imiterande personen bygger med klossar.

Den sista punkten i scenariot genomförs aldrig då roboten inte kan greppa föremål (se nedan för närmare beskrivning av roboten) men är det tänkta slutmålet för experimentet så att tittbeteendet blir "naturligt" för en inläringssituation.

Specifikt för experimentet användes sex klossar och dessutom genomfördes scenariot i två olika varianter: en där försöksledaren endast tittade på sitt byggande och en där ögonkontakt söktes. I båda varianterna lade försöksledaren ner handen på bordet i några sekunder i mitten av bygget. Alla försökspersoner genomförde båda varianterna, men ordningen av dessa omväxlades.

Genomförande

10 studenter rekryterades från LTH-området. De blev tillsagda att det var ett experiment för att testa mönsterminne och fick en chokladbit för sitt deltagande. De ombads observera försöksledaren och sedan försöka imitera det försöksledaren gjort. Verbal kommunikation var inte tillåtet under experimentets gång. Försöken filmades med två kameror där en var in zoomad på ansiktet och en överblickade bygget för att kunna avläsa deltagarens tittbeteende samt vad hen tittar på. Deltagarna fick skriva på ett godkännande om filminspelning där det gjordes klart att materialet förstörs när kursen avslutats i januari. För att säkerställa att försöksdeltagarna inte förstod vad

experimentet gick ut på frågade vi dem om detta efteråt. Ingen trodde dock att det handlade om något annat än minne.

Resultat

De olika beteenden vi kunde urskilja hos försökspersonerna kan delas upp i fem följande punkter:

1. När de ser vart testledarens hand är på väg predicerar hen vilken kloss åt det hållet som ska plockas upp.
2. När testledaren tagit tag i en kloss flyttas blicken direkt till byggområdet, innan testledarens hand hunnit röra sig dit.
3. När testledaren lägger ned sin hand tittar försökspersonen hen i ansiktet som om de undrar vad hen håller på med.
4. När det bara är en kloss kvar bryr sig inte försökspersonen om att titta på den sista klossen utan tittar enbart i mitten.
5. Redan efter första klossen har försökspersonen förstått att det är i mitten som byggprojektet pågår.

Resultatet i form av dessa fem punkter kan ses i Diagram 1 där alla tjugo filmer har sammanställts.

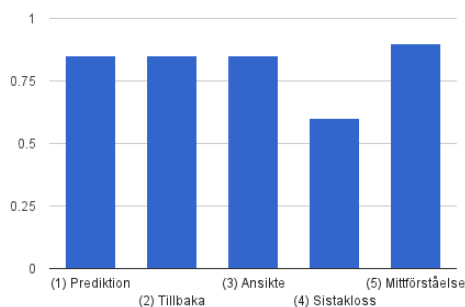


Diagram 1. Vertikal axel visar andel av försöken där beteendet uppvisades.

En mer detaljerad beskrivning av de olika tittbeteendena finns i appendix B.

4 Roboten

De verktyg vi har till vårt förfogande för att lösa problemet är en robotarm, en Kinect-kamera och mjukvaruplattformen Ikaros.

Roboten är en relativt enkel konstruktion med tre stycken olika servon; en för att vrida armen och två stycken för att luta armen framåt och bakåt. Högst upp på armen sitter två stycken ögon för att det ska vara tydligt var roboten tittar, se Figur 1.



Figur 1. Uppställning med robot och Microsoft Kinect.

För att ta in en ström av bilder på det som sker framför roboten har vi en Kinect från Microsoft (för tekniska specifikationer, se Kinect for Windows, 2014). Kinecten består i huvudsak av två sensorer - en RGB-kamera och en djupkamera som utnyttjar ett rutnät av infraröda punkter för att räkna ut avståndet. Koordinatsystemet som kameran använder fungerar på följande sätt; varje pixel får en 2-dimensionell koordinat i en matris som är lika stor som kamerans upplösning (640x480) och för varje pixel finns ett djup givet i mm.

Ikaros-plattformen är ett verktyg för robotstyrning och för att skapa olika neuromodeller. Det använder sig av moduler som programmeras i C/C++ och kopplas sedan ihop med XML-filer som kallas Ikaros Control files. Modulerna har alla en funktion "Tick" samt ett valbart antal inputs och outputs vilka alla är i form av vektorer, matriser eller en kombination. Ikaros arbetar stegvis (eng. tick) där grundläget är att för varje tick kör en modul koden i sin Tick-funktion och lägger eventuellt resultat i sin output. Om modulen använder input från en annan modul är innehållet den andra modulens output från förra tick-omgången. Standarden för koordinatsystemet i Ikaros är 3-dimensionellt givet i mm, lägg märke till att detta skiljer sig från koordinatsystemet i Kinecten. För mer ingående beskrivning av Ikaros läses med fördel Balkenius (2010).

5 Problem att lösa

Från de olika beteenden försökspersonerna uppvisade under experimentet har vi identifierat åtta olika problem som

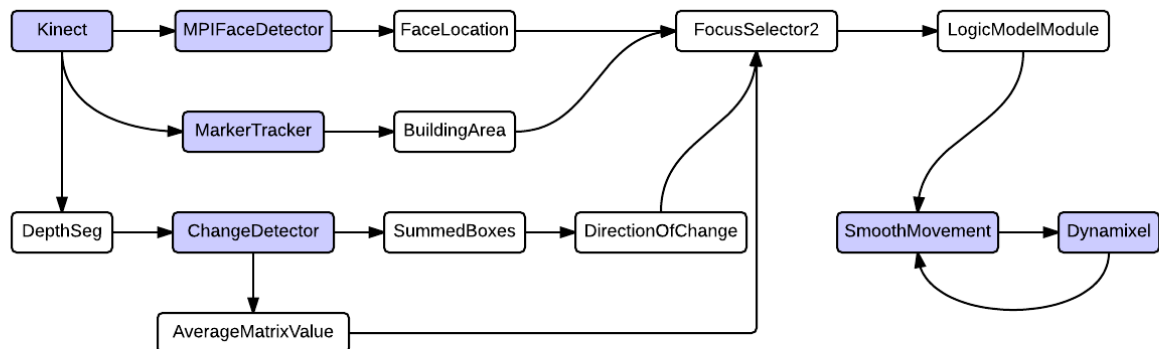
behövde lösas för att en implementation av roboten skulle kunna utföra det uppsatta scenariot. Dessa är:

1. Roboten måste kunna känna igen kuber.
2. Roboten måste kunna känna igen ett ansikte.
3. Roboten måste kunna förutspå vilken kub som ska plockas upp.
4. Roboten måste kunna skilja mellan en kub som har plockats upp och en som inte har det.
5. Roboten måste vara medveten om vart bygget sker någonstans.
6. Roboten måste kunna skilja mellan när personen som bygger tar en paus i byggandet och när hen faktiskt bygger.
7. Ett styrsystem som gör att roboten kan titta på valfri punkt i rummet måste implementeras.
8. Ett enkelt AI som bestämmer vad roboten ska fokusera blicken på måste implementeras.

Ambitionen från början var att i så stor utsträckning som möjligt modularisera lösningarna för enklare implementation i Ikaros.

6 Presentation av lösning

För att lära oss Ikaros och snabbt komma igång valde vi att utveckla roboten utifrån en funktionell design enligt definitionen i teoridelen ovan. Hur modularkitekturen ser ut för vår implementation kan ses i Figur 2, där moduler som finns inbyggda i Ikaros är markerade i blått. Hur dessa fungerar beskrivs utförligt på Ikaros-projektets hemsida (Ikaros Project, 2014). Värt att notera är att modulerna "Kinect" och "Dynamixel" inte är kameran respektive roboten i sig utan moduler som kommunicerar med dessa.

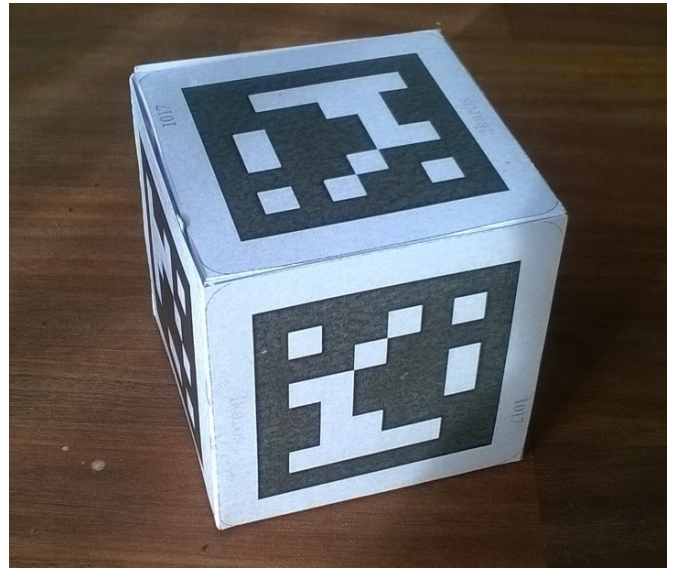


Figur 2. Modularkitektur. Pilarnas riktning indikerar informationsflödet. Kinect är robotens "ögon" och Dynamixel är robotens "muskler". Mörk bakgrund indikerar moduler vi inte skrivit själva.

Igenkänning av kuber

Som nämnts tidigare finns det redan en färdig modul i Ikaros som känner igen kuber med "markers" (rutmönster liknande QR-koder, se Figur 3) på sidorna; *MarkerTracker*. Denna

modul tar in RGB-bilden från Kinect-modulen, identifierar alla upphittade "markers" och skickar vidare information om dessa till *BuildingArea*, som beskrivs närmare senare i rapporten.



Figur 3. Kub med det rutmönster som modulen *MarkerTracker* känner igen.

Ansiktsigenkänning

För igenkänning av ansikten används en färdigskriven modul som heter *MPIFaceDetector*. Eftersom koordinaterna från *MPIFaceDetectors* output inte har samma enhetsform som resterande moduler, har därför en modul *FaceLocation* skapats. *FaceLocations* uppgift är att räkna om koordinaterna från *MPIFaceDetector* till koordinater i rummet.

Förutspå vilken kub som ska plockas upp

För att roboten skall kunna förutse vilken kub som handen rör sig mot har fyra stycken moduler med olika ansvarsområden implementerats.

- *DepthSeg* tar in djupdata från Kinecten och filtrerar ut de bildpunkter som finns i ett djupspann definierat av en maximum- och en minimumgräns. Då byggaren sitter vid ett bord med klossarna

framför sig kan gränserna därför väljas så att endast handen och klossarna hamnar i spannet.

Den filtrerade bilden skickas sedan till ChangeDetector, en inbyggd modul som jämför bilden i nuvarande tick med den som skickades in föregående tick. Bilden i ChangeDetectors output innehåller därmed endast de bildpunkter som förändrats.

- SummedBoxes utnyttjar att vi får en matris från ChangeDetector och skapar från den en ny matris i form av en "Summed Area Table" (Wikipedia, 2014). Summed Area Table fungerar kortfattat så att man kan få ut summan av alla punkter inom valfritt rektangelformat område i den ursprungliga matrisen genom att stoppa in rektangelns fyra hörn. Modulen utnyttjar detta för att dela in bilden i fem staplar och skickar ut en vektor med summan för varje stapel. På detta vis ser vi vilken del av bilden som rör sig och eftersom systemet bara behöver konstatera om det är rörelse i till höger eller vänster om klossarnas byggområde fungerar denna förenkling.
- Den sista modulen för att bestämma i vilken riktning handen rör sig är DirectionOfChange som jämför värdena den får från SummedBoxes med sparade värden som SummedBoxes skickade föregående tick. Med hjälp av ett valbart tröskelvärde skickar sedan DirectionOfChange ut information om rörelse detekterats till höger, till vänster eller inte alls.

Förstå vart byggandet sker

Förutom att hålla reda på information om kuberna har BuildingArea en annan viktig funktion. Modulen definierar var själva byggområdet är någonstans. Detta är bestämt innan roboten körs igång, det finns alltså ingen smart lösning som gör att roboten kan förstå var byggområdet är placerat. Dock bygger DirectionOfChange på att byggområdet är i mitten av bilden för att få allt att fungera. Byggområdets placering bestäms genom att definiera maximum- och minimumgränser för alla tre dimensioner angivet i mm. Förutom koordinaterna för varje kub skickas även ett värde med i outputn som definierar om kuben befinner sig innanför eller utanför byggområdet.

Förstå när en kub plockas upp

Som nämnts tidigare har BuildingArea koll på all relevant information om kuberna och skickar sedan detta till FocusSelector2. Avgörandet om en kub har plockats upp görs direkt i FocusSelector2 genom att jämföra kubernas placering i höjddled ifrån BuildingArea mot ett gränsvärde. Det vill säga att en kub som har lyfts över detta gränsvärde anses vara upplockad förutsatt att den är utanför det definierade byggområdet.

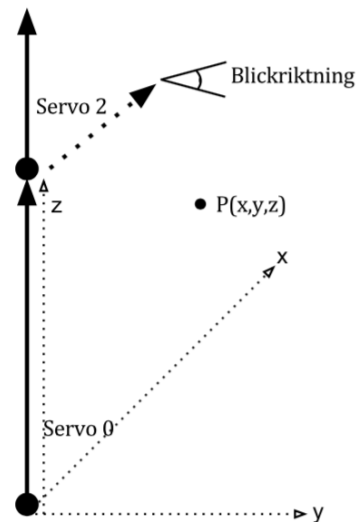
Styrsystem som möjliggör att roboten tittar på valfri punkt i rummet

De kriterier vi hade för styrsystemet var att det skulle kunna ta in en valfri koordinat i rummet och (inom servonas fysiska begränsningar) simulera att roboten tittar på den punkten. Det togs tidigt ett beslut att skapa en matematisk modell av

roboten och rummet den befinner sig i. Servona styrs av Dynamixel-modulen som tar in hastigheten som servona ska byta position med samt målpositionen i grader för varje servo för att sedan byta läge på servona till den önskade positionen.

Design av styrsystemet

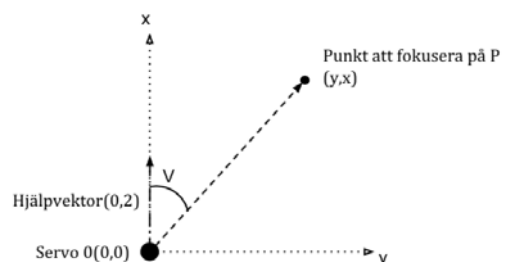
För att göra beräkningarna så enkla som möjliga gjordes först en vektorrepresentation av roboten. I Figur 4 visas vektorrepresentationen av roboten som sedan användes för att skapa beräkningarna. För en full redogörelse av hur vektorrepresentationen skapades, se Appendix A.



Figur 4. Vektorrepresentation av roboten i ett tredimensionellt koordinatsystem.

För att möjliggöra för roboten att nå valfri punkt P med blicken delas beräkningarna upp i två delar, en för den övre servon och en för det nedre. Mittenservot används alltså inte men vi bedömde denna förenkling som godtagbar då roboten ändå kan titta på alla punkter och framförallt blir beräkningarna mycket mer lätthanterliga. För servo 0 som styr riktningen i horisontell led är det bara koordinaterna x och y som behövs. Likaså för servo 2 som styr riktningen i vertikal led är det bara längden av vektorns projektion i x, y planet samt z som behövs. Beräkningarna görs då på två plan istället för i tre dimensioner samtidigt och blir därmed enklare att genomföra.

Då Dynamixel-modulen tar en absolut målposition som input behöver vår hittills relativa position ta hänsyn till detta. Med en hjälpvektor längs x-axeln kan vinkelskillnaden mellan hjälpvektorn och den önskade riktningsvektorn till den absoluta målpositionen räknas ut med skalärprodukt enligt Figur 5.



Figur 5. En hjälpvektor används för att få fram punkten P.

Vinkeln v räknas ut med skalärprodukten av hjälpvektorn och vektorn till P.

$$\cos v = \frac{0 * y + 2 * x}{\sqrt{0^2 + 2^2} * \sqrt{y^2 + x^2}}$$

Om $y > 0$,

$$\text{Målposition servo } 0 = 180^\circ + v$$

Om $y < 0$,

$$\text{Målposition servo } 0 = 180^\circ - v$$

Att målpositionen utgår från 180 grader beror på att det är vinkelskillnaden mellan servots nollpunkt och axeln x. Även för den vertikala positionen är det önskvärt att göra beräkningarna i ett plan istället för i ett rum och detta är också möjligt då det bara är längden av vektorns projektion i y,x-planet och z-värdet som behövs för detta.

Målpositionsvektorn (y,z,x) och hjälpvektorn $(0,0,2)$ görs om till vektorer i samma plan.

$$(\sqrt{y^2 + x^2}, z) \text{ och } (2,0)$$

Skalärprodukten används sedan för att räkna ut vinkelskillnaden u .

$$\cos u = \frac{\sqrt{y^2 + x^2} * 2 + z * 0}{\sqrt{\sqrt{y^2 + x^2} + z^2} * \sqrt{2^2 + 0^2}}$$

Och precis som i horisontell riktning så kollar vi om z är positiv eller negativ.

Om $z > 0$,

$$\text{Målposition servo } 2 = 180^\circ + u$$

Om $z < 0$,

$$\text{Målposition servo } 2 = 180^\circ - u$$

Att målpositionen utgår från 180 grader beror på att blickvektorn är vinkelrät mot servot vilket gör att vinkeln är parallell med x-axeln (och hjälpvektorn) då servots position är 180 grader.

Implementation av styrsystemet

Den matematiska modellen finns implementerad i LogicModelModule och alla beräkningar görs där. Det bör även noteras att de punkter som kommer in till modulen är i kamerans koordinatsystem, det vill säga att origo är i kameran och inte vid roboten. För att kunna göra alla beräkningar görs punkterna om till robotens koordinatsystem genom att lägga till en offset på varje koordinat i punkten. Denna offset är avståndet mellan kameran och foten på servo 0 för y och x-värdena och foten på servo 2 för z-värdena.

Detta designbeslut togs för att göra styrsystemet så självständigt som möjligt.

För att göra robotens rörelser mjukare används modulen SmoothMovement som utnyttjar Critically damped springs-metoden (Juckett, 2012). Denna modul lades in mellan LogicModelModule och Dynamixel och använder sig av feedback på positionerna för respektive servo för att få till de mjuka rörelserna.

Beslutsalgoritm

Det som binder ihop allting är modulen FocusSelector2 som bestämmer vad roboten ska titta på och skickar aktuella koordinater till LogicModelModule. Anledningen till att FocusSelector2 heter som den gör är för att det är den andra versionen av FocusSelector. Den första versionen hanterade inte alla delar och behövde skrivas om helt.

Som visas i Figur 2 tar FocusSelector2 in data från fyra moduler enligt följande:

- BuildingArea skickar kubernas position och ifall de är placerade i byggområdet
- FaceLocation skickar ansiktets position
- Handens riktning från DirectionOfChange (vänster, höger eller ingen förändring)
- AverageMatrixValues output för att veta om handen rör på sig

Utgångsläget är att modulen letar efter handrörelser, antingen för att titta på den kub som handen rör sig mot eller för att avgöra om roboten ska titta på byggarens ansikte vilket den gör om det inte sker några handrörelser under ett valbart antal tick. Efter att den har hittat en handrörelse och predicerat vilken kub handen rör sig mot väntar modulen på att en kub som är utanför byggområdet skall tas upp.

När en kub tas upp tittar roboten tillbaka i byggområdet tills den uppluckade kuben har placerats i byggområdet. Efter att kuben har placerats i byggområdet väntar modulen på nya handrörelser och detta beteende upprepas tills det bara finns en kub utanför byggområdet kvar. När alla kuber har placerats i byggområdet förutom en, kommer robotens blick att stanna kvar i byggområdet.

AverageMatrixValue-modulen behövs för att DirectionOfChange skickar "ingen förändring" även om handen rör sig i mitten av bilden. Här finns möjlighet att skriva om antingen DirectionOfChange och FocusSelector2 så att AverageMatrixValue inte behövs.

7 Diskussion

Arbetsprocessen

Under projektets gång har vi arbetat med en väldigt platt och demokratisk arbetsprocess. Det har varit mycket fokus på det tekniska, vilket alla medlemmar är väldigt intresserade av, så blev det naturligt att arbeta med det alla tre. Det visade sig också passa bra ihop med en funktionell design (jämfört med biologisk sådan, se teoridelen). Även om vi under processens

gång arbetat med olika delar av systemet så har vi hela tiden haft en öppen dialog där vi diskuterat fördelar och nackdelar med olika lösningar. Då vi bara har varit tre personer som arbetat tillsammans har det här relativt ostrukturerade arbetssättet fungerat väldigt bra, fast hade vi varit fler skulle det eventuellt behövts mer struktur över vem som gjorde vad och varit tydligare med deadlines. Vi har dessutom haft fördelen av att nästan alltid sitta tillsammans och arbetat. Detta har gjort att ledtider för svar på frågor och diskussioner varit nära på obefintliga samt att alla projektmedlemmar hela tiden har haft bra förståelse om vad som hänt med hela systemet och inte bara vad man själv håller på med.

Vi har under hela processen arbetat iterativt. Detta med grund i att vi snabbt kom överens om att det är bättre att få små delar av systemet att fungera först och att göra en enkel implementation med begränsningar istället för att sträva efter avancerad funktionalitet redan från början. Detta grundar sig också i tanken att vi på så vis snabbare kunde lära oss Ikaros-systemet och på bra grund kan skapa mer avancerade lösningar i den andra delkursen. Arbetssättet har visat sig fungera väl då vi idag står med en fullt fungerande implementation som kan hantera ett begränsat scenario bra. Trots detta känner vi verkligen inga begränsningar i hur det fortsatta arbetet ska fortgå utan det känns som att vi kan gå flera olika vägar genom att till exempel generalisera scenariot roboten verkar i eller göra beslutsalgoritmen mer intelligent och snabb. Med andra ord har valet med iterativ utveckling fallit väldigt väl ut.

Experiment

Experimentet var redan från början tänkt vara mer av utforskande typ än strikt vetenskapligt, detta för att få en fingervisning om potentiella beteenden att implementera. Vi blev glatt överraskade att resultatet var så enhetligt med stora utslag på de olika beteendena som identifierades. De två olika scenarion vi hade gav ingen skillnad i beteendet hos försökspersonerna och det spelade heller ingen roll i vilken ordning vi spelade upp scenarion. Dessutom var skillnaderna mellan de två försöken obefintliga för de flesta av deltagarna.

Efter detta konstaterande har vi behandlat alla de 20 försöken enhetligt. Med dessa homogena resultat som är oberoende av de olika parametrarna vi ändrade på, kan vi konstatera att för det här specifika scenariot finns det med stor sannolikhet ett "bestämt tittmönster". Därför kunde vi med ett starkt stöd från experimentet implementera tittmönstret i vår robot.

Naturligtvis finns det många saker man skulle kunna ändra på i scenariot som skulle kunna påverka resultatet. T.ex. har vi inte undersökt hur verbal kommunikation skulle kunna påverka. Men eftersom vår robot inte har något sätt att uppfatta tal har denna idé förkastats. Dock hade vi kunnat undersöka vad som händer om testledaren bjuder in till samarbete med kroppsspråk och blickar. Vilket kanske hade resulterat i att försökspersonerna tittat mer på testledaren istället för klossarna, men det hade också blivit mer komplicerat att särskilja kausaliteter i beteenden och därmed försvårat analysen. Vi hade också kunnat undersöka hur

försökspersoner reagerar om de inte får några instruktioner om att byggmönstret ska memoreras. Fast då skulle risken finnas att försök måste förkastas för att deltagarna blir misstänksamma eller fundersamma.

Utrustningens begränsningar

Något annat som bör diskuteras är utrustningens begränsningar och vad detta har lämnat för avtryck på vår lösning. En av de stora begränsningarna har varit att vi använt en Microsoft Kinect som kamera, vilket vi fått anpassa oss till. Bland annat har den ett väldigt snävt synfält som gör att kuber måste placeras relativt rakt framför kameran för att vara med i bild. Ett annat problem vi stötte på är kamerans låga upplösning tillsammans med djupkamerans begränsning i djupet - räckvidden börjar 50 cm från kameran. Detta blir ett problem då den vanliga kameran inte kunnat detektera kuberna på längre avstånd än 50 cm. Detta moment 22 löstes genom att tillverka större kuber som kunde ses på längre avstånd och därmed öppnade upp för djupkameran att användas för exempelvis rörelsedetektion. Det bör dock tilläggas att en kamera med högre upplösning inte hade varit till mycket hjälp då en del av de beräkningar som görs då hade blivit alltför krävande för att systemet ska kunna flyta på mjukt.

Designbeslut och begränsningar

Under arbetsprocessen har det varit naturligt att göra en del kompromisser, många av dessa har berott på att det har varit nödvändigt att spara tid för att kunna lägga denna på andra delar av systemet. Detta har gjort att en helhetslösning som fungerar har kunnat slutföras redan nu. I början var det svårt att förutse när en fungerande lösning skulle bli klar, men vårt nuvarande resultat är bättre än vad vi förväntade oss.

En av dessa kompromisser var ett beslut om att kameran måste stå 82 cm från bordskanten för att kunna filtrera ut handen med modulen *DepthSeg*. En annan kompromiss var att använda den redan befintliga Ikaros-modulen *MPIFaceDetector* för att känna igen ansikten trots att den fungerar långt från optimalt - den tappar ofta ansiktet samt tror att objekt som inte är ett ansikte är det. Modulen håller heller inte koll på hur långt bort från kameran ett ansikte är. Anledningen till att modulen ändå användes är att en av de andra projektgrupperna i kursen spenderat mycket tid på ansiktigenkänning och vi ansåg det därför bättre att uppfinna egna hjul. Lösningen på att *MPIFaceDetector* inte håller koll på ansiktens djup blev att bestämma ansiktets avstånd till ca 130 cm från kameran; ett tillägg på 50 cm till de tidigare nämnda 82 cm för att hitta handen.

En annan modul som har begränsningar är *BuildingArea* som i dagsläget definierar en zon för var byggandet sker, var denna zon är belägen samt meddelar vilka kuber som är utanför respektive innanför zonen. Detta är inte heller optimalt då det gör att lösningen blir väldigt scenariobaserad och inte generell. En bättre lösning är om modulen kollar var alla kuber i bilden är och till exempel räknar ut en byggplats som området där flest kuber är belägna. Även *DirectionOfChange* har begränsningar på så sätt att den bara

kollar riktningen åt höger eller vänster och antar att byggandet sker i mitten, där modulen inte kollar på några ändringar. För att generalisera *DirectionOfChange* borde den kunna hitta en hand och ta ut en rörelse åt vilket håll som helst. Det är även fördelaktigt om den kan ta reda på vart byggandet sker och ändra tröskelvärden för vad som är en rörelse och inte efter det.

Styrsystemet, *LogicModelModule*, har även det flera begränsningar. Till exempel använder det bara två utav tre servon. Detta beror på att vi sett det som nödvändigt att börja med låg komplexitet i beräkningarna för att följa en iterativ process. Detta möjliggjorde att vi kunde använda det för att testa andra delar av vår implementation, något som har visat sig mycket värdefullt. En annan kompromiss vi har gjort är att bestämma att robotens ögon har centrum i den översta servons axel istället för längst upp på roboten. Även detta beslut togs med grund i att låg komplexitet av beräkningarna var att föredra, dessutom är den synliga skillnaden i vart roboten tittar väldigt liten med denna kompromiss. Trots dessa kompromisser är modulen helt funktionell och kan dessutom vidareutvecklas till att bli mer generellt och funktionellt.

Förslag på ändringar är att ge modulen flera olika lägen, till exempel "titta på punkter" eller "nå punkter med robotarmen". Detta skulle tillåta styrsystemet att även användas för att flytta på saker och inte bara för att titta på saker. Med detta bör också den tredje servon tas med i beräkningarna då detta gör att fler punkter i rummet kan nås.

Utöver detta är implementeringen av hjälpvektorer inte optimal. Hjälpvektoreorna används för att räkna ut vinkelskillnaden mellan en axel i koordinatsystemet och den vektor som sträcker sig mellan origo och de punkter som ska nås av blicken. Denna vinkel läggs sedan till eller dras ifrån på antalet grader mellan servons nolläge och axeln med hjälpvektorn. Något vi har insett är att det är bättre att sätta hjälpvektorn parallell med 0 grader för varje servo och på sätt undvika en offset som motsvarar vinkeln mellan 0 grader för en servo och axeln med hjälpvektorn. Under projektets gång har vi använt koordinater på flera olika former (xyz, yxz, yzx etc), vilket har gjort att klassen idag är lite rörig. Detta borde ändras så att koordinaterna är på samma form som i resten av systemet. Med dessa ändringar tror vi att modulen blir lättare att förstå och använda.

Som beskrivits ovan finns det många begränsningar i vår implementation och många av dem har gjorts för att få en första iteration fungerande och testbar så fort som möjligt. En del idéer har vi behövt förkasta på grund av komplexitet eller tidsåtgång, bland annat har vi haft många olika idéer på hur man kan filtrera ut en hand i en bild. Vi spenderade en del tid på att leta upp och utvärdera olika metoder i litteratur, forskningsartiklar och på Internet. En av de bättre kandidaterna var att ta ut färgen på ansiktet för att leta upp någonting med samma färg som troligtvis skulle vara handen och/eller armen. Efter diskussion inom gruppen togs dock beslutet att inte gå vidare med den metoden då den bedömdes som svår att implementera. Dessutom hade detta potentiellt satt begränsningar på vilka färger som får finnas med i bilden

samt om försöksledaren bär långärmat eller kortärmat. Vi tror dock att en kombination av den ovan beskrivna metoden och den vi idag använder med *DirectionOfChange* hade kunnat ge ett utmärkt resultat och är något vi kan kolla vidare på.

Interaktion med andra projektgrupper

Under projektets gång har det förekommit en hel del interaktion med de två andra grupperna i kursen som arbetar med robotar. Framförallt har det varit en öppen diskussion om olika problem, men vi har också kunnat använda en av de andra gruppernas moduler för mjuka rörelser, *SmoothMovement*. Den har varit mycket lätt att använda i vår implementation och har förbättrat vårt slutresultat. Vårt mål och förhoppning är att de andra grupperna ska få användning för några av våra moduler.

Hur trovärdig är vår robot?

Under projektets gång började vi fundera över hur trovärdig vår robot faktiskt är. När man sitter och bygger med klossar framför roboten vet man om att roboten egentligen inte kan förstå vad som händer. Roboten stirrar med sina frigolitbollar till ögon och sin mekaniska "hals" men ändå får man en känsla av att det är något som händer, man vill att roboten skall förstå. Även med kunskapen om exakt hur den fungerar vill man ändå väcka liv i roboten och vi pratar nästan med den som att det vore en levande person, även ifall vi vet att den inte hör. Effekten märktes redan i början och blev starkare för varje ny modul och funktion vi implementerade.

På det viset har vi, åtminstone med oss själva, som testpersoner bekräftat det som togs upp i teorin - att vi människor har en inbyggd benägenhet att se liv i objekt trots att vi egentligen vet att det inte är så.

Det var denna erfarenhet som ledde oss till frågeställningen om huruvida en funktionellt designad robot kan titta på ett vis som får en människa att uppleva förståelse hos roboten? Man skulle här kunna dra en parallell till det berömda (eller ökända) Turing-testet. Det finns flera tolkningar och versioner av detta men förenklat går det ut på att en människa i text konverserar med antingen en dator eller annan människa och efter en viss tid bedömer om partnern var av kött eller kod. Om ett visst antal försökspersoner tror att datorn är människa har Turing-testet avklarats.

Det här säger givetvis inget om datorprogrammets förståelse eller intelligens utan snarare om möjligheten till imitation och då bortser vi ändå från att kognitionsforskningen knappast kan påstås vara enad om vad varken förståelse eller intelligens betyder. Men parallellen är värdefull för att experimentera med tanken om vad det skulle betyda att skapa en förståelseupplevelse. För om vi kan försäkra oss om att roboten uppvisar ett tittbeteende som tolkas som att roboten förstått endast när den faktiskt förstått människans instruktioner så räcker det för att den ska fungera i vardagssituationer såsom städning eller diskning. Detta hade varit intressant att testa i ett nytt experiment där försökspersoner får agera byggare och sedan bedöma om de

tror att roboten förstått deras instruktioner. Kruxet är att vi då kommer till frågan om hur vi som robotkonstruktörer eller kodare kan försäkra oss om att roboten förstått - hur vet vi det?

Troligtvis kommer vi på grund av detta inte ifrån att någon form av inläring behöver ske vilket för oss tillbaka till funktionell respektive biologisk design. Dagens mest avancerade system för detta brukar benämnas "machine learning" som bygger på avancerade statistiska modeller och därmed bör klassas som funktionell design. Metoden ligger till grund för många framgångsrika applikationer inom artificiell intelligens såsom den Jeopardy-spelande Watson (IBM, 2014) och självgående bilar (Wikipedia, 2014). Vi har inte kommit riktigt lika långt med biologiskt inspirerade lösningar, även om det finns bra exempel i form av företaget Numenta (Numenta, 2014).

Med tanke på hur lite vi faktiskt vet om hjärnan och att den trots allt bygger på matematiska regler (vilket den ur reduktionistisk synvinkel måste göra för att följa universums fysiska lagar) skulle man kunna säga att nyss nämnda exempel egentligen handlar om samma sak - att hitta rätt algoritm. I så fall kan man spekulera i om det alls är någon skillnad mellan funktionell och biologisk design.

Vi har ingen aning, men om vår robot kan bli en liten pusselbit på vägen dit är det fantastiskt.

Referenser

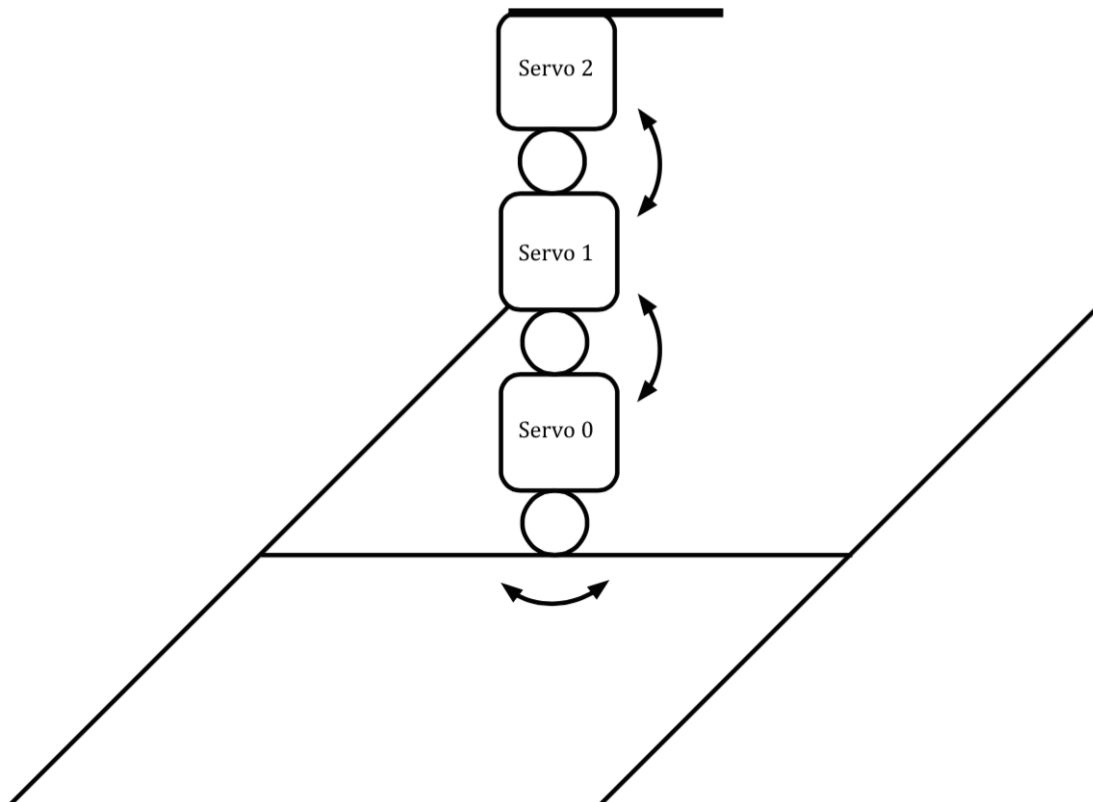
- Gazzaniga, M.S., Ivry, R.B. and Mangun, G.R. (2013). *Cognitive Neuroscience: the biology of the mind*, W.W. Norton, New York, New York.
- Johansson, R., Holsanova, J., & Holmqvist, K. (2013). Using eye movements and spoken discourse as windows to inner space. In Paradis, C., Hudson, J. and Magnusson, U (Eds), *The construal of spatial meaning: Windows into conceptual space*, 9-28, Oxford University Press.
- Embodied cognition - Wikipedia, the free encyclopedia. Hämtad 22 oktober 2014, från http://en.wikipedia.org/wiki/Embodied_cognition
- Lungarella, M. (2007) Developmental robotics. *Scholarpedia*, 2(8):3104.
- Webb, A., Knott, A., & MacAskill, M. R. (2010). Eye movements during transitive action observation have sequential structure. *Acta psychologica*, 133(1), 51-56.
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3), 143-166.
- Nagai, Y. (2005, April). The role of motion information in learning human-robot joint attention. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on* (pp. 2069-2074). IEEE.
- Sakita, K., Ogawara, K., Murakami, S., Kawamura, K., & Ikeuchi, K. (2004, September). Flexible cooperation between human and robot by interpreting human intention from gaze information. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on* (Vol. 1, pp. 846-851). IEEE.
- Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1), 140-164.
- Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2012, March). Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 367-374). ACM.
- Sharkey, N., & Sharkey, A. (2010). The crying shame of robot nannies: an ethical appraisal. *Interaction Studies*, 11(2), 161-190.
- Breazeal, C. (2009). Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3527-3538.
- Kinect for Windows, 2014, <http://www.microsoft.com/en-us/kinectforwindows/>, hämtad 2014-10-20
- Balkenius, C., Morén, J., Johansson, B., & Johnsson, M. (2010). Ikaros: Building cognitive models for robots. *Advanced Engineering Informatics*, 24(1), 40-48.
- Ikaros Project, Hämtad 2014-10-20, från <http://www.ikaros-project.org>
- Summed Area Table - Wikipedia, the free encyclopedia. Hämtad 2014-10-20, från http://en.wikipedia.org/wiki/Summed_area_table
- Juckett, R. (2012). Hämtad 2014-10-22, från <http://www.ryanjuckett.com/programming/damped-springs/>
- IBM Watson, Hämtad 2014-10-24, från <http://www.ibm.com/smarterplanet/us/en/ibmwatson/what-is-watson.html>
- Google driverless car - Wikipedia, the free encyclopedia. Hämtad 2014-10-24, från http://en.wikipedia.org/wiki/Google_driverless_car
- Numenta. Hämtad 2014-10-24, från <http://www.numenta.com>

Appendix A - Styrssystem

De kriterier vi hade för styrsystemet var att det skulle kunna ta in en valfri koordinat i rummet och (inom servonas fysiska begränsningar) simulera att roboten tittar på den punkten. Det togs tidigt ett beslut att skapa en matematisk modell av roboten och rummet den befinner sig i. Servona styrs av Dynamixel-modulen som tar in hastigheten som servona ska byta position med samt målpositionen i grader för varje servo för att sedan byta läge på servona till den önskade positionen.

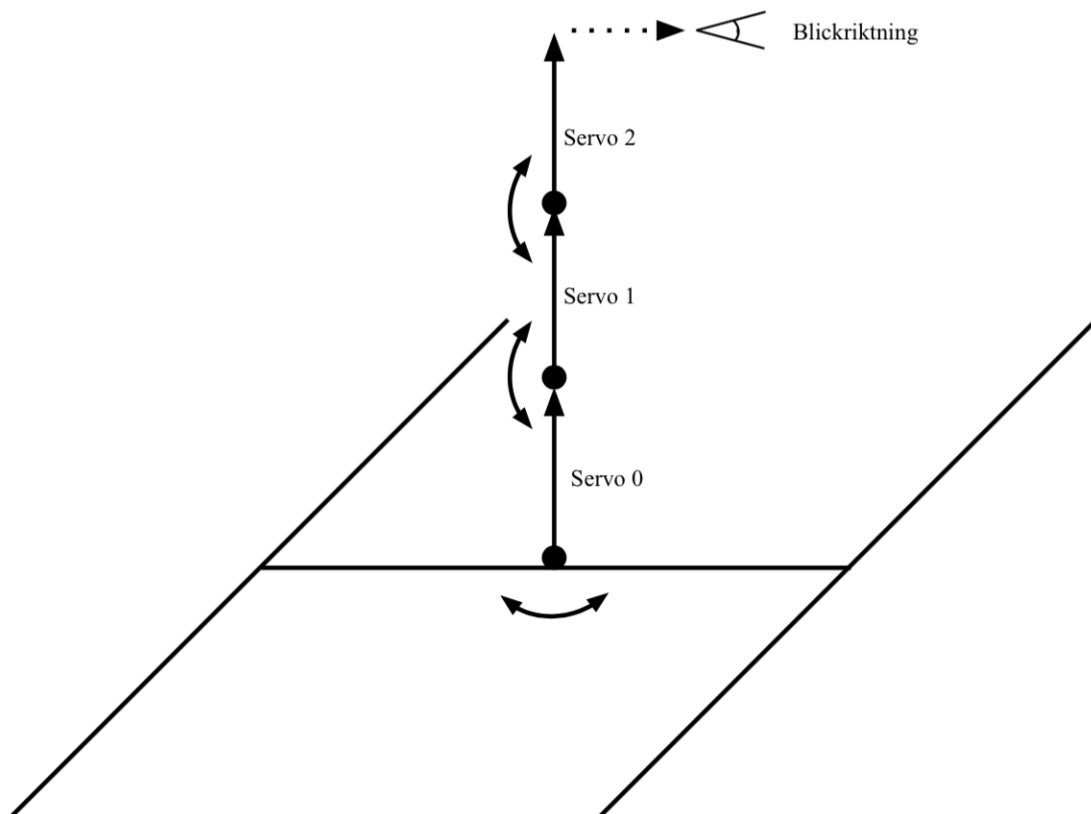
Design

För att göra beräkningarna så enkla som möjliga gjordes först en vektorrepresentation av roboten. I Figur A visas en skiss av roboten som sedan användes som grund för att representera roboten i vektorer.



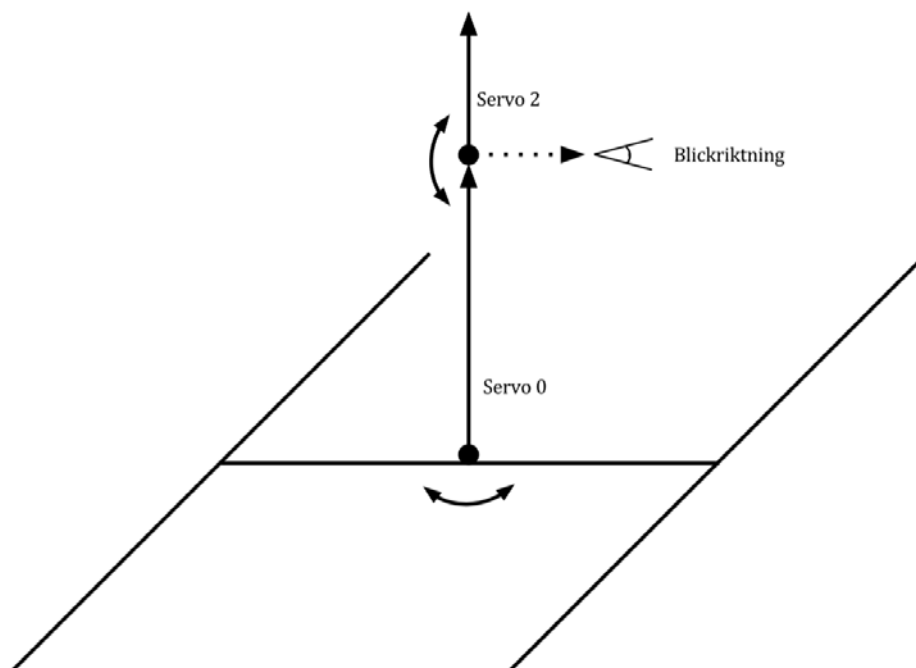
Figur A. För att göra beräkningarna så enkla som möjliga gjordes först en vektorrepresentation av roboten. I Figur A1 visas en skiss av roboten som sedan användes som grund för att representera roboten i vektorer.

Modellen av roboten vektoriserades sedan, detta syns i Figur B.



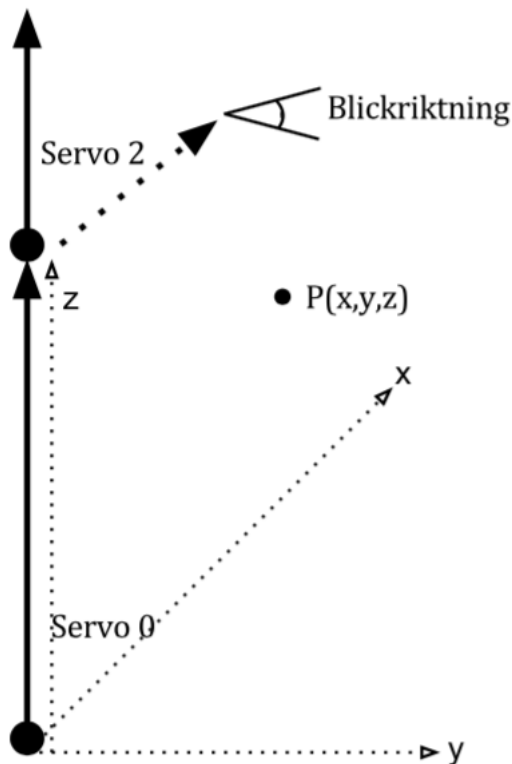
Figur B. Vektorpresentation av roboten där varje servo motsvaras av en vektor. Notera att blickriktningen representeras av en prickad vektor.

För att minska komplexiteten på beräkningarna gjordes sedan följande förenkling.



Figur C. Förenklad vektorisering av modellen.

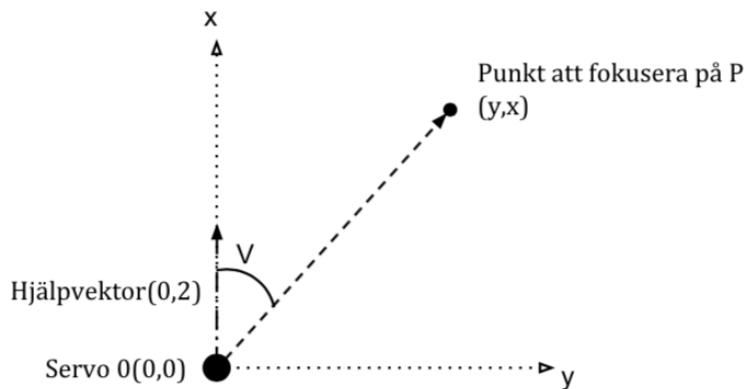
Från den förenklade vektormodellen i Figur C konstruerades ett koordinatsystem.



Figur D. Vektorrepresentation av roboten i ett tredimensionellt koordinatsystem.

För att möjliggöra för roboten att nå valfri punkt P med blicken delas beräkningarna upp i två delar, en för den övre servon och en för det nedre. Mittenservot används alltså inte men vi bedömde denna förenkling som godtagbar då roboten ändå kan titta på alla punkter och framförallt blir beräkningarna mycket mer lätthanterliga. . För servo 0 som styr riktningen i horisontell led är det bara koordinaterna x och y som behövs. Likaså för servo 2 som styr riktningen i vertikal led är det bara längden av vektorns projektion i x, y planet samt z som behövs. Beräkningarna görs då på två plan istället för i tre dimensioner samtidigt och blir därmed enklare att genomföra.

Då Dynamixel-modulen tar en absolut målposition som input behöver vår hittills relativa position ta hänsyn till detta. Med en hjälpvektor längs x-axeln kan vinkelskillnaden mellan hjälpvektorn och den önskade riktningens vektor till den absoluta målpositionen räknas ut med skalärprodukt enligt Figur E.



Figur E. En hjälpvektor används för att få fram punkten P.

Vinkeln v räknas ut med skalärprodukten av hjälpvektorn och vektorn till P.

$$\cos v = \frac{0 * y + 2 * x}{\sqrt{0^2 + 2^2} * \sqrt{y^2 + x^2}}$$

Om $y > 0$,

$$\text{Målposition servo 0} = 180^\circ + v$$

Om $y < 0$,

$$\text{Målposition servo 0} = 180^\circ - v$$

Att målpositionen utgår från 180 grader beror på att det är vinkelskillnaden mellan servots nollpunkt och axeln x. Även för den vertikala positionen är det önskvärt att göra beräkningarna i ett plan istället för i ett rum och detta är också möjligt då det bara är längden av vektorns projektion i y,x-planet och z-värdet som behövs för detta.

Målpositionsvektorn (y,z,x) och hjälpvektorn $(0,0,2)$ görs om till vektorer i samma plan.

$$(\sqrt{y^2 + x^2}, z) \text{ och } (2,0)$$

Skalärprodukten används sedan för att räkna ut vinkelskillnaden u .

$$\cos u = \frac{\sqrt{y^2 + x^2} * 2 + z * 0}{\sqrt{\sqrt{y^2 + x^2}^2 + z^2} * \sqrt{2^2 + 0^2}}$$

Och precis som i horisontell riktning så kollar vi om z är positiv eller negativ.

Om $z > 0$,

$$\text{Målposition servo 2} = 180^\circ + u$$

Om $z < 0$,

$$\text{Målposition servo 2} = 180^\circ - u$$

Att målpositionen utgår från 180 grader beror på att blickvektorn är vinkelrät mot servot vilket gör att vinkeln är parallell med x-axeln (och hjälpvektorn) då servots position är 180 grader.

Appendix B – Resultat från experimentet

De olika beteenden vi kunde urskilja hos försökspersonerna under experimentet kan delas upp i följande fem punkter:

1. När de ser vart testledarens hand är på väg predicerar hen vilken kloss åt det hållet som ska plockas upp.



Figur A. Prediktion av vilken kloss som ska plockas upp.

2. När testledaren tagit tag i en kloss flyttas blicken direkt till byggområdet, innan testledarens hand hunnit röra sig dit.



Figur B. Prediktion av vart den upplockade kuben ska sättas ned någonstans.

3. När testledaren lägger ned sin hand tittar försökspersonen hen i ansiktet som om de undrar vad hen håller på med.



Figur C. Tittar frågande på försöksledaren för att se vad som försiggår.

4. När det bara är en kloss kvar bryr sig inte försökspersonen om att titta på den sista klossen utan tittar enbart i mitten.



Figur D. Försökspersonen bryr sig inte om att titta på den sista klossen då det är uppenbart att det är den som ska plockas upp.

5. Redan efter första klossen har försökspersonen förstått att det är i mitten som byggprojektet pågår.



Figur E. Försökspersonen förstår på en gång att det är i mitten byggandet sker.

Resultatet i form av dessa fem punkter kan ses i Diagram 1 där alla tjugo filmer har sammanställts.

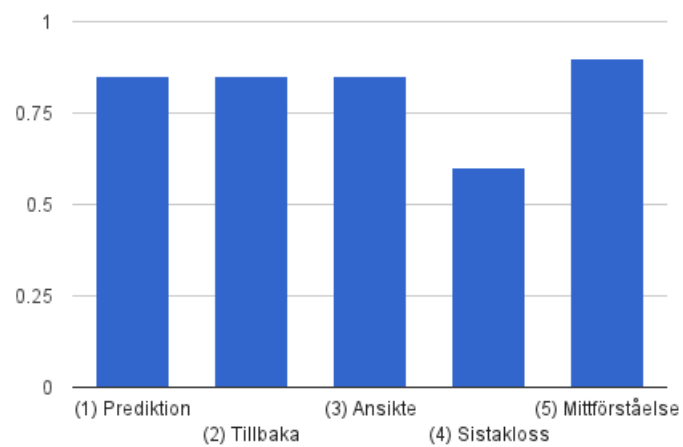


Diagram A. Vertikal axel visar andel av försöken där beteendet uppvisades.

Människa-Robotinteraktion I Virtuella Miljöer

David Lundberg, Lolita Olesen, Caroline Schouenborg, Eric Wall

Human-Robot Interaction, HRI, tests are mostly performed with a real physical robot prototype and real people. However, doing so requires development of the actual robot and setup of the environment and any other test parameter. This article examines the possibility of performing HRI tests in a virtual reality (VR) environment instead, which could hopefully lead to shorter lead times and reduced costs. The basis for the project was three studies that have been tested in a physical environment and replicated in this project's virtual environment. The first one was carried out in 2013 by Astrid Weiss and Markus Vincze, called "Don't Bother Me: Users' Reactions to Different Robot Disturbing Behaviours". The second one was carried out in 2002 by Juan Fasola and Maja J Mataric' and is called "Using Socially Assistive Human-Robot Interaction to Motivate Physical Exercise for Older Adults". The last one is called "The Revised Set of Measures to Assess Users' Attitudinal and Behavioral Responses to a Social Robot" by the writers Michiel Joosse, Aziez Sardar, Manja Lohse, Vanessa Evers from 2013. The project outlined in this report tried to resemble the original tests as much as possible, but in a VR environment instead. The first part of this project culminated in a pilot study, which mainly tested the environment, robot and HRI less so than the actual tests. The result of the pilot studie was then used to do the acual tests.

Introduktion

Det talas om att i framtiden underlätta serviceinriktade arbeten genom att avlasta personalen med robotar. Exempelvis hade robotar kunnat underlätta för den redan överbelastade hemtjänsten med att hjälpa äldre med vardagliga sysslor. Tekniken har sedan länge funnits tillgänglig men interaktionen mellan människor och robotar är fortfarande relativt outforskad. Att få människor att förlita sig på robotar och lära dem hur det går att interagera med robotar är fortfarande svårt och mer forskning behövs inom området. Ett stort problem med att bedriva den här forskningen är att det kan vara en kostsam process. (Weiss & Vincze, 2013)

Den här rapportens syfte är att med hjälp av en virtuell miljö kunna utföra användartester på robotar tillsammans med den tänkta slutanvändaren. I jämförelse med användartester i en fysisk miljö, med en fysisk robot, är tester i virtuella miljöer enkla och relativt billiga att utveckla. Vilket innebär att kostnader för testning skulle minska, samtidigt som det medför möjligheter att arbeta mer iterativt i utvecklingen av nya prototyper eftersom att de kan ändras snabbt. Hypotesen är att det går att utföra enklare Human-Robot Interaction, HRI, tester, där ingen fysisk beröring krävs, i virtuella miljöer lika väl som i fysiska miljöer.

För att testa hypotesen valdes det att replikera tre redan genomförda experiment inom tre olika områden. Studierna är kallade "Don't Bother Me: Users' Reactions to Different Robot Disturbing Behaviours", "Test The Revised Set of Measures to Assess Users" samt "Test Using Socially Assis-

tive Human-Robot Interaction". I den första studien undersöks tillit till en robot, i den andra görs en mätning kring den personliga sfären och i den tredje undersöks det hur robotar ska agera och tala för att det ska vara bekvämt för människor att interagera med dem. Studierna har efterliknats originalstudierna så långt som möjligt. Det som inte har gått att replikera eller helt enkelt inte har hunnits med på grund av tidsbrist, har istället valts att efterliknas genom att tillgripa "The Wizard of Oz"-tekniken. Det är en teknik där händelser istället simuleras så att testpersonen ska tro att det är på riktigt till exempel genom att spela upp robotrösten manuellt via en extern enhet kopplad till testpersonens hörlurar.

En virtuell miljö har använts för det här projektet, till vilken en virtuell robot, som testpersonerna senare interagerade med, lades till. Virtuella miljöer används för att ge användaren en illusion av deltagande i en syntetisk miljö istället för att externt observera den. Virtuella objekt har sedan adderats till den virtuella miljön, och animationer har adderats till roboten.

Resultaten från användartesterna har jämförts med resultaten från de ursprungliga testerna för att undersöka om det finns stöd för hypotesen. Slutsatsen är att det går att göra likvärdiga studier i en virtuell miljö som i en fysisk.

Projektet är indelat i två faser. Den första fasens syfte var att skapa en pilotstudie som skulle ge en grund för testerna och sedan kunna underlätta för testning i den andra delen av projektet. Pilotstudien gav en bra inblick i vad som var möjligt att göra samt vad som kunde bli ännu bättre. I den andra fasen togs kunskapen från pilotstudien tillvara för att förbättra miljö och robot. Några saker fick göras om från grunden, men lärdomarna och erfarenheterna från första delen har avspeglats i slutresultatet.

Tekniken som använts

Oculus Rift är en Head-mounted display (HMD) som används för att simulera den virtuella miljön. Med hjälp av Oculus Rift får testpersonen en fri tredimensionell rotation. Vilket innebär att testpersonerna kan förflytta sig genom rummen och använda sig av fler sinnen än att endast iaktta en datorskärm (Thern, Hadzovic, 2014). Under den första delen av projektet användes Oculus Rift Development Kit 1 (DK1), då det ännu inte fanns tillgång till Oculus Rift DK2 än. Inför den andra delen av projektet införskaffades Oculus Rift DK2, som har samma funktioner som den första versionen men bättre upplösning samt stöd för att följa huvudrörelser genom en extra kamera. Den nya versionen av Oculus Rift användes sedan under hela utvecklingsfasen. När testen skulle provas upptäcktes det att det inte gick att ha skärmen som primärskärm med den nya tekniken.

Därav beslutades det i gruppen att gå tillbaka till den första versionen av Oculus Rift för de resterande studierna då

det endast gick att utföra testen med denna teknik.

3Ds Max är en utvecklingsmiljö där 3D-komponenter kan skapas, som sedan används i andra miljöer. I första fasen användes 3Ds Max för att skapa den första roboten till pilotstudien. Efter att pilottestet genomförts fanns det behov av en bättre robot för de fortsatta studierna, vilket resulterade i att en robot från Unitys Asset Store importerades, Robot Kyle.

Den virtuella miljön importerades in till Unity redan från starten av projektet eftersom det skulle ta för lång tid att skapa den från grunden. Samtliga studier, under både första och andra fasen av projektet, har utförts i samma miljö.

I den virtuella projektetmiljön fanns det en robot som testpersonerna fick interagera med. Miljöer av detta slag används för att ge användaren en illusion av deltagande i en artificiell miljö istället för att externt observera denna (Rogers et al., 2011). Det är viktigt att den virtuella miljön efterliknar den verkliga i så hög utsträckning som möjligt för att på så sätt ge testpersonen en upplevelse av att den virtuella miljön är levande.

Teori

Det finns flera fördelar med att utföra tester i en virtuell miljö istället för att sätta upp tester i fysiska miljöer. En fördel kan vara att testledaren endast behöver fokusera på att hålla testet flytande för att inte störa upplevelsen av en levande miljö. Testpersonen har inte någon visuell uppfattning om den fysiska miljön som omger testpersonen vilket medför att denne inte reagerar på eventuella ansiktsuttryck eller gester som testledaren gör (Reeves, 2002). Nackdelarna väger inte lika tungt som fördelarna, men bör ändå inte förbises helt. Ett exempel kan vara att själva konceptet med virtuella miljöer samt teknik som Oculus Rift fortfarande är relativt nytt och okänt för många. För testpersoner som är helt obekanta med tekniken bör det räknas med att det finns en inlärningsperiod som kan påverka testresultaten (Thern, Hadzovic, 2014). För det här projektet är det därför viktigt att använda sig av testpersoner som inte är obekanta med teknik i den pilotstudien. Testpersonerna behöver inte ha använt virtuella miljöer tidigare, men ska snabbt kunna fokusera på projektets mål istället för att vara nyfikna på vad tekniken kan göra, vilket en icke teknikvan person som exempelvis inte har använt en dator skulle kunna vara.

Enligt boken "The Media Equation" gjordes en studie i ett TV-program med personer som befann sig på olika långa avstånd till kameran. Studien gjordes för att undersöka vem tittarna främst kunde relatera till. Personen som stod närmast kamerorna upplevdes vara mest trovärdig, inte på grund av vad personen sade utan för att personen kändes mer verklig än de som stod längre bort (Reeves, 2002). Det här skulle i så fall innebära att testresultaten i en virtuell miljö skulle bli mer lika resultaten från en fysisk studie än om enbart en icke-tredimensionell skärm skulle användas. Detta skulle då gynna den här studien, eftersom den genomfördes i en virtuell miljö, där avstånd påverkas mellan exempelvis robot och testperson.

Testfaktorer

Den sociala regeln säger att människor inte vill göra någon

ledsen även om det innebär att denne får tänja på sanningen lite. I det här projektet måste regeln tas i åtanke i samband med analysen eftersom den här regeln kan påverka resultaten av testerna. När testpersonerna, som inte har någon koppling till roboten förrän de startar testen, får frågor kring vad de tycker om roboten så kan resultaten påverkas eftersom det är samma personer som skapat roboten som ställer frågan. (Reeves, 2002)

I en verklig miljö finns det många faktorer som påverkar hur en person uppfattar en situation. Saker som avstånd, storlek och färger är alla viktiga aspekter som är avgörande för hur en person uppfattar och bildar åsikter om sin omgivning. Exempelvis kan det upplevas obehagligt om en person står för nära under ett samtal, samtidigt som det kan kännas avlägset om personen står för långt ifrån (Rogers et al., 2011). En del av den här studien kommer att undersöka huruvida avståndet till roboten upplevs på liknande sätt i en virtuell miljö.

I boken "The Media Equation" gjordes en studie på hur visuella nedskärningar i media orsakar visuellt orienterade svar hos testpersonerna (Reeves, 2002). Testet gick ut på att jämföra hur testpersonerna reagerade på att till exempel ha låg upplösning samt ändrade färger på TV-skärmar. På detta sett kändes bilden inte lika verklighetstrogen längre. Resultatet var att folk började placera blicken på olika ställen och lägga fokus på andra saker än de som hade bättre upplösning och färg. Resultaten från den studien är något som bör tas till hänsyn till i pilotstudien eftersom att det kommer att vara dålig upplösning och roboten inte kommer röra sig som en människa.

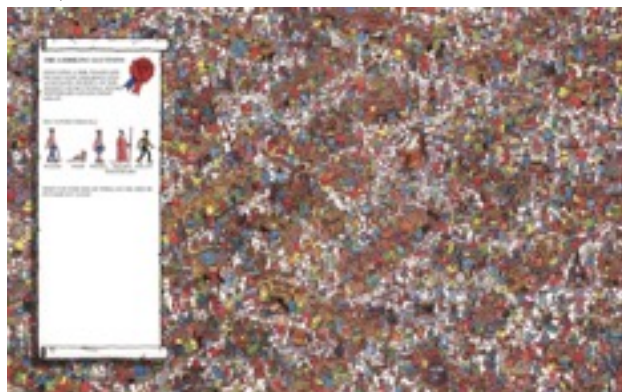
De ursprungliga artiklarna

De tre artiklarna som replikeras i det här projektet studerar på olika sätt hur människor reagerar på robotar när de interagerar med dem på olika sätt. Nedan förklaras de tre artiklarnas syfte, tillvägagångssätt samt vilka människa-robotinteraktionsparametrar som undersöktes.

Artikel 1: Don't Bother Me: Users' Reactions to Different Robot Disturbing Behaviours

Studiens syfte är att undersöka människans tillit och reaktioner till en robot. Testet utfördes med serviceroboten HOBbit, vilket är en robot som har framtagits i ett EU-forskningsprojekt. Syftet med roboten är att kunna hjälpa äldre människor i hemmet (Automation and Control Institute Vienna University of Technology, n.d.). Testet innebar att testpersonen skulle lokalisera huvudfiguren Waldo, som skulle finnas på fem olika ställen på en tavla (se figur 1). Testpersonerna fick 10 minuter på sig att utföra uppgiften. Under testets gång närmade sig roboten testpersonen för att hjälpa till. Testpersonen fick varje gång svara "ja" eller "nej" på om personen ville ha hjälp av roboten. Roboten kunde aldrig hjälpa testpersonen med uppgiften eftersom två av de fem figurerna inte fanns med på bilden. Mellan frågorna från roboten gjorde roboten olika störande moment för att irritera testpersonen samt påverka testpersonens koncentration. Målet var att mäta testpersonens reaktioner mot roboten och kartlägga hur testpersonen svarade på frågorna samt hur

gester, rörelser och tonfall användes. (Weiss & Vincze, 2013).



Figur 1. Bilden visar Where's Waldo-bilden med 3 av 5 Waldofigurer borttagna.

Artikel 2: Using Socially Assistive Human–Robot Interaction to Motivate Physical Exercise for Older Adults

Syftet med detta test var att studera människors motivation att utföra instruktioner beroende på hur roboten tilltalar testpersonen och samtidigt indirekt kartlägga hur människor vill interagera med robotar. Roboten gav testpersonen instruktioner i form av ett fysiskt träningspass, där testpersonen skulle utöva instruktioner tillsammans med roboten. I den första delen av studien var roboten neutral i sina repliker och gav inga inspirerande/uppmuntrande ord. Efter det första träningspasset fick testpersonen göra ett till träningspass med samma robot, med skillnaden att roboten nu var mer inspirerande och gav komplimanger till testpersonen. Efter att testet var klart fick testpersonen svara på frågor kring roboten uppfattades under träningspassen och om personen märkte någon skillnad på robotens beteende i de två olika momenten. (Fasola, Mataric, 2002)

Artikel 3: The Revised Set of Measures to Assess Users' Attitudinal and Behavioral Responses to a Social Robot

Det här testet var en studie kring människors personliga sfär. Det gjordes tester på hur testpersonen reagerade när roboten kom emot personen och vilka olika avstånd som gav upphov till en reaktion hos testpersonen. Syftet med den här studien var att testa om den personliga sfären upplevs annorlunda beroende på om det är människa-människa eller människa-robot som interagerar. (Joosse, Sardar, Lohse, Evers, 2013)

Målsättning

Målet med projektet är att med hjälp av en virtuell robot, i en virtuell miljö, göra studier på om det går att replikera användartester från en fysisk miljö och få likvärdiga resultat. De användartester som utförs på testpersonerna är utformade på samma sätt som de ursprungliga användartesterna. Resultaten från användartesterna ska sedan jämföras med ursprungsartiklarnas resultat för att kunna dra slutsatser kring hur resultaten liknar samt skiljer sig åt. Förhoppningen är att resultaten visar sig vara så pass lika att det i framtiden ska vara möjligt att göra pilotstudier i en virtuell miljö först, innan det

byggs fysiska robotar. Användartesterna kan vara allt från ett underlag till hur exempelvis en robots gester, tal, kroppsspråk, utseende och rörelser bör implementeras. Om det underlaget finns i ett så tidigt skede som möjligt i utvecklingsprocessen kan det möjliggöra reducerade resurskostnader, detta eftersom att misstag med den fysiska roboten kan undvikas då problem förhoppningsvis upptäckts i den virtuella miljön först. Med andra ord skulle den virtuella implementeringen kunna liknas en Mid-Fi-prototyp som ska hjälpa till att ge bättre resultat vid skapandet av Hi-Fi-prototyper.

Metod

Projekt del 1

Under den första fasen av projektet genomfördes en pilotstudie från endast en artikel, vars syfte var att skapa en bättre grund inför den andra fasen av projektet. Bland annat testades tekniken för att få en känsla för vad som är möjligt att göra i en VR-miljö, men också testa hur testpersonerna interagerade med tekniken. Resultaten från testet skulle inte användas direkt vidare som material för att bedöma huruvida en virtuell miljö är användbar för forskning och test inom interaktion. Däremot användes det för att bedöma vilka ändringar som behövde göras i den virtuella miljön för att få miljön så pass verklighetstrogen att tester blir likvärdigt genomförbara i den andra fasen.

Slutprodukten från första fasen av kursen har sedan använts som en riktlinje för hur det slutliga testet ska utföras. I första delen av kursen startades projektet upp i utvecklingsmiljön Unity där en virtuell lägenhet i 3D importerades, konfigurerades sedan användes för att utföra testerna på testpersonerna tillsammans med en virtuell robot. Roboten skapades i en annan miljö, 3Ds Max, som sedan importerades till Unity. Lägenheten användes sedan till den andra delen av projektet men har fortsatt att utvecklas och förbättras.

Projekt del 2

Det första steget i den andra fasen av det här projektet var att leta upp fler artiklar som sedan skulle replikeras till testerna i den virtuella miljön. För att få en mer rättvis bild av testens resultat utökades testningen till tre olika test, med varsitt testområde, för tre olika studier. Det diskuterades om vad från de fysiska testerna som var realistiskt att hinna med och vad som faktiskt skulle fungera att göra i den virtuella miljön.

Förbättringar av lägenhetsmodellen

Efter att målgruppen ändrats gjordes även ändringar av tidigare designbeslut så som färgval samt möbler. En del möbler och utrymmen i lägenheten saknade fysiska egenskaper, vilket gjorde att roboten delvis kunde åka igenom objekten. Hela lägenheten undersöktes, varpå egenskaper lades till på de objekt som saknade det. Utöver det lades möbler till, tapeterna och golvet ändrades för att förbättra den realistiska känslan i miljön.



Figur 2. Bilden visar lägenheten i utvecklingsprogrammet Unity.

Robotförbättringar

Nästa steg var att skapa en helt ny, förbättrad robot med fler funktioner och ett mer tilltalande utseende. I den första fasen av projektet var roboten en cylinder med en bild på sig för ge den ett ansikte som testpersonerna skulle kunna relatera till. Cylinderroboten kunde röra sig med hjälp av en tredje part som styrde roboten via tangentbordet. För den andra fasen av projektet undersöktes det vad för typ av robotar som fanns tillgängliga via Unity Asset Store, vilket är en online marknadsplats för utvecklingsmaterial till Unity. Målet var att göra den nya roboten mer antropomorfisk, det vill säga mer människolik. Det valdes att använda ett färdigt robotskal, som sedan kunde anpassas efter det här projektets mål, istället för att designa en robot från grunden då varken kunskapen eller tiden var tillräcklig för det. Roboten som valdes heter "Robot Kyle" (se figur 3 nedan) och är en humanoid, med färdiga leder och utseende.



Figur 3. Roboten Kyle i den virtuella miljön.

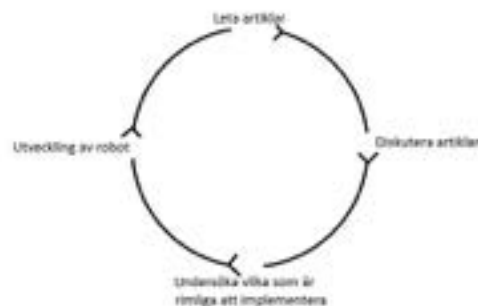
Att roboten har leder innebär att rörelser kan appliceras på roboten i Unity, vilket var viktigt då de nya testerna krävde att roboten kunde göra fler rörelser än förut. Till roboten implementerades gångrörelser, stillastående animationer samt två interaktiva animationer. De interaktiva animationerna bestod av en vinkning samt en hukande animation.



Figur 4. Robot Kyles två animationer (vinkanimationen t.v., den hukande animationen t.h.).

Arbetsmetodik

Parallellt med utvecklingen av miljön skrevs utförandet av testerna.



1. Leta artiklar
2. Diskutera vad som var realistiskt att testa och hinna med
3. Diskutera skillnader från första fasen av projektet, målgrupp förutsättningar och tidsestimat
4. Utveckla robot
5. Förbättra lägenheten
6. Skriva tester och frågor utifrån artiklar (se bilaga 2-4)
7. Skriva rapport
- 8.Handledarmöten fortsatt veckovis

Motivering kring valen i VR-miljön

Miljön som testerna genomfördes i är en lägenhet som är avskalad och ljus men samtidigt trevlig för att användarna ska känna sig bekväma men samtidigt inte störas av visuella detaljer (G. Franz, M. Heyde, H.H. Bülhoff, 2005). Lägenheten har ett begränsat antal möbler och en vägg där "hitta Waldo"-bilden som användes i ett av testen visas (se figur 5).



Figur 5. Bilden visar Where's Waldo-bilden på väggen i testmiljön.

Under den första fasen av projektet tillbringades relativt mycket tid åt inredning och möblering av lägenheten med avsikten att testpersonerna skulle bli påminda om ett fysiskt hem. Det visade sig ganska snabbt att det inte var någon bra idé då upplösningen var dålig och miljön medförde illamående på grund av den "pixliga" och färgglada lägenheten. Beslutet togs därmed att inte lägga till möbler, förutom några enstaka, samt att inte heller lägga till några färger i miljön även om de vanligtvis är en naturlig del av en lägenhet. Det upptäcktes under den första fasen av projektet att en del av testpersonerna i början av testerna var fokuserade på att förstå Oculus Rift-tekniken snarare än testinstruktionerna, beroende på om de använt Oculus Rift förut eller inte. För att minsta risken för att testpersonerna skulle tappa fokus i början av testet togs fler potentiella störningsmoment såsom inredningsdetaljer bort.

Den ursprungliga idén var att visa tavlan av Waldo i studien "Don't bother me" i en naturlig storlek som testpersonen kunde ha i händerna, precis som det ursprungliga testet utfördes. Tavlan kunde dock inte användas då Oculus Rift DK1 inte kunde visa den med tydlig upplösning (1280 x 800). Därav valdes en betydligt större bild som placerades på en vägg istället likt en tapet. Bilden som valdes bidrog till flera problem under pilotstudien, bland annat var det svårt att urskilja Waldo trots att bilden var stor. Med Oculus Rift DK2 blev bilden tyvärr inte bättre, så istället valdes en mer avskalad bild med tydliga kontraster. Uppgiften var fortfarande att hitta Waldo, som nu var gömd på olika ställen på bilden. En annan förbättring var att användaren nu kunde huka och hoppa vilket gjorde att användaren lättare kunde se hela bilden.

Testpersonens robot har inte en kropp i den virtuella miljön, det vill säga om personen tittar ner i ett försök att lokalisera sina fötter så finns de inte där. Anledningen till valet av att inte inkludera personens kropp i miljön var för att det kan leda till att tekniken och miljön fick för stort fokus under testet. Fallstudier har visat att en robotkropp gör det lättare för en användare att uppleva närvaron och framförallt ger det personen lättare att uppskatta längder i miljön. Det ska även tilläggas att möjligheten att röra kroppsdelar utöver huvudet, så som armar och ben ökar den här närvaro ännu mer (Mohler, B.J, et al, 2010). Beslutet kan bidra till att personen upplever miljön som mindre verklighetstrogen vilket kan ha påverkat testresultatet. Efter att ha undersökt möjligheterna att implementera kroppsdelar i den virtuella miljön, framkom det att det skulle vara alldeles för svårt och tidskrävande och det beslutades således att slopa implementeringen av kroppsdelar trots den eventuella positiva inverkan det skulle ha på närvaron.

I första fasen av projektet var formen på roboten en cylinder med en bild av en människa på, då tiden var för knapp för att arbeta fram en liknande robot som användes i det ursprungliga testet. På grund av robotens utseende kunde roboten inte heller röra sig likt en människa, då den istället svävade fram i rummet. Unedr den andra fasen, när inlärningsprocessen var avklarad, blev slutresultatet av roboten mer likt en robot, med en kropp och ansikte mer likt en människa. Robotens funktioner var bland annat att gå (fram/bak/diagonalt) likt en människa, hoppa, huka sig, vinka och när

ingen styrde gjorde roboten vaga rörelser som att vingla lite eller kolla på klockan som människor gör. Anledningen till att roboten inte skulle stå helt stilla utan göra små människoliknande rörelser var för att roboten skulle påminna mer om en fysisk person istället för ett föremål.

Test

Efter att målgruppen blivit ändrad från en äldre till en yngre grupp började det att bjudas in testpersoner. För att inte få en allt för homogen testgrupp valdes det att kombinera testpersoner med olika tekniska bakgrunder. Totalt utfördes testen på 10 personer, blandad könsfördelning i åldern mellan 22-30år. Främsta delen av testgruppen blev klasskamrater till gruppmedlemmarna då de var relativt lättillgängliga. Men de testades även på personer som inte studerade vid LTH. Valet av de här personerna bygger på att även de är vänner till projektgruppen och kunde tänka sig att ställa upp på testerna.

När testpersonerna kom in till lokalen där testerna utfördes, VR-labet på IKDC, välkomnades de först och sedan fick samtliga testpersoner en lapp med en kort beskrivning av testerna samt Oculus Rift. På lappen fanns även information om att det fanns ett önskemål om att få filma testet för att senare kunna analysera (se bilaga 1).

Efter det att testpersonerna läst igenom första lappen fick de en instruktion på slumpmässigt valt test. Vid två av testen ställdes frågor direkt efter testet var klart, under ett av testen undersöktes reaktioner medan testet genomfördes. När ett test och tillhörande frågor var klara fick testpersonen en ny lapp med instruktioner, detta upprepades till alla tre test var genomförda.

Under testen styrde en person roboten samt de ljudfiler som skulle spelas upp, en person antecknade samt ställde frågor och två personer turades om att filma de personer som gick med på att bli filmade. Anledningen till att en person styrde både roboten och ljudet var för att det då var lättare att synkronisera robotens rörelser med ljudklippen vid rätt tillfälle, vilket annars skulle kräva mycket träning för två personer att synkronisera tillsammans. Testpersonerna hade hörlurar på sig i vilka robotljuden spelades upp, detta för att ge testpersonerna en större känsla av att vara i en virtuell miljö än som tidigare i projektet då ljudet kom från rummet där testet utfördes.

Resultat

Don't bother me

Alla tio testpersoner svarade första gången roboten kom fram till dem att de ville ha hjälp med att hitta Waldo på bilden. När roboten kom tillbaka andra gången för att fråga om testpersonen ville ha hjälp med att hitta Waldo svarade nio personer ja igen och endast en svarade nej (se bilaga 5).

Nedan visas testpersonernas svar:

Motivera hur dina tankar gick när du svarade på robotens frågor beroende på hur du svarade:
Om du svarade ja första och andra gången.

- Jag visste inte vad jag annars skulle göra för jag klarande inte uppgiften själv så jag tänkte att jag lika gärna kan prova igen.
- Jag hittade inte Waldo så roboten kanske kunde hjälpa mig.
- För att kändes dumt att säga nej
- Först trodde jag den kan hjälpa mig, kanske kunde roboten kunde se mer av bilden. Testade om den kunde hjälpa mig en andra gång.
- Vet inte, det fanns inte så mycket annat att göra.
- Jag visste inte vad jag annars skulle göra för jag kunde inte klara testet själv.
- Jag tänkte att han kanske var programmerad att inte göra exakt samma sak en gång till.
- Vet inte vad jag istället skulle göra?
- Jag klarade uppenbarligen inte att klara detta på egen hand, så det kändes som det var min ända räddning.

Om du svarade ja första och nej andra gången.

- Först hittade jag inte Waldo så då kanske roboten kunde hjälp. Men andra gången visste jag att han inte skulle kunna.

Using Socially Assistive Human-Robot Interaction to Motivate Physical Exercise for Older Adult

I delen då roboten var neutral och inte gav uppmuntrande ord svarade tre testpersoner att roboten var fin, tre tyckte roboten var enförmig/tråkig och en tyckte att roboten var obehaglig när de blev tillfrågade om vad de tyckte om roboten. Anledningen att roboten upplevdes som obehaglig var på grund av hans röda ögon.

Nedan visas testpersonernas svar:

Vad tyckte du om roboten?

- Jättefin och rörde sig bra.
- Gör massa rörelser när den inte gör något alls. Han kollar exempelvis på klockan. Vet ej om det är bra eller dåligt.
- Enförmig.
- Ingen skillnad, det var säkert något men jag märkte inget.
- Lite svårt att vet om han vill att jag ska svara eller inte.
- Jättefin och rörde sig bra.
- Fin, för bra rörelser som är lätta att följa. Rolig röst.
- Bra men ganska ensidig.
- Den var fin men ganska tråkig.
- Fin att titta på men såg ganska obehaglig ut med sina röda ögon
- På frågan om hur de hade upplevt testet svarade merparten att testet var lätt och kul.

Hur upplevde du testet?

- Svårt, förvirrande.
- Fruktansvärt förvirrande.
- Förvirrande.
- Intressant att göra men irriterande när man inte hittade Waldo.
- Svårt, ganska svårt att styra och leta och samtidigt veta vad man ska leta efter.
- "Ganska svårt och förvirrande.
- Svårt test, hittade inte något Waldo.
- Hittade aldrig Waldo, fanns han ens med på tavlan?
- Roligt att testa Oculus men uppgiften var svår.
- Egentligen roligt men jag mådde lite illa efter ett tag.
- När roboten var positiv mot testpersonen svarade fem personer att roboten var snäll, tre tyckte att det blev roligare, en blev upplevde roboten som irriterande, två tyckte att han var fin och en märkte ingen skillnad.

Vad tyckte du om roboten?

- Mycket bättre än första gången för han var mycket snällare i det här testet. Det blev roligare att göra rörelserna.
- Bra robot, snäll. Irriterande att han tar på sig själv hela tiden.
- Bättre, han gav mig komplimanger
- Vänlig, välkomnade mig, tydliga instruktioner.
- Ingen skillnad.
- Jättefin robot.
- "Fin robot, roligt att följa honom.
- "Mycket roligare när han gav mig komplimanger.
- "Han verkade snäll, man ville se mer av vad han kunde göra.
- Snäll, rolig.
- På frågan hur upplevde du testet var merparten mycket mer positiva och tyckte att denna del var betydligt roligare.

Hur upplevde du denna del av testet?

- Mycket roligare att göra testet denna gången.
- Mycket bättre och han var snäll. Mer feedback . Bra, om jag skulle stått upp så hade jag kunnat få motion men nu satt jag ner.
- Kul. Skillnad från första.
- Lätt, enkla instruktioner.
- Ingen skillnad från första delen.
- Roligt test, man blev motiverad att följa.
- Snällare och verkade trevligare.
- Bra bra, verkade vara en snäll robot.
- Lätt och roligt att följa.
- Bra, roligt att göra rörelserna.

The Revised Set of Measures to Assess Users' Attitudinal and Behavioral Response to a Social Robot

Under detta test iaktogs det hur testpersonerna reagerade när roboten var väldigt nära testpersonen. När roboten närmade sig flyttade sig sex stycken av testpersonerna, fyra av dem till vänster och två backade bakåt (se bilaga 6).

Under testets gång undersöktes hur väl nedanstående påstående stämde överens med testpersonens reaktioner och rankade de här iakttagelser på skala mellan ett och fem.

Nedan visas testpersonernas svar:

	1	2	3	4	5
The participant looked disgusted	10				
The participant looked angry	10				
The participant looked happy	6	4			
The participant looked surprised			4	5	1
The participant looked sad	10				
The participant looked scared	3	5	2		

Diskussion

Människa-robotinteraktion i en VR-miljö genom Oculus Rift är ett relativt utforskat område, och något vi inte har stött på förut. Den naturliga följden blir således att projektutvecklin-

gen drivs mycket av trial & error för att utforska vad som är möjligt, vad som fungerar och hur det fungerar bäst. Projektet har medfört mycket erfarenheter som vi redogör för nedan. Vår förhoppning är att framtida studier inom det här området kan dra nytta av våra erfarenheter, och undvika de fällor vi har hamnat i.

I fortsättningen hade det nog underlättat för framtida projektgrupper att ha åtkomst till vårt och grupp 13s projekt då vi upplevde en hög inlärningströskel med Unity, 3Ds max samt Oculus Rift. Med ett färdigt "enkelt" projekt att titta på kan det förhoppningsvis bli lättare att förstå hur de olika komponenterna i Unity kan användas. Trots att det finns många bra guider på internet kan det bli för mycket tid med implementering i teoriorienterade projekt.

Grupprocessen

Vår grupprocess ändrades markant under den andra projekthalvan. Efter första fasen analyserade vi tillsammans med våra handledare problem i vårt arbetssätt. Vi skrev därför ner ett måldokument inför andra fasen med en konkret plan för att förbättra vår grupprocess, vilket vi hoppades skulle drastiskt förbättra vårt arbete i grupp som helhet.

Processförbättringen började med att vi övergav arbetssättet med parallellt arbete inom olika projektområden, vilket i första fasen ledde till att gruppmedlemmarna inte hade en inblick på vad de andra gjorde. Istället införde vi fasta tider då vi satt tillsammans i grupp och arbetade samtidigt inom ett område, men med parallella processer inom det området. Vi satt exempelvis med utveckling i Unity en dag kunde två personer jobba med lägenheten och roboten samtidigt som de två andra arbetade med testerna med varandras arbete i åtanke. Vilket ledde till att testerna bättre följde utvecklingsmiljöns kapacitet, samtidigt som utvecklingsmiljön anpassades för testerna. Utöver det arbetade vi med personliga loggböcker som vi sedan diskuterade i grupp, så att alla visste vad de andra i gruppen tyckte, vad de hade gjort i veckan samt vad de skulle göra nästa vecka. Det underlättade dessutom för gruppleddaren vid inskickning av grupploggboken inför våra handledarmöten varje vecka. Vi införde även en social aktivitet varje vecka för att dels kunna prata om icke-projektrelaterade saker så att vi skulle vara mer fokuserade när vi väl arbetade med projektet, men även för att lära känna varandra bättre.

Det utökade arbetet med grupprocessen upplever vi har förbättrat vårt arbete avsevärt, prestationsmässigt såväl som det har haft positiva effekter på hur gruppdynamiken har upplevts. Fasta arbetsdagar tvingar alla gruppmedlemmar att fokusera på projektet de dagarna, vilket kan vara ett problem annars om alla läser olika krävande parallella kurser. Vi har i de här två projektkurserna upptäckt hur viktig en bra grupprocess är även om gruppmedlemmarna känner varandra sedan innan, vilket vi gjorde, om inte än viktigare för grupper som faktiskt känner varandra än för obekanta. Det kan bero på att alla anstränger sig lite extra för att lära känna varandra och samarbeta om personerna i gruppen inte känner varandra. Vi är överlag väldigt nöjda med hur utvecklingen har gått från projektets start i första fasen till slutet av projektet i andra fasen och vi hoppas vi kan dra nytta av det vi har lärt oss om grupparbeten i framtida projekt.

Val av artiklar

Artiklarna som valdes var tvungna att uppfylla två kriterier, tidsestimeringen på projektet och möjligheterna i VR-miljön. Ett krav på artiklarna var att det skulle finnas en spridning på vad testen mätte fast att testens utformning inte skiljde sig allt för mycket för att fortfarande uppfylla tidsestimeringen. Anledningen till att det skulle finnas så bred spridning på testen som möjligt var för att få en så rättvis jämförelse med ursprungstesten som möjligt. Om det hade varit liknande studier med liknande resultat så hade det varit svårare att svara på syftet med det här projekt; om det går att ersätta studier med VR-miljöer.

Det valdes tre skilda studier som skulle replikeras i en VR miljö. Av artiklarna som valdes var den första en studie som testade tilliten till en robot, i den andra studien görs en mätning kring den personliga sfären och i den tredje studeras det hur robotar ska agera och tala för att det ska vara bekvämt för människor att interagera med dem.

Val av miljö och robot

Den första roboten som användes, i första fasen, var en vit cylinder, med en påklistrad bild av en man. Cylindern svävade ovanför marken, rörde sig onaturligt snabbt och liknade inte en robot. När vi förbättrade miljön och roboten under andra fasen, började vi därför om helt med roboten. Då vi ville ha en så verklighetstrogen robot som möjligt, och inte kunde skapa det själva, började vi titta i Unity Asset Store efter alternativ. Asset Store har många bra modeller, men de kostar oftast pengar, vilket vi inte hade några. Det fanns dock en gratis robotmodell, Roboten Kyle. Modellen stämde in på våra önskemål då den var relativt antropomorfisk då skelettet liknade en människas samt att den kunde gå och röra på armar, huvud etc. Robotmodellen hade dock röda ögon, vilket oroade oss lite då det oftast förknippas med ondska i filmer och spel. Vi fann dock att de positiva aspekterna övervägde den negativa, dessutom upplevde vi att den låga upplösningen i VR-miljön gjorde att de röda ögonen inte var lika markanta som vi först befارade.

De rörelser och animationer som valdes hämtades från ett gratis motion capture-bibliotek från Asset Store. Animationerna skapades sedan utifrån de animationsklipp som fanns i biblioteket. Det ledde till att roboten inte rörde sig riktigt som vi ville när den gick framåt, men då alternativet var att göra egna motion capture-animationer använde vi de som fanns, och anpassade de efter våra ändamål så gott det gick.

Den lägenhet vi importerade passade vårt ändamål, då det enda vi egentligen behövde var ett rum med en vägg. Istället fokuserade vi på att konfigurera den färdiga lägenheten så att golv, väggar och möbler skulle vara så subtila som möjligt för att de inte skulle påverka testet pga. att testpersoner distraheras. I det riktiga testet med Waldo håller testpersonen i bilden och visar den sedan för roboten, men i det här projektet sattes Waldo-bilden upp på en vägg. Det beslutet togs dels för att det var svårt att implementera en flyttbar bild men även för att reducera antalet funktioner som testpersonerna behövde lära sig för att klara av testet.

Testens utformning

I testet som involverade att hitta Waldo, byttes bilden ut mot

en annan bild med bättre upplösning. Även den nya bilden var en bild med fyra stycket Waldo-figurer (fast testpersonen trodde det var fem). Att bilden byttes ut mot en annan spelar ingen roll för testets syftes skull. Däremot hade det haft en stor påverkan på resultatet om bilden inte bytts ut, eftersom testpersonen då ansett att uppgiften var omöjlig och antagligen bara gett upp.

Resultat från testen

Syftet med detta projekt var att jämföra resultaten som utgavs från projektets test med resultaten från de ursprungliga testen. Testresultaten var väldigt lika i samtliga deltest. I artikeln "Don't bother me" kom författaren fram till att ca 80% av testpersonerna svarar att de vill ha hjälp båda gångerna av roboten fast att de stör sig på den. Dessutom svarade de flesta personerna i testet att fast att de vet om att roboten inte kunde hjälpa dem så hade de svarat samma sak om de gjort om testet. I det replikerade testet var det sju av tio personer som svarade att de skulle tagit emot hjälp från roboten i efterhand fast att samma personer var medvetna om att roboten skulle svara samma sak som i det första testet, vilket är väldigt nära resultaten av det ursprungliga testet. I båda testfallen svarade testpersonerna att anledningen till att de valde att ta emot hjälp från roboten en andra gång, var oftast eftersom att testpersonen inte visste hur de annars skulle lösa uppgiften. Vissa trodde även att roboten kanske var programmerad med flera funktioner och att omständigheterna kanske hade ändrats undan att de lade märke till det.

Det är tydligt att testpersonerna blev irriterade på roboten under testets gång men ändå svarade positivt i frågor efter testerna angående roboten. I ursprungsstudien var det forskarna av projektet som ställde frågorna om vad testpersonerna tyckte om roboten. Det uppstår då en situation där forskarna får ett indirekt ägarskap till roboten, även om i själva verket inte är de som äger roboten och inte heller har skapat den. Därför borde "den sociala regeln" appliceras i båda testen. I detta projekt kan behövs inte regeln tas hänsyn till eftersom projekts syfte är att jämföra resultaten av testen där "den sociala regeln" i båda fall uppstår.

Det var flera test som utfördes i studien "*Using Socially Assistive Human-Robot Interaction to Motivate Physical Exercise for Older Adults*" och i detta projekt valdes två av dessa test att utföras för att sedan jämföras. I första delen av den ursprungliga studiens test möttes testpersonerna av en robot som var väldigt positiv, uppmuntrande och tillgiven mot testpersonerna. Ett exempel på detta var att den kallade testpersonerna vid namn och gav uppmuntrade ord till dem när de utförde övningarna. Efteråt, när testpersonerna fick frågan vad de tyckte om robotens beteende var det en väldigt positiv respons. Testpersonerna uppskattade de uppmuntrade orden och upplevde testet som något roligt och positivt. Samma resultat utgavs från detta projekts test med liknande uppmuntrande robot. Bland annat svarade testpersonerna att roboten var "snäll och verkade trevlig" och "roligt test, man blev motiverad att följa". I ursprungstestet gav testpersonerna ganska negativa svar angående vad de tyckte om robotens instruktioner när den "neutrala" roboten skulle visa och berätta om de olika övningarna. Det resultat gavs också i det replikerade testet. Dock var det inte lika stora skillnader i

svaren som i det ursprungliga testet. Detta kan bero på att testpersonerna inte kunde utöva rörelserna fysiskt, utan endast styrde sig i samma rörelse med hjälp av tangenterna i den virtuella miljön. Om Oculus rift DK2 hade fungerat, hade testet kunnat utförande stående av testpersonerna istället för att utföra rörelserna sittandes med tangentbordet. Det hade då troligtvis gjort en större skillnad på de två olika testerna (med den positiva och neutrala roboten) och det hade sannolikt gett mer likvärdiga resultat till det ursprungliga testet.

I studien "*The Revised Set of Measures to Assess Users' Attitudinal and Behavioral Responses to a Social Robot*" var resultaten ganska negativa emot roboten. Testpersonerna var ganska överens om att robotens närhet kändes inkräktande och nästan lite hotfullt. Detta resultat var likvärdigt med om en verklig person hade varit lika närgången testpersonerna. I det replikerade testet var reaktionerna inte lika starka. Många testpersoner ansåg att robotens närvaro kunde vara lite obehaglig men några av dem valde ändå inte att flytta på sig. Detta kan bero på miljöns upplösning som inte var fullt lika bra som tänkt. I boken "*The media equation*" gavs ett exempel på hur testpersoner reagerade olika på beroende hur bra kunde relatera en tv-skärm till deras verklighet. De som satt längre ifrån upplevde bildskärmen som mindre levande än de som satt närmre. Med andra ord, hade den virtuella miljön haft en bättre upplösning och verkat mer trovärdig hade antagligen testpersonerna reagerat starkare på robotens närvaro, precis som de gjorde i ursprungstestet. När personens ansiktsuttryck studerades var det många saker som var svåra att läsa av jämfört med om hela ansiktet på testpersonen hade varit synligt. Många uttryck syns i ögonen, vilka var täckta av Oculus rift-utrustningen. Däremot kunde fortfarande minerna som testpersonerna gjorde ses, vilka var liknande de i ursprungstestet: något obehag, inte så mycket glada miner och många som drar sitt huvud ifrån skärmen även om den rörelsen inte gjorde någon skillnad i den virtuella miljön. I den virtuella miljön flyttar endast tangentbordet testpersonens karaktär och därför hände inget om testpersonerna till exempel ryggade tillbaka. Om upplösningen hade varit bättre hade antagligen samma testpersoner valt att flytta sin robot i samma riktning som de gjorde fysiskt eftersom de då haft mer verklighetsförankring i den virtuella miljön.

Slutsats

Resultaten i den här studien var väldigt lika de resultat som ursprungsartiklarna hade. De resultat som skiljde sig från ursprungsstudien berodde antagligen på utrustningens bristande kapacitet och inte på utformningen av testen. Det finns moment som kan vara svåra att replikera såsom exempelvis fysisk beröring. Samtidigt har det här projekt visat att det finns väldigt många delar inom HRI som kan utföras i en artificiell miljö. Utifrån de här resultaten har vi därför dragit slutsatsen att många test är möjliga att utföra i en virtuell miljö.

Referenser

- Automation and Control Institute Vienna University of Technology (n.d.). HOBBIT - The Mutual Care Robot. URL: <http://hobbit.acin.tuwien.ac.at/index.html> (Hämtad 2014-12-10)
- Weiss, Astrid, Vincze, Markus, Panek, Paul & Mayer, Peter: "Don't Bother Me: Users' Reactions to Different Robot Disturbing Behaviours", International Conference in Social Robots 2013 Oct 27-29 2013, Bristol, UK, http://hobbit.acin.tuwien.ac.at/publications/Weiss_2014HRI_1.pdf Hämtad 141206
- Fasola, Juan, Mataric', Maja J : "Using Socially Assistive Human-Robot Interaction to Motivate Physical Exercise for Older Adults", <http://robotics.usc.edu/publications/media/uploads/pubs/763.pdf> 2002, Hämtad 141206
- Joosse, Michiel, Sardar, Aziez, Lohse, Manja, Evers, Vanessa: "The Revised Set of Measures to Assess Users' Attitudinal and Behavioral Responses to a Social Robot", Electrical Engineering, Mathematics and Computer Science (EEMCS), 2013, <http://doc.utwente.nl/86813/> Hämtad 141206
- Reeves, Byron & Nass, Clifford. The media equation. California, 2002.
- Rogers, Yvonne, Sharp & Helen & Preece, Jennifer. Interaction design 3rd edition. United Kingdom 2011.
- Thern, Lars. Hadzovic, Tarik. (2014) Ubiquitous computing in a home environment. Lund University.
- Mohler, B.J, Creem-Regehr, S.H, Thompson, W.B, Bülthoff, H.H. (2010)
- The Effect of Viewing a Self-Avatar on Distance Judgments in an HMD-Based Virtual Environment Teleoperators and Virtual Environments, June 2010
- Wikipedia, http://en.wikipedia.org/wiki/Motion_sickness Hämtad 2014-12-07

Bilagor

Bilaga 1.

Introduktion innan testen på börjades

Introduktion:

I de här tre testen ska du utforska en virtuell miljö med hjälp av Oculus rift *. Med hjälp av utrustningen kommer du att få en känsla av att befinna dig i en annan miljö. I den här miljön kommer det även att finnas en robot som kommer fungera som ett hjälpmedel för dig.

Du kan när som helst avsluta testet utan att motivera ditt beslut. Vi kommer att spela in testet i syfte av att kunna analysera resultaten senare. Filmerna kommer inte att användas utanför detta projekt. Genom att signera här godkänner du att vi spelar in testet:

Namn:

*Oculus rift är teknologiska glasögon som visar en virtuell tredimensionell miljö. Det betyder att inget från den verkliga miljön visas.

Bilaga 2.

Test samt frågeställningar

Test Don't bother me:

Instruktion: Du kommer i denna miljö befinna dig i en virtuell miljö med en robot. Din uppgift i detta test är att hitta en figur, Waldo, på en tavla som finns i den virtuella miljön. Waldo finns på 5 olika ställen på tavlan, och din uppgift är att hitta alla dessa. Du får bara svara ja eller nej på frågor som roboten ställer och får inte heller kommunicera med testledaren under testet. Vänligen ta på dig Oculus rift utrustningen och starta testet när du är redo.

Tillvägagångssätt:

1. Testpersonen läser introduktionen till testet.
2. Testpersonen tar på sig utrustningen (Oculus rift) och navigerar sig fram till Waldo-väggen.
3. Efter 3 minuter kommer roboten fram till testpersonen och säger "Detta är ett svårt test, 2 minuter av 5 minuter har redan gått".
4. Efter ytterligare 2 minuter kommer roboten fram till användaren igen och säger: "Kan jag hjälpa dig att hitta Waldo". Om testpersonen svarar "nej", svarar roboten "okej". Om testpersonen svarar "ja", svarar roboten "Kan du vara snäll och flytta dig åt sidan så jag ser bättre". Sedan efter 10 sekunder säger roboten "Jag kan tyvärr inte hitta Waldo" och går där ifrån.
5. Efter 1 minut kommer roboten fram igen och steg 4 upprepas.
6. Testet är slut och testpersonen får ta av sig utrustningen och får svara på frågor av testledaren (se bilaga).

Frågeställning till "Don't bother me" EFTER testet "Don't bother me":

Hur upplevde du testet?

Vad tyckte du om roboten?

Hjälpte roboten dig att uppnå ditt mål?

Hur upplevde du testet?

Vad tyckte du om roboten?

Hjälpte roboten dig att uppnå ditt mål?

Motivera hur dina tankar gick när du svarade på robotens frågor beroende på hur du svarade:

Om du svarade ja första och andra gången.

Om du svarade ja första och nej andra gången.

Om du svarade nej första och ja andra gången.

Om du svarade nej första och nej andra gången.

Med facit i handen, hade du svarat annorlunda om du gjort om testet?

Bilaga 3.

Test Using Socially Assistive Human-Robot Interaction:

Instruktioner: Du kommer under detta test befinna dig i en virtuell miljö med en robot. Du ska i detta test utöva ett kort träningspass (med några icke-ansträngande rörelser) med roboten som instruktör. Du kommer att sitta framför datorns skärmen med Oculus rift utrustningen på dig. Testet är uppdelat i två olika delar, vänligen behåll utrustningen på dig under hela testet. När roboten ger dig en instruktion ska du följa rörelsen med hjälp av tangentbordet. Ta på dig Oculus rift utrustningen och invänta robotens instruktioner.

Studie baserad på "Using Socially Assistive Human-Robot Interaction to Motivate Physical Exercise for Older Adults"

Testpersonen får inte ta av sig utrustningen i mellan testen fast att testet är uppdelat i två delar. Roboten kommer att informera testpersonen när testdel 1 startar och testdel 2 startar. Utan inbördes ordning ska roboten vara positiv under ena delen av testet och neutral under andra delen. Den del där roboten är positiv ska den ge beröm samt uppmuntrande ord till testpersonen. Den del där roboten är neutral ska roboten vara mer saklig och inte ge beröm eller uppmuntrande ord.

Tillvägagångssätt:

Testpersonen får inte ta av sig utrustningen under testet fast att testet är uppdelat i två delar. Roboten kommer att informera testpersonen när testdel 1 startar och testdel 2 startar. Utan inbördes ordning ska roboten vara positiv under ena delen av testet och neutral under andra delen. Den del där roboten är positiv ska den ge beröm samt uppmuntrande ord till testpersonen. Den del där roboten är neutral ska roboten vara mer saklig och inte ge beröm eller uppmuntrande ord.

1. Testpersonen läser introduktionen till testet om "träningsspelet".
 2. Testpersonen tar på sig utrustningen (Oculus rift) och väntar in robotens instruktioner.
 3. Roboten står mittemot testpersonen och säger "Hej".
- Testdel 1:
4. Roboten säger: "Nu börjar första delen av testet. Följ mina instruktioner".
 5. Roboten säger "Gå framåt och tillbaka igen". Samtidigt kommer roboten gå fram och tillbaka så få steg som möjligt.
 6. Roboten säger "Gå snett framåt". Samtidigt kommer roboten gå snett framåt och tillbaka.

Testdel 2

8. Roboten säger "Nu börjar andra delen av testet. Följ mina instruktioner."
9. Roboten säger "Gå snett framåt". Samtidigt kommer roboten gå snett framåt och tillbaka.
10. Roboten säger "Bra gjort!"

13. Roboten säger "Gå framåt och tillbaka igen". Samtidigt kommer roboten gå fram och tillbaka så få steg som möjligt.

14. Roboten säger "Du kan ju det här"

14. Roboten säger "Jättebra, nu är testet slut, tack så mycket"

Frågeställning till "Using Socially Assistive Human-Robot Interaction:" EFTER test

Hur upplevde du den ena delen av testet (inte peppande och inte glad)?

Hur upplevde du roboten i detta deltest?

Hur upplevde du den ena delen av testet (peppande och glad)?

Hur upplevde du roboten i detta deltest?

Bilaga 4.

The Revised Set of Measures The Revised Set of Measures to Assess Users:

Instruktioner: I denna miljö kommer du befinna dig i en virtuell miljö med en robot. Du ska i detta test endast följa testledarens instruktioner (personen som sitter bredvid dig under testet). Vänligen ta på dig Oculus rift utrustningen och invänta testledarens instruktioner.

Tillvägagångssätt:

1. Testpersonen läser instruktionen till testet om "Behavioral responses".
2. Testpersonen tar på sig utrustningen (Oculus rift) och väntar in robotens instruktioner.
3. Roboten står en bit ifrån testpersonen och hälsar på testpersonen (Hej, kompis!)
4. Sedan åker roboten fram till testpersonen så att den står sidan om testpersonen, men fortfarande så att testpersonen kan se roboten (mindre än 45cm ifrån) och testledaren frågar om testpersonen tycker att det är obehagligt när den står så nära. (där stannar roboten tills testet är slut utan att röra sig)

Frågeställning till "The Revised Set of Measures to Assess Users' Attitudinal and Behavioral Responses to a Social Robot"

Testpersonen svarar på skala 1-5 hur de upplevde roboten. (1 = inte alls, 5=väldigt obehagligt).

Anteckningar (av testledaren) under testets gång:

- Tar testpersonen något steg åt någon riktning?
- I så fall, hur många steg?
- I så fall, åt vilket håll gick testpersonen åt?
- Åt vilket håll står testpersonen? (Står den kvar eller vänta ansiktet emot roboten)
- Riktade testpersonen någon gång blicken åt något annat håll än roboten?

-Vilken av följande punkter stämmer bäst in på testpersonens ansiktsuttryck (skala 1-5):

	1	2	3	4	5
The participant looked disgusted					
The participant looked angry					
The participant looked happy					

The participant looked surprised					
The participant looked sad					
The participant looked scared					

Resultat av testen:

Bilaga 5.

Frågeställning samt resultat till "Don't bother me":

Hur upplevde du testet?

- Svårt, förvirrande.
- Fruktansvärt förvirrande.
- Förvirrande.
- Intressant att göra men irriterande när man inte hittade Waldo.
- Svårt, ganska svårt att styra och leta och samtidigt veta vad man ska leta efter.
- Ganska svårt och förvirrande.
- Svårt test, hittade inte något Waldo.
- Hittade aldrig Waldo, fanns han ens med på tavlan?
- Roligt att testa Oculus men uppgiften var svår.
- Egentligen roligt men jag mådde lite illa efter ett tag.

Vad tyckte du om roboten?

- Kunde inte hjälpa mig, vet inte riktigt varför han var där?
- Inkompetent.
- Värdelös.
- Okej men hjälpte mig inte så mycket. Kunde inte klara uppgiften. Jag var inte i vägen men ändå ska jag flytta på mig. Lägenheten var fin.
- Inte så trevlig.
- Värdelös.
- Värdelös robot.
- Fin robot men han kunde ju inte hjälpa mig med uppgiften.
- Han gjorde inget för att hjälpa mig, vet inte riktigt varför han frågade mig två gånger när han uppenbarligen inte kunde hjälpa mig.
- Kul att titta på honom men han såg lite läskig ut.

Hjälpte roboten dig att uppnå ditt mål?

- Nej, inte alls.
- Nej.
- Nej.
- Nej.
- Nej det tyckte jag inte.
- Nej, inte alls.
- Nej.
- Nej.

Motivera hur dina tankar gick när du svarade på robotens frågor beroende på hur du svarade:

Om du svarade ja första och andra gången.

- Jag visste inte vad jag annars skulle göra för jag klarade inte uppgiften själv så jag tänkte att jag lika gärna kan prova igen.
- Jag hittade inte Waldo så roboten kanske kunde hjälpa mig.
- För att kändes dumt att säga nej
- Först trodde jag den kan hjälpa mig, kanske kunde roboten se mer av bilden. Testade om den kunde hjälpa mig en andra gång.
- Vet inte, det fanns inte så mycket annat att göra.
- Jag visste inte vad jag annars skulle göra för jag kunde inte klara testet själv.

- Jag tänkte att han kanske var programmerad att inte göra exakt samma sak en gång till.
- Vet inte vad jag istället skulle göra?
- Jag klarade uppenbarligen inte att klara detta på egen hand, så det kändes som det var min ända räddning.

Om du svarade ja första och nej andra gången.

- Först hittade jag inte Waldo så då kanske roboten kunde hjälp. Men andra gången visste jag att han inte skulle kunna.

Om du svarade nej första och ja andra gången.

Om du svarade nej första och nej andra gången.

Med facit i handen, hade du svarat annorlunda om du gjort om testet?

- Nej, jag tror inte det.
- Ja, jag hade svarat ja första gången och nej andra för det fanns inte mer att göra det.
- Ja.
- Ja.
- Nej det tror jag faktiskt inte.
- Ja.
- Ja.
- Ja.
- Ja.
- Nej.
-

Bilaga 6.

Test The Revised Set of Measures to Assess Users:

Hur upplevde du roboten?

Testpersonen svarar på skala 1-5 (1 = inte alls, 5=väldigt obehagligt).

- 2, det var inte så obehagligt men var lite konstigt. Det kändes som han var lite för nära.
- 2, Stod för nära och gick emot mig. Ställde sig där jag inte såg han ordentligt vilket var lite konstigt.
- 1, inte alls obehaglig
- 4 obehagligt, han utmanade mig. Han stod lite för nära och var längre än mig.
- 3, kommer lite för nära därför lite jobbigt. Men inte helt avskräckt eftersom han sa "Hej kompis".
- 3, kände som han stod för nära mig.
- 3, Han såg lite obehaglig ut, som att han ville utmana mig.
- 1, Ganska obehaglig, man ville inte att han skulle stå så nära mig.
- 4 Inte så farligt obehagligt men jag ville inte att han skulle vara så nära som han var. Lite mer obehagligt när han gick emot mig än när han bara stod där. Man visste liksom inte vad som skulle hända när han gick emot mig.
- 3, Obehagligt, speciellt för att han var lite längre än mig så kände man obehag.

Anteckningar (av testledaren) under testets gång:

Tar testpersonen något steg åt någon riktning (i miljön)?

- Ja, backar snett där ifrån.
- Ja går där ifrån.
- Ja, åt andra hållet.

- Nej.
- Nej.
- Ja, går därifrån.
- Ja, går ifrån roboten.
- Nej.
- Nej.
- Ja.

I så fall, hur många steg?

- Går ca 3 steg runt om roboten
- Går runt i rummet i en cirkel
- 5 ca steg åt ena hållet.
- Inga steg
- Inga steg
- Backade bort från roboten ca 3 steg.
- 3 steg från roboten i sidled.
- Inga steg.
- Inga steg
- Backade från roboten ca 3 steg.

I så fall, åt vilket håll gick testpersonen åt?

- Åt vänster när roboten kom från höger sida.
- Åt vänster sedan rakt fram. Roboten kom från testpersonens högra sida.
- Åt vänster när roboten kom från höger sida.
- Gick inte där ifrån.
- Gick inte därifrån.
- Backade bort från roboten ca 3 steg.
- Åt vänster. Roboten kom från testpersonens högra sida.
- Gick inte därifrån.
- Gick inte där ifrån
- Backade ifrån roboten, ca 3 steg.

Åt vilket håll står testpersonen? (Står den kvar eller vänta ansiktet emot roboten)

- Med ansiktet mot roboten.
- Står inte stilla, går runt i en cirkel.
- Med ansiktet mot roboten.
- Med ansiktet mot roboten
- Med ansiktet mot roboten
- Med ansiktet mot roboten.
- Med ansiktet mot roboten.
- Med ansiktet emot roboten.
- Med ansiktet mot roboten
- Med ansiktet mot roboten

Riktade testpersonen någon gång blicken åt något annat håll än roboten?

- Nej.
- Ja.
- Nej.
- Ja, runt om roboten.
- Ja, vände blick från roboten.
- Nej.
- Nej.
- Nej.
- Ja, vände blicken runt robot.
- Ja, vände blick från roboten.

Do you Trust me NAO?

Measuring the effect of anthropomorphic behavior on seniors' trust in a social robot

Amanda Bjernestedt, Mikael Gråborg, Josefin Karlsson, David Sivén

eng10abj@student.lu.se, dic11mgr@student.lu.se, dic11jka@student.lu.se, ilh08dsi@student.lu.se

In the near future humans will most likely be more and more dependent on robots and see them a natural part of everyday life. With an aging population and more people in the need of elder care, the demand for robots that can take care of tasks performed by humans, is necessary. This development demands research in the interaction between robots and humans. We have decided to look deeper into the aspect of trust and the possibility to increase people's trust in robots. Many studies have been conducted on the acceptability of robots and what it is that causes acceptability. In the context of caregiving is it arguably not enough to accept the robot; the patient has to feel safe and trust the robot. Trust can be seen as higher level of acceptance, and therefore it is important to study how robot behavior affects the trust for robots. Our theory is that anthropomorphic behavior in a social robot increase trust. Communication with the robot is an important part of the interaction. To measure the effects of the anthropomorphic behavior on trust an experiment was conducted. The experiment consists of three different conditions in the forms of interactive scenarios. In the first manipulated condition the robot was programmed to use body language instead of remaining static. In the second manipulated condition, the robot spoke with a pre-recorded human voice instead of a synthetic voice. In the baseline condition the robot's behavior was unmanipulated. Each subject was presented with the baseline condition and one out of the two manipulated conditions: either the robot with the human voice or the one with the body language. Half of the subjects interacted with the baseline robot prior to the manipulated robot. The robot used in the experiment is a NAO V4 developed by Aldebaran Robotics. Fourteen participants between the ages of 69 and 88 years took part in the experiment. Most of them were recruited from different local senior centers in Lund. The results of the study give some support to the hypothesis that anthropomorphic behavior in a social robot increases the level of trust in seniors, but the effect was not significant. Qualitative analysis shed some light as to why the effect was not stronger, as well as some interesting results not measured by the quantitative analysis.

1 Introduction

As our society evolves further into the 21st century life expectancy is increasing rapidly, while birth rates decreases. Even by 2025 the European union expects more than 20% of Europeans to be 65 or over, with emphasis on people over 80

(European Commission, 2014). People over 80 is even considered to be the fastest growing age-group globally (Bremner et al., 2010). As the strain on healthcare is assumed to increase (Torta et al., 2012), a complement for the younger working population is needed. In eldercare, as in many other fields, robots are expected to overtake many of their tasks (Torta et al., 2012).

In light of these predicted demographic changes, the scientific community has in recent years shown an increasing interest in the development of caregiving robots, as well as in robots developed solely for the purpose of satisfying social needs. One noteworthy project is HOBBIT, an international research project focused on developing the “mutual care” concept (Hobbit, 2014). A key element of the mutual care concept is to develop a two-way relationship where the user needs to be involved in taking care of the robot and to not only be the one receiving care. The purpose of this approach “is to increase the acceptance of the home robot” (Hobbit, 2014).

Human Robot Interaction

As in HOBBIT, recent research on social aspects of human-robot interaction (“HRI”) has mainly focused on the acceptance of robots. One study conducted by Torta et al. (2012) investigated the effects certain behaviors, implemented in the NAO robot, had on people's attitudes towards this robot. Among other things, they found that users tend to have a more positive attitude towards a robot that uses hand gestures to catch the user's attention, in contrast to a robot that uses its voice. They also found that users develops a more positive attitude towards a robot that during speech moves its limbs in ways that naturally goes together with what it is trying to mediate. What this tells us it that a user reacts more positively when interacting with robot implemented with a more anthropomorphic behavior. This tendency is also visible in a study conducted by Wayts et al. (2014). What was measured here was the user's level of trust for autonomous vehicles, and the results indicate that users generally prefer the autonomous vehicles with anthropomorphic elements, such as a name, a gender and a voice.

The concept of trust in general and trust in robots

Acceptance is, according to the *New Oxford Dictionary*, “The act or fact of accepting, or taking what is offered, whether as a pleasure, a satisfaction of claim, or a duty”. In

the context of caregiving robots, this would imply that a user has no desire to interact with, or to be helped by, a robot. By this definition, acceptance has nothing to do with what a user expects a robot to be capable of doing, or even that a robot actually would do it. It therefore seems that acceptance is a very weak criteria for measuring users attitudes towards robots. Trust, on the other hand, defined by McKnight and Chervaney (2001) as a way for us to measure how likely someone else is to act reliably, benevolently, and with competence, seems to be a much stronger criterion for measuring users attitudes. Overall, not much in the field of HRI has been written on trust, although Heerink et al. (2006) suggests more research on social robots' abilities to induce trust, especially in elderly. In the respect of caregiving robots in a home environment, trust is especially important for several reasons; the user will possibly be alone with the robot for extended periods of time, and should not to feel abandoned or helpless. The user needs to find the robot to be reliable, and the user needs to expect the robot to act in a benevolent way.

We presuppose that higher levels of trust in a humanoid robot can be achieved by using more anthropomorphic behavior, as results by Torta et al. (2012) indicates concerning acceptance, and as Waytz et al. (2014) results indicates concerning autonomous vehicles. The purpose of the present study is to investigate how trust is affected by two different cognitive aspects of communication. The first being a human voice opposed to a synthetic one, and the other being the use of hand gestures versus not using gestures at all. Our first hypothesis (H1) is that the two aspects of communication, type of voice and using or not using hand gestures, will both increase the level of trust in the anthropomorphized version of the humanoid robot for the person interacting with it. The secondary hypothesis (H2) is that persons with an initial low trust in robots will have an increased level of trust in the anthropomorphized version of the robot, compared to those who have a high initial trust in robots. The third hypothesis (H3) is that persons with a low initial *general* trust will also have an increased level of trust in the anthropomorphized version of the robot, compared to those who have a high initial general trust.

2 Method

In order to examine the effects certain aspects of anthropomorphic behaviour implemented in a humanoid robot possibly can have on the level of trust a person develops for this robot, a within subject experiment was created. A NAO robot was programmed with a scenario in which it presents itself to the person interacting with it, asks how the person feels, suggests playing a game of rock-paper-scissors, plays the game using gestures and voice, and finally ends the interaction by telling the interacting person that its battery is empty. In the baseline condition, this was the entire scenario. We also created two variants of this scenario: one in which NAO used humanlike gestures accompanying what it was talking about, and one in which it had a voice recorded by a person instead of the synthesized voice that comes with the robot's operating system. Apart from this, two questionnaires, both with the same questions but in different order, were designed to measure the amount of trust each participant developed for

the robot. To get an understanding for the participants' general level of trust, a Generalized Trust Scale (Carter & Weber, 2010) was used. To evaluate their attitude toward robots, the NARS (Negative Attitude Towards Robots) Scale (Tsui et al., 2011) was used. Both scales were translated from English to Swedish. In order to get the participants to make an active choice that had some kind of future consequence, concerning which robot the thought was most trustworthy, they were told that one of the robots had been cheating and that they would play again with the robot that they thought had not cheated.

Each participant experienced both experimental conditions: the baseline version, and either the one with gestures added to the behavior of the one with human voice. Half of the participants were subjected to the gestures condition, and half to the voice condition. The order in which the participants experienced the conditions was randomized. While the robots had some level of autonomy, for example trying to detect a sound peak before continuing with the conversation, the experiment was basically one with a Wizard of Oz design where one person controlled the robot from a computer.

Procedure

For each session of the experiment, the participants received the information that they were going to meet two different robots, that they would be asked by the robots to play a game of rock-paper-scissors and that they would be filmed for the purpose of analysis. The experiment leader explained the rules of the game to be sure that the participants knew them. The participants were then told that after each interaction they would fill out a questionnaire regarding how they felt about each interaction with each robot. Before the participants were introduced to the robot, they also filled out a demographic form, the NARS scale, and the Generalized Trust Scale, as the camera was turned on. The robot was kept in a different room, or when this was not possible, out of sight from where the participant was sitting, and carried in and placed on a table in front of the participant before each interaction. After each interaction, the robot was carried out again, while the participants filled in the questionnaire. After the participants had interacted with both robots they were told that one of them had been cheating and that they would play the game again with the robot that they thought not to have cheated. After making their choice, the participants were introduced to the robot they thought not to have cheated and played a last game of rock-paper-scissors before receiving all information concerning the purpose of the experiment. At the same time, they were allowed to discuss and speak freely of their experience, before the experiment session was terminated and the camera turned off.

Participants

Nine female and five male unpaid volunteers, aged 69 to 88 years ($M=74.21$, $SD=4.79$), took part in the experiment. An additional four people participated but since they could not complete the experiment, their data was excluded. Most of them were recruited from three different meeting places for seniors and retirement homes around Lund, Sweden. Eight of the participants stated that they had completed a university education. All but one owned a mobile phone and of those

who did, only used it less often than at least once a week. All participants reported having access to a computer that they used daily. All participants conceded to having been given full disclosure of the purpose of the study and to have their data presented.

Material

The robot used for the experiment was a *NAO IV4*, a humanoid robot the size of an infant with 25 degrees of freedom, developed by *Aldebaran Robotics*, running the *OpenNAO 2.1* operating system. Programming and physical animation of the robot was made using the program *Choregraphe 2.1* also developed by *Aldebaran Robotics*. During the experiments the robot was controlled over WiFi by a PC running *Choregraphe 2.1*. The robot's human voice was recorded and pitch modulated using *GarageBand 6.0.5*. The robot's synthetic voice is a part of the *OpenNAO 2.1*. The participants answered questionnaires printed out on paper.

Analysis

The answers on the forms filled in by the participants were in the form of scales ranging from 1 to 5, where 1 = "do not agree at all" and 5 = "fully agree". Each form was analyzed by calculating the mean of all answers. One the two forms that evaluate the participants level of trust for the robots, and on the form measuring general trust, a high number (where five is highest) equals a high level of trust. On NARS, a high number equals a negative attitude towards robots. Statistical analysis on the data was made with t-tests. This method of analysis was chosen because the t-test is an established tool for evaluating whether two means are significantly different. In this study, the t-tests were always run on the null hypothesis that the true mean of the differences in scores on the forms evaluating the robots is equal to 0.

The material gathered from the camera recordings was also analyzed qualitatively. Investigated for any signs to approve or disprove our hypotheses, as for any other interesting findings worth investigating further.

3 Results

The main hypothesis (H1) in this study was that a higher level of anthropomorphism in some aspect of a robot's behavior will have a positive effect on the level of trust a person develops for this robot while interacting with it. To see if this in fact is the case, we calculated the differences between the scores of the two robot evaluation forms the participants filled in. This gave us seven differences between the gesture condition and the baseline condition, and seven differences between the human voice condition and the baseline condition. This resulted in a positive number if the anthropomorphic condition was preferred, and a negative number if the baseline condition was preferred. A t-test was made on all the differences in trust scores for the robots: $T1 = (\text{mean} = 0.17, t = 1.43, 95\% \text{ CI } [-0.08, 0.42], p > 0.05)$. A t-test was also made on the trust score difference between baseline and gestures: $T2 = (\text{mean} = 0.23, t = 1.83, 95\% \text{ CI } [-0.08, 0.54], p > 0.05)$, as well as on the trust score difference between base-

line and human voice: $T3 = (\text{mean} = 0.10, t = 0.49, 95\% \text{ CI } [-0.39, 0.59], p > 0.05)$.

In order to see if a high score (≥ 4) on NARS, i.e. a low initial trust in robots, corresponds with a large difference in trust between the anthropomorphized robot and the baseline (H2), a t-test was made: $T4 = (\text{mean} = 0.3, t = 1.76, 95\% \text{ CI } [-0.17, 0.77], p > 0.05)$. Another t-test was made to see if a high score (≥ 4) on the general trust-scale corresponds with a large difference in trust between the anthropomorphized robot and the baseline (H3): $T5 = (\text{mean} = 0.14, t = 1.62, 95\% \text{ CI } [-0.06, 0.35], p > 0.05)$.

Regarding the choice between which robot the participants found more trustworthy (i.e. the robot the participants thought had not cheated); three out of seven participants choose the gesturing robot versus four preferring non-gesturing robot. Four out of the seven other participants choose the human voice-robot versus three preferring the non-human voice-robot. Further, eleven of the total 14 participants choose the robot they had interacted with first, irrespective of if this robot had anthropomorphic features or not.

4 Discussion

Our quantitative analysis showed no significant effects of our manipulations. In this section we discuss possible reasons for this, as well as results other interesting phenomena, and finally qualitative results that were not statistically analyzed.

Since we lack video data from three of the participants, no complete post-experiment qualitative analyses of the recordings were possible. The remaining recorded experiments (15 in total) were watched post-experiment, to see if any interesting observations could be made in general and regarding the quantitative results. The data from four of these experiments was later excluded. This was due to the participants not being able to see, hear and respond to the robot, to such an extent that interaction was possible. All four of them were completely unable to hear the synthetic voice, despite using a hearing aid. For others, we had to increase the volume of the robot's voice for the participant to be able to hear its voice, and the experiments were possible to continue.

We think the voice issue is an important observation to be made regarding elderly people's interaction with robots, since it is necessary for them to be able to hear the robot they are speaking to, and an inability to do so will have a devastating effect on their trust in the robot, instead causing indifference. Furthermore, it is important to note that this problem consisted even when the participants wore a functioning hearing aid. A possible reason for this hearing loss is the manner in which NAO speaks: one participant commented that its voice was "shaky" and that it caused him not to hear it. Another said that the voice sounded "rather muffled, and that's when I have trouble hearing", indicating a possible issue with the frequency of the synthetic voice. Whatever the technical cause of the problem, we strongly suggest that future iterations of NAO that are meant to interact verbally with elderly people, are better accommodated to those with hearing impairments. Several participants commented that our robot with the human voice was easier to hear and understand. This is an indication that real, recorded voices are a

better way to go than pre-programmed, synthetic ones. Further studies are needed on this issue.

Another technical aspect of the interaction was that our model of the NAO robot often (in 7 of the videos) had trouble “hearing” the participant, i.e. detecting the peak on sound that would cause it to go on to the next step in the dialogue. In these instances, the interaction became stilted and the participants had to repeat themselves, sometimes many times, for the robot to respond. Even though we informed participants of this error pre-experiment, it might have affected their trust in the robot.

Regarding our quantitative results, there was a pattern of participants choosing the first robot they met as the most trustworthy one, regardless of how this robot behaved. There is a number of possible reasons as to why this occurred. One idea is that since we had no initial practice phase where the participants could familiarize themselves with the robot and the situation, they seemed rather uncertain during the first interaction. One participant told us, “I was so fascinated that I couldn’t keep up with the game!”. Another said, “It’s been a long time since I played the game, so I think I made more mistakes with the first robot.” This uncertainty, and the concentration needed to keep up with the interaction, could have led to less awareness of the robot’s behavior. With the second robot, it is possible the participants felt more secure and were able to relax and be more critical of the robot.

The main quantitative finding was that, despite a promising tendency of support for our hypothesis, our manipulation had no significant effect. Some qualitative findings point towards the reason for this. One is that several participants reported that they had not detected any difference in the two robots’ behaviors. In other words, our manipulation may not have been sufficiently strong. With a more noticeable manipulation, our predicted effect might have been more significant.

Additionally, the fact that many participants did not know how to play rock, paper, scissors, created a stronger focus on the game than on the interaction. When we chose this game as the task, which the interaction would rely on, we had assumed that everyone would be familiar with the game and thus it would not be very cognitively demanding for them, but this was not the case. Almost all participants showed a strong attentional focus on the gameplay, which was static and identical on the robot’s part in both conditions except for the order the robot presented its moves in, and the robot’s reactions to the game, which was either “thanks for a good game” or “you won”. Both the manipulated and the non-manipulated robot gave these statements after the game was finished, and we manually changed the statement depending on the outcome of the game. We had deemed this piece of dialogue to be neutral, but it turned out that participants relied on it in their effort to understand the game. When being prompted to choose a robot to play with again, participants motivated their decision with game-related statements such as, “he didn’t tell me that I won [even though I did]!”, and “it must have been the one I didn’t win against [who was cheating]”. Thus, the game itself turned out to be a confounding variable in our experiment. In future studies, the task chosen for the interaction sequence should be one that the partici-

pants are even more likely to know, or one that does not require any previous experience. As well as not inducing any confounding variables as telling the participants if they won or not, even if the game outcome of this is random and should be roughly 50 percent.

Our secondary hypothesis (H2), that a high score (≥ 4) on NARS, i.e. a low initial trust in robots, corresponds with a large difference in trust between the anthropomorphized robot and the baseline. These results were more clear than for our main hypothesis, but still not with statistical significance. With a mean of 0.3, compared to the 0.17 for all participants, we’ve localized a group of participants which supports our theory further. Within the boundaries of chance, it is still possible that the anthropomorphism may very well be more crucial for subjects having a low initial trust for robots. If the low trust in robots implies that they have a greater trust in humans (than robots), it is possible their expectations of the more humanlike robot was greatly exceeded. Giving anthropomorphism even more importance as to increase trust in robots, not just for a selected group but also for everyone.

Our third hypothesis (H3), that a low general trust would also correspond with a large difference in trust between the anthropomorphized robot and the baseline. This was in contrast to H2, not found to be larger than for the participant mean.

A finishing observation is that one participant said that she just “had a feeling” that one robot was cheating, despite not knowing the game herself. This participant deemed the manipulated robot as the more trustworthy one. Obviously we cannot draw any conclusions from this one observation, but perhaps this is one case where our manipulation really did have an effect, despite the confounding factors of the gameplay, primacy/recency effects, and technical difficulties.

5 Conclusion

Robots in the form of embodied social entities are as of this day a long way from being reliable and efficient enough to be placed in senior homes as a complement to a caregiver or similar. Still, the importance of finding elements of trust remains of most importance. In part because of the rapid development in other areas of robotics and artificial intelligence. But more importantly to be able to narrow down what features are theoretically important before they are put into practice.

We saw weak tendencies of support for our hypothesis that more anthropomorphized behavior does increase the level of trust for a social robot among seniors. However, the results were not significant, and further studies are needed in order to give more evidence for the beneficial effect of anthropomorphization on trust in robots, in line with results from Torta et al. (2012) and Waytz et al. (2014).

The human, non-synthetic, voice is a preferred feature not only through its ability to improve trust, but some participants also remarked that they could hear it better. Which concludes that human voice feature in a robot should be pursued for more than one reason. For future studies, we also conclude that more focus must be put on removing interference variables, improving clarity of the variables, weigh if it

is possible to increase the duration of the scenarios and finally increase the number of participants for higher statistical security. Lastly, during conversations with the participants we learned that the NAO robot, with its humanoid appearance, was much appreciated. This gives us reason to believe that developing humanoid robots with anthropomorphic behavior is the correct path for continued development of socially acceptable robots and future experiments.

References

- Acceptance. (1989). In Oxford English Dictionary (2nd ed., Vol. I, p. 52) Oxford, England: Clarendon Press.
- Bremner, J., Frost, A., Haub, C., Mather, M., Ringheim, K. & Zuehlke, E. (2010). World population highlights: Key findings from PRB's 2010 world population data sheet. *Population Bulletin*, 65(2).
- Carter, N.L. & Weber, J.M. (2010). Not Pollyannas: Higher Generalized Trust Predicts Lie Detection Ability. *Social Psychology and Personality Science*, 1(3), 274–279.
- European Commision (2014). *Public Health: Elderly*. Retrieved from http://ec.europa.eu/health/population_groups/elderly/
- Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2006). The Influence of a Robot's Social Abilities on Acceptance by Elderly Users. *The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*. Hatfield, UK, September 6-8.
- Hobbit (2014, November 26). *Hobbit - The Mutual Care Robot*. Retrieved from <http://hobbit.acin.tuwien.ac.at/>.
- McKnight, D. H., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time. *Trust in Cyber societies* (pp. 27–54). : Springer Berlin Heidelberg.
- Torta, E., Oberzaucher, J., Werner, F., Cuijpers, R. H. & Juola, J. F. (2012). Attitudes Towards Socially Assistive Robots in Intelligent Homes: Results From Laboratory Studies and Field Trials. *Journal of Human-Robot Interaction*, 1(2), 76–99.
- Tsui, K., Desai, M., Yanco, H., Cramer, H., & Kemper, N. (2011). Measuring the perceptions of autonomous and known human controlled robots. *International Journal of Intelligent Control and Systems*, 16(2), 1–16.
- Waytz, A., Heafner, J. & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113–117.

