

# Röstbaserade Instruktioner

Vad som krävs för att göra sig förstådd.

Marcus Nilsson

# Abstrakt

Instruktioner i ljudform är något man sällan stöter på, såvida de inte ackompanjeras av video. Vanligast är dock text- och bildinstruktioner. Detta arbete utforskar om instruktioner i ljudform kan stå på egna ben, vilka begränsningar det finns och vad man bör tänka på för att skapa tydliga ljudinstruktioner.

Detta uppnås genom en iterativ process där jag testat ett flertal prototyper baserade på två olika användningsområden för instruktioner. Med bakning testar jag enkla instruktioner där ordning och precision inte är kritiskt avgörande. Med Lego testar jag komplicerade processer där varje stegs framgång hänger på att de tidigare stegen blivit korrekt utförda.

Oavsett hur avancerade instruktionerna är så spelar språket stor roll. Ju mer komplicerade processerna är desto viktigare är användarens språkkunskaper samt uppläsarens uttal och frasering. Det testerna visar är att endast enklare typer av instruktioner lämpar sig i röstbaserad form. Faktorerna som påverkar röstinstruktioners tillämplighet är processens komplexitet, språkkompatibilitet, diktion, samt förkunskaper.

# Innehållsförteckning

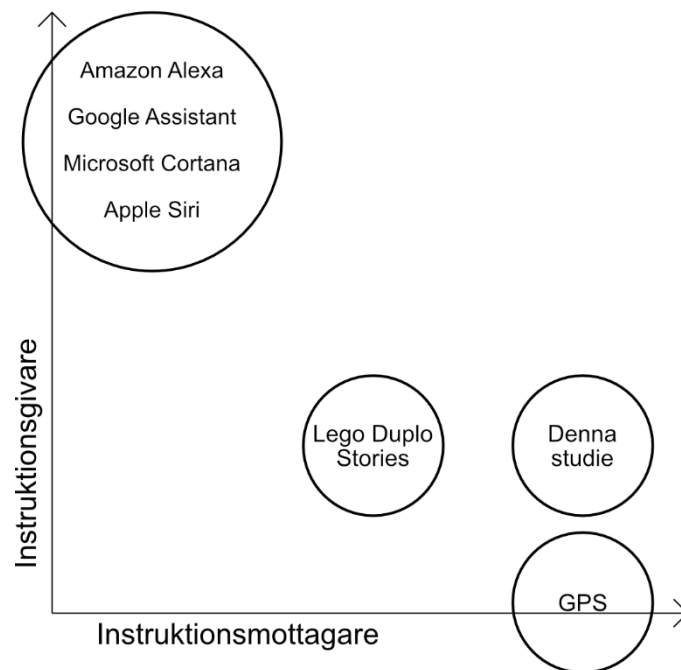
1	Introduktion .....	4
2	Forskningsfråga .....	6
3	Avgränsningar .....	6
4	Bakgrund .....	6
5	Metod.....	8
5.1	Rekrytering av testpersoner .....	8
5.2	Användartester .....	9
5.3	Etik i testprocessen .....	10
5.4	Wizard of oz.....	10
6	Designprocess .....	10
6.1	Prototyp 1 (Bakning) .....	12
6.2	Resultat av prototyp 1.....	14
6.3	Prototyp 2 (Lego) .....	16
6.4	Resultat av prototyp 2 .....	18
6.5	Prototyp 3 (Lego).....	19
6.6	Resultat av prototyp 3 .....	21
6.7	Prototyp 4 (Lego) .....	22
6.8	Resultat av prototyp 4 .....	22
6.9	Prototyp 5 (Lego).....	23
6.10	Resultat av prototyp 5 .....	25
7	Diskussion .....	26
7.1	Processens komplexitet .....	26
7.2	Språk.....	26
7.3	Diktion: Uttal, frasering, och takt .....	27
7.4	Förkunskaper .....	27
7.5	Sammanfattning.....	27
7.6	Vidareutveckling .....	28
8	Referenser.....	29

# 1 Introduktion

Att följa instruktioner är en del av vardagen för de flesta. Det kan vara enkla saker som att vika ihop en mjölkkartong eller något mer komplext, till exempel att montera möbler. Majoriteten av dessa instruktioner presenteras i text- eller bildform. Video är på uppgång och blir alltmer vanligt. Men det finns situationer när du inte kan interagera med en videospelare eller vända blad i ett instruktionshäfte då dina händer är upptagna. Ett bra exempel på detta är matlagning där dina händer kan vara för kladdiga för att ta i papper eller elektronik.

Röstbaserad interaktion är något som de senaste åren har utvecklats i rasande fart. Några av de största teknikföretagen inom mjukvara, Apple, Google, Microsoft, med flera, levererar produkter, kallade digitala assistenter, där människan styr elektroniken med sin röst. Det innebär att många har tillgång till sådan teknik i vardagen. Dessa produkter lyssnar till dina kommandon och frågor och uppfyller dem bäst de kan. Detta projekt ämnar utforska det omvända, om de kan ge oss människor instruktioner att följa.

## Människans roll i interaktionen



Figur 1: En uppskattning av människans roll i interaktionen i olika röstbaserade tjänster. (Storleken av cirklarna har ingen betydelse).

De befintliga proprietära tjänsterna Amazon Alexa, Apple Siri, Google Assistant, och Microsoft Cortana har fokus på att assistera utifrån kommandon från användaren. Alltså, datorn lyssnar på människan med syfte att fungera som en personlig sekreterare. Det viktigaste i deras tjänster är att maskinellt förstå människors tal, oavsett dialekt och uttal och kunna agera utefter det.

I andra sidan av spektrat ligger GPS. Akronymen står för Global Positioning System och är en satellitbaserad teknik för att avgöra position på jordens yta. I denna text syftar jag dock på den mer vardagliga tolkningen av akronymen: Ett namn för navigationssystem baserade på nämnd teknik som framförallt används i fordon istället för manuell kartläsning. Detta skiljer sig då avsevärt från de digitala assistenterna i att all fokus här ligger på att elektroniken instruerar användaren.

GPS levererar instruktioner till föraren med en kombination av tal och rörlig bild. Användaren kan dock välja att stänga av talinstruktionerna. Instruktionerna levereras baserat på din aktuella position och rutten kan uppdateras och anpassas i det fall användaren inte följer instruktionerna korrekt.

Denna studie skiljer sig från de digitala assistenterna i kommunikationsinriktning. Tekniktillämpningen som undersökts här har samma kommunikationsriktning som GPS men med några betydande begränsningar. Den skiljer sig i avseendet att det inte är multimodalt, utan enbart baserat på tal, och att instruktionsgivaren här inte kan anpassa instruktionerna efter hur de efterföljs.

Att enbart använda röstinstruktioner från dator till människa är inte vanligt. Under detta arbetes gång har Amazon dock i samarbete med Lego lanserat en ny tillämpning för Alexa som även ger ut instruktioner. Den heter Lego Duplo Stories och ämnar hjälpa barn i språkutvecklingen genom lek med leksaken Duplo (Lego, 2018). Barnet får med hjälp av rösten välja ett djur eller ett fordon ur en lista, varpå tjänsten läser upp en berättelse om det. Under historiens gång uppmuntras barnet att göra saker med leksaken, till exempel släppa ut en ko ur stallet, varpå tjänsten väntar på att barnet ska svara "jag är redo" för att sedan fortsätta berättelsen.

Kan rena röstinstruktioner användas till mer än lek? Att företag som Amazon och Lego intresserar sig för det tyder på att tekniken är mogen, men vad finns det för begränsningar? Vad finns det för förhållningsregler för att det ska fungera? Denna text ämnar undersöka dessa punkter.

Jag kommer använda mig av ett användarcentrerat tillvägagångssätt där jag involverar användare så mycket som möjligt. Användarcentrerad design hjälper designern att se förbi sina egna preferenser och fokusera på användarens behov (Saffer, 2010). Detta kommer jag åstadkomma genom många och frekventa tester av lo-fi prototyper. Saffer (2010) beskriver dem som grova imitationer av hur produkten skulle kunna fungera. De kräver

hjälp av en människa för att vara interaktiva men låter utvecklaren snabbt testa ett koncept.

## 2 Forskningsfråga

Vilka är de påverkande faktorerna och vilken inverkan har de på interaktionens tillämplighet i strikt röstbaserade steg-för-steg-instruktioner, för icke-professionella mottagare?

## 3 Avgränsningar

Arbetet kommer att begränsa sig till endast röstinstruktioner. Jag kommer inte att gå in på multimodala medier, trots att en kombination av tal, stillbilder, och rörliga bilder sannolikt skulle kunna underlätta interaktionen. Avgränsningen bidrar till att förstå interaktionen i röststyrning innan den kombineras med andra metoder.

Yrkesmässiga användningsområden utesluts då jag valt att fokusera på lekmän. Arbetet kommer inte att beröra handikapp. Dessa användningsområden är dock naturliga vidareutvecklingar för framtida arbeten.

## 4 Bakgrund

Detta projekt syftar till att undersöka vilka olika faktorer som har påverkan på hur det fungerar att instruera användaren verbalt. Resultatet hänger till stor del på dennes förmåga att uppfatta och kortsiktigt memorera information levererad på det viset. Ware (2008) beskriver att vi kan hålla cirka två sekunder av talinformation i verbalt arbetsminne, motsvarande tre minnesbitar av information. Detta kallar han ekokrets eller ekominne. Baddeley (2010) benämner det: Fonologisk krets. Tidsspannet Ware uppger tar inte hänsyn till det faktum att två sekunders tal kan innehålla olika mängd information beroende på talhastighet. Ett exempel där tal kan innehålla mer än tre bitar information på två sekunder är skärmläsare för blinda. Under Microsoft build 2017 konferens (Sneath, 2017) demonstrerar Saqib Shaikh hur han som blind programmerar. Skärmläsaren läser upp varje tecken han skriver eller varje menyalternativ han markerar. Uppläsningen är så snabb

att jag inte kunde urskilja ord, men Saqib kan navigera datorn i vad som kan anses normal hastighet för människor utan handikapp.

Baddeley (2010) föreslår att den fonologiska loopen har två huvudsakliga funktioner. Den första är en lagringsplats för tal som spontant förfaller efter cirka två sekunder. Den andra är en process som kan uppfriska det minnet genom verbal repetition, både via tal och genom den inre rösten.

Omedelbar återkallning av en kort ordsekvens blir grovt försämrad om orden har liknande uttal, exempelvis ”Map, Cat, Cap, Mat, and Can”. Liknande betydelser däremot påverkar inte återkallningsförmågan. Ytterligare försämring av förmågan att återge en ordsekvens är beroendet av ordens längd, där längre ord är svårare att komma ihåg. Kapaciteten att återge en orelaterad ordföljd ligger på cirka fem ord. Bildar orden en meningsfull mening växer kapaciteten till cirka 15 ord. (Baddeley, 2010).

Arbetsminnets funktion är beroende av det aktiva långtidsminnet på många sätt. Till exempel är förmågan att minnas ett telefonnummer avsevärt mycket bättre på individens modersmål än på ett annat språk (Baddeley, 2010).

Det korta hörminnet är inte nödvändigtvis en begränsning då Black, Carroll, och McGuigan (1986) visar att kortare instruktionsmanualer som låter läsaren härleda informationen snarare än att explicit säga allt, inte bara fungerar lika bra som de långa väldetaljerade instruktionerna, utan också är mer tidseffektiva.

Roediger och McDermott (2000) testade vår förmåga att dra slutsatser från liknande data. De läste upp ord som alla var relaterade till ytterligare ett ord, utan att nämna det ordet. I flera fall kom testpersonerna felaktigt ihåg även det associerade ordet. Till exempel kan orden säng, vila, vaken, e.t.c.. få folk att minnas ordet sova. Detta fenomen kan kraftigt minska om man visar en bild av det upplästa ordet samtidigt. Detta kan både vara en fara och en styrka vid verbala instruktioner då lyssnaren kan dra egna slutsatser som visserligen kan leda till fel men också kan göra instruktionerna kortare och effektivare, då längre mer ingående instruktioner presterar sämre än kortare (Black et al., 1986).

Dalton, Agarwal, Fraenkel, Baichoo, och Masry (2013) fann att enkla instruktioner är lättare att minnas i ljudform än som illustrationer när de ska utföras i steg. De demonstrerade också att denna form av instruktioner kan ges utan att vara distraherande. Komplexa instruktioner däremot resulterade i markant sämre resultat.

Talaren har stor betydelse för hur väl instruktionerna uppfattas. Crabtree, Miranda, och Beukelman (1990) fann att hur pass syntetisk en röst är inte påverkar lyssnarens inställning till den. Förhoppningen var att kunna använda syntetiska röster i detta arbete då det i en färdig produkt är billigare och flexiblere än förinspelade instruktioner. Det finns dock markanta skillnader i hörförståelse mellan syntetisk och naturlig röst. I en

undersökning av Nye, Ingemann, och Donald (1975) där testpersonerna fick lyssna på information uppläst av en människoröst och sedan svara på en enkät, tog det i genomsnitt fyra och en halv minut. När samma test gjordes med en syntetisk röst ökade tiden det tog att svara på frågorna med en minut och 45 sekunder.

Venkatagiri (2004) visade att en människoröst har 95% procent begriplighet i test i rum med dålig akustik (eko). Datorröst under samma förhållande nådde endast 68% begriplighet. Att lyssna på text-till-syntetiskt-tal-röster i bullriga miljöer såsom i bilar eller på flygplatser, kontor, och i klassrum kräver opropotionerligt mycket kognitiva resurser av lyssnaren. Detta kan negativt påverka aktiviteter lyssnaren utför under tiden (Venkatagiri, 2005).

Tal hastighet påverkar hur väl lyssnare hinner bearbeta vad som sägs. Högre hastighet ger mindre utrymme för lyssnaren att reflektera (Smith & Shaffer, 1995). Detta är viktigt att hålla i åtanke även under de bästa röst- och ljudförhållandena då det är viktigt att lyssnaren fullständigt förstår instruktionerna.

En del av prototyperna i denna studie kräver beskrivning av legobitar. För detta behövs lättförstådd terminologi. Lego är en leksak bestående av byggklossar av plast som med hjälp av utbuktningar i ett rutnät på toppen, ofta kallade ploppar, kan fästas i motsvarande urgröpningar i botten. Detta låter användaren fritt skapa sammanhängande konstruktioner från lösa delar. I en studie av Boerger och Henley (1999) låter de deltagarna parvis montera legobyggsatser. De båda deltagarna i varje par är separerade med en skiljevägg så att de inte kan se vad den andra gör. En av dem har Legos bildbaserade instruktionsmanual och den andra har legobitarna. 11 av 16 deltagare i deras tester beskriver bitarna med hjälp av antalet ploppar i bredd och djup. Detta mönster uppstod utan att de blivit instruerade i terminologi.

## 5 Metod

### 5.1 Rekrutering av testpersoner

Testernas utformning påverkade mina möjligheter att engagera testpersoner. Målet var att göra individuella tester på flertalet personer för att få en bild av hur prototyperna uppfattas. Det fanns ingen budget att betala testpersonerna, vilket ledde mig till att skala ner testerna till en omfattning som folk frivilligt skulle kunna ställa upp på.

Jag valde att förlägga alla testerna i Niagarabyggnaden på Malmö Universitet eftersom det är en samlingsplats med bra rotation på människor men där många stannar längre stunder. Personerna som rör sig här har eller strävar mot högre utbildning. Jag utgår därför ifrån att dessa testpersoner bör klara

uppgifterna bättre än om urvalet representerar hela befolkningen. Har testpersonerna problem så kommer framtida användare också ha det.

Kriterierna för att delta var mycket öppna; endast att de aldrig deltagit eller kan ha sett tidigare test. En testperson som sett eller deltagit i tidigare test skulle antagligen bete sig annorlunda än en ny person.

Saffer (2010) varnar för att vi människor ofta undermedvetet väljer att interagera med folk som liknar oss, något som i designforskning kan få oss att gå miste om värdefulla åsikter. För att inte falla i den fällan skapade jag en regel som lyder: Om tvekan uppstår måste personen tillfrågas. Detta fungerade i denna situationen då målgruppen för testet var mycket öppen.

## 5.2 Användartester

Användartesternas syfte var att få inblick i hur personer agerar utifrån röstinstruktioner och hur väl de fungerar. Detta för att kunna avgöra om de kan ersätta traditionella text- och bildinstruktioner i situationer där dessa inte är optimala. Detta med utgångspunkten att text- och bildinstruktioner för samma ändamål fungerar då de för närvarande är standard.

Forskningsmetoden bygger på aktiviteter (Saffer, 2010) som engagerar testpersonerna, vilket framhäver känslor och låter mig förstå hur de tänker.

Testerna var upplagda som en iterativ process, där jag testade ofta och ändrade prototypen efter varje test baserat på resultaten.

En av de mest givande metoderna i designforskning enligt Saffer (2010) är att noggrant observera vad människor gör. Användartesterna byggde till störst del på observationer där jag var noga med att lägga märke till mönster i deltagarnas uppträdande. Mönster kan vara likheter i beteende, i vad testpersonerna uttrycker, svar på frågor, och andra händelser som repeteras genom testerna (Saffer, 2010). En tumregel är att när man börjar lägga märke till flertalet mönster har man antagligen gjort nog med tester för att kunna dra meningsfulla slutsatser (Saffer, 2010).

För dokumentation togs det bilder på det testpersonerna gjorde. En ljudinspelning pågick under hela testet för att kunna föra en normal konversation och för att anteckningar inte var möjligt, då prototypen krävde manuellt handhavande från min sida under testets gång. Efter varje test skrev jag ner kortfattade anteckningar rörande vad som hänt under testet, vad som gått bra eller dåligt och om det berodde på instruktionerna, språket eller testpersonens personlighet. Detta baserades till stor del på mina observationer men också på vad testpersonerna sade. Senare transkriberades inspelningarna, där utfyllnadsord och läten till största del utelämnades.

Testpersonerna fick väldigt få instruktioner inför testet. De fick endast veta att de skulle ignorera min närvaro och låtsas att det var en röststyrd app de använde.

Miljön i Niagarabyggnaden där testerna utfördes erbjöd bra med bordsytor att arbeta på och tillgång till ett grundläggande kök, vilket var relevant i ett av testerna. Det är en allmän men lugn plats, utan buller och distraktioner.

### 5.3 Etik i testprocessen

I enlighet med Vetenskapsrådets (2017) riktlinjer togs inga bilder eller ljudupptagningar förrän jag fått testpersonernas uttryckliga tillstånd. Jag såg också till att berätta om testets längd och vilka uppgifter de skulle få utföra i förväg, så att testpersonerna skulle kunna göra ett informerat beslut om de ville delta eller ej.

För testet där de skulle baka såg jag även till att informera om allergener bland ingredienserna.

### 5.4 Wizard of oz

För att uppnå resultat i testerna behövdes inget fungerande system. All respons från produkten skulle ske i ljudform. Användaren skulle alltså varken se eller röra vid något gränssnitt under testet. Jag utnyttjade wizard-of-oz-principen som innebär att i testet simulerar produkten för att låta användaren ha en äkta upplevelse, innan systemet finns på riktigt (Buxton, 2007).

## 6 Designprocess

Testerna består av två huvudprototyper, där den andra utvecklas i fyra iterationer. Den första går ut på att deltagaren ska baka muffins, styrd endast av verbala instruktioner. Den andra styr användaren på samma vis, men med en mer komplicerad uppgift, legobygge.

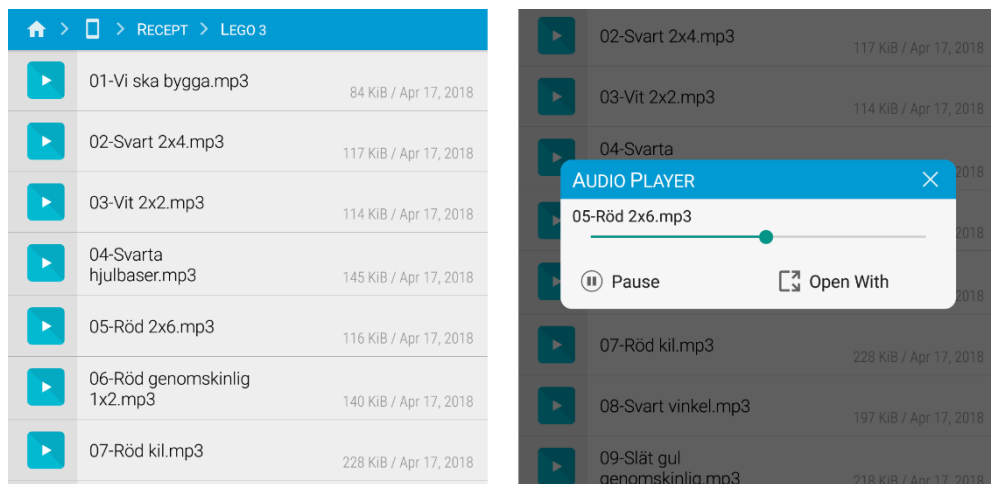
Prototyperna utformades som en röststyrd app, baserat på Wizard of oz principen (Buxton 2007), vilket innebär att jag manuellt fejkade appens funktioner för att ge testpersonen en autentisk upplevelse. Plattform och utseende är inte viktigt då testpersonerna endast interagerar med ett verbalt gränssnitt.

För de verbala instruktionerna hade jag två alternativ, datorgenererad röst och att spela in eget tal. Ur synvinkeln att detta någon gång kan bli en riktig produkt är en syntetisk röst fördelaktig då den ger mycket mer flexibilitet i utformningen av instruktionerna. Det tillåter skaparen av instruktionerna att ändra dem utan att spela in nya klipp. Ett problem med syntetiska röster är dock urvalet, särskilt på svenska. Valet föll på en engelsk röst då jag prioriterade tydlighet. Enligt en undersökning av Education First (2017) är svenskar näst bäst i världen på engelska av länder som inte har språket som

modersmål. “Very high proficiency” är orden de använder för att beskriva dem. Antagandet var därför att det skulle fungera lika bra på engelska som på svenska.

Wu, Huang och Wu (2009) undersökning av röstinstruerad navigation belyser vikten av timing i leveransen av instruktioner. Om instruktionerna kommer vid olämpliga tillfällen kan det leda till missförstånd eller att lyssnaren inte är mottaglig för information i den stunden. Därför levereras instruktionerna först när lyssnaren aktivt ber om dem.

Prototypen bygger på Wizard of oz principen (Buxton, 2007) genom manuell styrning av uppspelningen från min sida med en vanlig filhanterare i en smartphone, baserat på vilka kommandon testpersonen gav. Till telefonen var ett par trådlösa hörlurar kopplade som både spelade upp instruktionerna och spelade in vad testpersonen sade.



*Bild 1, 2: Applikationen FX File explorer som användes för att spela upp ljudinstruktionerna.*

Den simulerade appen skulle ha tre kommandon för att navigera instruktionerna:

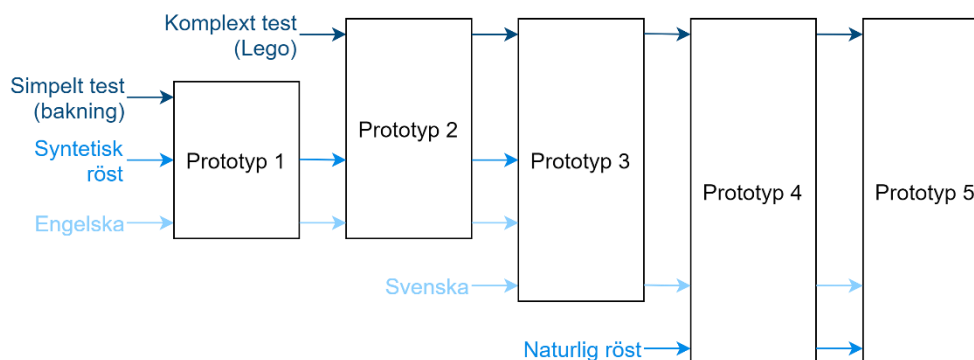
- Nästa steg.
- Repetera.
- Backa ett steg.

Jag var mycket strikt med att inte agera utan att ha fått ett tydligt kommando. Dock var jag inte strikt med att det skulle vara exakt efter instruktionerna, då modern röststyrning enligt min erfarenhet klarar av att tolka kommandon sagda på olika sätt. Till exempel “gå bakåt” istället för “backa ett steg”.

För varje steg i instruktionerna var målet att inte ha mer än tre minnesbitar av information då det är gränsen för vad man kan hålla i verbalt arbetsminne

(Ware, 2008). Det var även viktigt att det var tillräckligt mycket information i varje steg för att inte skapa osäkerhet.

Här följer en graf över de fem prototyperna som testades, vad uppgiften var, vilket språk, och huruvida rösten var mänsklig eller syntetisk för var och en.



Figur 1: De fem prototypernas upplägg.

## 6.1 Prototyp 1 (Bakning)

Första prototypen ämnade att vara så enkel som möjligt för att identifiera de grundläggande faktorerna som påverkar hur väl de röststyrda instruktionerna fungerar. Ett chokladmuffinsrecept av Swah (2015) valdes som går att tillaga i mikrovågsugn. Jag testbakade det för att försäkra mig om att det passade inom mina ramar för tid och svårighetsgrad, men även för att avgöra vilka redskap deltagarna skulle behöva då testerna inte skulle utföras i ett fullutrustat kök. En av fördelarna med bakning är att det lärs ut mellan årskurs 6 och 9 som en del i ämnet hem- och konsumentkunskap (Skolverket, 2018). Därför har testpersonerna med största sannolikhet tidigare erfarenhet inom området.

Då testet inleddes var alla ingredienserna framställda på arbetsytan. Risken finns att detta kan ha påverkat resultatet något då man kan lista ut vilken ingrediens instruktionerna syftar på med hjälp av uteslutningsmetoden. Dock bör man säkerställa att man har ingredienserna innan man påbörjar även ett vanligt recept, så det antas inte ha någon betydelse för resultatet.

I enlighet med Wares (2008) teori om tre minnesbitar delade jag upp instruktionerna i ingrediens, mängd, och handling. Efter ett antal omskrivningar valde jag att skriva instruktionerna som kompletta meningar med så få utfyllnadsord om möjligt. Ursprungligen ville jag punkta upp varje steg med ingrediens först, följt av mängd och sist handling, eftersom det är ordningen du behöver informationen i. Exempel på detta: Sugar, one and a quarter tablespoon, add. Men det blev osammanhängande då det inte följer förväntad meningsbyggnad. Så valet föll på en normal men kortfattad mening. Exempel: Add one and a quarter tablespoon of sugar.

### Instruktionerna:

1. We'll start with the cake mix.
2. Put 1 and a half tablespoons of flour into a small bowl.
3. Add 1 and a quarter tablespoon of sugar.
4. Add 3 quarters of a tablespoon cocoa powder.
5. Add a quarter of a teaspoon baking powder.
6. Stir until mixed properly.
7. Add half a tablespoon of melted butter.
8. Add 1 and a half tablespoons of milk.
9. Add a splash of vanilla extract.
10. Stir with a small whisk or fork until smooth.
11. Pour into cup.
12. Put the cup aside to prepare the frosting.
13. Put 1 and a half tablespoon of butter into a bowl.
14. Using a spatula, mix the butter well until smooth.
15. Add half a tablespoon of cocoa powder.
16. Stir until all lumps are gone.
17. While stirring, gradually add icing sugar until creamy and thick. 2 or 3 tablespoons.
18. If the mixture is too thick to spread, add a dash of milk to soften it up.
19. Put the frosting aside.
20. Cook the cup in the middle of the microwave for 1 minute at medium heat.
21. Smear the frosting onto the cupcake.
22. Finish with sprinkles.

Medellängden per instruktion är tre sekunder, detta är dock mätt på de avrundade värdena i filinformationen, samt att varje fil har ett ögonblick av tystnad i början och slutet av filen. Detta sänker medellängden något. Största utstickarna är steg 17 på sju sekunder och steg 18 på fem sekunder. Steg 17 går även tekniskt sett över trebitarsgränsen för information. Dock bygger första delen på att fortsätta med handlingen från det föregående steget, varpå jag teoretiserar att man inte behöver hålla det i minnet.

Detta test skedde i köket på plan 2C i Niagarabyggnaden på Malmö Universitet. Köket är endast utrustat med vask samt kopiösa mängder mikrovågsugnar. Jag tillhandahöll ingredienser och fler redskap än nödvändigt så att testpersonerna kunde välja fritt utefter sin tolkning av instruktionerna.

Inför testet instruerades deltagarna om att det kommer fungera som en röststyrd app, vilka kommandon de kunde ge samt att de skulle ignorera min närvaro. De tillfrågades även om det gick bra att jag fotograferade och spelade in ljud för dokumentationssyfte, vilket endast en hade invändningar mot.

De tillfrågade fick veta att de skulle baka muffins och att det hela skulle ta cirka 15 minuter. Receptet säger 4 minuter men det är väldigt optimistisk enligt min egen testbakning samt att det ska finnas tid för frågor efteråt. Tidsuppskattningen stämde bra men varje test i sin helhet tog cirka 30 minuter med efterarbetet inräknat. Det var mycket svårt att hitta villiga testpersoner.

## 6.2 Resultat av prototyp 1

Deltagare	3
Röst	Syntetisk
Språk	Engelska
Antal som lyckades	2 (varav 1 efter tillrättavisning)



*Bild 3, 4: Två av deltagarna under bakningstesterna.*

I samtalet efter testet var person A mycket positiv till att få instruktionerna levererade på kommando, att inte behöva hålla reda på var i receptet man är, samt att inte behöva korsreferera mellan instruktioner och ingredienslista. Hon sade sig vara ovan vid att baka, och dålig på att följa recept. Metoden hjälpte henne att fullfölja instruktionerna, vilket hon inte brukar göra med textinstruktioner. Det gick bra och resultatet såg bra ut. Den enda oklarheten var vid steg 13 där hon inte visste om det skulle vara smält smör eller ej. Detta var ett återkommande problem även för de andra testpersonerna. Hon förutspådde att om man är dålig på engelska så skulle det nog vara svårt att följa instruktionerna, men hon tyckte själv att språket i instruktionerna var på grundläggande nivå.

Person B sade sig också vara ovan vid att baka, men klarade det mindre bra. Jag var tvungen att ingripa i steg två då han höll på att ta en och en halv matsked bakpulver istället för mjöl. Han bad då om att repetera steget varpå han tog mjölet men bytte till en tesked. Jag ingrep igen och förklarade att tablespoon betyder matsked, varpå han tog en vanlig sked, inte ett matskedsmått. Jag valde att inte rätta det då det inte är helt tokigt, bara oprecist.

Anledningen till att jag ingrep trots att jag hade planerat att inte göra det var för att testet syfte var att testa röstinstruktioner, inte engelskkunskaper. Detta var uppenbart språkproblem.

Efter mina ingrepp gick det bra. Han tog dock mycket mer vaniljpulver än vad jag och de andra testpersonerna gjorde vid instruktionen "a splash of vanilla extract".

Person B föredrog textinstruktioner före röstinstruktioner, men tyckte att det hade fungerat bra ändå. Han tyckte att pappersrecept lät honom ta det lugnare och gav mer kontroll.

Person C var mer van i köket, mer när det gällde matlagning än bakning dock. Han valde att göra två istället för en muffins med mängden smet han hade, vilket gav ett bättre resultat. Han tyckte att röstinstruktionerna fungerade helt okej och var varken för eller emot metoden. De två invändningar han hade var att rösten var lite för datoriserad och att han inte kunde förbereda sig eftersom man inte kan se det förväntade slutresultatet i förväg.

Alla hade problem andra gången smöret kom in i bilden, steg 13 i instruktionerna. Första gången skulle det smältas nämndes det specifikt i instruktionerna, i steg 7. Vid steg 13 skulle det inte smältas, därför sa rösten inget om det. Detta skapade förvirring och tyder på att man inte bara kan säga vad som ska göras i alla lägen. Om en ingrediens ska användas på ett nytt sätt så måste det förtydligas.

Testerna påvisade även att måttenheter som är subjektiva bör undvikas. De kan ge väldigt olika resultat, som i fallet med vaniljpulvret. Relaterat är oprecisa instruktioner som vad "medium heat" innebär för tillagningen. Mikrovågsugnar följer inte några standardiserade effektnivåer. Vad "medium heat" syftar på kan skilja flera hundra watt mellan olika modeller. Detta skapade tveksamheter hos alla testpersoner. Dock lär detta påverka alla former av instruktioner i lika stor grad.

Dessa tre test är inte mycket för att dra slutsatser men det belyser intressanta mönster (Saffer, 2010) som senare även gick att finna i de andra testerna. Oavsett lyckades alla med sina muffins när vissa språkbekymmer klarats upp. Det tycks finnas ett samband mellan hur bra engelskkunskaper deltagarna har och hur lätt det var att följa instruktionerna. Dock finns det fler variabler som kan påverka resultatet, till exempel hur bekväm testpersonen är i köket.

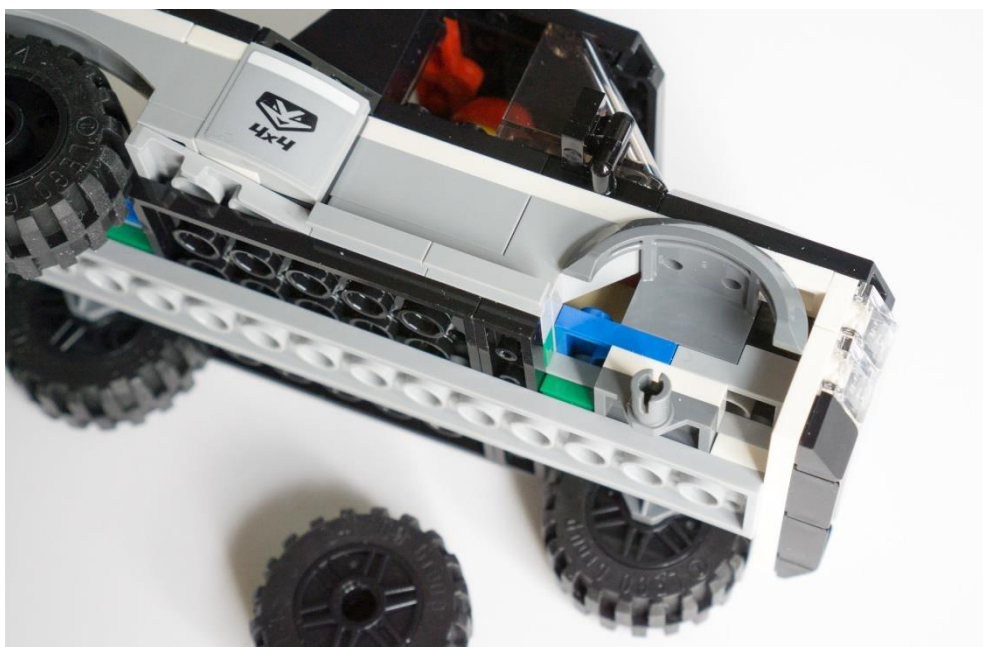
## 6.3 Prototyp 2 (Lego)

Bakningstestet påvisade färre svagheter än väntat. Det var svårt att bedöma resultatet då det inte går att objektivt utvärdera en muffins. Jag behövde finna ett mer utmanande test som tydligare skulle visa resultatet och som inte var lika förlåtande för misstag.

Valet föll på Lego. Med Lego är det viktigt att alla instruktioner följs korrekt. Placeras en bit fel så kan det leda till stora konsekvenser för efterkommande bitars placering. Lego låter sig även monteras och plockas isär utan märkbart slitage, vilket gör testprocessen effektivare.

Lego har även haft lång tid på sig att utveckla sina instruktioner, och de är skapade för att fungera från låg ålder. Det gör deras instruktioner till en väldigt bra utgångspunkt för röstinstruktionerna. De saknar dock beskrivningar helt då de är bildbaserade. Det enda som kan räknas som text i en legomanual är siffror som markerar ett antal eller numrerar stegen.

Röstinstruktionerna byggs därför på termer från Boerger och Henleys (1999) testdeltagares naturliga språk som uppstod under deras tester. När man bygger en legobyggsats enligt deras instruktioner blir det även uppenbart att skaparna har haft instruktionerna i åtanke hela vägen. Bitar som inte syns i den färdiga konstruktionen är ofta i kontrasterande färger som inte passar med det estetiska färgtemat. Min tolkning är att detta är ämnat att förtydliga instruktionerna. Det ger även mig fler sätt att beskriva varje bit.



*Bild 5: Här syns några av de olika färgerna som gömts inuti som vanligtvis döljs av hjulet.*

Alla testerna gjordes på samma legofyrhjuling, en liten byggsats på 19 bitar. Fyrhjulingen kom från en större byggsats som valdes för att den innehöll flera fordon i olika storlekar. Jag tog tiden på varje del när jag byggde dem enligt de medföljande instruktionerna första gången. Valet baserades på de tiderna. Jag ville hålla det kort för att lättare rekrytera deltagare till testerna.



*Bild 6: Fyrhjulingen som byggdes i testerna.*

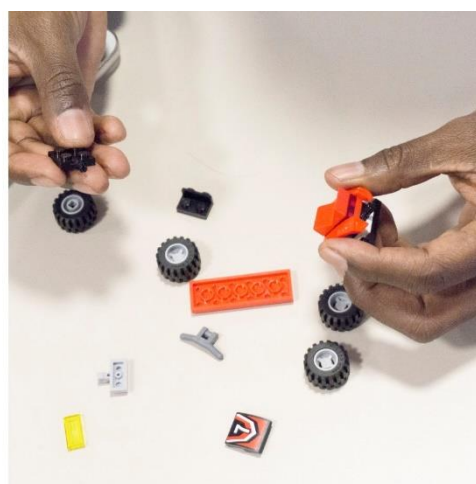
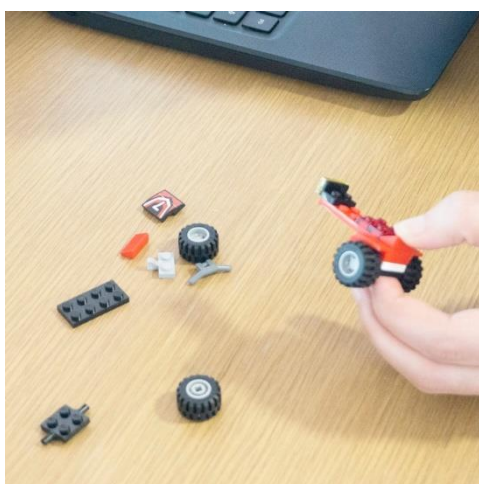
Instruktionerna för prototypen baserades på de bildinstruktioner som följde med byggsatsen. Jag bytte till en annan engelsk röst än den jag använde i bakningstestet, denna gången med brittisk dialekt eftersom den amerikanska jag hade använt i bakningstestet inte längre var tillgänglig. Båda dialekterna var del av samma tjänst. Även för dessa instruktioner utgick jag från Wares (2008) gräns på tre bitar information i arbetsminnet. Den var dock mycket svårare att hålla sig till i detta fallet. I stegen där det inte var möjligt valde jag att gå över tre-bitars-gränsen för att uppnå tillräckligt mycket information för att placera en bit per steg. Jag undvek ord som inte tillför något viktigt. Att beskriva legobitar med antalet ploppar i bredd och djup verkar falla naturligt. 11 av 16 deltagare i Boerger och Henleys (1999) tester antog denna taktik när de skulle kommunicera en helt bildbaserad lego-instruktionsmanual via endast tal. Detta mönster uppstod inte bara vidspritt utan även väldigt tidigt i interaktionen mellan testdeltagare. Ordningen på måtten baserade jag på hur engelsktalande utifrån min erfarenhet refererar till träreglar och liknande; kortsidan först.

### Instruktionerna:

1. We are building a small four-wheeler.
2. Grab a black 2 by 4 plate.
3. Put a white 2 by 2 on the middle of it.
4. Black wheelbases on each side of the white piece.
5. Cover with red 2 by 6 plate.
6. Red translucent 1 by 2 on the edge.
7. On top of it, a red 1 by 2 wedge with the slope pointing into the vehicle.
8. Black 1 by 2 piece with a 90-degree angle on the opposite side, with the mount points facing out.
9. Smooth yellow translucent 1 by 2 on the side of the angle piece.
10. Red translucent 1 by 2 behind the angle piece on the red base.
11. Red 1 by 2 wedge next to it with the slope pointing backwards.
12. Red piece with wheel arches on top of it and the angle piece.
13. Grey 1 by 2 with a claw on top of that, claw facing the rider.
14. Sloped black piece with a 7 on top of that.
15. Handlebars in the claw.
16. Attach the wheels.

## 6.4 Resultat av prototyp 2

Deltagare	6
Röst	Syntetisk
Språk	Engelska
Antal som lyckades	0



*Bild 7, 8: Två deltagare under sina respektive test. Vänster: 2 av 9 placerade bitar är på rätt plats. Höger: 1 av 8 placerade bitar är på rätt plats.*

Totalt deltog 6 personer i testet av prototyp 2. Generellt gick det snett väldigt tidigt. Flertalet blandade ihop bitarna mellan steg 4 och 2. De tog också dubbelt av steg 6, så att de hade en röd genomskinlig bit på båda sidorna av basplattan. Detta medförde att efterföljande bitar hamnade fel då beskrivningen ofta är baserad på tidigare placerade bitar.

Alla nämnde att det var svårt att förstå instruktionerna, både med referenser till uttalet och språket.

Person nummer 3 skiljde sig från de andra i test 2. Han var utbytesstudent och pratade engelska, men inte som modersmål. Han hade aldrig hållit i en legobit förut och visste inte hur Lego fungerar. Efter att han suttit fast en stund på steg tre, där man sätter en bit på en annan för första gången, gick jag emot regeln att inte ingripa och förklarade hur legobitar låser fast i varandra. Detta hjälpte lite men efter att ha kämpat i 25 minuter gav han upp. Detta tyder på att förkunskaper är mycket viktigare än jag trott. Detta syntes inte i bakningstestet. Tyvärr hade han inte tid att prova att bygga den efter Legos bildinstruktioner.

I slutändan klarade inte en enda person att bygga klart fyrhjulingen. Ofta blev det mycket fel. Utfallet belyser problemområden att göra ytterligare iterationer av prototypen på, språk, förkunskaper, och vilka följder ett fel i tidigt stadie har för slutresultatet.

## 6.5 Prototyp 3 (Lego)

Prototyp 3 är en utveckling av prototyp 2. Eftersom deltagarna i både prototyp 1 och 2 (bakningstestet samt det första legotestet) hade klagat på språket valde jag att även skapa en version på svenska. Jag ändrade också specifika instruktioner som hade varit svåra att förstå för deltagarna av prototyp 2 (instruktion 4, 8, 9, 11, och 12). Ordningen förblev som enligt bildinstruktionerna. Mot bakgrund av deltagarnas svårigheter att klara uppgiften i prototyp 2 dubbelkollade jag att benämningarna följer vedertagna termer Lego-hobbyister använder, hämtade från TheBrickBlogger (2010), vilket de gjorde.

Engelska instruktioner:

1. We are building a small four-wheeler.
2. Grab a black 2 by 4 plate.
3. Put a white 2 by 2 on the middle of it.
4. Black wheelbases on both sides, next to the white piece.
5. Cover with red 2 by 6 plate.
6. Red translucent 1 by 2 on the edge.
7. On top of it, a red 1 by 2 wedge with the slope pointing into the vehicle.
8. Black 1 by 2 angle bracket on the opposite side, with the studs facing out.

9. Smooth yellow translucent 1 by 2 on the side of the angle piece, forming a front light.
10. Red translucent 1 by 2 behind the angle piece on the red base.
11. Red 1 by 2 wedge next to it with the slope pointing backwards, leaving a 2 by 2 gap for the rider.
12. Red piece with wheel arches on top of the red translucent piece and the angle bracket.
13. Gray 1 by 2 with a claw on top of that, claw facing the rider.
14. Sloped black piece with a 7 on top of that.
15. Handlebars in the claw.
16. Attach the wheels.

Deltagarna fick själva välja vilken version de ville ha, men blev varnade att uttalet på den svenska rösten var mycket sämre än den engelska.

Jag översatte så likt som möjligt, för att testerna skulle vara jämförbara. Det var dock svårare att vara lika kortfattad på svenska. Notera att i denna texten har jag skrivit siffrorna på två olika sätt för att framkalla en naturligare mening när den läses upp av den syntetiska rösten. Till exempel, *två* då den annars endast uttalar siffran 2 som *två*.

Svenska instruktioner:

1. Vi ska bygga en liten fyrhjuling.
2. Ta en svart 2 gånger 4 platta.
3. Sätt en vit 2 gånger tvåa mitt på den.
4. Svarta hjulbaser framför och bakom den vita biten.
5. Täck över med röd 2 gånger 6 platta.
6. Röd genomskinlig 1 gånger tvåa på kanten.
7. Ovanpå den, en röd 1 gånger två stor kil, sluttningen pekar inåt fordonet.
8. Svart 1 gånger tvåa vinkel på motsatt sida, med prickarna utåt.
9. Slät gul genomskinlig på sidan av vinkelbiten, blir en framlampa.
10. Röd genomskinlig 1 gånger tvåa bakom vinkelbiten på den röda basen.
11. Röd 1 gånger två stor kil bakom den förra biten med sluttningen bakåt.
12. Röd bit med hjulbågar ovanpå den röda genomskinliga och vinkelbiten.
13. Grå 1 gånger tvåa med klo ovanpå föregående bit, med klon pekandes bakåt.
14. Svart 2 gånger tvåa med en sju på ovanpå föregående.
15. Styret sätter du i klon.
16. Sätt på hjulen.

## 6.6 Resultat av prototyp 3

Deltagare	8
Röst	Syntetisk
Språk	Svenska (7) eller engelska (1)
Antal som lyckades	1 (3 kom nära)



*Bild 9, 10: Två deltagare under sina respektive test. Vänster: 7 av 10 placerade bitar är på rätt plats. Höger: 2 av 5 placerade bitar är på rätt plats.*

I den andra legotestet (prototyp 3) deltog 8 personer, varav endast en valde den engelska versionen.

Gemensamt för alla i detta test var att det gick fel på steg 6. De placerar nästan alltid en röd genomskinlig bit på varje sida, trots att instruktionen inte nämner två. Det tyder på att de skapar egna grupperingar av bitarna som inte finns i instruktionerna. Det kan bero på flera saker. Dessa är mina teorier:

- De vet inte vad biten ska föreställa (bakljus).
- Tidigare i instruktionerna behandlas ett par.
- Symmetri tilltalar.

De två sistnämnda punkterna styrks av Roediger och McDermotts (2000) teori om vår förmåga att dra slutsatser av liknande data. Här leder testpersonernas slutledningsförmåga dem i fel riktning.

Efter att den biten hamnat fel går det för det mesta lavinartat fel. Om inte instruktionerna är kristallklara så fungerar det inte.

I detta test var det tre personer som kom nära att lyckas med uppgiften, med endast en eller två bitar fel och ett resultat som liknande facit. Det var även en som klarade det felfritt. Den sistnämnda och en av dem som kom nära, var de första två som utnyttjade att man kunde backa tillbaka i instruktionerna. Den som löste uppgiften backade tillbaka hela vägen till start, flera gånger. Han var bestämd att klara det då han ansåg sig bra på logiskt tänkande.

Det största klagomålet i samtalet efteråt var hur svår den syntetiska rösten var att förstå, på grund av felaktig betoning. Några tyckte att placeringen av bitarna som beskrevs i relation till föregående bit var svår att förstå.

Generellt lyckades testpersonerna så pass mycket bättre i detta testet än föregående att språket måste ha varit den största faktorn till förbättringen.

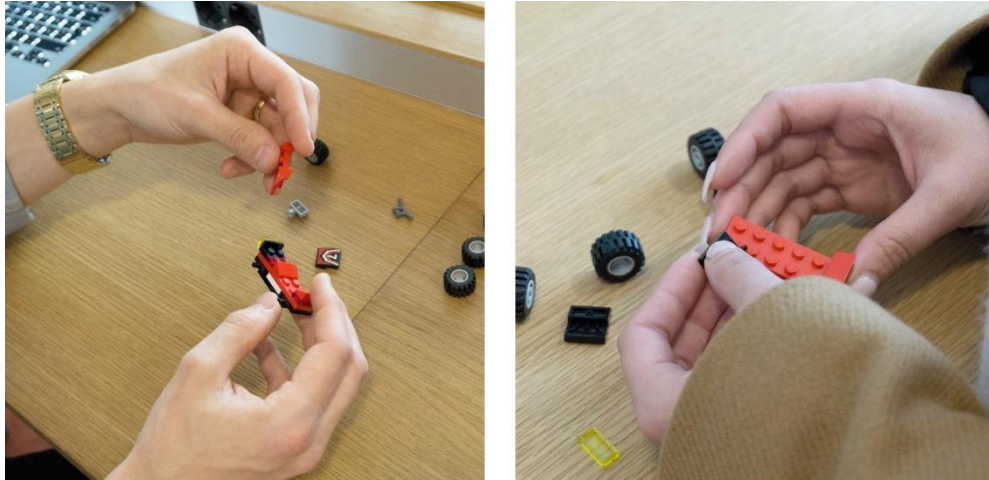
## 6.7 Prototyp 4 (Lego)

I fjärde iterationen av prototypen slutade jag tillhandahålla instruktioner på engelska och höll mig enbart till svenska. De tidigare testerna visade att om instruktionerna inte var på ditt modersmål så var det mycket svårare att följa dem. Det knyter an till Baddeleys (2010) beskrivning av hur korttidsminnet tar stöd av långtidsminnet, exempelvis med bättre minne på det egna modersmålet.

Den datorgenererade rösten ersattes av en mänsklig. Här läste en välartikulerad bekant in instruktionerna istället. Instruktionernas omfattning och ordalydelse ändrades inte från prototyp tre.

## 6.8 Resultat av prototyp 4

Deltagare	5
Röst	Verklig
Språk	Svenska
Antal som lyckades	3 (1 kom nära)



*Bild 11, 12: Två deltagare under sina respektive test. Vänster: 10 av 10 placerade bitar är på rätt plats. Höger: 4 av 5 placerade bitar är på rätt plats.*

I detta test deltog 5 personer, varav 3 klarade det, och en hade ett minimalt fel. Den som inte klarade det hade inte lekt med Lego som liten, och tyckte inte om pilliga saker. De andra fyra har alla byggt Lego i sin barndom. Detta styrker resultatet från första testet som visade på hur viktigt det är med förkunskaper.

Nu när det gick så pass bra för deltagarna var det flera som uttryckte att det var skönt att ha händerna fria att arbeta med byggsatsen, att inte behöva referera till instruktionerna hela tiden.

Förbättringen av andelen som klarade det i detta testet tyder på att rösten spelar stor roll.

## 6.9 Prototyp 5 (Lego)

I denna iteration ändrade jag om ordningen på vissa delar baserat på vad tidigare deltagare hade sträckt sig efter redan innan de hört nästa instruktion. Samma person som läste upp instruktionerna i prototyp 4 läste upp dessa reviderade instruktioner.

Fokus för ändringarna i denna version låg på att beskriva deras syfte tydligare.

### Instruktioner:

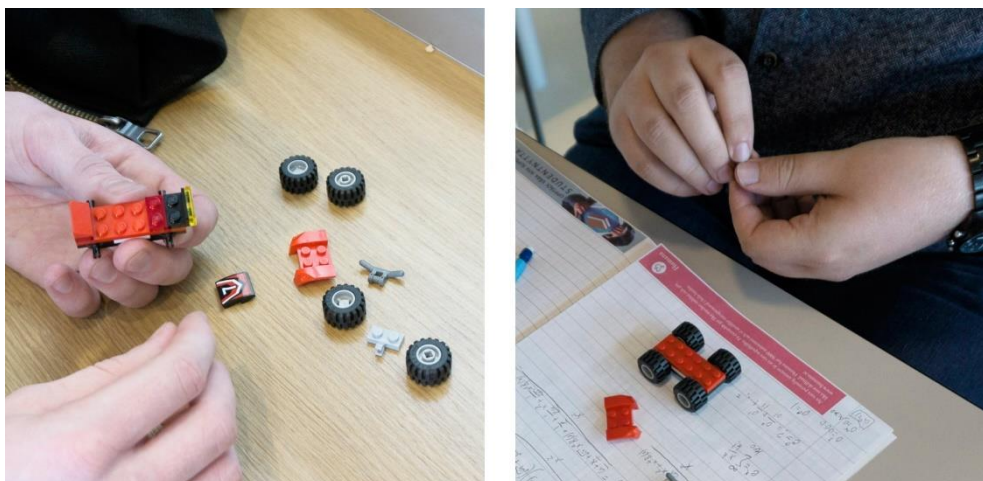
1. Vi ska bygga en liten fyrhjuling.
2. Ta en svart 2 gånger 4 platta.
3. sätt en vit 2 gånger tvåa mitt på den.
4. Montera hjulen på de svarta hjulbaserna.
5. Hjulen framför och bakom den vita biten.
6. Täck över med röd 2 gånger 6 platta.
7. Röd genomskinlig 1 gånger tvåa på kanten blir en baklampa.
8. Ovanpå den, en röd 1 gånger 2 kil, slutningen pekar inåt fordonet.
9. Svart 1 gånger 2 vinkel på motsatt sida, med prickarna utåt.
10. Slät gul genomskinlig på sidan av vinkelbiten blir en framlampa.
11. Röd genomskinlig 1 gånger tvåa bakom vinkelbiten, på den röda basen.
12. Röd 1 gånger 2 kil bredvid med slutningen bakåt.
13. Röd bit med hjulbågar längst fram med slutningen framåt.
14. Grå 1 gånger tvåa med klo ovanpå den, ett steg från fronten, med klon pekandes bakåt.
15. Svart 2 gånger tvåa med en sjua täcker den gråa.
16. Styret sätter du i klon.
17. Färdig.

Den avgörande skillnaden i formulering av denna version är att man sätter hjulen på hjulbaserna direkt, innan de monteras på resten av fordonet. Detta gör att man har färre alternativ för var man kan placera hjulkonstruktionerna, och därmed större chans att det blir rätt. Det lägger också en tydligare grund för kommande bitar.

Den röda, genomskinliga biten från steg 7 har genom testerna varit den absolut mest förvirrande biten. Vid det steget är fordonet symmetriskt och det finns två sådana bitar, varpå många fortsatte på det symmetriska spåret och satte dit en på varje sida. Detta är än en gång fenomenet Roediger och McDermott (2000) iakttar apropå mönsterbaserad slutledningsförmåga. Detta problemet hade aldrig uppstått i bildinstruktioner då det går att se vad målet för biten är. För att hjälpa deltagaren att bilda en vision av vad varje bit föreställer och vad som är fram och bak på fordonet beskrivs nu den röda genomskinliga biten som baklampa.

## 6.10 Resultat av prototyp 5

Deltagare	6
Röst	Verklig
Språk	Svenska
Antal som lyckades	4



*Bild 13, 14: Två deltagare under sina respektive test. Vänster: 9 av 9 placerade bitar är på rätt plats. Höger: 8 av 8 placerade bitar är på rätt plats.*

6 personer deltog i testet. Detta gick till stor del lika bra som testet av prototyp 4. Någon gjorde fel på steg 12 vilket resulterade i att resten blev tokigt. En norsk testperson fick backa tillbaka för att få den genomskinliga röda biten som blev fel så ofta rätt. Detta var antagligen en följd av att instruktionerna inte var på personens modersmål.

Deltagare 4 var extra intressant här. Personen var relativt ny på svenska och hade inte lekt med Lego som barn. Likt deltagaren i test 1 med liknande förutsättningar kom testpersonen ingenvart. Detta belyser än en gång hur viktigt språket och förkunskaperna i den här typen av instruktioner är.

Det är inte tydligt om ändringarna i manuset för detta test gjort skillnad då deltagarna lyckades mycket bra även i testet av prototyp 4. Däremot styrker de fortsatt framgångsrika resultaten i denna prototyp slutsatserna från prototyp 4 om att rösten har stor betydelse, då det är samma uppläsare.

Trenden genom alla testerna har varit ständig förbättring, framförallt genom matchning mot användarens språk, både i uttal och modersmål. I denna sista

prototyp klarade samtliga deltagare med skandinaviskt modersmål av uppgiften.

## 7 Diskussion

Bild och text är de vedertagna sätten att leverera instruktioner. Röst har fördelarna att du inte behöver slita blicken från det du arbetar med och inte behöver hantera manualen med händerna. Detta är extra användbart i situationer där du inte har rena händer, till exempel vid matlagning, byte av cykelkedja, och liknande. Det är även användbart i situationer där du inte kan släppa det du arbetar med utan att det brakar samman, eller på platser där vädret försvårar användningen av en textbaserad manual. Röstinstruktioner fungerar i vissa lägen, med ett antal begränsningar och förhållningsregler.

### 7.1 Processens komplexitet

Testerna av instruktioner i ljudform visade skillnad i resultat beroende på hur komplex uppgiften var. Deltagarna vars uppgift var att baka muffins klarade det bättre än de som fick den mer komplexa legobyggsatsen. Detta är i enlighet med Dalton et al.:s (2013) resultat. Anledningen till att legoprototypen var så pass mycket mer komplex är att komponenterna i mycket större grad var beroende av varandra. Ett fel i bakningen hade inte lika stor tendens att orsaka ytterligare fel. En felplacerad bit i legobyggsatser resulterade i en snöbollseffekt, då felet snabbt växte med varje påföljande steg.

Ytterligare grund till komplexiteten i legoprototypen var det väldigt specifika slutresultatet. Med receptet hade deltagarna chansen att visualisera slutresultatet då de visste att det var en muffins de bakade, ett relativt generiskt objekt. Allt de visste om legobyggsatsen var att det skulle bli en fyrhjuling. Den är mer detaljrik och därför svårare att visualisera i legoformat.

### 7.2 Språk

Språk spelar stor roll i hur väl användaren förstår instruktionerna. Det gick avsevärt mycket sämre i de komplexa legotesterna där instruktionerna gavs på deltagarnas andraspråk engelska, än vad det gjorde när de gavs på modersmålet svenska. Testerna motbevisade således antagandet att engelska inte skulle utgöra en språkbarriär för universitetsstudenter i Sverige, som statistiskt sett har mycket goda färdigheter inom engelska. (Education First, 2017) De som inte har engelska som förstaspråk erfar markanta svårigheter i hörförståelsen av engelska syntetiska röster jämfört med de som har engelska

som modersmål. (Venkatagiri, 2005) Svårigheter uppstod även under testerna med svenskt tal där svenska inte var deltagarnas modersmål.

### 7.3 Diktion: Uttal, frasering, och takt

Deltagarna i recepttesterna hade inga större problem med uttalet på den datorgenererade rösten. Nye et al. (1975) konstaterade en väsentlig skillnad mellan mänsklig och syntetisk röst, till den mänskliga röstens fördel. Legotesterna i denna studie leder till samma slutsats. Svårare instruktioner kräver alltså bättre uttal och frasering i instruktionerna.

En del av talhastigheten är takten med vilken varje nytt meddelande levereras. Detta påverkar uppfattningsförmågan hos lyssnaren (Smith & Shaffer, 1995). När instruktionen kommer, spelar också in i mottagarens mottaglighet (Wu et al., 2009). I denna studie kunde deltagarna avvakta tills de var redo för nästa instruktion, vilket de uttryckte uppskattning över. De hade också möjlighet att repetera tidigare delar av instruktionen när de så behövde, vilket förbättrade förutsättningarna att lyckas med uppgiften.

### 7.4 Förkunskaper

Roediger och McDermotts (2000) slutsatser förutsätter att det krävs viss förkunskap för att kunna associera baserat på given information. I de komplexa testerna deltog ett fåtal som aldrig hade lekt med Lego förut. De hade avsevärt mycket större problem med uppgiften jämfört med alla andra kategorier. Det visar att viss erfarenhet krävs för att röstinstruktioner ska fungera. De som hade lekt mycket med Lego som barn visade också mindre osäkerhet under testerna än de som inte hade det.

### 7.5 Sammanfattning

Alla dessa tester har varit avskalade och utförda av personer med hög utbildning. De representerar vad man kan förvänta sig i bästa fall. Slutsatsen av studien är att röstinstruktioner kan stå på egna ben i enkla uppgifter och är användbara i synnerhet där händerna är upptagna, såsom vid matlagning.

Begränsande faktorer med röstinstruktioner är att de är svåra att använda effektivt i uppgifter som kräver precision, exempelvis montering av byggsatser. Komplexiteten av en uppgift sjunker dock avsevärt vid goda förkunskaper.

Förhållningsregler att beakta är kopplade till språkbruket. Om instruktionerna ges på lyssnarens modersmål kan man förvänta sig färre missförstådda instruktioner, även när lyssnarens andraspråkskunskaper är goda. Slutligen är uttal och frasering också kopplat till hur väl lyssnaren förstår instruktionerna. Testpersonerna i denna studie hade lättare att förstå den naturliga mänskliga rösten än de syntetiska rösterna, som alla hade onaturlig takt och betoning.

## 7.6 Vidareutveckling

Vidareutveckling av projektet skulle kunna ske på flera fronter. Ett område är att rikta in sig på olika handikapp. Personer med dyslexi eller koncentrationssvårigheter skulle troligtvis kunna bli hjälpta av denna teknik. Det finns sannolikt rum för att hjälpa blinda också men man bör vara försiktig med att leta efter ett problem till en lösning istället för tvärtom. Professionellt bruk är även intressant baserat på upptäckten att förkunskaper spelar stor roll även i enklare uppgifter. Röstinstruktioner är potentiellt mer tillgängliga för yrkesgrupper så som bilmekaniker, där användaren är erfaren men inte vet precis hur varje bilmodell är konstruerad.

En annan riktning man kan ta är att expandera appens talförmåga och låta den svara på frågor om varje steg. Till exempel genom att kunna fråga vad "transparent" betyder och få svaret "genomskinlig". Det skulle kunna finnas flera användargenererade förklaringar att välja mellan för att anpassa instruktionerna till ditt sätt att tala.

Sist men inte minst är att kombinera röstinstruktioner med andra medier. Vissa testdeltagare uttryckte en önskan att se ingredienslistan i skriven form innan man börjar i bakningstesterna, något som känns som en självklarhet för en recept-app med talfunktionalitet. Ett behov de inte uttryckte men som blev uppenbart i observationerna är att få en förklaring av saker, till exempel vad ett matskedsmått är eller hur ett Lego-hjulhus ser ut. Detta är saker man skulle man kunna visa på skärm tillsammans med röstförklaringar.

## 8 Referenser

- Baddeley, A. (2010) Working memory. *Current Biology*, 20(4) 136-140. doi:10.1016/j.cub.2009.12.014
- Black, J. B., Carroll, J. M., & McGuigan, S. M. (1986). What kind of minimal instruction manual is the most effective. *SIGCHI Bulletin*, 18(4), 159-162. doi:10.1145/1165387.275623
- Boerger, M. A., Henley, T. B. (1999). The Use of Analogy in Giving Instructions. *The Psychological Record*, 49(2), 193-209. doi:10.1007/BF03395316
- Buxton, B. (2007). *Sketching user experiences*. San Francisco, CA: Diane Cerra.
- Crabtree, M., Mirenda, P., & Beukelman, D. (1990). Age and gender preferences for synthetic and natural speech. *Augmentative and Alternative Communication*, 6(4), 256-261. doi:10.1080/07434619012331275544
- Dalton, P., Agarwal, P., Fraenkel, N., Baichoo, J., & Masry A. (2013). Driving with navigational instructions: Investigating user behaviour and performance. *Accident Analysis & Prevention*, 50, 298-303. doi:10.1016/j.aap.2012.05.002
- Lego. (2018, 2 Maj). Lego® Duplo® launches interactive experience on Amazon Alexa. *Lego*. Hämtad från <https://www.lego.com/en-us/aboutus/news-room/2018/may/lego-duplo-alexa>
- Nye, P. W., Ingemann, F., & Donald, L. (1975). Synthetic speech comprehension: a comparison of listener performances with and preferences among different speech forms. *Haskins Laboratories: Status report on speech perception SR-41*, 117-126.
- Roediger, H. L., & McDermott, K. B. (2000). Tricks of Memory. *Current Directions in Psychological Science*, 9(4), 123-127. doi:10.1111/1467-8721.00075
- Saffer, D. (2010). *Designing for interaction* (andra upplagan). Berkeley, CA: New Riders
- Skolverket. (2018). *Läroplan för hem- och konsumentkunskap LGR11*. Hämtad från <https://www.skolverket.se/laroplaner-amnen-och-kurser/grundskoleutbildning/grundskola/hem-och-konsumentkunskap>

Smith, S. M., & Shaffer, D. R. (1995). Speed of Speech and Persuasion: Evidence for Multiple Effects. *Personality and Social Psychology Bulletin*, 21(10), 1051–1060. doi:10.1177/01461672952110006

Sneath, T. (2017, 15 Maj). *Microsoft Build 2017 conference* [video fil]. Hämtad från <https://www.youtube.com/watch?v=iWXebEeGwn0&t=175s>

Swah. (2015, 9 Mars). *Microwave chocolate cupcakes for two* [Blogginlägg]. Hämtad från <http://loveswah.com/2015/03/microwave-chocolate-cupcakes-for-two/>

Swedish research council (2017). Good Research practice [Elektroniskt resurs]. Hämtad 2018-01-18 från <https://publikationer.vr.se/en/product/good-research-practice/>

TheBrickBlogger. (2010, 30 November). Learn to speak Lego – Basic terms [Blogginlägg]. <http://thebrickblogger.com/2010/11/lego-disctionary-basic-term/>

Venkatagiri, H. S. (2004). Segmental Intelligibility of Three Text-to-Speech Synthesis Methods in Reverberant Environments. *Augmentative and Alternative Communication*, 20(3), 150-163. doi:10.1080/07434610410001699726

Venkatagiri, H. S. (2005) Phoneme Intelligibility of Four Text-to-Speech Products to Nonnative Speakers of English in Noise. *International Journal of Speech Technology*, 8(4), 313-321. doi:10.1007/s10772-006-0449-1

Ware, C. (2008). *Visual thinking for design*. Burlington, MA, USA: Elsevier Inc.

Wu, C. F., Huang, W. F., & Wu, T. C. (2009). A study on the design of voice navigation of car navigation system. *In International Conference on Human-Computer Interaction*, 141-150. Springer, Berlin, Heidelberg.