

inf2260 Team Old School - Rapport

28. november 2011

Innhold

1	Introduksjon	3
1.1	Formålet med testingen	3
2	Deltakere	3
3	Anvendte metoder	4
3.1	Experimental research	4
3.1.1	Hypoteser, variabler og vilkår	4
3.1.2	Prototyping: Low fidelity - High fidelity/Wizard of Oz	5
3.2	Evaluering av Low Fidelity	5
3.3	Intern Hi-fi-evaluering førte til endring	6
3.4	Designstruktur - Basic design	6
3.5	Within-group	6
3.6	Feltstudie, brukbarhetstest, spørreskjema	7
4	Prototypeverktøy	8
4.1	Verktøyet: Android-basert	8
5	Funn under pilottesting	8
5.1	Affordance	8
5.2	Motorikk	9
6	Innhentede data	10
6.1	Oppgavene	10
6.1.1	Tidbruk	10
6.1.2	Feilantall	10
6.2	Spørreskjemaene	10
6.3	Årsaker til bias	11

6.3.1	Forutsette årsaker	11
6.3.2	Faktisk opplevde årsaker under testing	11
7	Diskusjon	12
7.1	Pålitelighet og gyldighet	12
7.1.1	Pålitelighet	12
7.1.2	Gyldighet	12
7.2	Prototypediskusjon	14
8	Pilottesting	14
8.1	Pilottesting vs. virkeligheten	14
8.2	Så mye pilottesting som mulig	15
9	Analyse	15
9.1	Hypotesene	15
9.2	Brukbarhet	15
10	Konklusjon	16
11	Referanser	16
12	Samtykkeskjema - konfidensialitet	17

1 Introduksjon

Team Old Schools tema er «Design for eldre». Vår løsning er ment å hjelpe målgruppen med å mestre hverdagslig tv-bruk, som i varierende grad er en utfordring for mange. Flere av gruppens medlemmer kjenner personlig til behovene en del eldre har i denne sammenheng, og vår ide kommer direkte fra dette: vi har laget en prototype på en fjernkontroll med berørings skjerm. Visjonen er at denne løsningen gjør to ting: bidra til å bringe målgruppen med i den teknologiske utviklingen, og bruke nettopp dagens teknologi til å legge til rette for målgruppens fortsatte mestring av et dagligdags gjøremål. Rent konkret skal brukeren gå fra utstyr som er av, til riktig kanal med godt volum, uten hodebry og fallgruver, og med færrest mulig nødvendige handlinger. Brukeren skal også kunne endre tv-innstillinger og manipulere favoritter under disse omstendigheter. Vi har testet prototypen med målgruppa, både tidlig og sent i prosessen. I denne rapporten beskriver vi prosessen og metodene vi har benyttet. Deretter analyserer vi funnene våre.

Vi har en god del materiale vi ikke rekker å diskutere eller sågar plassere i rapporten. De legges i en egen undermappe ved siden av rapportfilen i den endelige leveringen. Det er mye der det hadde vært interessant å jobbe videre med, og ikke minst kunne presentere i rapporten, men dessverre er det rett og slett ikke tid.

1.1 Formålet med testingen

Vi har laget tre oppgaver deltakerne skal utføre. Gjennom de to første ønsket vi å sammenlikne en typisk analog fjernkontroll og vår digitale fjernkontroll. Vi har formulert dette i hypotesesettet, se nedenfor. Med den tredje oppgaven gir vi deltakeren grunnlag for å besvare to spørreskjemaer. Disse skjemaene skal gi oss en indikasjon på hvordan deltakeren opplever vår digitale fjernkontrolls brukbarhet og graden av tilfredshet den kan gi. Oppgave tre introduserer også funksjonalitet som fordrer en høyere mestringsgrad. Denne delen av oppgaven kan gi oss en indikasjon på hvor god affordance vi har gitt den funksjonaliteten (Sharp, Rogers, Preece, s. 33). Vi vil også få en pekepinn på om den typen funksjonalitet er noe målgruppen vår kan tenke seg å ta i bruk. Det er snakk om å slide grensesnitt-elementer fram og tilbake på en berørings skjerm.

2 Deltakere

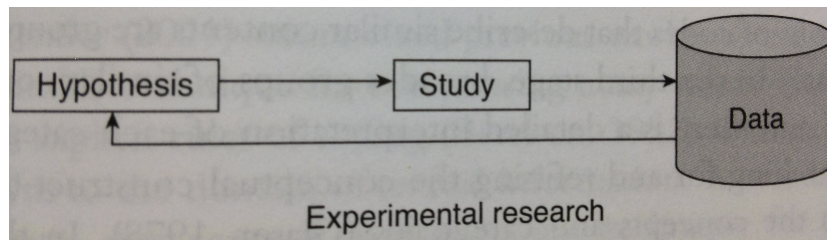
- Gunnar Gahnstrøm - gunnarga

- Karolina Matyja - karolama
- Kristine Larsen - kristwla
- Jens Peter Nilssen - jensni

3 Anvendte metoder

3.1 Experimental research

Vi baserer oss ikke på eksisterende data i dette prosjektet. Snarere formulerer vi en hypotese og samler inn data som vi bruker til å bekrefte eller avkrefte den. Streng experimental research er gjerne et ekte eksperiment (www.experiment-resources.com), men vi kommer kort inn på hvorfor vårt eksperiment ikke er ekte.



3.1.1 Hypoteser, variabler og vilkår

- Hypoteser
 - H0: Det er ingen forskjell på brukervennlighet mellom analog fjernkontroll og vår digitale fjernkontroll.
 - H1: Analog fjernkontroll har bedre brukervennlighet enn vår digitale fjernkontroll.
 - H2: Vår digitale fjernkontroll har bedre brukervennlighet enn analoge fjernkontroller.
 - Variabler
 - Uavhengig
 - * Fjernkontrolltypen.
- Vi forestiller oss at de avhengige variablene (se nedenfor) vil få forskjellige målinger mellom de to vilkårene (se nedenfor).

- Avhengige
Deltakeren skal utføre alle trinn i konkrete oppgavelister. Per oppgave måler vi de avhengige variablene
 - * tid
 - * antall feil.
- Kontrollerte variabler
 - * Oppgavelistene for vanlig fjernkontroll og for vår fjernkontroll er identiske. Altså kontrollert innenfor hver oppgave per deltaker og mellom hver deltaker.
 - * Vi ville sørge for at vanlig fjernkontroll var en kontrollert variabel mellom hver deltaker. Mer om dette under Ekstern gyldighet.
 - * I den forbindelse ønsket vi også å sørge for identisk kanal-layout mellom deltakerne, altså at kanalene i oppgavelistene befant seg på samme knapp på den vanlige fjernkontrollen for hver deltaker. Dette forkastet vi raskt, siden å ordne det ville kreve omfattende tidbruk hos deltaker og ikke minst et lokalt inngrep, i form av omrokking av flere deltakeres vanlige kanallayout på deres TV/dekoder.

- Vilkår

1. Typisk analog kontroll
2. Vår digitale kontroll

3.1.2 Prototyping: Low fidelity - High fidelity/Wizard of Oz

Vi startet med papirskisser og lagde skjermbilde-mockups av dem. Senere utviklet vi en nærmest selvstendig prototype, hvis hensikt var å være fullt testbar med deltakerne.

3.2 Evaluering av Low Fidelity

En tidlig brukerbasert test. Vi gjennomførte intervjuer med personer fra målgruppen vår, der skjermbilde-mockup-ene var tema. I vedleggene Spørsmål-Brukerundersøkelse.doc og Resultaterfralow-fidelity.doc er spørsmålene og dataene i stikkordform fra disse intervjuene. Dette er en form for formativ testing, der vi fortsatt hadde muligheten til å endre kurs på bakgrunn av deltakernes innspill og synspunkter. Vi var her mer opptatt av hvordan

grensesnitt-utkastet ble oppfattet enn av å få synspunkter på mulighetene det ga for oppgaveløsning.

Vedlegg gg1.png viser en av prototypene som var med på den formative testen. Vedlegg work-in-progress.png viser prototypen slik den framsto på bakgrunn av dataene fra den formative testen.

3.3 Intern Hi-fi-evaluering førte til endring

Etter at vi som «eksperter» gikk gjennom vår high fidelity-prototype uten primære brukere, så vi at vi gir for få muligheter til ytterligere mestring. Fjernkontrollen presenterte i realiteten brukeren for et begrenset antall kanaler som alltid var synlige på skjermen. Den hadde også samme prinsipp som en ordinær fjernkontroll, nemlig knapper. Vår interne evaluering fortalte oss at det finnes bedre måter å representere ekstrautvalget av kanaler og kategorier på, i vårt eksempel ved bruk av «slide»-funksjonen på en berøringsskjerm. Vedlegg skjermdump-video-slidefunksjon.png er et skjermbilde fra en videonutt som viser slide-funksjonen i bruk.

3.4 Designstruktur - Basic design

Vårt eksperiment er et basic design. Vi har én uavhengig variabel.

3.5 Within-group

- Fordeler
 - Blant vår målgruppe er det noe vrient å få tak i relevante deltakere. Denne metoden krever få deltakere.
 - Vi undersøker forskjellene mellom vilkårene med de samme deltakerne.
- Ulemper
 - Medfører læringseffekt.
 - Krever mer tid per deltaker, som utsettes for testing av alle vilkårene.
 - Økt fare for utmattelse og frustrasjon.

3.6 Feltstudie, brukbarhetstest, spørreskjema

Vi gjennomførte testingen ute i felten. Det vil si hjemme hos deltakerne. Deltakerens naturlige omgivelser for bruk av ordinær fjernkontroll dannet omstendighetene for testing av både den og vår egen prototype. Tidligere i høst gjennomførte vi en formativ brukbarhetstest. Brukbarhetstesten i feltstudien er summativ - prototypen var av så høy fidelitet som vi hadde muligheten til å oppnå. Vi ønsket å stadfeste dens brukbarhet sammenliknet med en ordinær fjernkontroll. Vi ønsket også å få et inntrykk av deltakernes opplevelse av en av grensesnittmulighetene en berøringsskjerm tilbyr, nemlig å slide grafiske elementer.

Den siste øvelsen i forsøkene besto av at deltakerne fylte ut to spørreskjemaer.

- System usability scale. Se vedlegg sustabell.png. (Jeff Sauro).
- Parvise motsetninger. Se fire vedlegg fra og med parviskjema-1.jpg (Toftøy-Andersen, Wold, s. 149).

Skjemaene representerer metoder for måling av brukervennlighet og tilfredsstillelse. Begge begreper er nokså komplekse, derfor er det nødvendig at man tester dem på riktig måte. Ikke minst gjelder det å dele opp i mindre deler som er lettere å identifisere. Vi er spesielt interessert i brukervennlighet, at prototypen er lett å bruke. Derfor teller vi hvor mange feil deltakeren gjør under brukertesting. Med feil menes individuelle handlinger som ikke bringer deltakeren nærmere gjennomføringen av et trinn i oppgavelistene. En kan regne med at det som er lett å bruke gir få feil. Når det gjelder tilfredshet har vi valgt å bruke de to spørreskjemaene nevnt ovenfor. Begge to tar for seg spørsmålet om hvor tilfreds brukeren er når det gjelder å bruke vår prototype. Det er viktig å merke seg at selv om alle tester kunne ha gått bra, så bør man bruke denne typen spørreskjema uansett, for å se hva slags personlig forhold brukeren har fått til prototypen. System usability scale-skjemaet er et universaldokument som kan brukes av de fleste som vil ha en indikasjon på hvor tilfreds deltakeren er for eksempel med et fremtidig produkt. Det er formet på en måte som passer til absolutt alle systemer, slik at det gir høy pålitelighet til resultater vi får. Deltakeren besvarer 10 spørsmål ved å rangere dem etter en Likert-skala (en.wikipedia.org - Likert scale).

Gi poeng fra 1 til 5, hvor 1 er sterkt uenig og 5 er enig. Hvis brukeren ikke forstår spørsmålet eller er usikker på hva h*n skal svare, kan vedkommende skrive 3 poeng, som er midt på skalaen og ikke dytter resultatet i den ene eller den annen retning. Under framstilling av vår egen utgave av

SUS-skjemaet, kom vi i skade for å reversere skalaen, slik at vi fikk speilvendte resultater da vi regnet dem ut. Det tok oss omlag en time å finne ut av det. Resultatene omtales kort under Diskusjon. Skjemaet for parvise motsetninger er litt enklere enn SUS-skjemaet, men det gjør det tydeligere hva som er negativ og hva som er positivt. På dette skjemaet har brukeren mulighet til å velge mellom to motstridende egenskaper. Hvis de er usikre kan de bare krysse av for at de ikke vet.

4 Prototypeverktøy

4.1 Verktøyet: Android-basert

Prototypeverktøyet er nettbrettet Xoom fra produsenten Motorola. (www.motorola.com)

Prototypen er i sin helhet framstilt ved å designe en brukerflate og programmere funksjonalitet i programmeringsspråket java.

All utvikling har foregått i utviklingsverktøyet Eclipse, med installert ekstrafunksjonalitet for å støtte android-programmering. (developer.android.com/)

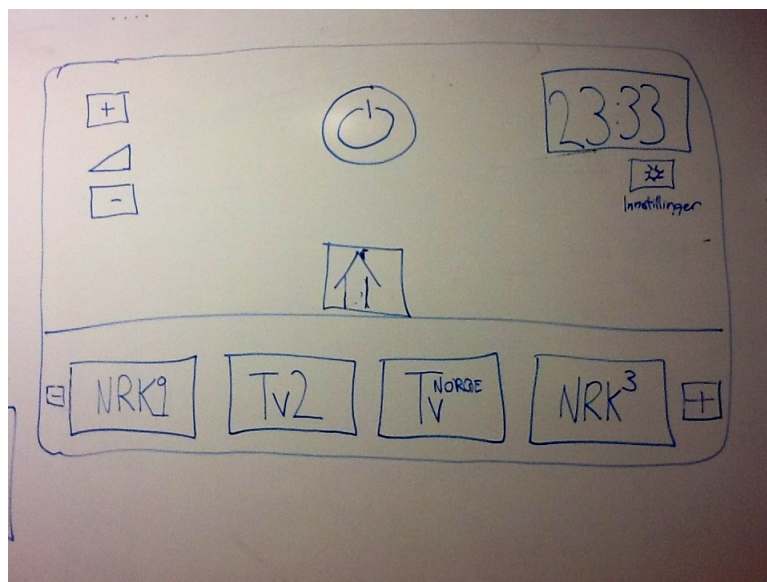
5 Funn under pilottesting

Gruppens medlemmer var selv pilottestere.

5.1 Affordance

Vi hadde lenge en prototype som begrenset brukeren til muligheten til å velge mellom fire kanaler. Vi så at dette var i mot en av hensiktene med prosjektet, som er å øke målgruppens muligheter/sannsynlighet for mestring.

Her følger et lite utdrag fra gruppens arbeidslogg, som illustrerer et vendepunkt når det gjelder dette. Den 3. november la vi følgende bilde og tekst inn i loggen vår:



« Funksjonen nederst fungerer som en slide funksjon, hvor brukeren har større mulighet for valg av kanaler i stedet for at han/ hun er begrenset til noe bestemt på forhånd. Vi mener det vil gi større følelse av mestring i den forstand at brukeren kan få til ting han ikke bare ser men også forstår. Vi mener det vil være også greit å ha en knapp "home" som automatisk flytter slide'en til de fire mest populære og favrite kanaler.»

Vi arbeidet en del med forskjellige metoder for å indikere mulighetene for å slide katalogremsene. Det kan være vanskelig å forstå at favorittremsa og kategori-remsa kan skyves/rulles.

5.2 Motorikk

Gjennom å forsøke å gå i brukerens sko og simulere forskjellige svekkelser, både fysiske og mentale, kunne vi forestille oss noen av utfordringene brukerne kan få med vår prototype. Noen av testene viste oss at en «tilstrekkelig reduksjon» av brukerens motorikk (vi har ikke hatt anledning til å bestemme eller tallfeste hva som er en tilstrekkelig reduksjon, derfor det vage begrepet), kan gi det utslag at handlingen på vår fjernkontroll ikke utføres i det hele tatt eller at det tar lang tid å få til. For eksempel når deltakeren treffer knappen, men ikke fjerner fingeren straks og heller lar fingeren skli av knappen. Knappen er programmert slik at den da ikke oppfatter handlingen som et trykk, men ikke desto mindre gir den visuell tilbakemelding. Dette kan utbedres programmatisk, men tar tid. Derfor har vi valgt ikke å gjøre det. Testdeltaker b1-333333 var i nettopp denne situasjonen. H*n skjelver på hendene, og

bruker tid og flere forsøk på å treffe knappene.

6 Innhentede data

6.1 Oppgavene

Vi hadde tre oppgaver deltakerne skulle gjennomføre. Oppgave en og to handler om forskjellene mellom ordinær fjernkontroll og vår prototype. Se vedleggene oppgave-1-effektivitetsmåling.doc og oppgave-2-effektivitetsmåling.doc. Oppgave tre handler om brukbarhet via affordance. Se vedlegg oppgave-3-brukbarhetstest.doc.

6.1.1 Tidbruk

Tabell 1 viser tidbruk per deltaker for hver oppgave de utførte. Bokstavene a og b i deltakerreferansen står for henholdsvis mann og kvinne. Tallet etter bindestreken indikerer den aktuelle testansvarlige for den deltakeren.

Tabell 1: Tidbruk per deltaker for hver oppgave

deltaker	oppgave 1	oppgave 2	oppgave 3
b1-123456	5:00	0:40	3:40
b1-654321	2:11	0:34	3:28
a1-654321	1:52	0:53	3:49
b1-333333	5:15	2:53	9:57
totalt	14:18	5:00	20:54
gjennomsnitt	3:34	1:15	5:15

6.1.2 Feilantall

Tabell 2 viser hvor mange feil hver deltaker gjorde for hver oppgave.

6.2 Spørreskjemaene

Se kapitlet om Feltstudie, brukbarhetstest, spørreskjema for diskusjon om spørreskjemaene.

Tabell 2: Antall feil per deltaker for hver oppgave.

deltaker	oppgave 1	oppgave 2	oppgave 3
b1-123456	21	0	16
b1-654321	5	0	2
a1-654321	3	(3)	1
b1-333333	9	13	14
total	38	16	33
gjennomsnitt	9,5	4	8,25

6.3 Årsaker til bias

6.3.1 Forutsette årsaker

- Tilfeldig
 - Testansvarlig forklarer ikke oppgaven godt nok.
 - Deltakeren oppfører seg ekstraordinært, f.eks mer fokusert (ikke nøytral) eller ikke fokusert i det hele tatt (ikke interessert). Resultat blir dårligere eller bedre på unaturlig grunnlag.
 - Prototypeverktøyet kræsjer.
 - Miljøet rundt deltakeren skaper stressende og/eller distraherende atmosfære.
- Systematisk
 - Oppgavelisten er ikke godt nok lagt opp.
 - Uvøren bruk av stoppeklokke.
 - Ulike omstendigheter/miljø rundt deltakeren og gjennomføring av brukertesting.

6.3.2 Faktisk opplevde årsaker under testing

- Den analoge fjernkontrollen som skulle være fremmed for og likestille alle deltakerne, og slik redusere bias, virket ikke (dette gjaldt oppgave 1, se vedlegg oppgave-1-effektivitetsmåling.doc)
 - Deltakerne brukte forskjellige fjernkontroller

- Kun én deltaker brukte ukjent fjernkontroll
- Testansvarlige har ikke samme forståelse av testopplegget.
- Testansvarlig hadde ikke øvd nok på gjennomføring (knapp tid).
 - Wizard of Oz var synlig for deltakeren
 - Deltaker ba om og fikk veiledning.
- Fysiske omstendigheter på teststedet gjorde god gjennomføring vanskelig.
 - Wizard of Oz ble synlig for deltakeren
- Første gang deltaker er med på slikt
 - Usikker
 - lett distraherert
- Hos to av deltakerne var en kanal fra oppgavesettet ikke tilgjengelig i kanalpakken.
 - Testansvarlig hos hver av disse deltakerne løste dette på forskjellige måter - problematikken var ikke forutsett.

7 Diskusjon

7.1 Pålitelighet og gyldighet

7.1.1 Pålitelighet

Deltakerne gjennomførte oppgavene i tre lister. Underveis ble antall feil de gjorde registrert i et skjema av testansvarlig, og tiden det tok å utføre hele oppgavelisten ble tatt.

7.1.2 Gyldighet

- Intern gyldighet

Testing hjemme hos hver deltaker gir et svakere grunnlag for intern gyldighet enn laboratorietesting (www.umdj.edu) gjør. Testmiljøene

varierer fra hjem til hjem, med tanke på det rent fysiske som blant annet møbler. Deltakeren er hjemme og ikke i et laboratorium kontrollert av oss. Dette åpner for at distraksjoner (f.eks. vandrende kjæledyr) og innfall (f.eks. om å lage kaffe) gir bias.

At alle deltakere skulle teste med samme verktøy og med samme oppgaveliste er en selvfølge. For å teste vår H_0 trengte vi to artefakter, en ordinær fjernkontroll med trykknapper og vår prototype med berøringsskjerm. For å redusere potensiell bias fra deltakernes egne, kjente fjernkontroller, anskaffet vi en Logitech Harmony 300i. Dette er en fullt funksjonell fjernkontroll som er i stand til å kontrollere nærmest alle elektroniske apparater på markedet i dag og et stykke bakover i tid. En kobler fjernkontrollen til sin pc, og laster ned apparatets styringskoder direkte til fjernkontrollen fra produsentens servere. Under pilottesting var det en smal sak å stille den inn til å kontrollere tv-ene på IFI2. Under utføringen av den reelle testingen var produsentens servere nede. Vi ble altså forhindret fra å bruke den artefakten som skulle likestille deltakerne når det gjelder analog fjernkontroll. Dette svekker oppgave 1s interne gyldighet ytterligere. Med deltaker b1-123456 fikk vi testet med en for deltakeren helt fremmed fjernkontroll. Slik kom vi i det minste nærmere eksperimentets ønskede interne gyldighet. Deltakeren brukte tid og mange forsøk på å løse oppgavene sine. De andre måtte bruke sine egne, og også de hadde problemer med å løse oppgavene sine. På den ene siden er oppgave 1s interne gyldighet svak, på den andre har vi en indikasjon på at forskjellene H_0 refererer til er signifikante.

Vi anser pilottestingen vi gjennomførte som en faktor for intern gyldighet. Artefaktene ble utprøvd og ikke minst ble oppgavelistene justert og utbedret, blant annet for å unngå tvetydighet som kunne gi forskjellig utslag blant deltakerne. Vi øvde også på gjennomføring av forsøkene for å forene hver testholders forståelse av opplegget. Ett eksempel er god utføring av Wizard of Oz-metodikken.

På grunn av en noe lav tilgjengelighet på relevante testdeltakere har det ikke vært aktuelt for oss å bruke randomisering som en faktor for å øke intern gyldighet.

- Ekstern gyldighet

Ting som kan tale for at vårt forsøk har høy ekstern gyldighet er det faktum at forsøkene ble gjennomført hjemme hos deltakerne, altså i de omgivelser det er naturlig for deltakerne å utføre aktivitetene i. Vi har

også truffet vår målgruppe generelt godt med deltakerne våre. Når det er sagt, så har vi observert variasjoner innen deltakermassen. De to viktigste er de forskjellige nivåene på kognitiv og motorisk svekkelse. Dessuten finner vi variasjoner som springer ut i fra faktorer som alder, personlighet og preferanser. I målgruppens yngste segment har vi en deltaker som ikke er kognitivt eller motorisk svekket, men som er noe uvant med teknologi og sågar har en viss skepsis overfor sin egen potensielle mestring av nyvinninger. Av de eldre deltakerne var det et par som søkte bekreftelse for rett valg fra testansvarlig, før de ville utføre trinnene i oppgavelisten.

7.2 Prototypediskusjon

Vi startet med papirskisser av våre forestillinger om fjernkontroll på berørings skjerm. Disse skissene ble overført til skjerm bilde-mockups, som vi brukte som grunnlag for uformelle intervjuer med representanter fra målgruppen (formativt). Se notater fra disse intervjuene i vedlegg Resultaterfralow-fidelity.doc.

Derneft utviklet vi en hifi-prototype for å forberede tester av funksjonalitet og affordance.

Vi havnet på Wizard of Oz-gjennomføring, fordi det å implementere kontakten mellom vår fjernkontroll og TV-er og dekodere var en større arbeid enn vi har tid til å takle i løpet av et semester.

8 Pilottesting

8.1 Pilottesting vs. virkeligheten

I vårt pilottestlokale var det lett å skjule sin utførelse av Wizard of Oz-grepene. Det viste seg å være en utfordring å skjule dette like godt under selve testingen. Hovedårsaken er hva det er praktisk å ta med seg for en testansvarlig av hjelpemateriale - f.eks. kamerastativ, hva som finnes på stedet - f.eks. egnede møbler å legge notatblokk og stoppeklokke på, hvor deltakeren sitter og hvilke konsekvenser det har for hvor den testansvarlige kan plassere seg.

8.2 Så mye pilottesting som mulig

Vi er helt enige om at så lenge det er ressurser til det, bør man utføre så mye pilottesting som mulig, for å se til at selve forsøket blir gjennomført på en så bra måte som mulig. Er det ikke en og samme person eller gruppe som utfører alle forsøkene, vil god pilottesting redusere sjansen for at forsøkene utføres på forskjellige måter, noe som kan gi en rekke forskjellige biaser. En får også luket bort eventuelle misforståelser individuelle testansvarlige har om opplegget og en får trent på å gjennomføre forsøkene riktig på generelt grunnlag og i å hente seg inn dersom noe uforutsett skulle skje.

Ikke minst får en kvalitets sikret sitt eksperimentdesign. Utvetydig og konsis oppgaveliste, en hensiktsmessig rekkefølge for gjennomføring av de forskjellige elementene i eksperimentet, og så videre. En må sørge for at eksperimentet gir en de dataene en trenger, og skaper minst mulig rom for bias.

9 Analyse

9.1 Hypotesene

Tallene fra oppgave 1 og 2, som vist i tabell 1 og 2 under Innhentede data, indikerer at H_0 kan forkastes. Oppgave 1 og 2 er identiske, men 1 ble utført med ordinær fjernkontroll og 2 med vår prototype. Det tok altså deltakerne samlet sett ni minutter og atten sekunder og i gjennomsnitt to minutter og nitten sekunder mer å utføre oppgaven med den ordinære fjernkontrollen enn med vår prototype. Når det gjelder antall feil, så ser vi at deltakerne samlet sett begikk godt over dobbelt så mange og i gjennomsnitt så vidt over dobbelt så mange feil med ordinær fjernkontroll som med vår prototype. På bakgrunn av dette forkaster vi også alternativ hypotese H_1 , og finner at alternativ hypotese H_2 støttes. Vi understreker ordet «indikerer». Vi hadde bias, blant annet med tanke på den ordinære fjernkontrollen vi ønsket å teste med, og kan ikke generalisere basert på våre data. Se Intern gyldighet for mer om dette.

9.2 Brukbarhet

Det går ikke fram av tabellen, men tiden brukt og feilene begått i oppgave 3 stammer i hovedsak fra trinn 5 i oppgavelisten. Der bes deltakeren finne en kanal som ikke er synlig i skjermbildet uten at brukeren slider kanalremsa. Dette indikerer at slide-løsningen har svært lav affordance. Det er også en god

del videomateriale som støtter dette via deltakerens kommentarer underveis og etterpå.

Forskjellene i antall feil i oppgave tre skyldes i hovedsak de forskjellige deltakernes måte å takle oppgavene på. b1-123456 prøvde seg mye fram selv, og gjorde mange feil. b1-654321 og a1-654321 stoppet fort opp, og kom videre etter hjelp fra testansvarlig. I lys av at de lave feiltallene oppsto i testene som kun den ene testansvarlige holdt, kunne det være interessant å undersøke om det er andre faktorer enn deltakernes oppgavetakling som ligger bak forskjellene. Det er det ikke tid til.

10 Konklusjon

Vi har ikke kunnet generalisere på bakgrunn av våre funn, til det har vi for få deltakere og for mye bias.

Men, vi har utført en brukbarhetstest som del av et uekte eksperiment, og har fått indikasjoner på at målgruppen er interessert i løsningen vår fordi den har potensial til å gi dem tilbake litt av mestringsfølelsen deres. Vi har også funnet mangler ved løsningen, som det ville være svært interessant å arbeide videre med. Dette dreier seg om affordance i et spesifikt grensesnittelement: slide-funksjonen man bruker for å se flere kanaler på fjernkontrollen.

I løpet av prosjektet fant vi flere eksempler på at «problemet» med mange unødvendige knapper eller mange kontroller på bordet ikke bare gjelder for eldre, men at det også sees på som et allment problem. Blant fjernkontrollbrukere er det altså en rekke forskjellige målgrupper, som må forholde seg til tv-er, dekodere, forsterkere og dvd-spillere. Man kunne ønske å utvikle en universal løsning for alle disse målgruppene. For oss er det dog ikke tvil om behovet for tilpasninger for forskjellige segmenter. Vårt produkt henvender seg til et av segmentene, nemlig eldre med svakt fallende kognitive evner. Som beskrevet tidligere i rapporten så ønsker vi ikke å fornye den gamle kontrollen. Vi vil heller utforske nye alternativer som bygger på nyere teknologi, som berøringsskjermer. En annen interessant vei å gå er talegjennkjenning, men det hører hjemme i et annet forsøk. Det er ikke sikkert vi må spesielt langt inn i framtiden før det er slutt på dedikerte fjernkontroller for hvert apparat. Kanskje dekkes også disse behovene av for eksempel mobiler?

11 Referanser

- Toftøy-Andersen, Elin og Wold, Jon Gunnar, Praktisk brukertesting, Cappelen Damm Akademisk, 2011. SUS ss 147-149, parvise motset-

ninger ss 149-150.

- Sauro, Jeff, Measuring Usability with the System Usability Scale (SUS), <http://www.measuringusability.com/sus.php>
- University of Medicine and Dentistry New Jersey, Internal and external validity, http://www.umdnj.edu/idsweb/shared/internal_external_validty.htm
- Sharp, Helen og Rogers, Yvonne og Preece, Jenny, Interaction Design beyond human-computer interaction, John Wiley & Sons Ltd, 2007.

12 Samtykkeskjema - konfidensialitet

I vedleggsmappen ligger et ikke utfylt samtykkeskjema. Vi overleverer kopier av de utfylte skjemaene til Alma Culén dersom det er behov for dette. Dette går bra all den tid Almas navn faktisk står på samtykkeskjemaene, der hun er oppført som prosjektansvarlig.