

Master's thesis

2019

Master's thesis

Jørgen Hoff Amundsen

**NTNU**  
Norwegian University of  
Science and Technology  
Faculty of Architecture and Design  
Department of Design

Jørgen Hoff Amundsen

## Taleinteraksjon: Den nye fjernkontrollen?

En *Research Through Design*-studie for å undersøke hvordan *voice-first*-interaksjon påvirker oppfatning av pragmatisk og hedonisk kvalitet i TV-produkter.

June 2019





Norwegian University of  
Science and Technology

## Taleinteraksjon: Den nye fjernkontrollen?

En *Research Through Design*-studie for å undersøke hvordan *voice-first*-interaksjon påvirker oppfatning av pragmatisk og hedonisk kvalitet i TV-produkter.

**Jørgen Hoff Amundsen**

Master of Interaction Design

Submission date: June 2019

Supervisor: Thomas Porathe

Norwegian University of Science and Technology  
Department of Design



## Masteroppgave for Jørgen Hoff Amundsen

**Tittel** Taleinteraksjon, den nye fjernkontrollen?

**Title** Voice Interaction, The New Remote Control?

RiksTV er en av Norges største formidlere av TV-innhold og er nysgjerrige på å hvordan ny teknologi vil være med på å forme morgendagens TV-tjenester. Målet med denne oppgaven er å gi selskapet et bedre informasjonsgrunnlag og en større forståelse av hvordan taleinteraksjon vil kunne bidra til å danne bedre brukeropplevelser. Oppgaven vil, kvalitativt, sammenlikne taleinteraksjon med interaksjon via tradisjonell fjernkontroll, for å avdekke hvilke spesifikke scenarier som kan løses best via tale og hvilke scenarier som kan løses best via fjernkontroll. I tillegg vil det undersøkes om det finnes scenarier der en kombinasjon av de to interaksjonsformene vil være gunstig. Oppgaven vil også utforske hvordan dette kan variere med målgrupper definert etter parametere som alder og funksjonsnivå.

### Gjøremål:

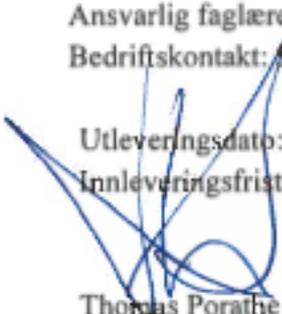
- Litteraturstudie av tematikk
- Danne parametere (hva er betyr bra og hva betyr dårlig?)
- Etablere hvilke målinger/studier som kan gi svar på disse parametere?
- Utføre studier/målinger (Intervjuer, Observasjoner, Rollespill, Mock-ups, testing)
- Analyse av resultater
- Måle resultater opp mot parametere
- Diskutere hvordan resultatene påvirker problemstillingen

Oppgaven utføres etter "Retningslinjer for masteroppgaver i Industriell design".

Ansvarlig faglærer: Thomas Porathe  
Bedriftskontakt: Sindre Hallandvik Langenes

Utleveringsdato: 11. Januar 2019

Innleveringsfrist: 7. Juni 2019

  
Thomas Porathe  
Ansvarlig faglærer

Trondheim, NTNU, dato

  
Ole Andreas Alsos  
Instituttleder

# **Taleinteraksjon: den nye fjernkontrollen?**

En *Research Through Design*-studie for å undersøke hvordan *voice-first*-interaksjon påvirker oppfatning av pragmatisk og hedonisk kvalitet i TV-produkter.

Jørgen Hoff Amundsen

Juni, 2019

# Sammendrag

Denne masteroppgaven utforsker, gjennom utvikling og evaluering av en forskningsprototype, hvordan *voice-first* interaksjon kan påvirke brukeropplevelsen i TV-produkter. Først utforskes *state of the art* innen multimodale systemer og noen sentrale begreper innenfor taleinteraksjon presenteres for leseren. Deretter beskrives et omfattende litteraturstudie, som kartlegger hva som bidrar til en god brukeropplevelse og hvordan en brukeropplevelse kan evalueres. Videre beskrives Marc Hassenzahls reduksjonistiske modell for brukeropplevelse, bestående av to dimensjoner: pragmatisk kvalitet og hedonisk kvalitet. Pragmatisk kvalitet omhandler produktets evne til å effektivt gjennomføre brukerens funksjonelle mål. Hedonisk kvalitet på den andre siden, dreier seg om produktets evne til å bringe frem positive følelsesresponsen hos brukeren.

Videre beskrives hvordan en *Research Through Design* tilnærming ledet utviklingen av en forskningsprototype, gjennom en brukersentrert iterativ designprosess. Avslutningsvis, for å besvare forskningsspørsmålet, blir prototypen sammenliknet med en tradisjonell TV-løsning i et komparativt studie med ti deltagere. En AttrakDiff semantisk differensial, som måler deltagerens subjektive responser til de to systemene, i tillegg til ytelsesmål og kvalitative intervjuer, gir positive resultater i favør *voice-first* prototypen. Resultatene viste at *voice-first* kan gjøre navigasjon enklere og mer effektiv og dermed bidra til en høyere grad av pragmatisk kvalitet. Når det gjelder hedonisk kvalitet, viste prototypen seg i stand til å frembringe mer positive følelser, noe som indikerer at *voice-first* interaksjon har en positiv innvirkning på den totale brukeropplevelsen i TV-produkter.

# Abstract

This master thesis explores, through the development and evaluation of a research prototype, how voice-first interaction can affect the user experience of TV-products. Firstly an investigation of *the state of the art* in multimodal systems I conducted. Then a comprehensive literature review is conducted in order to answer the question of what constitutes a good user experience and how user experience can be measured. Further, Marc Hassenzahls reductionist model of user experience is explained. This model separates product quality into two dimensions; pragmatic quality and hedonic quality. Pragmatic quality refers to a products ability to assist users in fulfilling functional goals, whereas hedonic quality refers to a products ability to evoke a positive emotional response.

Following a Research Through Design approach, a prototype was developed through a user centered iterative design process. Further, in order to answer the research question, a controlled laboratory study with ten participants was conducted, comparing the prototype to a traditional TV-solution. An AttrakDiff semantic differential questionnaire, measuring the users responses to the two systems, as well as performance metrics and interviews, gave positive results in favor of the voice-first prototype. Results showed that voice-first interaction can make navigation easier and more efficient, thus contributing to a higher degree of pragmatic quality. Regarding the hedonic quality, the prototype also elicited more positive emotional response among the participants, suggesting that voice-first interaction have a positive impact on the overall user experience of TV-products.

# Forord

Denne masteravhandlingen er skrevet ved Institutt for design ved Norges teknisk-naturvitenskapelige universitet våren 2019.

Først og fremst vil jeg takke RiksTV, ved Sindre og Sebastian, for deres åpenhet og samarbeidsvillighet gjennom de siste ni månedene. Denne oppgaven kunne ikke blitt utført uten deres interesse og tilrettelegging. Videre vil jeg takke til alle som har deltatt i brukertester og forsøk. Til min hovedveileder Thomas, takk for din tid, fleksibilitet og entusiasme. Din faglige ekspertise og erfaring har vært til stor hjelp gjennom hele prosessen. Jeg vil også takke Trond Are for veiledning i D9-prosjektet som ledet opp til dette. Takk til Emil og Trygve for fem år med mye gøy og gode opplevelser, men også mye slit og stressende tider. Deres samarbeid og vennskap har gjort gjennomføringen mye lettere. Takk også til Sindre og Sturla for at dere alltid er der for meg. Takk til Nina og Geir for alle bidrag. En ekstra stor takk også til Alf-Henrik for uvurderlig korrekturlesning.

Til slutt vil jeg gi en enorm takk til min familie. Mamma, Pappa og Anders. Deres støtte og motivasjon er helt avgjørende for meg. Kjære Rikke, takk for at du er den du er og for alt du har bidratt med i denne prosessen.

Jørgen Hoff Amundsen,  
Oslo, Juni 2019

# Innholdsfortegnelse

<b>1 Innledning</b> .....	<b>1</b>
1.1 <i>Motivasjon</i> .....	2
1.2 <i>Forskningsspørsmål</i> .....	3
1.3 <i>Kapittelguide</i> .....	4
<b>2 Bakgrunn</b> .....	<b>6</b>
2.1 <i>Hva er taleinteraksjon?</i> .....	6
2.2 <i>Muligheter og utfordringer med taleinteraksjon</i> .....	8
2.3 <i>Multimodale grensesnitt: tale og skjerm</i> .....	11
2.4 <i>Taleinteraksjon i TV</i> .....	13
<b>3 Teoretisk rammeverk</b> .....	<b>15</b>
3.1 <i>Hva er UX?</i> .....	15
3.2 <i>Reduksjonisme vs. holisme</i> .....	22
3.3 <i>Syn på evaluering</i> .....	23
3.4 <i>Marc Hassenzahl: Den hedonisk/pragmatiske modellen</i> .....	24
<b>4 Metodisk tilnærming</b> .....	<b>31</b>
4.1 <i>Research through Design</i> .....	31
4.2 <i>Min tilnærming til RtD</i> .....	34
4.3 <i>Evalueringemetoder</i> .....	35
<b>5 Designprosessen</b> .....	<b>38</b>
5.1 <i>Brukerinnsikt</i> .....	38
5.2 <i>Dramatiseringsworkshop</i> .....	41
5.3 <i>Utvikling av prototype</i> .....	43
5.4 <i>Scenarioer</i> .....	45
5.5 <i>«Happy path»</i> .....	46
5.6 <i>Digital prototyping</i> .....	47
5.7 <i>Iterativ utviklingsprosess</i> .....	50

5.8 Endelig forskningsprototype .....	52
<b>6 Komparativ studie .....</b>	<b>55</b>
6.1 De to løsningene .....	55
6.2 Deltagere .....	55
6.3 Målinger .....	55
6.4 Prosedyre .....	56
<b>7 Resultater .....</b>	<b>58</b>
7.1 AttrakDiff .....	58
7.2 Loggdata .....	62
7.3 Intervjuer .....	62
<b>8 Diskusjon .....</b>	<b>65</b>
8.1 Implikasjoner .....	65
8.2 Svakheter .....	67
8.3 Replikabilitet og validitet .....	69
<b>9 Konklusjon.....</b>	<b>71</b>
9.1 Bidrag .....	71
9.2 Fremtidig arbeid.....	71
<b>Referanser .....</b>	<b>73</b>
<b>Appendix A: Intervjuguide .....</b>	<b>80</b>
<b>Appendix B: Workshopplan .....</b>	<b>81</b>
<b>Appendix C: Forsøksguide.....</b>	<b>83</b>
<b>Appendix D: Resultater ytelsesmål .....</b>	<b>84</b>



# 1 Innledning

Trykk på «på»-knappen, trykk på «meny»-knappen, trykk tre ganger på «til venstre»-knappen, trykk på «ok»-knappen, trykk fire ganger på «til høyre»-knappen, trykk to ganger på «ned»-knappen, trykk tre ganger på «høyre knappen», trykk på «ok»-knappen, trykk en gang på «ned»-knappen, trykk på «ok»-knappen, trykk åtte ganger på «ned»-knappen, trykk to ganger på «ok»-knappen.

Instruksjonen over beskriver fremgangsmåten for å finne siste episode av Game Of Thrones med Smart Boks 2, Dagens TV og streamingprodukt fra RiksTV. Alternativt kan man se for seg å si: «Sett på siste episode av Game of Thrones».

Talestyring i forbindelse med TV og streaming er i startfasen, og flere av markedets store aktører investerer stort i teknologien. Som en av landets største formidlere av TV-innhold, er det interessant for RiksTV å utforske denne teknologien og undersøke hvordan den kan bidra til å gjøre deres produkter bedre. Å gjøre et skifte mot talestyringssystemer er en kostbar og tidkrevende prosess som krever en rekke beslutninger om alt fra hvilket rammeverk man bygge på, hvilke teknologier man bør benytte seg av og hvilken funksjonalitet som skal implementeres. Disse valgene gjøres enda mer komplekse av en teknologi i rask utvikling, der *state-of-the-art* endres kontinuerlig, og en standard i bransjen mangler. For å kunne gjøre disse valgene, er man avhengig av en god forståelse av hvor nytteverdien i et slikt system ligger og hvilke spesifikke brukscenarier som bør prioriteres. Tar man den lange veien via utvikling og implementering, vil det være en risiko for at man utvikler noe brukeren faktisk ikke er interessert i.

Gjennom en *Research-Through-Design*-tilnærming, tar denne mål av seg om å avdekke hvilken nytte talestyring har i TV-sammenheng, uten å faktisk utvikle teknologien. En brukersentrert utviklingsprosess resulterer i en forskningsprototype, som videre brukes i et komparativt forsøk, der den sammenliknes med en tradisjonell TV-løsning, basert på interaksjon via en fjernkontroll. Resultatene tilsier at talestyring i TV-sammenheng både kan bidra til å øke den pragmatiske kvaliteten, dvs. gjøre det enklere og raskere å utføre oppgaver, men det kan i tillegg tilføre en hedonisk dimensjon, som innebærer at det enklere kan frembringe positive følelsesresponses.

## 1.1 Motivasjon

### 1.1.1 Taleinteraksjon i streaming

Talestyring i forbindelse med *streaming* er allerede tatt i bruk av store aktører som Netflix og HBO, via rammeverkene til Google, Apple og Amazon. Disse tjenestene tilbyr enkle funksjoner som å spille av innhold, kontrollere lydnivå, og spoling ved hjelp av naturlig tale. Som en av landets største formidlere av TV-innhold, kreves det at RiksTV tar stilling til denne teknologien og hvordan den kan bidra til å gjøre deres produkter bedre og mer attraktive.

Talestyring er en ung teknologi, og det finnes lite empiri eller standarder for hvordan den bør tas i bruk. Mye av dokumentasjonen baserer seg på enkle sjekklister fra de største aktørene i markedet. Store investeringer kreves for å utvikle en god talestyringsteknologi. Dersom man mislykkes med å forstå hva brukerne har behov for og hvilke aspekter de setter pris på, vil en slik investering være kritisk for et selskap som RiksTV. På den andre siden vil det å være sent ute med å adoptere ny teknologi kunne koste dyrt, dersom konkurrentene lykkes med dette. Dersom man klarer å forstå hva som gir verdi i taleinteraksjon vil, dette kunne bidra til en mer fornøyd kundebase og en styrket posisjon i marked for RiksTV.

### 1.1.2 Personlig motivasjon

Før oppstarten av dette masterprosjektet hadde jeg allerede utviklet en sterk interesse for taleinteraksjon, og en stor tro på at dette vil spille en vesentlig rolle i interaksjonen mellom menneske og maskin i årene som kommer. Gjennom et internship hos Making Waves, sommeren 2018, ble jeg introdusert for taleinteraksjon i forbindelse med utvikling av en talestyrt matoppskrift. Dette ga innblikk i hvordan tematikken behandles i bransjen og en forståelse av hvor høyt det står på agendaen for mange av kundene. Denne interessen ble ført videre da jeg høsten 2018 skrev en artikkel om den samfunnsmessige påvirkningen fra taleinteraksjon (Amundsen, 2018a). I emnet «Design 9» (D9), ved NTNU høsten 2018, ble samarbeidet med RiksTV opprettet. I dette prosjektet ble taleinteraksjon i TV-sammenheng undersøkt i form av en mulighetsstudie (Amundsen, 2018b). Jeg ønsket, gjennom dette masterprosjektet, å videreføre innsikten fra D9 i utvikling og forskning på en prototype. Målet er å kunne bidra til videre utvikling for interaksjonsformen.

I tillegg ønsket jeg gjennom masterprosjektet å danne en bedre forståelse av brukeropplevelse. Brukeropplevelse, eller UX (User eXperience) er et begrep som blir brukt hyppig både i bransjetekster og faglitteratur, men en felles oppfatning av hva det

innebærer uteblir. Jeg ønsket å danne meg en forståelse av hva dette begrepet faktisk innebærer, og hvordan det blir behandlet i akademia og i industrien. For å kunne gjøre en vurdering av hvordan en brukeropplevelse kan forbedres, kreves en forståelse av hva det egentlig innebærer.

Mitt ønske om å gjøre en akademisk undersøkelse av mitt eget fagfelt har medført at formatet på oppgaven skiller seg noe fra masteroppgavene som er levert på instituttet de siste årene. Fokuset har i denne oppgaven vært på å forstå metoder, teori og forskning, fremfor å bruke hele prosjektet på design. Systemet som designes i oppgaven er utviklet hovedsakelig for å generere innsikt, og det vil bli fokusert på å presentere denne innsikten fremfor å presentere systemet i seg selv.

## 1.2 Forskningsspørsmål

I denne masteroppgaven undersøkes hvordan taleinteraksjon kan bidra til å danne bedre brukeropplevelser for TV-brukere. Ved å utvikle en prototype som presenterer brukerne for en potensiell forestilt fremtid, ønsker jeg å generere kunnskap om brukernes preferanser og teknologiens nytteverdi i denne sammenhengen. Forskningsspørsmålet i oppgaven er formulert på følgende måte:

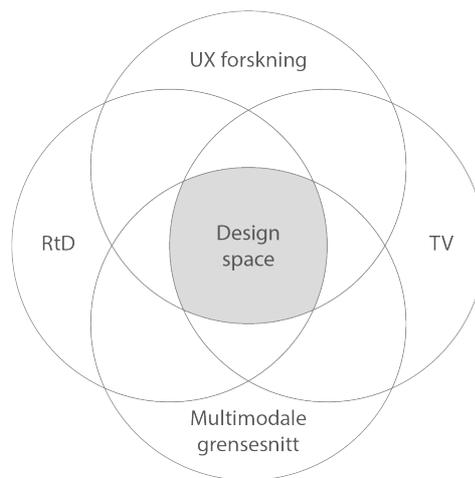
*Hvordan kan taleinteraksjon bidra til å skape bedre brukeropplevelser i TV-sammenheng?*

For å besvare det sentrale forskningsspørsmålet, ble det dannet tre underforskingsspørsmål som ble brukt til å kontekstualisere og lede prosjektet:

1. Hvilke utfordringer og svakheter finnes ved tradisjonelle TV-løsninger?
2. Finnes det en felles enighet innenfor akademia om hva som definerer en god brukeropplevelse? Finnes det metoder som kan gi reelle svar på om en brukeropplevelse er god?
3. Hvordan kan man, gjennom utvikling av en forskningsprototype, skape ny innsikt om en fremtidig situasjon eller teknologi?

### 1.2.1 Design Space

Jeg plasserer arbeidet innenfor det jeg karakteriserer som et *Design Space*, eller et «mulighetsrom». Mulighetsrommet illustrerer hvordan bidraget forholder seg til andre relevante forskningsområder. Mulighetsrommet dannes av: ønsket om å forstå UX og hvordan det kan evalueres; RtD som metode, som brukes til å forstå hvordan en forskningsprototype kan generere innsikt om en fremtidig situasjon; ønsket om å forstå taleinteraksjon i multimodale grensesnitt; og til slutt hvordan TV-produkter kan tjene på å introdusere taleinteraksjon. Det er innenfor dette mulighetsrommet forskningsspørsmålene utforskes og forskningsbidraget plasseres.



*Figur 1: Illustrerer hvordan denne oppgaven forholder seg til andre relevante og bredere forskningsområder*

## 1.3 Kapittelguide

**Kapittel 2:** Presenterer *bakgrunnen* leseren trenger for å forstå hva talestyring er, brukskonteksten, dagens situasjon og retninger for fremtiden. Først presenteres en oversikt over hva talestyring er og hvordan det fungerer. Videre presenteres en oversikt over viktige aktører innen taleinteraksjon og et utvalg brukssituasjoner og domener der talestyring suksessfullt har blitt tatt i bruk. Til slutt presenteres noen ulike syn på hvordan taleinteraksjon vil prege HCI i årene som kommer.

**Kapittel 3:** Presenterer mitt *teoretiske rammeverk*. I dette kapitlet gis en oversikt over ulike oppfatninger innen akademia om hva brukeropplevelse (UX) er og hvordan det kan

evalueres. Videre introduserer jeg Marc Hassenzahl modell om hedonisk – og pragmatisk produktkvalitet, og hvordan denne modellen har veiledet design- og evalueringsprosessen.

**Kapittel 4:** Beskriver min *metodiske tilnærming* i dette prosjektet. Tilnærmingen baserer seg på prinsippene fra «Research through design» (RtD). Videre presenteres de konkrete evalueringsmetodene som legges til grunn for den komparative studien.

**Kapittel 5:** Presenterer *designprosessen* på en kronologisk og historisk lineær måte. Her beskrives involveringen av brukere på de ulike stegene i prosessen og den iterative utviklingen av forskningsprototypen.

**Kapittel 6:** Beskriver den *komparative studien* som ble utført med forskningsprototypen og en tradisjonell TV-løsning. 10 deltagere utførte syv oppgaver med begge løsningene. Vurderinger ble gjort utfra en AttrakDiff undersøkelse, i tillegg til loggdata og et avsluttende intervju..

**Kapittel 7:** Legger frem *resultatene* produsert av forsøket.

**Kapittel 8:** Presenterer en *diskusjon* av resultatene i lys av det teoretisk rammeverket. I tillegg gjøres en kritisk refleksjon over av metodene som er brukt og validiteten i funnene.

**Kapittel 10:** Presenterer en *konklusjon* med hensyn til forskningsspørsmålene der oppgavens forskningsbidrag understrekes og forslag til videre arbeid fremmes.

## 2 Bakgrunn

Dette kapitlet presenterer bakgrunnsinformasjonen leseren trenger for å forstå hva talestyring er, brukskonteksten, dagens situasjon og retninger for fremtiden. Selv om denne oppgaven ikke konsentrerer seg om de teknologiske aspektene ved talestyring, men på selve brukeropplevelsen, sees det som nødvendig å gi leseren et bilde på hvordan teknologien fungerer og hvordan den har blitt tatt i bruk. Dette kapitlet peker på noen sentrale temaer innenfor taleinteraksjon og multimodale grensesnitt. Først presenteres en oversikt over hva talestyring er og hvordan det fungerer. Videre presenteres noen sentrale muligheter og utfordringer som interaksjonsformen presenterer. Videre diskuteres *voice-first* og *screen-first* som to retninger innenfor taleinteraksjon med hver sine fordeler og ulemper. Til slutt presenteres en oversikt over hvordan taleinteraksjon har blitt brukt i TV-sammenheng.

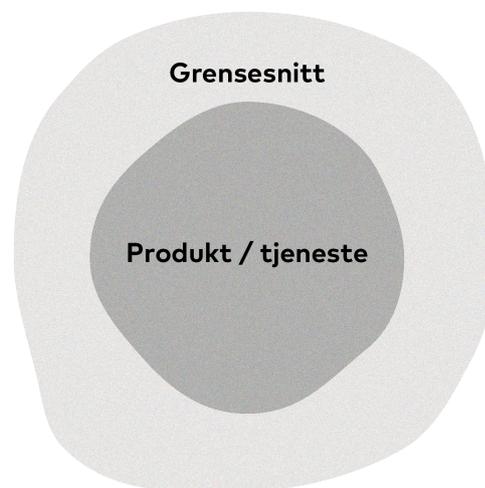
### 2.1 Hva er taleinteraksjon?

Taleinteraksjon vil kort fortalt si at man benytter stemmen som den sentrale interaksjonsformen. Innenfor *Human Computer Interaction* (HCI) har taleinteraksjon vært et forskningstema siden 50-tallet, og utopien om å bestå den berømte *Turing*-testen, som innebærer at man får teknologien til å fremstå som en fullverdig menneskelig samtalepartner, har siden da vært høyt oppe på forskningsagendaen (Bowden, 2006). Vi har også sett flere eksempler i populærkulturen som beskriver en fremtid der mennesker snakker naturlig med datamaskiner. Vi har ikke kommet fullt dit enda, men vi nærmer oss raskere enn noen gang.

Vi har sett en eksponentiell vekst i den kommersielle bruken av taleteknologi de siste årene. Fra introduksjonen av Apples *Siri* gjennom smart-telefoner og nettbrett, til den nyere introduksjonen av såkalte smarthøytalere fra blant annet Google og Amazon. De ulike smarthøytalerne har forskjellig funksjonalitet og utvider stadig antall tilgjengelige kommandoer, men kjernefunksjonene er imidlertid de samme: Du kan spørre om informasjon, spille av media eller fullføre oppgaver med ulike tilkoblede tjenester og enheter. Altså er det ikke produktene og tjenestene man når gjennom et talegrensesnitt som har forandret seg så mye, men taleinteraksjon kan ses som en annen måte å interagere på, for å løse de samme underliggende problemene som andre brukergrensesnitt.

Flere eksperter spår at talebaserte grensesnitt vil kunne revolusjonere måten vi bruker teknologi på i løpet av de neste årene, og potensielt medføre et skille på linje med

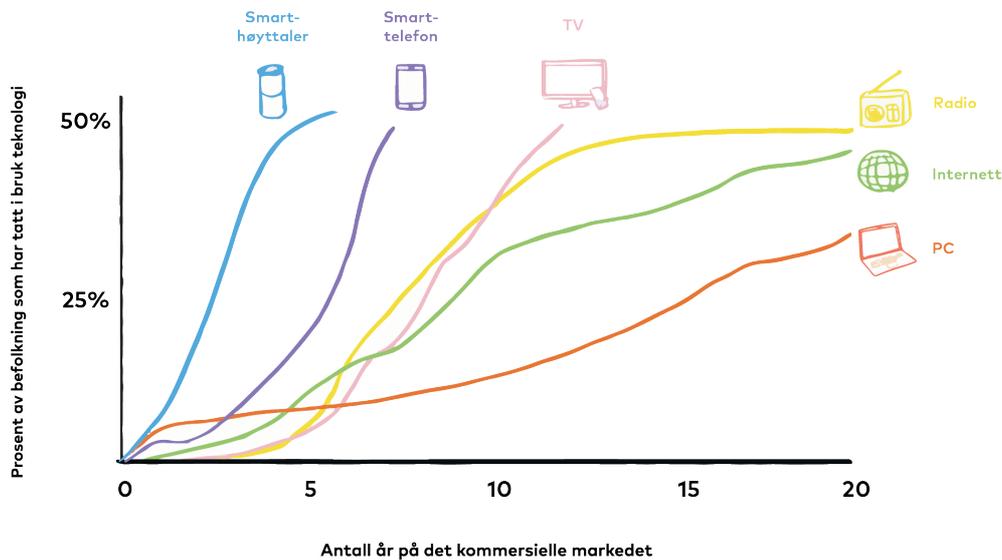
overgangen fra kommandolinjen til grafiske grensesnitt (Fichter & Wisniewski, 2017). Hirschberg & Manning (2015) krediterer denne utviklingen til en rask progresjon innen NLP (Natural Language Processing), som de igjen forklarer med følgende fremskritt: “En stor økning i databehandlingskraft, tilgjengelighet av store mengder språkdata, utvikling av svært vellykkede maskinlæringsmetoder og økt forståelse av strukturene i menneskelig språk og hvordan det brukes i sosiale sammenhenger” (Hoi, 2018, s.82 ). Det spås at markedet for taleassistenter vil fortsette å øke de neste årene (McTear et al., 2016). En av faktorene som brukes til å forklare dette er den såkalte «*Cycle of increasing returns*». Ettersom brukere viser større aksept for taleinteraksjon og bruker det mer, vil dette bidra til gjøre teknologien bedre ved å generere større mengder av læringsdata (McTear et al., 2016).



*Figur 2: Taleinteraksjon kan sees som en ny måte å nå de samme produktene og tjenestene som før.*

For å understreke hvor raskt denne utviklingen går kan man se på adaptjonsraten til smarthøytalere på det amerikanske markedet sammenliknet med adaptjonsraten til andre store teknologiske nyvinninger (Activate Analysis, 2018). Den blå kurven i Figur 3 viser at 50 % av den amerikanske befolkningen har adaptert til smarthøytalere innen fem år etter lansering. Til sammenlikning er dette tre til fire ganger raskere enn adaptjonen av radio, internett og PC. Det er mange faktorer som spiller inn på en slik graf og det blir ikke riktig å tolke verdiene isolert, for eksempel fordi nettopp veksten av internett og medfølgende globalisering har gjort det mulig å markedsføre og distribuere produkter på en helt annen

måte enn ved lanseringen av for eksempel radioen og tv-en. Allikevel indikerer grafen at taleinteraksjon på kort tid har blitt en viktig teknologi i forbrukermarkedet.



Figur 3: Prosent av den amerikanske befolkningen som har adaptert teknologiske nyinnvinninger, etter antall år på det kommersielle markedet (Activate Analysis, 2018).

## 2.2 Muligheter og utfordringer med taleinteraksjon

Den digitale utviklingen går raskere enn noen gang og brukere stiller stadig nye krav til hastighet, effektivitet og bekvemmelighet. Taleinteraksjon har vist seg å kunne gi et høyere komfortnivå og en mer effektiv oppgaveløsning i mange sammenhenger (McTear et al., 2016). Enten det er å for å finne informasjon, handle eller utføre oppgaver, vil det, eksemplifisert med fjernkontroll-eksempelet i introduksjonen, i mange tilfeller være raskere å si noe enn å utføre komplekse handlinger via et tradisjonelt grafisk grensesnitt. Samtidig oppstår det utfordringer når en skjerm fjernes. Noen sentrale muligheter og utfordringer med taleinteraksjon vil diskuteres i den påfølgende delen.

### 2.2.1 Fleksible inngangspunkter

En av de største fordelene med et talegrensesnitt er mulighetene det gir for å gjøre navigasjon i datasett flat ved å tillate fleksible inngangspunkter (Seo, 2018). Det vil si at man ikke trenger å filtrere eller sortere seg frem til ønsket objekt eller resultat, men man kan spørre direkte om det man er ute etter. Med dette fjerner man behovet for hierarkiske menystrukturer, og man kan nå funksjoner som i tradisjonelle grensesnitt ville vært begravet flere nivåer nede i en applikasjon.

Samtidig kan fleksible inngangspunkter være den største utfordringen i et talegrensesnitt. I motsetning til grafiske grensesnitt, er talegrensesnitt ofte ugjennomsiktige og gir liten eller ingen indikasjon på funksjonalitetene eller arkitekturen i applikasjonen. Følgelig er et av de største problemene med talegrensesnitt i dag at brukere ikke oppdager funksjonaliteten (Seo, 2018). Den vanligste løsningen for å gi brukeren innsikt i taleapplikasjonenes funksjonalitet er å sekvensielt ramse opp alt brukeren kan gjøre i en velkomsthilsen (McTear et al., 2016). Dette er dog ikke en optimal løsning, da mennesker sliter når den må holde mer enn kun noen få ting i korttidshukommelsen samtidig (Ware, 2012). Denne utfordringen refereres til som *oppdagelsesproblemet*, eller *The Discoverability problem* (Seo, 2018).

### **2.2.2 Usability-utfordringer i talegrensesnitt**

Don Norman hevder at for at brukere skal kunne interagere suksessfullt med et system, må de (1) finne ut hvilke handlinger de kan gjøre for å nå målene sine, og (2) forstå resultatene av disse handlingene (Norman, 2013). Han beskriver disse behovene som *The Gulf of Expectation* og *The gulf of Evaluation*. Han går videre med å trekke frem noen sentrale elementer i et brukergrensesnitt som kan bidra til å bygge broer disse såkalte «Gulfene». Tre av disse elementene som er særlig relevant for et talegrensesnitt er *Constraints*, *affordances* og *feedback* (Norman, 2013). Et grensesnitt eksklusivt basert på lyd, som både *input-* og *output-*medium, har problemer med å tilby alle disse tre elementene.

*Constraints* dreier seg om å gi brukeren en forståelse av omfanget av funksjonalitet i et system (Norman, 2013). Som beskrevet i det ovennevnte *oppdagelsesproblemet*, er det vanskelig å gi brukeren oversikt over den funksjonaliteten som er mulig å oppdrive i et talesystem. Tilsvarende er det vanskelig å indikere hva som *ikke* er mulig å gjennomføre. Dagens teknologi klarer ikke å svare på alle mulige ytringer, så risikoen for at brukeren vil spørre etter funksjonalitet som ikke finnes er stor. Dette kan resultere i frustrasjon for brukeren.

*Affordances* dreier seg om et produkts evne til å få brukeren til å forstå hvordan det skal brukes (Norman, 2013). At man kan snakke til et talegrensesnitt er gitt, men hvordan man bør ordlegge seg er åpent for tolkning. Mennesker har mentale modeller for hvordan en mellommenneskelig samtale fungerer (Norman, 2013). Dersom en bruker har samme forventning til interaksjon med et talegrensesnitt, vil brukeren bli skuffet. For å kunne tilby brukeren *affordances* om hvordan et talegrensesnitt kan brukes, er man avhengig av å kommunisere talegrensesnittets tilkortkommenhet og mangler. Dette er en utfordring fordi

dagens idealer innenfor taleinteraksjon er å imitere mellommenneskelig interaksjon så langt det lar seg gjøre. Denne skeuomorfismen minner om måten man tok i bruk referanser til fysiske elementer i tidlige grafiske grensesnitt, for å gi brukerne en forståelse av hvordan det brukes (Brautigam, 2017). Problemet med skeuomorfismer i talegrensesnitt er at det ikke holder det de lover.

*Feedback* refererer til et systems evne til å kommunisere at en handling eller operasjon har blitt utført eller pågår (Norman, 2013). I et talegrensesnitt er man avhengig av å gi denne typen bekreftelse gjennom lyd. Dette er ikke et problem i seg selv, og lyd blir i mange tilfeller brukt som et virkemiddel for å gi feedback til brukere i andre typer grensesnitt, i form av for eksempel en lydeffekt når en epost blir sendt. Når det gjelder handlinger som krever en kontinuerlig bekreftelse på at de pågår, blir dette fort en utfordring i talegrensesnitt. Et eksempel er å sette på en alarm, som er en nøkkelfunksjon i mange smarthøytalere. Å sette på en alarm gjennom tale er praktisk, men å måtte spørre hvor lenge det er igjen før alarmen ringer er mindre praktisk.

### **2.2.3 Opprettholdelse av kontekst**

Som nevnt innledningsvis vil denne oppgaven vil ikke behandle de teknologiske aspektene ved taleinteraksjon direkte, da dette ligger utenfor oppgavens omfang. Allikevel sees det som nødvendig å peke på en sentral utfordring NLP som begrenser noe av mulighetsrommet for utformingen av talegrensesnitt, nemlig å skille mellom kontekstuavhengige og kontekstavhengige kommandoer.

Kontekstuavhengige kommandoer er enkle. “Finn filmer med skuespiller X” som en kontekstuavhengig kommando, bør returnere alle filmene med skuespiller X. Men hva om denne kommandoen var kontekstavhengig? Med kontekst som en variabel kan to identiske kommandoer bety noe veldig forskjellig. La oss si at brukeren først spurte: “Finn en film med skuespiller X,” og deretter sa: “Finn en komedie”? Resultatene burde være komedier med skuespiller X. Dette er umulig å gjøre uten kontekst.

Opprettholdelse av kontekst er ikke et løst problem, men det skjer sakte, men sikkert. Google og Amazons talegrensesnitt er begge kontekstuelte oppmerksomme til en viss grad (Armour, 2018). Spør man om været på torsdag, vil de svare nøyaktig. Følger man opp med “Hva med fredag” vil de forstå° at du fortsatt snakker om været. Men dersom kompleksiteten overgår dette, vil de begge ha vanskeligheter. I denne oppgaven vil

utfordringene med kontekst ikke bli behandlet direkte, men brukt som en begrensende faktor for kompleksiteten i dialogene som utformes.

## 2.3 Multimodale grensesnitt: tale og skjerm

Da en talestyrt applikasjon for TV vil medføre at brukeren blir presentert for mer enn en interaksjonsflate, både skjerm og tale, vil det per definisjon være et multimodalt grensesnitt. Et multimodalt grensesnitt vil si et grensesnitt der det benyttes mer enn en interaksjonsform. Nielsen Norman Group (Whitenton, 2017) hevder at dersom man klarer å kombinere taleinteraksjon og skjerm på en god måte, vil man kunne bedre brukeropplevelsen i svært mange systemer, ved å benytte seg av styrkene til begge interaksjonsformene.

Tale er et effektivt for *input*-medium, fordi den tillater, som diskutert i 2.2, fleksibilitet og fjerner behovet for komplekse navigasjonsstrukturer. Skjerm på den andre siden et effektivt *output*-medium, fordi det tillater å vise store mengder informasjon samtidig og dermed reduserer belastningen på brukerens korttidshukommelse (Whitenton, 2017). Visuell scanning er raskere enn den sekvensielle oppramsingen av informasjon som kreves i et rent talegrensesnitt. En skjerm kan også effektivt gi visuelle signaler som kan antyde både *affordances*, *feedback* og *constraints* og på den måten bidra til å løse problematikken med *The Gulf of Evaluation* og *The Gulf of Execution*.

Samtidig som det åpenbart er klare fordeler ved å kombinere disse to interaksjonsformene, oppstår det også utfordringer når man kombinerer to ulike interaksjonsparadigmer. For eksempel bør man ta i betraktning at brukerne har forutinntatte ideer om hvordan skjermbaserte produkter opereres og av den grunn kan slite med å omstille seg til å bruke enheten på en annen måte (McTear et al., 2016). Videre, dersom man gir brukeren et valg mellom to ulike interaksjonsformer for å nå samme funksjonalitet, kan dette resultere i økt kognitiv aktivitet hos brukeren og potensielt økt frustrasjon (Whitenton, 2017).

Multimodale grensesnitt som kombinerer skjerm og taleinteraksjon har blitt utforsket i de senere årene, og ført til en rekke ulike produkter og systemer som kan deles inn i to kategorier: *Screen-first* og *voice-first* (Whitenton, 2017). I et *screen-first*-grensesnitt er utgangspunktet en applikasjon som primært er designet for skjerm, som blir supplert med talefunksjonalitet med mål om å gjøre brukeropplevelsen bedre. I et *voice-first*-grensesnitt

er rollene snudd. Her er applikasjonen i utgangspunktet designet for tale, mens skjermen er et supplement for visuell informasjonsformidling.

### **2.3.1 Screen-first**

Inntil nylig har de fleste kombinasjonene av tale og skjerm basert seg på *screen-first*-interaksjon, eksemplifisert av smarttelefoner, der et talestyringssystem er lagt til i et eksisterende grafisk brukergrensesnitt, i form av et taleassistent, som Apples Siri eller Googles Assistant. Disse systemene viser imponerende talegjenkjennelse og språkbehandling, men den totale brukeropplevelsen svekkes på grunn av den grunnleggende oppdelingen mellom tale og berøringsbasert funksjonalitet (Whitenton, 2017). I de fleste tilfeller kan taleassistenten bare utføre det første trinnet i en oppgave, og eventuelle påfølgende trinn krever at brukeren går over til fysisk interaksjon. For eksempel vil Siri utføre en nettsøk som svar på en talekommando, men brukeren må da trykke på skjermen for å velge et søkeresultat eller få tilgang til en nyhetshistorie. Google Assistant krever også skjermberøring for å bevege seg utover det første trinnet i de fleste tilfeller.

### **2.3.2 Voice first**

Voice-first refererer til et system som primært aksepterer brukerininput via talekommandoer, men som kan supplementere audiell output med å vise informasjon på et integrert skjermdisplay. Den kommersielle suksessen til smarthyttalere, kombinert med ovennevnte usability-utfordringer med tale som både inn- og ut-modalitet, har nylig ført til fremveksten av en ny produktkategori: smarthyttalere med skjerm (Whitenton, 2017). *Echo-show* fra Amazon og *Nest* fra Google er essensielt videreføring av deres respektive smarthyttalere, der det er lagt til en skjerm for å utvide *output*-funksjonaliteten. Sammenlignet med tradisjonelle skjermbaserte enheter er disse skjermene langt mindre i stand til å utføre grunnleggende funksjoner, som eksempelvis å bruke en nettleser eller skrive inn meldinger, men fungerer som en overflate for presentasjon av informasjon.



*Figur 4: Google Nest og Amazon Echo-show. Eksempler på voice-first interaksjon. (Google, 2019., Amazon, 2019).*

Det disse produktene tilbyr er en fundamentalt forskjellig interaksjonsstil som kan beskrives som "voice-first", og som baserer seg nesten utelukkende på *input* gjennom tale, i stedet for å tillate tale som en sekundær interaksjonsform. Essensielt representerer voice-first en ny tilnærming til kombinasjonen mellom tale og skjerm. Der talefunksjonalitet tidligere har blitt implementert i et eksisterende grafisk grensesnitt, er voice-first fundamentalt motsatt. Utgangspunktet er en ren taleinteraksjon, og skjermen blir introdusert for å supplere den opprinnelige taleinteraksjonen. Selv om det teknisk sett er en berøringsskjerm de er supplert med, inneholder disse skjermene sjelden knapper eller menyer. I stedet for å oppfordre brukere til å trykke eller sveipe, vises ofte foreslåtte verbale kommandoer på skjermen, for eksempel «Prøv, 'bla til høyre'».

Nielsen Norman Group hevder at voice-first så langt har vist seg å være den mest effektive måten å kombinere tale og skjerm på, men det krever presisjon å mestre en slik kombinasjon (Whitenton, 2017). For eksempel vil det å bevisst hindre funksjonaliteten til en skjerm, kun for å oppnå "ren" voice-first interaksjon, kunne legge unødvendige begrensninger på et produkt og gjøre det mindre brukbart i tillegg til å øke brukerens kognitive belastning og frustrasjon (Whitenton, 2017).

## **2.4 Taleinteraksjon i TV**

Å benytte taleinteraksjon i TV-sammenheng er ikke en ny idé. Flere av de største aktørene innen streaming har allerede adoptert teknologien og utforsket ulike løsninger for blant

annet talestyrt søk i innhold, avspillingsfunksjoner, justering av lydnivåer og valg av undertekster. Eksempelvis er fjernkontrollen til fjerdegenerasjons Apple TV utstyrt med taleassistenten Siri. Smart-TV-leverandører som Sony, Sharp og Philips støttes av Google Assistant via tilsvarende fjernkontroller (Google, 2018). Amazon Alexa på sin side, støtter smart-TV-er fra blant andre LG og Hisense (Amazon, 2018). I tillegg kan smart-høytalere som Google Home og Amazon Echo utnyttes i forbindelse med smart-TV-er og castingenheter som Chromecast. Tabellen under sammenlikner den innebygde funksjonaliteten som finnes i Google Assistant og Amazon Alexa rettet mot TV og streaming (Google 2018, Amazon 2018).

Felles for de ovennevnte produktene er at de baserer seg på screen-first interaksjon. Taleinteraksjon er lagt til som en tilleggsfunksjonalitet i det eksisterende TV-grensesnittet, for eksempel i form av en implementasjon i fjernkontrollen. Lite utforskning har blitt gjort på voice-first interaksjon i TV-sammenheng, noe som validerer valget om å fokusere på dette i denne oppgaven.

## 3 Teoretisk rammeverk

For å kunne gjøre en vurdering av effekten taleinteraksjon har på brukeropplevelsen i TV-sammenheng, trengs en tydelig forståelse av hva brukeropplevelse er og hvordan det kan evalueres. Brukeropplevelse, eller UX (User eXperience), har flere ulike definisjoner og er et begrep som har hatt en progressiv utvikling siden opprinnelsen i HCI-miljøet på sekstitallet (Simonsen, J. G., 2017). Én ting er hva som definerer en god brukeropplevelse, en annen ting er hvordan man kan kvantifisere og måle om en brukeropplevelse er god. Her finnes det også motstridende modeller. Fokusskiftet fra usability til UX har gjort dette til et sentralt fokus i produktdesign og evaluering (Vermeeren et al., 2010). Selv om interessen for UX er høy både i industri og akademia, mangler fortsatt en systematisk forskning på hvordan UX kan evalueres og måles (Hassenzahl et al., 2006). Mange metoder eksisterer for å gjennomføre tradisjonell usability-evaluering, men UX-evaluering krever et nytt fokus på mindre håndgripelige aspekter (Krömker & Schulze, 2010). Fremveksten av nye interaksjonsformer som taleinteraksjon og multimodale grensesnitt har også bidratt til å gjøre evaluering av UX enda mer komplekst.

I dette kapittelet presenteres de underliggende teoriene og modellene om UX som er utnyttet i dette prosjektet. Kapittelet har til hensikt å danne et rammeverk og presentere verktøy som kan bistå i utvikling og evaluering av en forskningsprototype. Det vil først gis en beskrivelse av begrepets bakgrunn og hvordan det har blitt introdusert og behandlet i akademia. Videre vil det presenteres noen sentrale tilnærminger til begrepet, før jeg vil beskrive Marc Hassenzahls reduksjonistiske syn på opplevelse og hans foreslåtte oppdeling mellom hedonisk og pragmatisk produktkvalitet som blir lagt til grunn for prosjektet.

### 3.1 Hva er UX?

Innenfor HCI dukker begrepet «User Experience» sporadisk opp i akademisk litteratur på 80-tallet (Simonsen, 2017). Tidlige tekster om usability utviste tanker om at de sentrale bestanddelene av usability, som produktivitet og effektivitet ikke var den eneste hensikten et produkt eller en tjeneste skulle tjene (Simonsen, 2017). Whiteside og Wixon (1987) hevdet at det sentrale var den opplevelsen en person har i det øyeblikket produktet eller tjenesten brukes. Den moderne drivkraften for akademisk studie av "User Experience" er hevdet å være en artikkel av Norman, Miller og Henderson Jr. (1995, s. 155) ved CHI'95 som en del av en beskrivelse av designpraksis på Apple Computer Inc. på nittitallet: *“critical aspects of human interface research and application at Apple or, as we prefer to*

*call it, the 'User Experience'*”. Det ser ut til at begrepet var i intern bruk hos Apple så tidlig som i 1993 (Simonsen, 2017).

Populariseringen av begrepet fra nittitallet og frem til i dag har ført med seg en rekke forsøk på å løfte begrepet inn i den akademiske litteraturen, og samtidig konkretisere og definere begrepet. Alben (1996 s. 12) beskriver brukeropplevelse som «alle aspekter av hvordan folk bruker et interaktivt produkt: hvordan det føles i hendene, hvor godt brukeren forstår hvordan det fungerer, hvordan brukeren føler seg i brukssituasjonen, hvor godt det tjener deres formål, og hvor godt det passer inn i hele brukskonteksten». Forsøkene på å konkretisere og definere begrepet har fortsatt frem til dag, men en allment akseptert forståelse av UX mangler fortsatt (Hassenzahl, 2008). Hassenzahl (2008) mener en slik felles forståelse er nødvendig, fordi spriket mellom bruken av begrepet i akademia og industrien er stort. I industrien dukker begrepet svært hyppig opp som en viktig del av markedsføringen av nye produkter. Hassenzahl (2008) hevder dog at begrepet i denne sammenhengen som oftest blir brukt som et synonym for usability og brukersentrert design. I akademia på den andre siden legges det stor vekt på å skille mellom tradisjonell usability og UX.

ISO, den internasjonale organisasjonen for standardisering, gir i ISO 9241-210 følgende definisjon på brukeropplevelse: «...*alle aspekter ved en brukers oppfatninger og responser som resultat av bruk eller forventet bruk av et produkt, system eller tjeneste...*” (ISO, 2010). I henhold til ISO-definisjonen innebærer brukeropplevelsen brukers følelser, tro, preferanser, oppfatninger, fysiske og psykologiske responser, oppførsel og prestasjoner som oppstår før, under og etter bruk. Denne definisjonen er, ifølge Simonsen (2017), for vag til å gi en klar indikasjon på hva UX er.

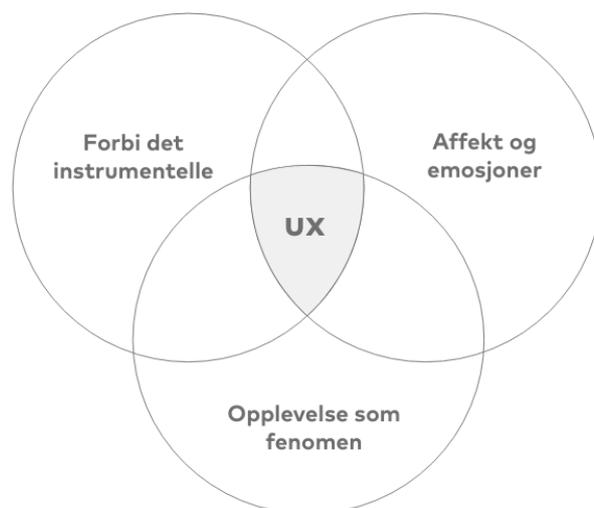
Det er gjort flere forsøk på å kartlegge de ulike synspunktene på definisjoner av User Experience. For eksempel undersøkte Law, Roto og Väänänen-Vaino-Mattila (2009) 270 UX-forskere og utøvere, og fant at respondentene ikke bare hadde svært varierte personlige oppfatninger om hva brukeropplevelsen er og handler om, men også hadde svært varierte kriterier for hva en fornuftig definisjon bør inneholde (for eksempel mente noen respondenter at en definisjon "burde ta opp hva brukeropplevelsen er, snarere enn hva som forårsaker brukeropplevelse").

På samme måte tok Effie Law, Arnold Vermeeren, Marc Hassenzahl, & Mark Blythe (2007), dokumentert gjennom Law et al. (2007) og Blythe et al. (2007), for seg tolv fagfelleverderte artikler omhandlende «*User Experience*», skrevet av 22 ulike forskere,

undervisere og utøvere, fra ulike europeiske land. Funnene fra artikkelstudiene ble videre diskutert i en felles workshop med de ulike bidragsyterne. Målet med workshopen var å danne et manifest som kunne fungere som en referansemodell for fremtidig arbeid med UX. Analysen viste et stort sprik i de ulike bidragsyternes tanker om grunnleggende forutsetninger som ligger bak UX, posisjoneringen av UX i forhold til andre fagfelt og meninger om hva som skal til for å forbedre design og evaluering av UX (Blythe et al., 2007).

Obrist et al. (2012) spurte 70 deltagere i en interessegruppe for User Experience «Hvilke teoretiske røtter vi bygger på, hvis noen, i User Experience-forskning?». Svarene inneholdt 56 ulike teoretiske perspektiver av ulikt omfang, bredde og vitenskapelig opprinnelse (f.eks. "aktør-nettverksteori", "kritisk teori" og "kognitiv belastningsteori"), som tilhører ni tradisjonelle vitenskapelige disipliner (antropologi, kunst, kommunikasjon, design, utdanning, markedsføring, filosofi, psykologi og sosiologi).

Mye av dagens litteratur på området virker å ha akseptert at en felles helhetlig definisjon av brukeropplevelse sannsynligvis ikke er mulig å etablere, og enkelte mener det heller ikke er hensiktsmessig (Simonsen, 2017). Det er allikevel flere temaer og diskusjonspunkter som går igjen i litteraturen. Gjennom en analyse av 28 akademiske artikler, relatert til UX, peker Hassenzahl og Trachinsky (2006) på det de omtaler som tre tydelige «threads», eller tråder, som de hevder tegner et godt bilde av hva UX innebefatter. En tråd dreier seg om å adressere menneskelige behov forbi det instrumentelle, neste tråd legger vekt på affekt og emosjoner i interaksjon og den siste tråden dreier seg om opplevelse som fenomen. Forfatteren hevder at ingen av de tre trådene fanger begrepet fullstendig, men at de alle bidrar til vår forståelse av brukeres interaksjoner med teknologi, samtidig som de deler flere av de samme argumentene (Hassenzahl og Tractinsky, 2006). Denne oppdelingen legges til grunn for utforskningen av UX i denne oppgaven.



Figur 5: De tre trådene som utgjør UX: forbi det instrumentelle, flølelser og affekt, og brukeropplevelse som fenomen (Hassenzahl, 2006).

### 3.1.1 Forbi det instrumentelle

Denne delen av UX-litteraturen tar et oppgjør med det oppgave-orienterte fokuset innenfor HCI-miljøet, der gjennomføringen av funksjonelle mål har vært hovedfokus i evalueringen av kvalitet. Her uttrykkes et ønske om å berike modellene for produktkvalitet med ikke-instrumentelle aspekter, som humor og skjønnhet, for å danne en mer komplett og holistisk interaksjon (Hassenzahl & Trachinsky, 2006). En sentral del av denne tråden er å differensiere UX fra den tradisjonelle begrepet forbundet med brukskvalitet i interaktive produkter: Usability. I den internasjonale standarden ISO 9241, defineres usability som:

*«I hvilken grad et produkt kan brukes av bestemte brukere for å oppnå bestemte mål med 'efficiency', 'effectiveness' og 'satisfaction' i en spesifisert brukskontekst» (ISO, 2010).*

På norsk har vi ikke et eget ord for *efficiency* og *effectiveness*, vi har bare et ord for effektivitet. For å forklare forskjellen på de to begrepene kan man si at *effectiveness* handler om å løse oppgavene, mens *efficiency* handler om å løse oppgavene ved hjelp av minimalt med ressurser (Jordan, 1998). Patrick Jordan beskriver *effectiveness* som kapasiteten et produkt har til å generere eller aktivere output, mens *efficiency* måles som innsatsnivået som investeres i å fullføre en handling eller en oppgave. *Satisfaction* dreier seg om den subjektive tilfredstillelsen brukeren har ved bruken av et produkt, og kan i så måte sees som det aspektet som knytter UX og usability sammen.

Ortony et al.(1988, s. 118) definerer *satisfaction* som «Å være fornøyd med gjennomføringen av en forventet ønskelig begivenhet» (forf. Oversettelse). Med andre ord, hvis folk har forventninger om utfallet av å bruke et bestemt produkt og disse forventningene blir bekreftet, vil de føle seg fornøyd. I motsetning til tilfredsstillelse krever humor eller glede ingen forventninger. Det er definert som å være fornøyd med en ønskelig begivenhet i seg selv (Ortony et al., 1988). Hassenzahl hevder at det ikke nødvendigvis bør ligge forventninger til grunn for å kunne oppleve et produkt som godt.

Hassenzahl hevder dog at hvis man ser nærmere på de faktiske måling av *satisfaction*, ser det ut til at mange av de nåværende tilnærmingene tester brukerens anerkjennelse av designmål i stedet for faktisk brukertilfredshet. For eksempel inneholder Software Usability Measurement Inventory (SUMI, Kirakowski & Corbett, 1993) fem underskalaer (*efficiency, hjelpsomhet, kontroll, lærbarhet og positive følelser*). Bortsett fra *positive følelser*, hevder Hassenzahl (2001) at alle de andre underskalaene kartlegger *effectiveness*- eller *efficiency*-krav for usability. Som et annet eksempel trekker han frem «*End User Computing Satisfaction Instrument*». Denne modellen beskriver *satisfaction* som noe som oppnås dersom underkategorier som innhold, nøyaktighet av informasjon, enkelhet i bruk og aktualitet innfris. Igjen, hevder Hassenzahl, peker alle disse underdimensjonene på opplevd *efficiency* eller *effectiveness* av produktet. *Satisfaction*, slik disse modellene beskriver det, er bare konsekvensen av å anerkjenne kvaliteten som er utformet "inn i" et produkt, noe som fører til en enkel likning: Hvis brukerne oppfatter produktet som *efficient* og *effective*, vil de bli fornøyd. Hassenzahl hevder at denne tilnærmingen neglisjerer bidraget fra ikke-oppgaveorienterte aspekter som moro og i brukskvalitet (Hassenzahl, 2001).

Law og Hornbæk (2007) hevder at god usability er nødvendig, men ikke tilstrekkelig for at en bruker skal bli fornøyd, men god UX derimot er ensbetydende med fornøyde brukere. Plassert på denne måten, er usability underordnet av UX. Men andre hevder at UX bare er en utvidelse av usability for å imøtekomme mindre håndgripelige kvalitetsegenskaper som følelser og moro (Law og Hornbæk 2007). Sentrale utfordringer innenfor denne tråden er å forstå og definere ikke-instrumentelle behov, hvordan ikke-instrumentelle behov kan oversettes til produktkvalitet og hvilke produktsegenskaper er knyttet til hvilke behov (Hassenzahl & Trachinsky, 2006). Bevan (2008) hevder at det å forstå hvordan ulike aspekter av UX relaterer seg til usability vil bidra til et bedre evalueringsgrunnlag.

### 3.1.2 Affekt og emosjoner

«Affekt og emosjoner» omhandler den delen av UX-litteraturen som er interessert i å forstå rollen følelser spiller i interaksjonen med et interaktivt produkt. Ifølge Hassenzahl og Tractinsky (2006), har det siden opprinnelsen vært et stort fokus på å forhindre negative følelser innenfor HCI miljøet, men UX bringer med seg et større fokus på potensialet teknologien har for å fremskape positive følelser. Generelt sett hevder de det er to perspektiver på hvordan følelser behandles i UX: En retning ser på følelser som en konsekvens av bruken av et produkt. Den andre, inspirert av Don Normans (2004) «visceral level of design», ser på følelser noe som kan føre til eller påvirke bruken av et produkt.

Forlizzi og Battarbee (2004) trekker frem følelsene som kjernen i enhver menneskelig opplevelse og som en essensiell del av UX. Fra et psykologisk synspunkt har følelser tre grunnleggende funksjoner: å forme våre planer og intensjoner, å organisere handlingene knyttet til planene, og å evaluere utfall (Forlizzi & Battarbee, 2004). Forfatterne hevder at om man ser følelser fra et designsynspunkt, danner de gapet som eksisterer mellom mennesker og produkter i verden. Følelse påvirker hvordan vi planlegger å samhandle med produkter, hvordan vi faktisk samhandler med produkter, og oppfatninger og utfall som omgir disse interaksjonene. Følelse tjener som en ressurs for å forstå og kommunisere om hva vi opplever.

Desmet et al. (2005) peker på følelsenes evne til å påvirke brukernes oppfattelse av et produkt, ved å introdusere begrepet «Design for Wow». Konseptet med *wow* er beskrevet som en kombinasjon av fascinasjon, positiv overraskelse og tiltrekning. Forfatterne hevder at begrepet *wow* er mer enn et enkelt verbalt utrop som henger sammen med begeistring. Opplevelsen av *wow* er rikere og mer komplekse enn som så, og den fremkaller en komplisert og sammensatt vurdering og har en opplevelsesmessig, fysisk og atferdsmessig påvirkning. På den måten kan design for wow resultere i produkter som folk ikke bare er begeistret for å se - men også begeistret for å bruke og å eie.

### 3.1.3 Opplevelse som fenomen

I tråden «Opplevelsen som fenomen», behandles opplevelse som et dynamisk, komplekst og unikt fenomen. Her vektlegges de kroppslige og psykologiske aspektene ved opplevelse i interaksjon med produkter (Hassenzahl, 2006). Florizzi & Battarbee (2004) introduserer et rammeverk som deler bruker-produkt-interaksjoner inn i tre former: flytende, kognitiv og uttrykkende. Flytende interaksjoner omhandler automatiske og lettlærte interaksjoner.

Disse interaksjonene krever lite eller ingen fokus for å gjennomføres. De lar oss heller fokusere på konsekvensene av interaksjonen eller helt andre ting. Et eksempel på en slik interaksjon er å lage seg en kopp kaffe om morgenen. Kognitive interaksjoner krever tankevirksomhet og fokus på produktet du interagerer med. Denne typer interaksjoner kan resultere i læring eller frustrasjon. Kognitive interaksjoner fører til en endring hos brukeren og ofte i brukskonteksten i tillegg. Med uttrykkende interaksjoner menes de interaksjonene som bidrar til å danne et bånd mellom bruker og produkt. I uttrykkende interaksjoner kan gjerne brukeren endre eller modifisere produktet for å passe bedre med eget selvbylde. Eksempler på slike former for interaksjon er å restaurere et gammelt møbel eller å tilpasse en bil med nye deler.

Wright et. al. (2004) tar en holistisk tilnærming når de forsøker å definere brukeropplevelse. Heller enn å dele opp og studere du ulike komponentene som utgjør en opplevelse isolert, hevder de at for å forstå brukeropplevelse som fenomen, må man forstå sammenhengen mellom de ulike komponentene og hvordan de påvirker hverandre. Basert på Dewey og Bakhtins opplevelsespsykologi, presenter de et rammeverk som deler opp brukeropplevelse i fire sammenflettede og gjensidig avhengige tråder (Wright et. al, 2004). Den kompositoriske tråden omfatter substansen i en produkt eller tjeneste, den sanselige tråden omfatter de sansene som engasjeres i brukssituasjonen, den emosjonelle tråden omfatter de følelsesmessige aspektene i en brukssituasjon. Sistnevnte kan både referere til egne følelser som mestring, glede, skuffelse eller følelser man kan relatere til gjennom empati. Til slutt, den spatio-temporale tråden omfatter hvordan opplevelsen av tid og sted påvirker en situasjon. Wright et.al. konkluderer med at en opplevelse er like mye et resultat av hva brukeren tar med seg inn i en brukssituasjon, som det er et resultat av produktet eller artefakten som tas i bruk. Det denne oppfatningen innebærer er at det ikke er mulig å designe en opplevelse, men at det er mulig, gjennom en god forståelse av brukeren, å designe for en opplevelse.

Florizzi og Battarbee (2004) diskuterer den temporale dimensjonen i opplevelse. De er her opptatt av å skille mellom opplevelse, som den kontinuerlige samtalen mennesker fører med seg selv i bevisst tilstand, og én opplevelse, som oppstår i møtet med et produkt eller tjeneste. Forfatterne stiller her spørsmål ved om en brukeropplevelse innebærer forventninger før bruken av et produkt, refleksjoner etter bruken av et produkt, eller bare den tiden man faktisk interagerer med produktet. Viktige spørsmål i denne tråden er hvordan man kan skille en opplevelse fra en retrospektiv bedømmelse av en opplevelse. Dette er et viktig spørsmål, fordi det er slike retrospektive bedømmelser som blir

representasjonen av en opplevelse. De påvirker fremtidige opplevelser og danner grunnlaget for kommunikasjonen om en opplevelse (Hassenzahl og Trachinsky, 2006) .

### **3.2 Reduksjonisme vs. holisme**

I de ulike studiene som undersøker oppfatninger og definisjoner av UX i faglitteraturen trekkes det frem to sentrale tilnærminger: en holistisk og en reduksjonistisk (Law et al., 2007; Blythe et al., 2007; Obrist et al., 2012; Law et al., 2009; Hassenzahl og Trachinsky, 2006). Denne todelingen er forankret i den gamle filosofiske debatten om holisme mot reduksjonisme (Law et al., 2014). Blant forskerne med en holistisk tilnærming argumenteres mot abstraherte modeller av opplevelse og det legges vekt på meningen av opplevelse som fenomen og den subjektive unikheten i en opplevelse. Denne siden av litteraturen problematiserer dekonstruksjon av opplevelse, og mener at en opplevelse kun kan forstås som summen av dens komponenter.

På den andre siden har tilnærmingene som baserer seg på eksperimentell psykologi en tendens til å dekonstruere opplevelse i underkomponenter (for eksempel motivasjon, tillit, aversjon, hedonikk, moro osv.). Disse ulike komponentene kompletteres ytterligere av prosesser, som for eksempel adresserer det tidsmessige aspektet av erfaring eller dynamikken i psykologiske tilstander ( Law et al., 2007). Det holistiske perspektivet fokuserer på den detaljerte analysen av erfaring, og gir rik innsikt i spesifikke interaksjoner. Disse modellene baserer seg gjerne på svært få, kvalitative - men detaljerte - eksempler på opplevelse. Den reduksjonistiske retningen ser snarere etter mer generelle prinsipper og mekanismer som er basert på "gjennomsnittsmålinger" fra kvantitative studier. Styrken til en tilnærming basert på aggregert kunnskap er den potensielle enkelheten av de resulterende modellene (Law et al, 2007). Samtidig medfører dette - per definisjon - en reduksjon. Denne reduksjonen har kostnader. For eksempel blir det stilt spørsmål ved om en slike abstraherte gjennomsnittsmoeller faktisk kan fungerer for ekte brukere og deres ekte opplevelser.

Reduksjonistiske tilnærminger forsøker å forenkle kompleksiteten i en opplevelse, for å kunne beskrive og operasjonalisere begrepet. For eksempel ved å måle ulike individuelle komponenter som spiller en rolle for den totale brukeropplevelsen. En holistisk tilnærming forsøker å sammenflette ulike aspekter som følelser, affekt, estetikk, hedoniske aspekter, for å forstå det store bildet. Den holistiske tilnærmingen frastår fra å dele en opplevelse opp i underkomponenter og mener at en opplevelse kun kan forstås som en helhet.

### 3.3 Syn på evaluering

Å definere evalueringsprinsipper og tilhørende metoder for UX er også et hyppig debattert tema (Law et al., 2014). Law et al. (2014) hevder at evaluering og definisjon bør henge sammen, og at disse to grunnleggende problemene må løses parallelt. Da det som illustrert over er stor uenighet og sprikende meninger rundt definisjonen av UX, er det naturlig nok like stor uenighet om hvordan UX kan evalueres. Roto et al. (2010, s.8) beskriver at «...ingen generelt akseptert helhetlig måling av UX eksisterer, men UX kan evalueres på mange forskjellige måter...» (forf. oversettelse). Det er en viss enighet om at det finnes måter å vurdere UX på, men at en opplevelse ikke kan måles i sin helhet (Blythe et al., 2007).

HCI-feltet, der UX er forankret, har arvet teoretiske konsepter, epistemologiske antagelser og metoder fra et mangfold av disipliner. Fra ingeniørfag, hvor måling er sterkt omfavnet, til humaniora, hvor måling kan betraktes som naivt eller overforenklet, spesielt når begrepene som skal måles, er udefinerte og åpne for tolkning (Law et al., 2014). Da UX utnytter en rekke lite håndgripelige opplevelseskvaliteter, som glede, avsky, overraskelse og kjærlighet, for å nevne noen få, er uenighet rundt målingen av UX uunngåelig (Law et al., 2014).

I samsvar med de to like tilnærmingene til opplevelse, som holistisk og reduksjonistisk, er det også to motstridende tilnærminger til hvordan hvordan UX bør evalueres. Det sentrale skille mellom de to tilnærmingene er legitimiteten ved å bryte ned opplevelse i mindre komponenter for å studere komponentene enkeltvis. UX-forskere kan, i følge Law et al. (2014) grovt deles inn i to leire, som kan betegnes som "kvalitativ designbasert UX-forskning" og "kvantitativ modellbasert UX-forskning" (Law, 2014). I den «modellbaserte» leiren legges det vekt på å representere UX på en bestemt, sammenlignbar og generaliserbar måte, mens det i den «designbaserte» leiren legges vekt på å artikulere rike legemlige (*Rich embodied*) opplevelser med kontekst (Boehner et al. 2007) og det utvises en sterk motvilje mot å redusere menneskelige erfaringer og følelser til tall (Hornbæk & Law, 2007).

Forlizzi og Battarbee (2004 s.265) hevdet for eksempel at: "... følelsesmessige responser er vanskelig å forstå, og enda vanskeligere å kvantifisere.". Tilsvarende skriver Swallow et al. (2005 s.91) at: "... reduksjonistiske tilnærminger kan være nyttige for eksperimentell analyse, men de kan miste noen av innsiktene som er tilgjengelige i tilnærminger som

motstår slik reduksjon ... kvalitative data gir en rikdom og detalj som kan være fraværende i kvantitative målinger" (forf. overs.).

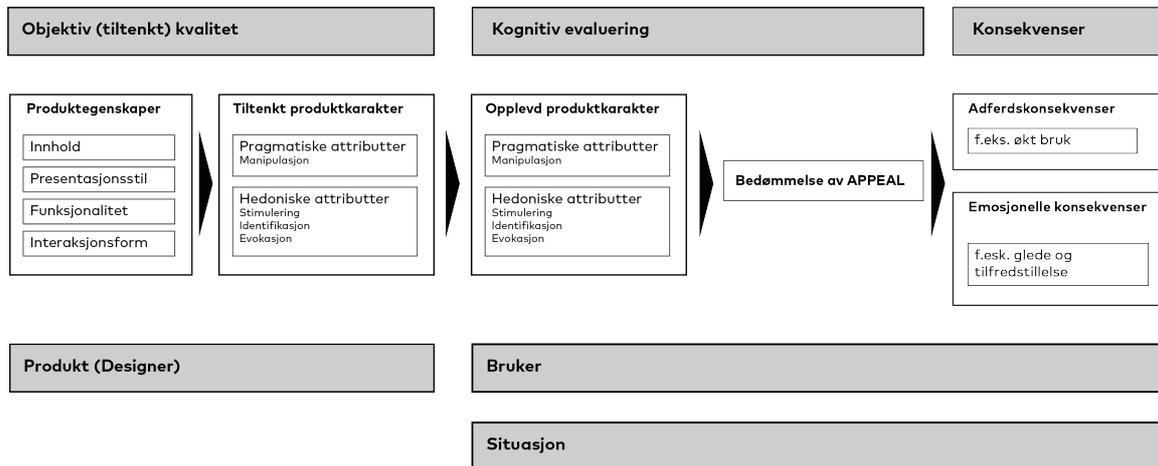
På den andre siden hevder Hassenzahl (2008) at menneskers individuelle erfaringer med teknologi er mye mindre unike og varierende enn det en fenomenologisk tilnærming foreslår. Tractinsky (i Roto et al., 2010) mener at som en kompleks konstruksjon bør UX studeres med vitenskapelige metoder og at det er nødvendig å utvikle kvantitative tiltak og måleinstrumenter for å teste og forbedre UX. I motsetning til dette blir noen svært negative holdninger til målinger og reduksjonisme uttalt av Höök (Roto et al., 2010, s.17): «...Spørsmålet er om måling av sluttbrukeropplevelsen som noen enkle, målbare variabler, virkelig hjelper oss til å forstå brukeropplevelsen... Etter min mening er det for mange reduksjonistene der ute som skader forskningen på dette området ved å late som at vi kan gi målinger og metoder som gjør at noen kan vurdere UX-verdien av et designet system...» (forf. overs.).

I denne oppgaven sees det som hensiktsmessig, på tross av de motstridende meningene, å følge en reduksjonistisk tilnærming til evaluering av UX. Begrunnelsen for dette er todelt. For det første vil en holistisk evaluering kun være mulig å gjennomføre med en prototype som har nådd en tilstrekkelig grad av oppløsning til å kunne evalueres med ekte brukere over tid i en reell kontekst. Dette er ikke mulig å gjennomføre i dette prosjektet, da resurser for utvikling og tid til gjennomføring og evaluering av studiene ikke er tilstrekkelig. Som illustrert over vil enkelte hevde at en holistisk tilnærming kunne gitt et større mangfold i datamaterialet og på den måten bidratt til en dypere innsikt og et bedre svar på problemstillingen. På den andre siden kan det hevdes at en reduksjonistisk tilnærming, der de ulike deltagerens oppfatninger brytes ned til kvantifiserbar data, vil kunne gi en tydeligere indikasjon på om prototypen som tas frem, faktisk foretrekkes fremfor en tradisjonell TV-løsning.

### **3.4 Marc Hassenzahl: Den hedonisk/pragmatiske modellen**

Den tyske design-psykologen Marc Hassenzahl foreslår en modell for UX som deler produktkvalitet inn i to hovedkategorier: pragmatisk kvalitet og hedonisk kvalitet, og videre antyder at disse to kategoriene har like stor innvirkning på brukertilfredsheten en bruker opplever i et produkt (Hassenzahl, 2008). Hassenzahls modell definerer nøkkelelementene i brukeropplevelse og deres funksjonelle sammenheng. Modellen tar særlig hensyn til spørsmål som brukeropplevelsens subjektive natur, persepsjon av et

produkt og følelsesmessige responser til et produkt i varierende situasjoner (Hassenzahl, 2003). Modellen består av tre separate lag: objektiv produktkvalitet (tilsiktet av designeren), subjektiv oppfatning og evaluering av kvalitet ved bruk, og følelsesmessige og atferdsmessige konsekvenser (Hassenzahl, 2001).



Figur 6: Hassenzahls pragmatisk/hedonisk UX modell. (Gjengitt fra Hassenzahl, 2001 og Hassenzahl, 2003).

### 3.4.1 Produktkarakter

I følge Hassenzahl har produkter bestemte egenskaper (innhold, funksjonalitet, presentasjonsstil osv.) som er valgt og kombinert av en designer for å formidle en bestemt *tiltenkt produktkarakter*. En produktkarakter er en høynivå beskrivelse som oppsummerer et produkts attributter, for eksempel «nyskapende», «interessant», «nyttig» eller «forutsigbar». Karakterens funksjon er å redusere kognitiv kompleksitet og å utløse bestemte strategier for håndtering av produktet (Hassenzahl, 2001). Når brukere kommer i kontakt med et produkt, utløses en prosess. Først oppfattes produktets egenskaper, basert på disse konstruerer hver enkelt bruker en personlig versjon av produktkarakteren – en *opplevd produktkarakter*. Videre fører den *opplevde produktkarakteren* til *konsekvenser*: en vurdering av produktets tiltrekningskraft, eller *APPEAL*.

Hassenzahl sammenlikner oppfattelsen av den opplevde produktkarakteren, og den påfølgende evalueringen av et system med kognitiv evaluering (*Cognitive appraisal*). Innenfor psykologien forklarer *Cognitive appraisal theory* menneskers atferd- og følelsesresponser i en situasjon (Scherer, K. R., 2005). Teorien bygger på at mennesker oppfatter en situasjon basert på sine subjektive tolkninger av en situasjonen og situasjonens

betydning for dem (Scherer, K. R., 2005). Dette innebærer at mennesker kan ha ulik oppfatning av samme situasjon fordi den kan ha ulik betydning for dem. I følge Hassenzahl kan denne evalueringen ha to utfall. På den ene siden kan den føre til atferdskonsekvenser, i form av for eksempel økt bruk av et produkt eller økt kvalitet på arbeidet som utføres gjennom produktet. På den andre siden kan det føre med seg emosjonelle konsekvenser, som følelsen av glede, eller å ha det gøy (eller frustrasjon eller skuffelse). Atferdsmessige og følelsesmessige konsekvenser kan være relatert til hverandre. For eksempel fant Igarria et al. (1994) at opplevd glede og bruksmengde i software systemer var korrelert. Konsekvensene av en bestemt produktkarakter er imidlertid ikke alltid det samme, de avhenger av den gitte brukssituasjonen. Samsvaret mellom tiltenkt og opplevd kvalitet i et produkt kan være lavt (Kurosu & Kashimura, 1995), noe som indikerer forskjeller i hvordan designere tenker på et produkt og hvordan brukerne oppfatter det.

Hassenzahl hevder at den opplevde produktkarakteren er en kognitiv struktur som representerer produktegenskaper og relasjoner som angir samspillet mellom egenskapene. Den tillater brukeren å gjøre slutninger om produktet utover det som kan oppfattes på overflaten. For eksempel kan et produkt med et enkelt brukergrensesnitt også vurderes som enkelt å bruke, selv om brukeren ikke har noen praktisk erfaring med grensesnittet. Hassenzahl skiller disse egenskapene i fire kategorier utfra produktets hovedfunksjoner: Produkter gjør det mulig for folk å manipulere sine omgivelser, stimulere personlig utvikling og uttrykke identitet. Videre kan et produkt evokere minner og dermed ha en symbolsk verdi.

### **3.4.2 Pragmatiske egenskaper**

Å manipulere av omgivelsene krever relevant funksjonalitet (dvs. verktøy) og måter å få tilgang til denne funksjonaliteten (dvs. usability). Hassenzahl kaller denne gruppen av produktegenskaper *pragmatiske egenskaper*. Typiske pragmatiske egenskaper er "tydelig", "støttende", "nyttig" og "kontrollerbart". Et pragmatisk produkt er primært instrumentelt. Det brukes til å oppfylle eksterne eller interne atferdsmål. Hvis for eksempel noen ber deg om å spikre en spiker i en vegg for å sette opp et bilde, bruker du et verktøy for å gjøre det. Fra et pragmatisk perspektiv er de eneste kravene til verktøyet at det i prinsippet kan brukes til å spikre inn en spiker, og at du kan finne ut hvordan du gjør det.

### **3.4.3 Hedoniske egenskaper: stimulans, identifikasjon og affeksjon**

De øvrige produktegenskapene ansees av Hassenzahl som hedoniske. Mens pragmatiske egenskaper legger vekt på oppfyllelsen av brukerens atferdsmål, legger hedoniske

egenskaper vekt på brukerens psykologiske velvære. Typiske hedoniske egenskaper er "fremragende", "imponerende", "spennende" og "interessant". De hedoniske egenskapene til produktene kan videre deles inn i tre kategorier: stimulans, identifikasjon og affeksjon.

- **Stimulans:** Stimulans dreier seg om menneskers iboende ønske om å tilegne seg kunnskap og kompetanse. For å kunne bidra i dette må produkter være stimulerende. De må tilby nye inntrykk, muligheter og innsikt. Stimuleringen fra ny, interessant eller til og med spennende funksjonalitet, innhold, presentasjon eller interaksjonsstil vil også indirekte kunne bidra til måloppnåelse. Det kan øke oppmerksomheten, kompensere for manglende motivasjon til å oppfylle eksternt oppgitte mål, eller fasilitere for nye løsninger på problemer.
- **Identifikasjon:** Mennesker uttrykker seg selv gjennom fysiske eiendeler. Denne selvuttrykkende funksjonen er fullstendig sosial. Enkeltpersoner ønsker å bli ansett på bestemte måter av andre. Å bli sosialt anerkjent og å utvise makt til andre er et grunnleggende domene for menneskelige motiver (Schwartz & Bilsky, 1987). For å oppfylle dette behovet må et produkt kommunisere identitet. Generelt foretrekker mennesker produkter som kommuniserer fordelaktige identiteter til andre.
- **Affeksjon:** Dreier seg om et produkts evne til å fremkalle minner. I denne sammenhengen representerer produktet tidligere hendelser, relasjoner eller tanker som er viktige for den enkelte brukeren (Prentice, 1987). Suvenirer er for eksempel en hel produktkategori som bare gir symbolsk verdi ved å holde minner om en hyggelig reise levende. Et mer teknologirelatert eksempel er vintage dataspill. Spillene tilbyr verken kompleksitet eller slående grafikk. Verdien kommer fra å utløse minner fra de gode gamle dager, da disse spillene var spennende og fengslet folk i timevis.

#### 3.4.4 APPEAL

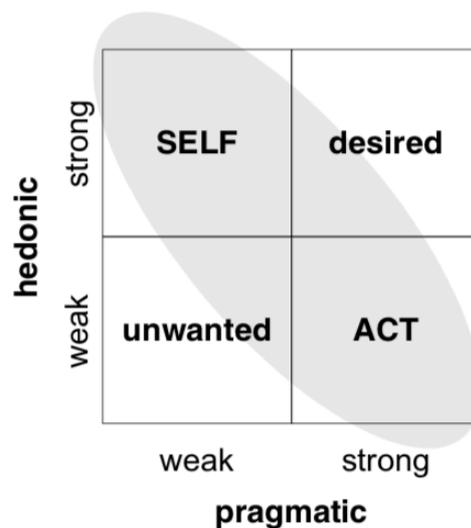
Hassenzahl introduserer begrepet «*APPEAL*», som kan oversettes til «grad av tiltalenhet» (Hassenzahl, 2001). Hvis et produkt er i stand til å utløse positive følelsesmessige reaksjoner, er det tiltalende. Tiltalende produkter har egenskaper som god, sympatisk, hyggelig, attraktiv, motiverende, ønskelig og innbydende. Et produkts *APPEAL* er situasjonsavhengig. For eksempel kan enkeltpersoner vurdere et produkt som tiltalende fordi målene som oppnås av produktet, er av høy relevans for dem i en bestemt situasjon.

Imidlertid kan andre personer (eller til og med samme person) vurdere det samme produktet som mindre tiltalende, fordi de er mer interessert i å kommunisere en gunstig identitet til andre enn å oppnå atferdsmål. Kort sagt sammenfører *APPEAL* opplevelser og følelser overfor et produkt i en bestemt situasjon til en målbar enhet. På grunnlag av deres oppfatninger, danner brukerne en vurdering av produktets *APPEAL*. Med andre ord kan brukeren vekte og kombinere oppfattet EQ og HQ i en enkelt dom. *APPEAL* (eller mangel på) manifesterer seg i en global vurdering av produktene (for eksempel god vs dårlig).

### **3.4.5 ACT og SELF produktkarakterer**

Hassenzahl ser på pragmatiske og hedoniske egenskaper som uavhengige av hverandre. Sammensetningen av disse egenskapene utgjør produktkarakteren (Hassenzahl, 2003). Ved å ta i betraktning at folks oppfatning av pragmatiske og hedoniske egenskaper kan være svake eller sterke, danner han fire typer produktkarakterer. Det understrekes at produkter kan være pragmatiske eller hedoniske av ulike årsaker (Hassenzahl, 2003). For eksempel kan et verktøy av et bestemt merke være hedonisk fordi dette verktøyet kommuniserer profesjonalitet (dvs. kommuniserer identitet). Andre verktøy kan være hedoniske fordi de er innovative og stimulerer brukeren til å gjøre spennende nye ting.

Kombinasjonen av svake pragmatiske og svake hedoniske egenskaper gjør et produkt overflødig og uønsket (Hassenzahl, 2003). Det er et tegn på at et produkt ikke er i stand til å tilfredsstille hverken pragmatiske eller hedoniske behov hos potensielle brukere. Kombinasjonen av sterke pragmatiske og sterke hedoniske egenskaper gjør et produkt er *desired*, eller ønsket. En kompromissløs kombinasjon av begge er det ultimate designmålet. Det er vanskelig å skape en perfekt balanse mellom disse egenskapene, og de fleste produkter har en overvekt av enten det ene eller det andre. Hassenzahl kaller et primært pragmatisk produkt (dvs. "sterk pragmatisk / svak hedonisk") et *ACT*-produkt, og et primært hedonisk produkt (dvs. "svakt pragmatisk / sterkt hedonisk produkt") et *SELF*-produkt.



Figur 7: Forholdet mellom pragmatisk og hedonisk kvalitet (Hassenzahl, 2003).

Et ACT-produktet er sterkt knyttet til brukerens atferdsmål og avhenger sterkt av brukerens målsetninger. På den andre siden er SELF-produktet knyttet til brukern selv, gjennom for eksempel idealer, minner og relasjoner. Verdien av SELF-produkter er mye mer stabil enn verdien av ACT-produkter, fordi sannsynligheten for at enkeltpersoner forandrer det de trenger fra et produkt for å tilfredsstille seg selv, er mye lavere enn sannsynligheten for at atferdsmålene endres. Videre hevder Hassenzahl at båndet mellom et *SELF*-produkt og brukeren generelt er mye sterkere enn båndet mellom et *ACT*-produkt og brukeren. Først når de atferdsmessige målene som oppnås via *ACT*-produktet har høy personlig relevans, kan det forventes et sterkt bånd mellom et *ACT*-produkt og brukeren. Hassenzahl mener dette understreker betydningen av hedoniske egenskaper, og at kun produkter som gir brukeren muligheten til å relatere seg til det på et personlig nivå vil bli verdsatt over tid (Hassenzahl, 2003).

### 3.4.6 Situasjonen: Handlingsmodus og målmodus

Hassenzahl understreker gjentatte ganger viktigheten av ulike situasjoner for å forstå brukeres bedømmelse av *APPEAL* (Hassenzahl og Roto, 2007). En brukssituasjon kombinerer den opplevde produktkarakteren med et bestemt sett med ambisjoner, for eksempel bestemte atferdsmål eller behov for stimulering. Disse situasjonene vil naturligvis kunne variere i stor grad, noe som gjør det utfordrende å forutsi de følelsesmessige reaksjonene eller vurderingen av *APPEAL* i en gitt situasjon. Som en løsning på dette problemet foreslår Hassenzahl å fokusere på brukerens mentale tilstand

ved å definere ulike bruksmoduser (se Hassenzahl, Kekez, & Burmester, 2002). Spesifikt skiller han mellom et «handlingsmodus» eller «action mode» og et «målmodus» eller «goal mode».

Når man bruker et produkt har man alltid atferdsmål, og man gjør tiltak for å oppfylle disse målene. I målmodus er måloppnåelse i fokus. Det aktuelle målet har en viss betydning og bestemmer alle handlinger. Produktet er derfor bare et middel for å nå målet, og individer prøver å være effektive. I handlingsmodus er handlingen i fokus. Den gjeldende handlingen bestemmer målene underveis og målene er "flyktige". Bruk av produktet kan være et mål i seg selv. I handlingsmodus spiller effektivitet en mindre rolle. I prinsippet anser Hassenzahl bruksmodusene som psykologiske tilstander, og alle produkter kan oppleves i begge tilstander. Oppfattelsen av et produkt karakter som primært pragmatisk eller hedonisk vil ikke bli påvirket av bruksmodus, men vurderingen APPEAL og følelsesmessige reaksjoner avhenger sterkt av hvordan produktet passer inn bruksmodusen (Hassenzahl, 2003).

### **3.4.7 AttrakDiff**

I en studie rettet mot utprøving av deler av modellen, nærmere bestemt den kognitive evalueringen, konsentrert, undersøkte Hassenzahl et al. (2000) om brukerne oppfatter EQ og HQ uavhengig og hvordan bedømmelse av APPEAL er avhengig av oppfatningen av de to forskjellige kvalitetsaspektene. Studien tok form av en semantisk differensial bestående av motstridende 28 ordpar. Antonymene er plassert i hver sin ende av en syvstegs skala, og brukeren vurderer sin opplevelse av produktet ved å markere en verdi på skalaen. Studien konkluderer med at brukerne oppfatter EQ og HQ aspekter uavhengig av hverandre, og disse egenskapene ser ut til å bidra jevnt til den overordnede vurderingen av et produkts APPEAL. Denne undersøkelsen (Hassenzahl et al., 2000) la grunnlaget for det som senere har blitt til en anerkjent og standardisert spørreundersøkelse (AttrakDiff), som brukes for å vurdere brukeropplevelsen i interaktive systemer (Walsh, 2014).

## 4 Metodisk tilnærming

I det følgende kapittelet presenterer jeg *Research through Design* (RtD) som metodisk tilnærming og videre hvordan denne metodikken har guidet utviklingen og evalueringen i dette prosjektet. Som diskutert i det teoretiske rammeverket i denne oppgaven er det svært ulike syn på evaluering av UX. Metodene som legges til grunn i dette prosjektet er bestående både av kvalitative og kvantitative metoder. Utviklingen av prototypen baserer seg på en brukersentrert iterativ designprosess, der brukere involveres gjennom hele prosessen. Den komparative studien som konkluderer oppgaven, evalueres med en standardisert spørreundersøkelse (AttrakDiff) og ytelsesmål.

### 4.1 Research through Design

Research Through Design, vil enkelt forklart si at en artefakt eller en prototype, brukes til å gi innsikt i en alternativ eller fremtidig situasjon eller en teknologi. I dette prosjektet er innsikten som genereres mer interessant enn selve prototypen som utvikles. Jeg ønsket i dette prosjektet å utvikle en prototype som presenterer brukeren for en potensiell forestilt fremtid, og bruke denne til å generere kunnskap om brukernes preferanser og teknologiens nytteverdi. RtD er en forskningstilnærming som er i fremvekst innen HCI. Tilnærmingen fokuserer på å skape og utnytte artefakter og prototyper for å generere ny kunnskap om et fenomen eller et felt. Med tanke på mitt forskningsspørsmål er min tilnærming sterkt knyttet til prinsippene fra RtD. I den påfølgende delen presenteres et kort overblikk over RtD som metode, videre beskrives min spesifikke RtD-tilnærming som blir fulgt i dette prosjektet.

RtD som tilnærming har hatt en voksende interesse innenfor HCI i de siste tiårene, men ble først introdusert innenfor kunst og design. Sir Christopher John Frayling presenterte navnet i 1993, i en artikkel der han problematiserte at design og forskning ble behandlet som to separate enheter (Frayling, 1993). Frayling understreket i sin artikkel hvilken betydning forskning har hatt, vil ha og kan ha på praksis og undervisning innenfor design og kunst. Han presenterte tre måter hvor han forklarer hvordan forskning kan forenes med design: forskning på design, forskning gjennom design og forskning for design. Forskning på design omhandler forskning på den menneskelige aktiviteten design. Forskning for design omhandler forskning som har til formål å fremme utviklingen av design (Zimmerman & Forlizzi, 2014, s. 169). Forskning gjennom design, eller RtD (Research through Design), er snarere en forskningspraksis der fokuset ligger på forbedre verden gjennom å bruke

design som et middel for å «forstyrre eller forandre dagens tilstand» (Zimmerman & Forlizzi, 2014, s. 169). Det er en tilnærming til å gjennomføre forskning og handler i følge Frayling om å ta et problem utenfor design og bruke design til å løse problemet og generere ny kunnskap (Durrant & Price, 2015). Forskeren spekulerer i fremtiden og bygger sitt designarbeid på en grundig forståelse av brukergruppen og anvendelsen av tilgjengelig teknologi (Zimmerman & Forlizzi, 2014). Zimmerman & Forlizzi (2014, s. 167) hevder at RtD er «en tilnærming til å gjennomføre vitenskapelig forskning som benytter metoder, praksis og prosesser fra designverden med det formål å generere ny kunnskap» (forf. overs.). Designere kan på denne måten dra nytte av deres ferdigheter ved å studere et problem gjennom design. Denne prosessen innebærer som regel å lage prototyper eller forskningsprodukter som kan bidra til å skape kunnskap om et fenomen og utforske alternative framtidsscenarioer.

Gaver (2014) viser essensen av RtD ved å påpeke forskjellene mellom forskning gjennom vitenskap og forskning gjennom design. I sin artikkel gjør han et poeng ved å polarisere vitenskap og design som noe som trenger å «forsvare sin kunnskap på den ene siden og at sitt produkt virker på den andre» (Gaver, 2014, p. 148). Han forklarer hvordan vitenskapen ikke bryr seg om hvor interessante eller nyskapende resultatene er, eller om hvilke problemer forskningen løser. Vitenskapen er opptatt av å bevise at kunnskapen som er samlet og kommunisert er sann, ved å forklare og forsvare prosessen og de konseptuelle, teoretiske og praktiske avgjørelsene som er tatt. Det viktigste er om forskningen er reviderbar, objektiv, generaliserbar og basert på kausale forklaringer, og at den kan forklare fenomener og forutsi nye. På den andre siden, mener han, er design fundamentalt opptatt av å lage noe nyskapende som fungerer, og ikke av å forklare verden slik den er. Videre introduserer han konseptet «epistemological ambiguity», et konsept som beskriver «den flytende strømmen mellom usikkerhet og spekulasjon som design tillater» (Gaver, 2014, p. 153).

Når et produkt utvikles, forklarer han, utforsker designeren kontekster, utvikler et design space, kommer med forslag, velger, raffinerer og lager et fungerende design, og lærer om designet ved å plassere det i hendene til brukerne. Gjennom denne prosessen av sammensveising mellom ideer, materialer, teknologier, situasjoner, mennesker og kulturer generer designeren kontinuerlig ny kunnskap (Gaver, 2014, p. 162). Ifølge Gaver (2014) er det nettopp denne kunnskapsgenereringen, gjennom design praksis, som utgjør RtD. RtD bør, ifølge Gaver (2014) ikke ansees som et forsøk på å bringe vitenskapens prinsipper inn i design, men som en selvstendig tilnærming som bruker en kreativ utviklingsprosess som verktøy for å lære om mennesker, teknologier og om verden. Når jeg ser tilbake på mitt

forskningsspørsmål, er en slik alternativ tilnærming et verdifullt verktøy og en nødvendighet i min forskningstilnærming. Med en slik tilnærming har jeg kunnet utvikle en prototype for å generere ny innsikt i hvordan interaksjonen mellom menneske og maskin generelt og mellom menneske og TV og streamingtjenester spesielt, vil kunne se ut i fremtiden.

RtD som designforskningsmetode er som ovennevnt relativt nytt innen HCI, og spørsmål om kvaliteten på forskningsbidragene diskuteres fortsatt (Gaver, 2012; Höök et al., 2015). Forskere på feltet stiller spørsmål ved hva slags kunnskap RtD genererer, prototypens eller artefaktens rolle, om det er en bestemt måte å utføre RtD-prosjekter på og hvordan den genererte kunnskapen skal dokumenteres og kommuniseres. Det finnes ingen entydig svar på disse spørsmålene, og ulike forskere har ulike meninger om kunnskapsbidragene RtD som metode kan tilby (Gaver, 2012; Koskinen et al., 2011; Zimmerman & Forlizzi, 2014, Zimmerman, 2007). Gaver (2012) hevder det er positivt at det ikke finnes noen standardisert form for Rtd, og at mangfoldet i tilnærminger er fruktbart. Koskinen et al. (2011) undersøker hvordan RtD, eller «Constructive design», som han kaller det, har blitt utført med suksess i de siste tiårene. De skisserer et rammeverk bestående av tre varianter av design research som de kaller «Lab», «Field» og «Showroom». I den følgende seksjonen vil disse tre variantene bli beskrevet med et kort overblikk, før jeg vil argumentere for hvorfor Lab-tilnærmingen egnet seg best som utgangspunkt for dette prosjektet.

#### **4.1.1 Lab, Field, and Showroom**

##### ***Lab***

Lab-tilnærmingen har sine røtter i naturvitenskapen, men er også mye brukt i design. Koskinen et al. (2011) hevder at det kontrollerte miljøet som tilbys av Lab-tilnærmingen, kan være gunstig når man tester konkurrerende hypoteser eller alternative forklaringer. Lab-tilnærmingen gir forskeren muligheten til å kontrollere bestemte variabler lettere enn i en naturlig setting, og dermed se på en ting om gangen (Koskinen et al., 2011, s. 55). I en laboratorium-setting kan en foreta detaljerte målinger og observasjoner av deltakerens oppførsel og responser (Koskinen et al., 2011, s. 56). Fordelene nevnt ovenfor kommer med en pris, og Koskinen et al. (2011, s. 63) stiller spørsmål ved om slike kontrollerte studier kan fortelle oss noe om den virkelige verden.

### ***Field***

Målet med en Field-tilnærming er å se hvordan mennesker interagerer med et design i en naturlig kontekst. Ifølge Koskinen et al. (2011), kontekstualiserer Field tilnærmingen den design forskningen som er utført. De kaller prosessen ved å kontekstualisere og forstå den virkelige verden «design etnografi», fordi metodene som brukes for å skape forståelse finner sted ute i felten. Field-tilnærmingen har mye til felles med måten sosialforskere arbeider på, men tilføyer en ekstra komponent ved å plassere et produkt, en ting eller en prototype ut i verden for å generere ny kunnskap (Koskinen et al., 2011, p. 75). Forfatterne forklarer videre at prototyper vanligvis blir utviklet for å skape en dialog mellom brukerne og forskerne, og for å fokusere på hva brukerne gjør i stedet for hva de sier. Prosessen når en designidé fremgår av feltarbeid kan noen ganger virke mystisk for ikke-designere. Dette aspektet ved Field tilnærmingen påpekes av forfatterne som et potensielt problem. Et annet problem de peker på er at kunnskapen skapt fra feltforskning gjerne er lokal og situasjonsbestemt, og ikke ikke ukritisk kan brukes andre sammenhenger (Koskinen et al., 2011, s. 81).

### ***Showroom***

I showroom-tilnærmingen oppretter og utvikler designforskere "provoserende gjenstander som tvinger folk til å tenke, legge merke til, og å revurdere et visst aspekt av verden" (Zimmerman & Forlizzi, 2014, s. 173). Tilnærmingen er først og fremst påvirket av kritisk design, en tilnærming som utfordrer vanlige antagelser og status quo gjennom design av artefakter (Zimmerman & Forlizzi, 2014). Tilnærmingen er avhengig av en debatt for å fremkalle forandring i hverdagen til mennesker, og en vanlig måte å presentere objekter opprettet i designprosessen er gjennom prototyper, fotografier, video, bøker og artikler (Koskinen et al., 2011).

## **4.2 Min tilnærming til RtD**

I dette prosjektet, har jeg fulgt en Lab-tilnærming. Utgangspunktet for prosjektet er å utforske om talestyring, som en spesifikk interaksjonsform, er egnet for å samhandle med TV og strømmetjenester, sammenliknet med andre interaksjonsformer. Denne problemstillingen vil utforskes gjennom å utvikle en prototype, for så å sammenlikne denne i en komparativ studie med en tradisjonell TV-løsning.

Prototypen utvikles gjennom en brukersentrert iterativ designprosess, der brukere involveres gjennom hele prosjektet. Kvalitative metoder som intervjuer, observasjoner og en workshop legges til grunn for utformingen av systemet. Teknologiske begrensninger vil i stor grad bli utelatt i studien, for å oppnå et system som gjenspeiler en ideell fremtid, der NLP fungerer feilfritt. På denne måten kan prototypen, i tråd med ideene i RtD, gi et bilde på hvordan en fremtidig situasjon eller teknologi vi arte seg. For å kunne oppnå en prototype med tilstrekkelig oppløsning til å brukes i en komparativ studie, vil mye av funksjonaliteten simuleres i form av Wizard of Oz. De ulike metodene beskrives nærmere i kapittel 5.

Evalueringen av prototypen vil skje i form av et kontrollert komparativt laboratorieforsøk. I forsøket måles prototypen opp mot en tradisjonell TV-løsning, med brukeropplevelse som variabel. Forsøket utføres i form av en kooperativ evaluering, der testleder aktivt leder testpersonen gjennom et sett med predefinerte scenarioer (Bevan, 1995). Scenarioene beskriver en bestemt oppgave som løses gjennom å interagere med systemet. Fordelen med et slikt forsøk er at de tillater god kontroll over oppgavene som utføres av testpersonene, samtidig som det er mulig å dekke et stort spekter av ulike situasjoner i interaksjonen (McTear et al., 2016). I et labriatorieforsøk har man derfor muligheten til å kontrollere ulike variabler og se hvordan en bestemt endring i et system kan påvirke opplevelsen. Scenariene gjør det også enklere å sammenlikne resultatene til de to ulike systemene (McTear et al., 2016). For å forsøke å gjenskape en så realistisk kontekst som mulig, vil testen bli utført i de respektive deltagerens TV-stue.

Å følge en Field-tilnærming og ta frem en prototype som brukerne kunne bruke i en naturlig setting i hjemmet, over tid, ville sannsynligvis kunne gi et mer nyansert bilde av den opplevde brukskvaliteten og kunne kanskje ha sagt noe mer om hvordan taleinteraksjon egner seg i denne sammenhengen. For å kunne la brukere interagere fritt, er man avhengig av å ha nådd en tilstrekkelig høy oppløsning på produktet som skal testes. I dette prosjektet vil ikke dette ikke la seg gjøre. Det kan også argumenteres for at de kvalitative resultatene som en Field-tilnærming resulterer i er vanskeligere å sammenlikne og dermed kunne gjort det vanskelig å besvare forskningsspørsmålet. En showroom-tilnærming ansees ikke som passende i dette prosjektet.

### **4.3 Evalueringsmetoder**

Multimodale grensesnitt, som beskrevet i 2.1, er veletablerte, men det finnes ingen konsensus om hvordan denne typer grensesnitt evalueres. Selv om det eksisterer flere

etablerte designmetoder som delvis kan oversettes og anvendes til å vurdere multimodale grensesnitt, finnes det få metoder som er spesifikt utviklet for å vurdere multimodale grensesnitt (Möller et al., 2010). Dette kan ha en sammenheng med at definisjonen for hva som kan beskrives som et multimodalt grensesnitt er veldig bred og rommer en rekke ulike sammensetninger av interaksjonsformer. Gibbon (2000) gir følgende begrunnelse for hvorfor evaluering av multimodale grensesnitt er utfordrende: Det er vanskelig å gjøre opptak av tester og skape repliserbare resultater, evalueringskriteriene er svært varierende og ofte uklare fordi multimodale grensesnitt kan være svært forskjellige og kvalitative aspekter spiller en sentral rolle. På tross av at disse påstandene ble presentert for nesten 20 år siden, hevder Kühnel (2012) at utfordringene fortsatt ikke har blitt tilstrekkelig løst. Dybkjær et al. (2004) hevder at *state of the art* innen multimodale grensesnitts usability og evaluering fortsatt ikke er kartlagt, men at den teknologiske utviklingen og markedsvæksten i feltet, særlig med tanke på taleinteraksjon, setter dette høyt på forskningsagendaen. Selv om det finnes få evalueringsmetoder, har flere metoder fra andre interaksjonsformer vist seg å være hensiktsmessige for multimodale systemer.

Tradisjonelle kriterier som har blitt brukt i evaluering kan deles inn i objektive og subjektive målinger. Objektive målinger regnes som ytelsesmål fra brukernes interaksjoner med et system. Eksempler på slike mål kan være antall feil som oppstår, eller hvor lang tid det tar å nå et ønsket mål. Subjektive mål dreier seg om brukernes subjektive meninger om kvaliteten på et system, og beregnes gjerne i form av spørreundersøkelser eller intervjuer (McTear et al., 2016). Disse objektive målene baserer seg gjerne på ISO-standardens definisjoner av *efficiency* og *effectiveness* (ISO, 2010). Her finnes en rekke ulike metoder som er spesielt utviklet for evaluering av grafiske grensesnitt samt talegrensesnitt. Den mest brukte rammeverket for å gjennomføre objektive evalueringer av talegrensesnitt er *PARAdigm for DIalogue Evaluation System* (PARADISE) (McTear et al., 2016). Dette rammeverket beregner ytelse basert på en vektet funksjon av oppgavefullføring, effektiviteten i dialogen, kvaliteten på dialogen og brukertilfredsstillelse (Walker et al., 1997). I grafiske grensesnitt regnes *effectiveness* gjerne ut fra en binær evaluering av suksessscore (Albert og Tullis, 2013). Dersom en bruker klarer å gjennomføre oppgaven som blir etterspurt, får man 1, dersom man ikke klarer det får man 0. Et mål som er mye benyttet for å angi *efficiency* i grafiske grensesnitt er tid-per-oppgave eller handlinger-per-oppgave (Albert og Tullis, 2013). Kühnel (2012) hevder ytelsesmål absolutt har en betydning for brukertilfredshet, men det blir for enkelt å se det som en direkte sammenheng.

I denne oppgaven vil ikke de spesifikke ytelsesmålene for talegrensesnitt kunne utregnes, da testingen baserer seg i stor grad på en prototype, hvor den relevante teknologien ikke ligger til grunn. Det vil være meningsløst å vurdere språkgjenkjenning på en prototype som ikke benytter denne teknologien på en reell måte. Det er allikevel valgt å inkludere en tidsmåling og en suksesscore, da dette vil kunne gi indikasjoner på hvordan et tiltenkt system ville prestert. Basert på forfatterens egne oppfatninger og erfaringer om hva som er mulig å få til med dagens teknologi, vil tidsaspektet ved taleinteraksjon gjengis så realistisk som mulig i prototypen. Dette målet vil derfor kunne gi en indikasjon på om interaksjon via tale går raskere enn interaksjon via en tradisjonell fjernkontroll. Det vil også dannes et manus som beskriver nøyaktig hvilke ytringer som vil kunne lede til korrekte handlinger i prototypen, på denne måten vil det også være mulig å indikere feilprosent. Den eksakte metoden for disse utregningene kan sees i del 6.3.

Subjektive målinger gjøres gjerne ved hjelp av spørreundersøkelser og intervjuer som gjennomføres enten etter endt oppgave eller etter endt test. I Möller et al. (2010) diskuteres etablerte spørreskjemaer utfra deres hensiktsmessighet for å vurdere multimodale kvalitetsaspekter. De undersøkte spørreundersøkelsene var AttrakDiff (Hassenzahl et al. 2003), System Usability Scale (SUS) (Brooke 1996), Software Usability Measurement Inventory (SUMI) (Kirakowski og Corbett 1993) og spørreskjemaet for subjektiv vurdering av talesystemgrensesnitt (SASSI) (Hone and Graham 2000). Möller et al. (2010) konkluderer med at AttrakDiff er den spørreundersøkelsen som er best egnet til å vurdere multimodale grensesnitt. Andre studier har også vist at AttrakDiff er godt egnet for å undersøke den opplevde kvaliteten i et interaktivt produkt utfra en oppdeling mellom hedonisk kvalitet og pragmatisk kvalitet (McTear et al., 2016). AttrakDiff legges, i tråd med den anvendte teoretiske modellen, til grunn for evaluering i det komparative forsøket i denne oppgaven. I kapittel 6.3 beskrives AttrakDiff, og hvordan den ble brukt i evalueringen, i detalj.

# 5 Designprosessen

I dette kapittelet presenteres designprosessen som ledet frem til den endelige forskningsprototypen. På en kronologisk og historisk lineær måte beskrives trinnene i prosessen for å gi leseren innblikk i hvordan prototypen ble til. Her beskrives involveringen av brukere på de ulike stegene i prosessen og den iterative utviklingsprosessen som ledet frem til den endelige forskningsprototypen som brukes i den komparative studien som presenteres i kapittel 6. Å beskrive designprosessen og utformingen av prototypen har vært en utfordring, da prosessen inneholder mer enn de eksplisitte metodene og aktivitetene som har blitt utført. Flere avgjørelser har mest sannsynlig blitt tatt basert på ubevisste prosesser som jeg ikke har vært klar over, ettersom prosessen med å lage en prototype inneholder en blanding av mine ferdigheter som designer, min forståelse av brukerkonteksten, den teoretiske tilnærmingen, resultatene fra brukerinnsikten og gjennomgang av relevant litteratur. Som nevnt tidligere har spørsmålet om hvordan man dokumenterer, beskriver og kommuniserer kunnskapen generert gjennom en designforskningsprosess blitt omtalt i RtD-fellesskapet (Gaver, 2014; Höök et al., 2015; Zimmerman, Stolterman, & Forlizzi, 2014 ). Mine refleksjoner rundt denne problemstillingen tas opp i kapittel 8.

## 5.1 Brukerinnsikt

### 5.1.1 Målgrupper

I utgangspunktet var en av hensiktene i oppgaven å undersøke hvordan aksepten for taleinteraksjon i TV varierer mellom ulike målgrupper. Underveis i prosjektet ble det besluttet at det ikke fantes tilstrekkelig tid til å evaluere dette. Utviklingen har derfor tatt utgangspunkt i målgruppen til RiksTV. RiksTV ønsker å nå ut til et bredt spekter av brukere. Deres TV-tilbud inneholder flere ulike plattformer for utnyttelse av tjenestene, noe som gjør at de treffer brukere i flere ulike aldersgrupper og livssituasjoner. Allikevel har de flest kunder i 50-årene, og dette utgjør altså kjernemålgruppen. Utover dette har de kunder fra 20-årene til 70-årene. Dette innbefatter altså alt fra barnefamilier til enslige eldre. Målgruppen som legges til grunn i utviklingen av dette prosjektet er derfor mennesker i alderen 25-60 år.

### 5.1.2 Brukerintervjuer

Det har blitt utført ti intervjuer med brukere i målgruppen. Dette ble gjort med mål om å kartlegge brukernes erfaringer med TV-tjenester og mer spesifikt for å identifisere problemområder med eksisterende løsninger. Intervjuobjektene ble ikke avgrenset til

brukere av RiksTVs tv-produkter, men omfattet brukere med en vilkårlig TV-leverandør. Dette begrunnes med at mye av hovedfunksjonaliteten i produktene til de norske TV-tilbyderne er ganske lik. I tillegg gjorde dette det noe enklere å finne intervjuobjekter. En tredje konsekvens var å potensielt få ytterligere innsikt som kunne vært oversett dersom man begrenset objektene til kun RiksTV-brukere.

I intervjuene ble det stilt åpne spørsmål om TV-vaner, hvilke produkter som brukes, hva som er utfordrende med dagens løsninger og hva brukerne er fornøyd med. Videre ble det spurt om erfaringer og tanker forbundet med taleteknologi (Se appendix A). Det å fastslå at talestyring er den rette løsningen på utfordringene i denne konteksten er et kunstig utgangspunkt for et intervju. Det ble derfor innledningsvis ikke lagt noen føringer på at taleassistenter skulle være tema. Utover i intervjuene ble talestyring introdusert som et alternativ.

### **5.1.3 Observasjoner**

For å sikre en triangulering ble det utført fire observasjoner av brukere i naturlige omgivelser, mens de benyttet seg av TV-en. Det ble i denne omgang kun observert brukere av RiksTVs produkter. For å få et tydeligere bilde av problemområdene som ble identifisert i intervjuene, ble det gitt et utvalg oppgaver til observasjonsdeltagerne. Disse oppgavene var konstruert slik deltagerne ble tvunget til å utsettes for noen av utfordringene fra brukerintervjuene. Deltagerne ble underveis i observasjonen bedt om å fortelle om hva de tenkte og hvilke vurderinger de gjorde seg underveis.

### **5.1.4 Funn fra intervjuer og observasjoner**

Innsikten fra intervjuer og observasjoner ble systematisert for å se hvilke tilbakemeldinger og holdninger som gikk igjen. Dette ble gjort ved å ekstrahere konkrete tilbakemeldinger fra de ulike intervjuene og skrive disse ned på post-it lapper. Videre ble de tilbakemeldingene som gikk på samme tema plassert nært hverandre og forsøkt kategorisert. Dette dermed et slags «heatmap», der de tilbakemeldingene som ble gitt flest ganger ble størst. De kategoriene som ble vist mest interesse eller som av andre grunner sees på som relevante presenteres her.

#### ***Leting etter innhold***

Brukerne bruker mye tid på navigasjon for lete etter innhold i underholdningsportalen. Dette gjelder først og fremst i forbindelse med RiksTV-boksen og i mindre grad i Appen.

Flere sier at dette ofte kan føre til frustrasjon, og at man noen ganger ender med å se på direktesendt TV fordi man ikke gidder å lete mer.

### ***Programguide***

Flere sier de synes det er for mye trykking for å finne det programmet de er ute etter. En bruker sier hun savner de gamle tv-boksene der man kunne nå programguiden ved å trykke på én knapp. De fleste er riktignok positive til programguide da de liker å se en oversikt over programmene som går og i tillegg ikke alltid husker hvilket nummer kanalene ligger på i kanallisten.

### ***Sliter med å velge***

Flere brukere opplever det som vanskelig å bestemme seg for noe å se på når de bruker streaming eller filmleie. Enkelte sier de unngår å bruke disse tjenestene fordi de aldri klarer å bestemme seg for noe å se på basert på informasjonen som gis. Noen sier de ønsker seg et anbefalingssystem som kunne si noe mer om kvaliteten på innholdet.

### ***Sette på opptak***

Selv om opptak etter hvert er i ferd med å bli overflødig, fordi innhold som regel blir lagret automatisk i underholdningsportalen, viser det seg at det er en viktig funksjon for flere. Noen synes dog at det kan være vanskelig å huske hvordan man gjør det.

### ***Liten erfaring med taleteknologi***

De fleste intervjuobjektene har ingen eller svært liten erfaring med taleteknologi. Mange har fått med seg trenden, men er avventende med å ta det i bruk. Det er varierende tiltro til påliteligheten i taleteknologien.

### ***Spørsmål***

Brukerne mente det kunne være fint å kunne spørre en taleassistent om informasjon knyttet til TV. For eksempel «hva går på tv i kveld?», «når begynner Farmen?», «hvilken kanal sender kampen?» og «har det kommet en ny episode av Folkeopplysningen siden forrige uke?».

### ***Fjernkontrollen***

Flere synes det er tungvint at de må bruke flere forskjellige fjernkontroller i forbindelse med TV-en. Ellers mener de fleste at fjernkontrollen er forståelig å bruke når man har lært

seg den. Videre mente samtlige at det var et (lite) irritasjonsmoment å måtte reise seg for å hente fjernkontrollen.

### ***Positive til talestyring***

Mange synes det høres fristende ut å bare kunne si hva man ønsker å se på TV og ser potensialet for talestyring i denne konteksten. Det var dog flere som dro frem at en slik løsning er nødt til å fungere veldig bra før det ville være aktuelt å ta det i bruk.

### ***Avslapning på kveldstid***

Brukerne bruker tv i egen stue, stort sett i forbindelse med avslapning. Det er lite stress og dårlig tid forbundet med tv-opplevelsen. Noen sier de ofte kan være litt slitne og har behov for ro. De fleste ser på TV etter kl. 18:00.

### ***Mer direkte enn streaming***

De fleste intervjuobjektene sier de ser mer på direkte-TV enn streaming. De som bruker streaming foretrekker å gjøre det gjennom Apple TV. Brukere vet oftere hva de skal se på når de bruker streaming enn når de ser på direkte-TV. To intervjuobjekter sier du kun ser TV direkte eller ved opptak.

## **5.2 Dramatiseringsworkshop**

### **5.2.1 Planlegging og gjennomføring**

For å få et innblikk i hva brukerne ser på som gode løsninger når det kommer til taleinteraksjon, ble det holdt en dramatiseringsworkshop. Workshopen bestod av en dramatiseringsøvelse der fire personer, delt inn i par, gikk sammen og dramatiserte to gitte scenarioer. Den ene personen spilte talestyrt TV og den andre spilte seer. Scenarioene som ble dramatisert ble valgt ut på bakgrunn av funn fra intervjuer og observasjoner.

**Scenario 1:** “Du setter deg ned for å finne noe å se på. Du vet ikke på forhånd hva du ønsker å se, men du vet at du kanskje kunne tenkt deg en komedie som minner litt om en film du har sett før.”

**Scenario 2:** “Du skal se en fotballkamp mellom Norge og Island, men aner ikke hvilken kanal den går på. Du har dårlig tid fordi du vet at kampen snart begynner.”

Parene fikk 15 minutter til å komme opp med de dialogene de mener er best for å løse de gitte situasjonene gjennom tale. Etter dette presentere hver av gruppene sine løsninger, ved å dramatisere for hverandre. Avslutningsvis ble det holdt en liten diskusjon der deltagerne og jeg drøftet de ulike løsningene og utvekslet erfaringer. Se appendix C for detaljert workshopplan.

## **5.2.2 Funn fra Workshop**

### ***Rett på sak***

Begge deltagerparene løste scenarioene via noen få enkle steg, ved å kutte gjennom menyer og kanallister. Når man ønsket å se på fotballkampen, spurte man taleassistenten direkte om å sette på kampen mellom Norge og Island, og «taleassistenten» skjønte hvilken kanal den skulle sette på. Deltagerne var alle enige om at dersom interaksjonen med TV-en kunne fungere slik som løsningene deres, ville det gjøre opplevelsen mye bedre. De påpekte riktignok, i likhet med flere av intervjuobjektene, at det må være pålitelig og fungere hver eneste gang.

### ***Skjermen brukes aktivt***

Når «TV-en» ble spurt om å komme med noen forslag til filmer av brukeren, svarte den med å vise titlene på skjermen, fremfor å ramse opp forslag via tale. Dette er i tråd med tidligere funn som tilsier at oppramsing av mer enn noen få alternativer gjennom tale kan være vanskelig å prosessere for brukeren. Å bruke TV-en aktivt for å presentere alternativer kan være en måte å omgå denne utfordringen.

### ***Enveiskommunikasjon***

Et viktig funn var at personen som dramatiserte taleassistenten i begge gruppene svarte veldig lite. Totalt hadde personene som spilte seer 14 taleinteraksjoner, mens personen som spilte taleassistent bare hadde 5 totalt. Da deltakerne ble spurt om hvorfor de valgte å svare så lite, hadde de flere ulike forklaringer. En av begrunnelsene var at det ville være irriterende om taleassistenten skulle snakke mye når man ser på TV. En deltager mente at grunnen til at taleassistenten svarte så lite med tale var at den på en måte ville svare via skjermen, altså at de svarene man trengte dukket opp på skjermen, så da trenger ikke assistenten å si det i tillegg.

### ***Kort og konsist***

Et annet funn var at selv om deltagerne hadde frie tøyler når det kommer til teknologiske muligheter, valgte de allikevel å ordlegge seg med korte konsise meldinger. Det virket altså ikke som om de ønsket seg en dialog, men heller en kommandobasert interaksjon. Dette kan selvsagt være valgt som ble gjort på bakgrunn av at workshopen var basert på talestyring, og at det kan være vanskelig å fri seg helt fra tanken på hva som er teknisk mulig å få til. Det er allikevel flere ting som tyder på at brukerne ønsker seg en minimalistisk interaksjon med taleassistenten i TV-sammenheng.

## **5.3 Utvikling av prototype**

### **5.3.1 Hva er en prototype?**

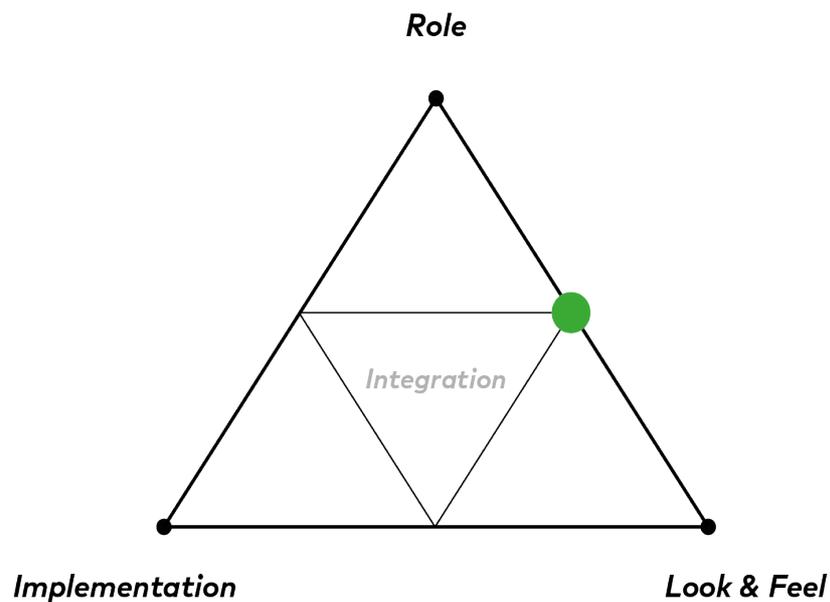
Prototyper har hatt en lang og viktig historie i HCI og har vist seg å være en produktiv metode for å teste teorier, konsepter og interaktive systemer på en menneskesentrert måte. Bruken av prototyping og prototyper for å fremkalle tilbakemeldinger fra mennesker spiller en svært viktig rolle for å besvare spørsmålet om hvordan nye teknologier kan gjøre produkter og tjenester enklere og mer tilfredsstillende å bruke (Odom et al., 2016).

Prototyping og prototyper i HCI og interaksjonsdesign spiller flere roller, fra åpen utforskning til provokasjon av kritiske refleksjoner og testing eller validering av hypoteser (Wensveen og Matthews, 2015). En prototype kan være en manifestasjon eller bare en utforskning av et teoretisk konsept som ikke skal dømmes for sin faktiske funksjonalitet eller nåværende tilstand, men heller dens potensiale (Wensveen og Matthews, 2015). Prototyper kan også være et steg på veien mot et realisert kommersielt produkt som brukes til å teste spesifikke behov og krav. Prototyper kan defineres som «fysiske manifestasjoner av idéer eller konsepter» eller som «representasjoner av et design som er gjort før endelige systemer eksisterer» (Odom et al., 2016)(forf.overs.). Oppløsningen på en prototype kan variere fra enkle løsninger i papir til mer gjennomarbeidede digitale løsninger. Men uavhengig av dette forblir de referanser til fremtidige produkter, systemer eller tjenester. På denne måten er prototyper plassholdere for noe annet, de er en etablering av et fremtidig utfall (Lim et al. 2008).

### **5.3.2 Role, Look and Feel og Implementation**

Houde and Hill (1997) understreker betydningen av hvilke spørsmål prototypen faktisk stiller og dermed hva prototypen prototyper. De presenterer en modell som deler prototyper inn i tre dimensjoner: *role*, *look* and *feel* og *implementation*. Hver dimensjon tilsvarer en

klasse spørsmål som er sentrale for utformingen av et interaktiv system. *Role* refererer til spørsmål om funksjonen, eller rollen, et produkt eller system tjener i en brukers liv - hvordan det er nyttig for brukerne? *Look and feel* angir spørsmål om den konkrete sensoriske opplevelsen av å bruke et produkt. Hva brukerene ser, føler og hører når de bruker det. *Implementation* omhandler de teknologiske aspektene ved produktet som skal utvikles. Her kan det for eksempel utforskes hvilken teknologi eller hvilke komponenter som er egnet for å utvikle et produkt. Som en fjerde kategori beskriver de *Integration*, som beskriver prototyper der alle de tre øvrige kategoriene implementeres i en felles prototype.



*Figur 8: De tre hjørnene beskriver de tre dimensjonene av prototyper beskrevet i Houde og Hill's modell (1997): role, look and feel og implementation. Integrering okkuperer midten av modellen, de denne type prototype undersøker alle tre dimensjonene. Den grønne markør*

### 5.3.3 Min tilnærming til prototyping

Dette prosjektet utforsker hvordan taleinteraksjon kan bidra til å gjøre brukeropplevelsen for TV-brukere. Jeg er her interessert i rollen og funksjonen denne nye interaksjonsformen tjener. I tillegg er jeg interessert i hvilke følelsesmessige reaksjoner og betraktninger brukerne gjør i møte med prototypen. Det vil etterstrebtes å utvikle en prototype som ser realistisk ut, og som gir brukeren en følelse av å interagere med et reelt system. I så måte vil prototypen stille spørsmål som havner i kategoriene *Role* og *Look and feel*. Når det gjelder de teknologiske aspektene ved taleinteraksjon, vil ikke prototypen stille spørsmål ved disse. I dette tilfellet ønsker jeg ikke å la meg begrense av hva som er teknologisk

mulig, både når det kommer til NLP kapasitet eller hastighet og timing. Den grønne prikken i figur 8 viser hvordan jeg plasserer prototypen som utvikles i dette prosjektet på triangelet til Houde og Hill (1997).

## 5.4 Scenarioer

Basert på brukerinnsikten generert gjennom intervjuer, observasjoner og workshops, ble det definert et sett med scenarioer. Disse scenarioene ble behandlet som funksjonalitetskrav under utviklingen av prototypen. Scenarioene baserer seg i stor grad på dagens TV-funksjonalitet, og er i så måte ikke fremtidsscenarioer. Det er interaksjonsformen som skal utforskes. Det tas selvsagt forbehold om at bruksmønstre vil endres i takt med teknologisk utvikling, og at det finnes en viss sammenheng mellom ny teknologi og nye bruksmønstre. Underholdning i form av levende bilder som i TV, film og serier har dog vist seg å være et konstant medium gjennom den teknologiske utviklingen som har skjedd fra introduksjonen av fjernsynet på 60-tallet, og til dagens smarttelefoner, så det er grunn til å tro at mennesker også fremover vil oppsøke denne typen underholdning, dog kanskje i nye formater og ved nye interaksjonsformer.

- **Scenario 1: Finne underholdning**

Få bistand til å finne underholdning. Brukere opplever at de sliter med å finne innhold i TV-tjenesten. Dette skyldes dels at det er vanskelig å søke, filtrere og sammenlikne innhold. Et annet aspekt er at brukerne mener det kan være vanskelig å bestemme seg for noe å se på og at de ofte blir sittende å diskutere eller overveie ulike alternativer uten å bli enig. Sistnevnte problem ansees som noe som ikke vil endres ved en ny interaksjonsform.

- **Scenario 2: Navigere i kanaler**

Enkelte brukere hevdet at de kunne slite med å huske hvor de ulike kanalene ligger plassert i den numeriske kanalrekken. Noen hevdet også at det å bruke en TV-guide kunne innebefatte litt for mange operasjoner for brukeren, slik at det kan være utfordrende å huske.

- **Scenario 3: Opptak**

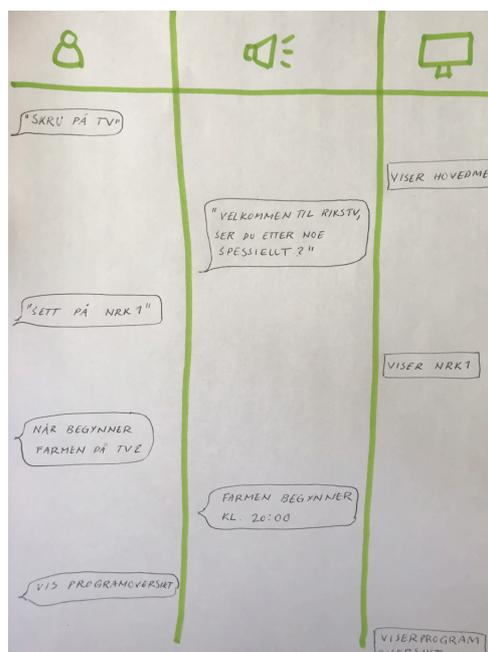
Selv om opptak etterhvert er i ferd med å bli overflødig, fordi innhold som regel blir lagret automatisk i underholdningsportalen, viser det seg at det er en viktig funksjon for flere. Noen synes dog at det kan være vanskelig å huske hvordan man gjør det. Det kan kanskje av noen tolkes som en motsigelse at et scenario

som baserer seg på en utdøende tv- trend blir inkludert. Det argumenteres for at selv om fremtidens TV- løsninger ikke nødvendigvis vil inneholde opptaksfunksjoner, så er de en viktig del av dagens bruksmønster. Ønsket med studien er som tidligere presisert ikke å redefinere funksjonaliteten i TV-opplevelsen, men å undersøke en alternativ interaksjonsform.

## 5.5 «Happy path»

Som et første steg i utviklingen av prototypen, ble det skissert en «happypath» for de tre ulike scenarioene. En happypath er en høynivå beskrivelse av en dialogen med systemet, uten å ta høyde for potensielle feil eller misforståelser. Det ble her fokusert på å lage den enkleste mulige veien til målet.

Prosessen foregikk ved å lage et diagram med tre kolonner, en for ytringer fra brukeren, en for tale-responser for systemet og en for skjermresponser fra systemet. Disse interaksjonsbeskrivelsene eller «skriptene» ble analysert og videreutviklet sammen med tre brukere gjennom en miniworkshop. *Rapid prototyping* (Kühnel, 2012) egnert seg i denne tidspunktet, da det tillot å itterere raskt over ulike dialogflyter og teste disse med en gang, gjennom å dramatisere de for hverandre. Miniworkshopen resulterte i et sett med dialogflyter som var tydelige og intuitive for alle deltagerne. Disse dialogflytene ble ikke brukt direkte, men ga et bilde på hvordan interaksjonen i det multimodale systemet kunne fungere.



Figur 9: Beskrivelse av papirprototyping

## 5.6 Digital prototyping

Neste steg i prosessen var å danne et verktøy som kunne brukes til å teste ut dialogflytene på en reell måte. Her ble mangfoldige ulike verktøy undersøkt, med mål om å finne et verktøy som kunne kombinere tale input og output, med visuell output via skjerm.

### 5.6.1 Dialog Flow og React

Dialoge Flow er et verktøy som er utviklet av Google og som brukes for å lage applikasjoner for deres Google Assistant. Etter noen runder med utprøving av dette verktøyet viste det seg at det vanskelig lot seg kombinere via skjerm. I et tidligere prosjekt har jeg løst dette ved å etablere en egen backend som router taleinput fra Dialoge Flow til en frontend i React. Dette fungerte ikke optimalt og var svært tidkrevende. Dette ble også gjort i samarbeid med utvikler Anders Lysne, som studerer Dataingeniør ved NTNU. Etter litt prøving og feiling besluttet jeg derfor å se meg om etter et annet verktøy som kunne løse oppgaven.

### 5.6.2 Adobe XD

Adobe XD er et prototypingsverktøy som ble introdusert i 2016. Siden oppstarten har det i all hovedsak dreid seg om utvikling av skjermbaserte prototyper. I likhet med verktøy som Sketch og Invision, tillater Adobe XD at man skisserer opp troverdige skjermbilder, før man deretter kan gjøre disse skjermbildene om til klikkbare prototyper ved å legge inn markørfelt som linker til et annet skjermbilde. Ved å trykke på dette markørfeltet vil man bli videresendt til neste skjermbilde. Denne typen verktøy har blitt svært utbredt blant interaksjonsdesignere, fordi det tilbyr en enkel og tidseffektiv måte å teste ut idéer og konsepter.

I november 2018 introduserte Adobe XD muligheten for prototyping av taleinteraksjoner. I stedet for å legge inn markørfelt som man kan trykke på for å nå neste skjermbilde, kunne man nå ved hjelp av en innebygget NLP algoritme, legge inn ytringer som blir oppfattet av systemet og sender deg videre til neste skjermbilde. Man har også muligheten til å legge inn ulike ytringer på ett skjermbilde, slik at man avhengig av hva brukeren sier, blir sendt ulike nye skjermbilder. Når man så lander på neste skjermbilde, kan man legge inn forhåndsskrevne tilbakemeldinger som blir lest opp. På denne måten kan man danne taleinteraksjoner mellom bruker og prototype, uten programmering. Det finnes flere ulike verktøy for prototyping av taleinteraksjon, men ingen av de lar seg kombinere med skjerm. På grunnlag av dette ble det besluttet å bygge prototypen i Adobe XD.

### ***Utfordringer med XD***

Dette verktøyet ble testet ut i form av en skisse av en «happypath». Dette viste seg å fungere ganske bra, og man kunne nå målet kun ved hjelp av taleinteraksjoner. Baksiden var et man måtte holde nede mellomromstasten på tastaturet for å aktivere mikrofonen.. For å unngå at mikrofonen skal lytte til enhver tid, legger man inn en form for aktivering, slik at mikrofonen slår seg på kun når det er nødvendig. I en test vil det være lite hensiktsmessig å be brukeren trykke på en tast før han snakker, da selve ideen er å teste ut hvordan brukeren vil oppleve å interagere utelukkende via tale. Dersom man må forholde seg til en knapp mister man noe av grunnlaget for vurdere interaksjonen, da denne knappen vil bidra til å danne et interaksjonsmønster om minner mer om et tradisjonelt, der man trykker på noe for å få noe til å skje. Jeg ønsket å unngå trykking fullstendig.

En annen utfordring var hvordan jeg kunne presentere en prototype for brukeren i en mest mulig troverdig kontekst. Jeg ønsket at brukeren skulle kunne sitte i sofaen og få prototypen presentert på en vanlig TV, for å bevare mest mulig av de naturlige omgivelsene.

En tredje utfordring var at NLP-algoritmen i XD ikke forstår norsk. Det var dermed et krav at brukerne både måtte snakke engelsk og bli besvart på engelsk. Det ble bestemt å bygge prototypen på engelsk, da det forventede engelsknivået på brukerne som skal teste prototypen ansees som høyt nok til å kunne gi fullverdige resultater.

### **5.6.3 Adobe XD + Wizard of Oz**

Begge de to ovennevnte problemene ble løst etter noen runder med prøving og feiling. Adobe XD har en forhåndsvisningsfunksjon som gjør det mulig å kjøre prototypen på en smarttelefon, mens telefonen er koblet til datamaskinen som kjører programmet. Hvis man setter det opp på denne måten har man i tillegg mulighet til overstyre hvilket skjermbilde som vises på telefonen, ved å velge skjermbilder fra et oversiktlig «skjermbilde-tre». Dette løste dermed utfordringen med at man behøver å trykke på noe for å aktivere mikrofonen. Nå kan jeg som testkjører styre hvor brukeren skal bli sendt, basert på hva som blir sagt. Denne testformen kan minne veldig om «Wizard of Oz» (WoZ) testing. I en Wizard Of Oz-test tror brukerne at de interagerer med et ekte system, mens det i virkeligheten er et menneske (*Trollmannen*) som simulerer hele eller deler av teknologien (McTear et al., 2016).

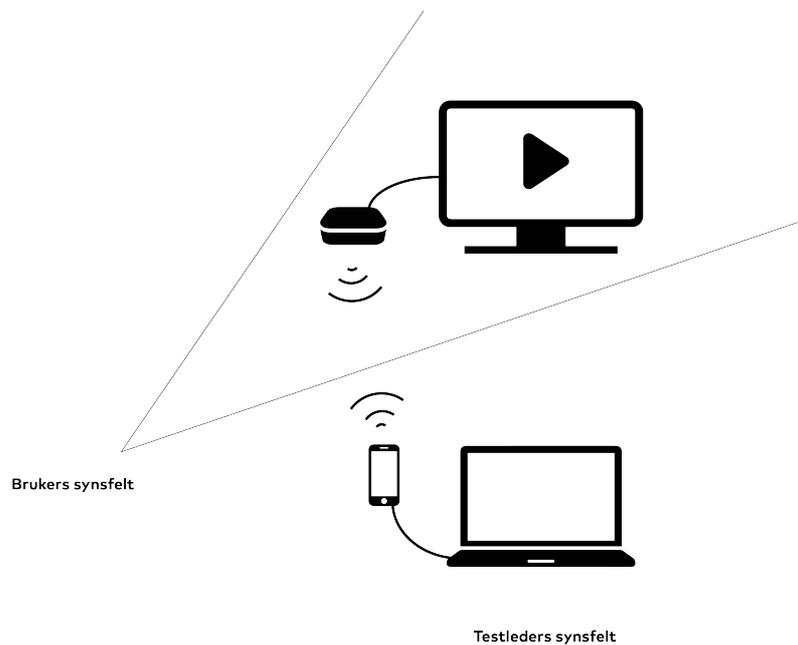
For at en slik testmetode skal fungere, må den være godt planlagt med tydelig definerte scenarier. Trollmannen kan ikke improvisere, men må forholde seg til de definerte

scenarioene. For å oppnå dette må den tiltenkte systemoppførselen være spesifisert tydelig, og *trollmannen* må være trent opp til å gjennomføre testen (McTear et al., 2016). For å få brukeren til å tro at de interagerer med et ekte system, må trollmannen kunne velge systemets respons i sanntid slik at det presenteres riktig for brukeren (McTear et al., 2016). På tross av kompleksiteten er WoZ en god måte å evaluere ut et system eller en teknologi før implementasjon, og kan derfor være svært kostnadsbesparende (McTear et al., 2016).

En svakhet ved WoZ er at man under en test kan bli revet med og ende opp med å inkludere funksjonalitet og forståelse som ikke var planlagt på forhånd. Dersom dette skjer vil ikke testen kunne gi indikasjoner på om man er på rett spor eller ikke. Det er essensielt å være bevisst de potensielle misledende effektene som kan oppstå som et resultat av den overdrevne kapasiteten til en menneskelig trollmann, sammenliknet med reell teknologi. Det er derfor viktig at trollmannen holder seg til «manus» fremfor å benytte sin egen kunnskap og intuisjon. Dette er ikke enkelt å gjøre, og derfor er WoZ-tester gjerne automatisert så langt det lar seg gjøre, for å begrense trollmannens frihet (McTear et al., 2016).

En fordel ved å kombinere en relativt høyoppløslig prototype med WoZ er at man har fysiske begrensninger for hva som er mulig å vise. Man har fasilitert for de scenarioene man har i prototypen, og selv om det er et menneske som sender deg fra ett skjermbilde til det neste, vil det være umulig for dette mennesket å finne opp nye skjermbilder underveis.

For å beholde denne funksjonaliteten og samtidig få kjørt prototypen på en TV, ble det besluttet å dele skjermen på smarttelefonen via Apple Air Play til en Apple TV som igjen var koblet opp til en TV. På denne måten kunne jeg altså styre hvilke skjermbilder og talemeldinger som ble vist på TV-en, samtidig som brukeren slapp å trykke på noe for å kunne snakke til prototypen.



*Figur 10: Beskrivelse av hvordan prototype settes opp. Testleder styrer skjerm og taleresponser uten for brukerens synsfelt.*

## 5.7 Iterativ utviklingsprosess

Første utkastet av prototypen ble laget som en del av en utforskning av talefunksjonaliteten i Adobe XD. Utkastene ble hyppig testet og iterert. På bakgrunn av innsikten fra intervjuer og observasjoner i tillegg til workshopen og litteraturstudiet, ble det fokusert på å bruke skjermen som en «flate» som støtter opp under taleinteraksjonen. Taleresponsene fra systemet ble holdt til et minimum, og bestod av enkle og tydelige svar.

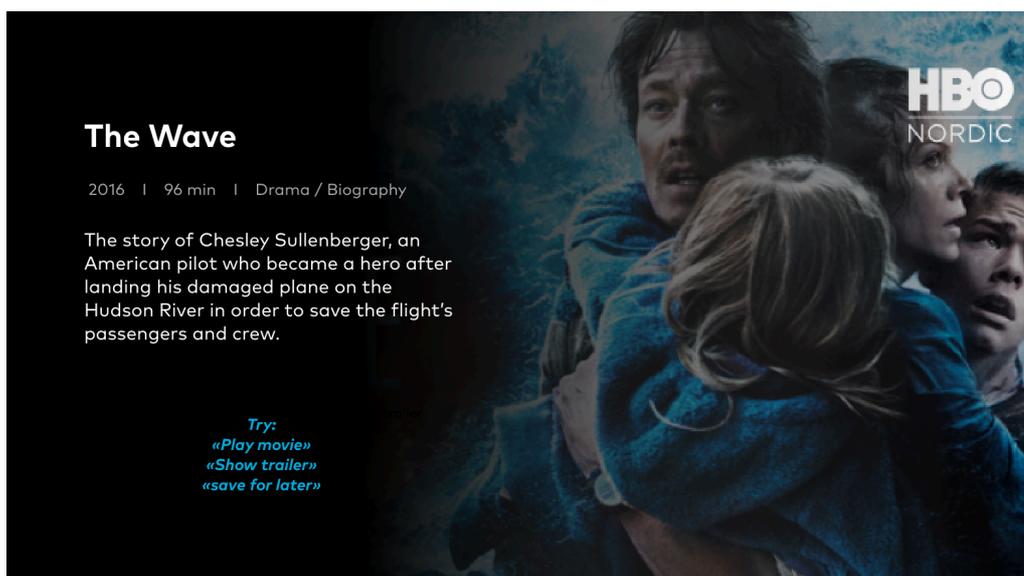
De ulike detaljene i den første iterasjonen av prototypen ble testet jevnlig, men etterhvert som den begynte å ta form var det nødvendig med en mer helhetlig evaluering. Dette ble gjennomført i form av en test der tre deltagere fikk prøve å interagere med systemet i RiksTVs lokaler på Økern i Oslo. Testen foregikk på en uformell måte, der det ble ført en dialog mellom deltakerne og meg som testleder gjennom hele prosessen. På dette tidspunktet var antall oppgaver som var mulig å gjennomføre begrenset til kun tre. Man kunne søke etter filmer med en bestemt skuespiller, velge å få mer informasjon om en av søkeresultatene, og eventuelt spille av filmen.



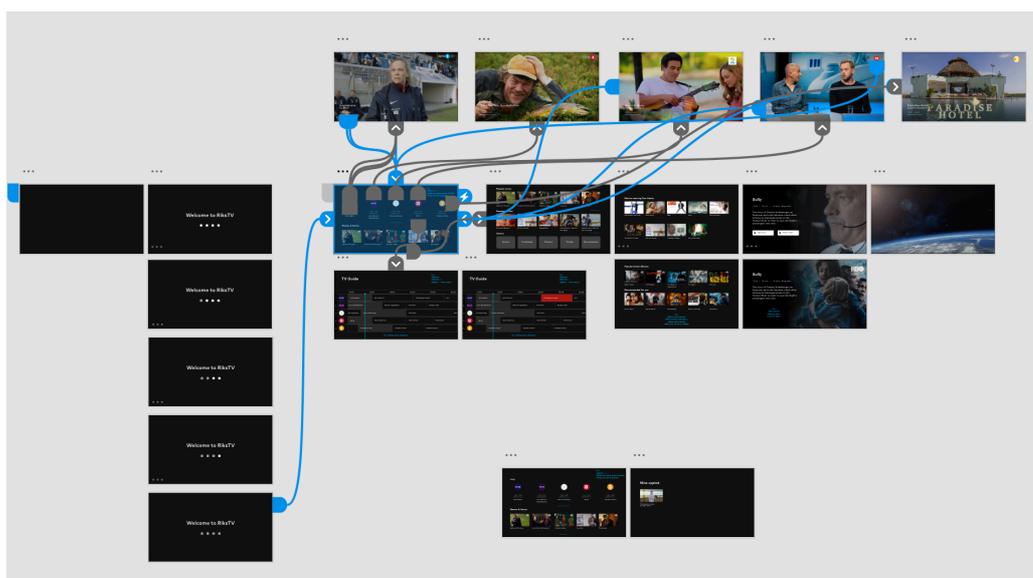
*Figur 11: RiksTVs testlokaler på Økern i Oslo.*

Resultatene fra testene viste at det var vanskelig å forholde seg til et system som var så «skriptet» altså at det er så forutbestemt hva man skal si. Testleder måtte nøye gå gjennom funksjonaliteten med testdeltagerne og informere om de ulike valgene som var mulig å ta. Testdeltagerne kommenterte at de ikke følte at de hadde noen innflytelse og dermed følte at de testet et fungerende system. Utover dette ga testen indikasjoner på at det tekniske oppsettet fungerte tilstrekkelig. Det ble besluttet å gå videre med det samme oppsettet på prototypen, men at funksjonaliteten måtte utvides i stor grad, for å gi brukerne større frihet og minimere testleders involvering.

En viktig tilbakemelding fra første test var et oppsettet minnet mye om tradisjonelle TV-løsninger, og at det til tider var fristende å ha noe å trykke med. Et konkret grep for å forsøke å komme vekk fra dette var å introdusere «try» i stedet for å presentere valgmuligheter i knapplignende form. Inspirert av løsningene til både Google og Amazons smarthøytalere med skjerm. Dette er en subtil tekst, plassert gjerne i ett av hjørnene på skjermen, som kommer med noen forslag til hva man kan spørre om. Dette er særlig et godt tilskudd når man begynner med å bruke tjenesten, da man enda ikke har bygd opp forståelse av funksjonaliteten. Disse meldingene kan hjelpe brukeren med å forstå hva man kan spørre om. Tanken er ikke at man skal ramse opp all mulig funksjonalitet, men å gi noen eksempler som kan gi brukeren en forståelse av mulighetene.



Figur 12: Try: Den blå teksten ned til venstre indikerer forslag til hva brukeren kan spørre etter.



Figur 13: Skjermdump fra utviklingen av prototype i Adobe XD

## 5.8 Endelig forskningsprototype

De begrensede tidsrammene i masterprosjektet legger føringer for hvor mye tid man kan bruke på en utviklingsprosess. På et tidspunkt må man akseptere at det man har laget er tilstrekkelig, og gå videre. Når prototypen var tilstrekkelig i stand til å utføre de definerte

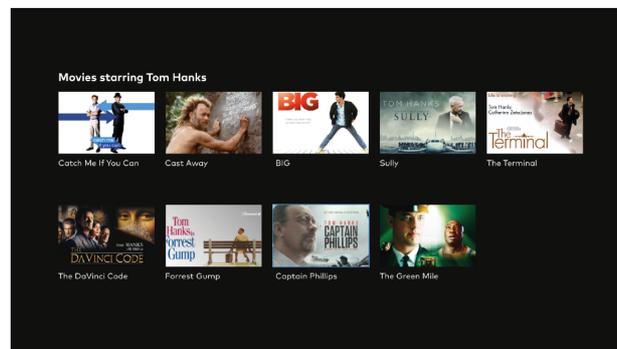
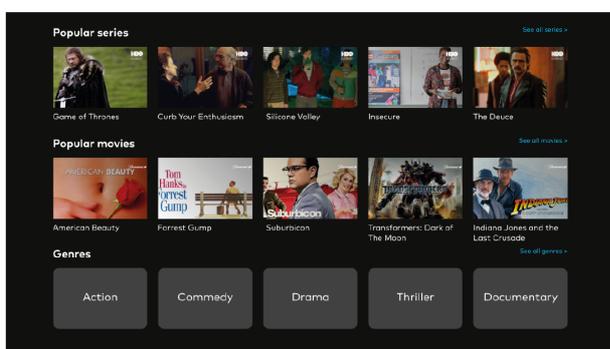
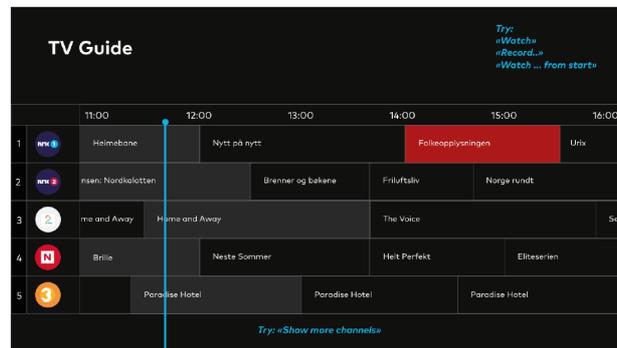
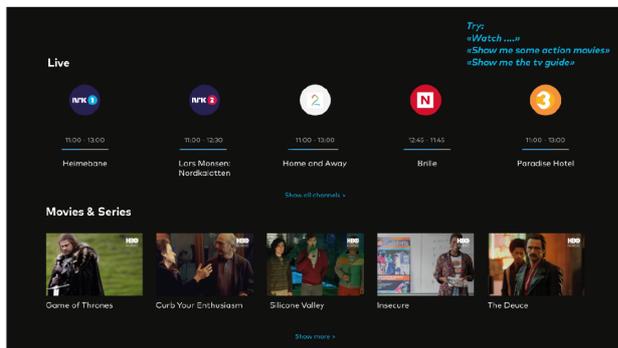
oppgavene, ble det besluttet å stoppe utviklingen. Den endelige forskningsprototypen beskrives i de påfølgende avsnittene.

Når man slår på TV-en får man en velkomsthilsen, før man blir sendt direkte til en «oversiktsmeny» som viser et utvalg tv-kanaler, med tilhørende informasjon om hva som spilles på kanalen. I tillegg får man se et lite utvalg av filmer og serier. Herfra har man muligheten til å hoppe direkte til et direktesendt tv-program, enten etter kanal eller program. Man har også muligheten til å spørre om å se filmer og serier. Dersom man spør om «film og serier» vil man bli sendt til en film- og seriemeny som inneholder et utvalg populære filmer og serier. Man kan også gjøre raffinerte søk ved å for eksempel spørre om å se filmer med en bestemt skuespiller, regissør eller en bestemt sjanger. Disse søkene kan gjennomføres uavhengig av hvor du befinner deg i tjenesten på det gitte tidspunktet. Dersom du ser på direkte TV, skal du allikevel kunne søke direkte om «Finn en film med Tom Hanks», og du vil bli presentert med søkeresultatet.

Fra «oversiktsmenyen» har man også muligheten til å se en «TV-guide» eller en «oversikt over TV-kanaler», presentert med en tidslinje som viser når programmene starter og slutter. Fra denne tv-menyen kan man enten velge å undersøke hva som går på de ulike kanalene og deretter bestemme seg for noe å se på, eller man kan velge å sette på et opptak for et program som begynner senere. Argumentet for å allikevel beholde en menystruktur er at det skal være mulig å gå «Hjem» for å få oversikt. Menystrukturen vil dog holdes til et minimum, med kun følgende sider: «oversiktsside», «oversiktsside for filmer og serier» og «oversiktsside over kanaler/kanalliste».

Forskningsprototypen som ble brukt i den komparative studien inneholder følgende funksjonalitet:

- Man kan sette på TV-en
- Man kan si hvilken kanal, eller hvilket program man vil se på.
- Man kan se en oversikt over tilgjengelige kanaler.
- Man kan sette på opptak av ønsket program.
- Man kan se en oversikt over opptak som er gjort.
- Man kan spørre etter en bestemt film, skuespiller, TV-serie eller sjanger
- Man kan se en oversikt over populære filmer og sjangere
- Man kan gå «tilbake» til den forrige siden man var på
- Man kan spørre etter når et TV-program begynner.



Figur 14: Skjermbilder fra prototype: Øverst til venstre: Hovedmeny, Øverst til høyre: Kanaloversikt med optak, Nederst til venstre: Hovedmeny for filmer og serier, Nederst til høyre: Søkeresultat

## 6 Komparativ studie

I denne delen beskrives det komparative forsøket som ble utført med forskningsprototypen og en tradisjonell tv-løsning, i form av RiksTV's SmartBoks2. Forsøket ble utført for å undersøke hvordan taleinteraksjon påvirker brukeropplevelsen i TV-sammenheng.

### 6.1 De to løsningene

De to systemene som ble testet var (a) SmartBoks 2 (SMB2) (en tradisjonell TV-dekoder med tilhørende fjernkontroll fra RiksTV) og forskningsprototypen, basert på taleinteraksjon, utviklet i dette prosjektet.

### 6.2 Deltagere

Ti personer (fem kvinner og fem menn) deltok i studien. Samtlige deltakere ble rekruttert via Facebook, med eneste krav om at de skulle være i målgruppen 25-60 år og at de skulle ha ingen eller svært liten erfaring mer RiksTVs produkter. Gjennomsnittsalderen blant deltagerne var 43 (fra 26 til 59). Erfaring med talestyring (målt ved en femstegs spørreundersøkelse) varierte fra lav (syv deltagere) til ingen (tre deltagere).

### 6.3 Målinger

Spørreundersøkelsen AttrakDiff ble brukt til å vurdere den oppfattede pragmatiske kvaliteten, den hedoniske kvaliteten og attraktiviteten i de to ulike produktene. Pragmatisk kvalitet refererer til produktets brukervennlighet, det vil si hvordan brukeren kan oppnå sine mål ved å bruke produktet. Hedonisk kvalitet refererer til brukerens psykologiske behov. AttrakDiff deler videre hedonisk kvalitet i to underdimensjoner: stimulering og identifikasjon, som refererer til, henholdsvis, produktets potensial til å gi en positiv følelsesmessig opplevelse, og produktets potensiale for å gi en følelse av eierskap (Hassenzahl, 2003). Hassenzahl indikerer at hedoniske og pragmatiske kvaliteter er uavhengige av hverandre, men begge bidrar likt til vurdering av attraktivitet (Hassenzahl, 2003).

AttrakDiff evaluerer brukeropplevelsen ved hjelp av en semantisk differensial, bestående av ordpar med motstående adjektiver (for eksempel "forvirrende - klar", "god - dårlig"). I denne studien ble det brukt en norsk versjon av spørreskjemaet, oversatt av forfatteren. Adjektiver ble evaluert på en syv-punkts skala, fra -3 til 3, der 0 indikerer nøytralitet (Hassenzahl, 3004). Spørreskjemaet består av totalt 28 ordpar, fordelt på fire dimensjoner:

pragmatisk kvalitet (PQ), hedonisk-identitetskvalitet (HQ-I), hedonisk stimuleringskvalitet (HQ-S) og tiltrengelighet, eller *Attractiveness* (ATT).

AttrakDiffs resultater fremhever problematiske UX-aspekter, men sier ikke noe om bestemte problemer. For eksempel, hvis deltakeren vurderer PQ-dimensjonen under null, kan det antas at deltakeren ikke er fornøyd med systemets pragmatiske kvaliteter. Allikevel er det vanskelig å fastslå hva som forårsaket denne misnøyen. For å analysere problemer i dybden, ble det utført et intervju etter gjennomføring av begge systemene. Intervjuet ble utført i et åpent format der deltakeren fikk fortelle fritt om opplevelsen av de to systemene. Intervjuene ble filmet, for å kunne analyseres i ettertid.

Det ble i tillegg gjort opptak av alle forsøk. Loggdataen fra opptakene ble brukt til å regne ut ytelsesparameterne tid-per-oppgave og suksesscore. Tid per oppgave ble målt ved å ta tiden fra brukeren ble presentert med en oppgave, til oppgaven var utført. Suksesscore ble regnet ut ved å gi deltagerne en score på 1 dersom de gjennomførte oppgaven uten behov for veiledning og 0,5 dersom de gjennomførte oppgaven med noe behov for noe veiledning og 0 dersom deltagerne ikke klarte å gjennomføre oppgaven.

## 6.4 Prosedyre

Forsøkene fant sted i hver enkelt deltagers dagligstue. Både SMB2 og prototypen ble satt opp hjemme hos hver enkelt testdeltager, for å simulere mest mulig naturlig brukssituasjon. Hver enkelt deltaker gjennomførte forsøket separat. For å unngå læringseffekt av å se noen andre gjennomføre før seg selv, ble ingen deltager eksponert for forsøket før de selv gjennomførte det. Etter en introduksjon og instruksjoner gitt av testleder, ble hver deltager instruert til å utføre følgende syv oppgaver:

- Skru på TV
- Sett på bestemt program
- Finn ut når bestemt fremtidig program starter
- Ta opp program
- Finn film med bestemt sjanger
- Finn film med bestemt skuespiller
- Spill av arkivert opptak

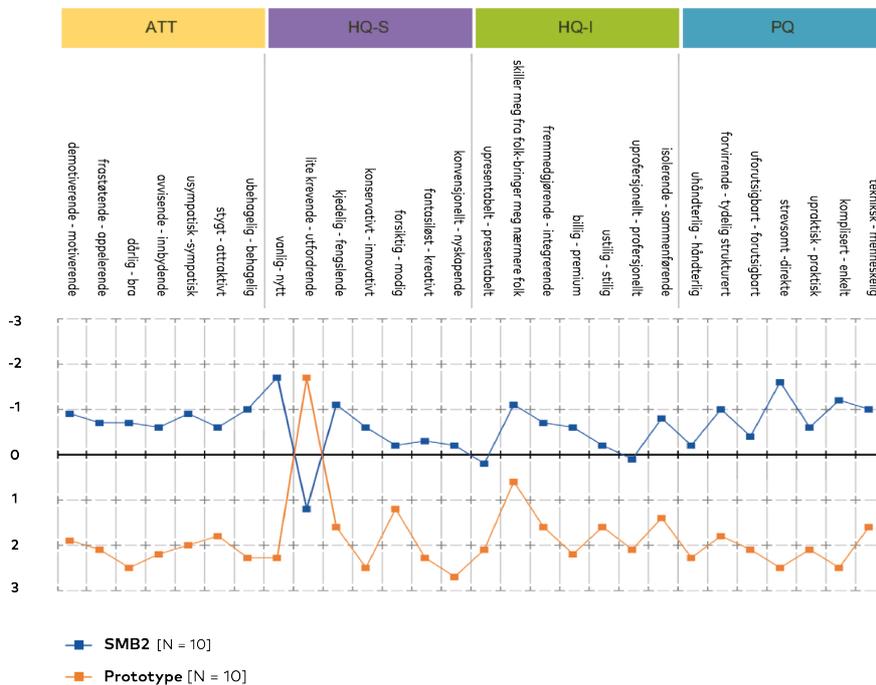
For å unngå gjensidig påvirkning av de to testene, utførte halvparten av deltakerne oppgavene med SMB2 først, mens resten utførte oppgavene med forskningsprototypen

først. Etter gjennomføring av oppgavene med ett system, fylte deltagerne ut en AttrakDiff spørreundersøkelse, deretter ble prosedyren gjentatt med det andre systemet. Til slutt fylte deltagerne ut en kort spørreundersøkelse om tidligere erfaring med teknologien og generell demografi, før et avsluttende intervju om den subjektive opplevelsen. Hvert enkelt forsøk varte i omtrent 60 minutter. Interaksjonstid med hvert enkelt system tok mellom to og ti minutter.

# 7 Resultater

## 7.1 AttrakDiff

Figur 15 viser gjennomsnittsverdiene for de 28 ordparene. Utregningene er gjort på bakgrunn av de ti deltageres individuelle besvarelser for de to testede systemene. Den blå grafen representerer SMB2 og den grønne grafen representerer prototypen. En plassering av plottet så langt som mulig mot høyre (på en skala fra -3 til 3) er tegn på en positiv opplevelse. Gjennomsnittsberegningene er beregnet ut fra de ti deltageres respektive evalueringene av de to systemene gjennom en semantisk differensial, bestående av 28 ordpar. Ordparene er delt inn i de fire kategoriene *pragmatisk kvalitet* (PQ), *hedonisk kvalitet identitet* (HQ-I), *hedonisk kvalitet stimulering* (HQ-S) og *Attractiveness* (ATT). PQ består av ordpar som angår brukernes oppfattelse av systemets funksjonelle kvalitet. HQ-I består av ordpar som angår systemets evne til å frembringe en følelse av eierskap og identitet. HQ-S består av ordpar som vurderer systemets evne til å frembringe positive følelsesmessige responser. ATT består av ordpar med helhetlige karakteristikker av systemets kvalitet.

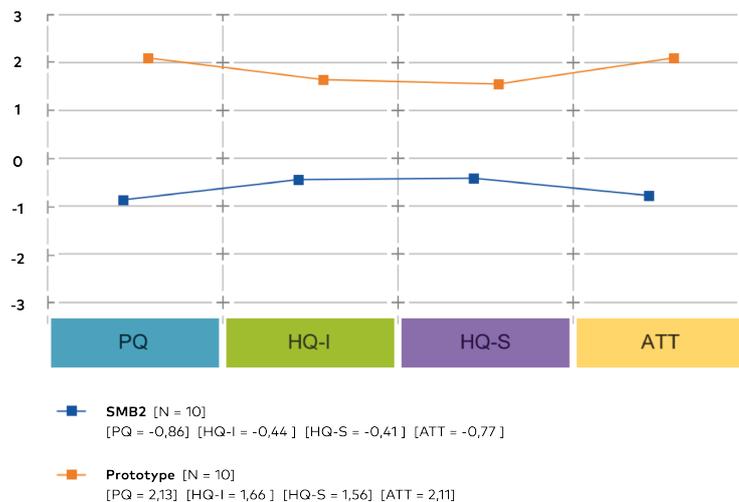


Figur 15: Gjennomsnittsverdier fra de 10 deltageres evalueringer av de 28 ordparene for SMB2 og prototypen. Y-akse = sjustrinns skala (-3 til 3).

Ordparet «skiller meg fra folk - bringer meg nærmere folk» viser en utpreget lav score for begge systemene. TV er ikke et medium som i seg selv har en sosial mellommenneskelig effekt, noe som gjenspeiles i den lave scoren på dette området. TV-opplevelsen blir først sosial når man ser på TV med andre mennesker, da kan en sosial situasjon oppstå i form av for eksempel en diskusjon om hva man skal se på eller diskusjoner om innholdet som vises. Da testen ble utført individuelt, var det ikke lagt til rette for mellommenneskelig kontakt, noe som kan være med på å forklare den lave scoren. I så måte kan man si at testen ikke fanger opp eller vurderer det sosiale potensialet ved systemene.

Ordparet «lite krevende – utfordrende» viser et totalt motsatt trend enn de øvrige ordparene. Slik skaperne av spørreundersøkelsen har konstruert dette ordparet, er «utfordrende» ment som en positiv betraktning, fordi det kan bidra til stimulering i form av en læringseffekt. «Lite krevende», i den andre enden av skalaen, er ansett som en negativ produkttegenskap. Som illustrert i (6.2 brukerinnsikt) er interaksjonen som oppstår når man bruker TV-en gjerne rettet mot å raskest mulig finne det innholdet man leter etter eller å finne inspirasjon til ting å se på. I denne situasjonen kan det hevdes at det ikke er behov for å bli utfordret, og at deltagerne i studien har ansett en lite krevende interaksjon som en fordel fremfor en svakhet.

Figur 16 viser middelveidene for de fire AttrakDiff-dimensjonene. En høyere plassering på y-aksen (på en skala fra -3 til 3) viser en bedre erfaring med systemet. Figuren viser at talestyringsprototypen scorer høyere i alle dimensjonene. De to kategoriene der det skiller mest mellom de to systemene er *Pragmatisk kvalitet* og *Attractiveness*. I disse to kategoriene scorer taleprototypen tydelig høyt, samtidig som SMB2 scorer lavt. For de to kategoriene, er resultatene mindre ekstreme, men prototypen scorer allikevel høyere i begge kategoriene. Dette indikerer at taleprototypen lykkes bedre med å gi deltagerne en positiv emosjonell respons og en følelse av eierskap. Det er også verdt å merke seg at SMB2 scorer under null i alle kategoriene, noe som indikerer at produktet ikke svarer til brukernes forventninger.



Figur 16: Gjennomsnittsverdier for de fire AttrakDiff kategoriene. Y-akse = sjustrinns skala (-3 til 3).

Figur 17 viser forholdet mellom hedonisk kvalitet og pragmatisk kvalitet. Y-aksen i figuren representerer hedonisk kvalitet (HQ), og x-aksen representerer pragmatisk kvalitet (PQ). De to HQ-kategoriene HQ-I og HQ-S er her slått sammen til en felles verdi. De ulike kvadratene representerer ulike «karakter-regioner». Disse regionene gjenspeiler Hassenzahls oppdeling mellom ACT-produkter, SELF-produkter, overflødige og ønskelige produkter. Den lille fargede indikatoren representerer den konkrete gjennomsnittsmålingen, mens det gjennomsiktige rektangelet beskriver konfidensintervallet i målingen. Et mindre rektangel medfører en større sikkerhet på at målingen er representativ for målgruppen. Beregningene av konfidensnivå ble gjort gjennom AttrakDiffs nettbaserte beregningssystem (AttrakDiff, 2018). Konfidensintervallene er basert på en 95% sikkerhet, med T verdi = 2.262.

Prototypen befinner seg i karakter-regionen *desired*, eller «ønskelig», noe som innebærer at forholdet mellom pragmatisk og hedonisk kvalitet er relativt symmetrisk, og at testdeltagerne vurderer begge de to dimensjonene godt over middels. Gjennomsnittsmålingen for PQ er 1,61, mens gjennomsnittsmålingen for HQ er 2,13, på en skala fra -3 til 3. Konfidensnivået er beregnet til henholdsvis 0,5 for PQ og 0,37 for HQ. Dette innebærer at verdien for PQ 95 % sikkert ligger mellom 1,63 og 2,63. Verdien for HQ ligger 95% sikkert mellom 1,24 og 1,98.



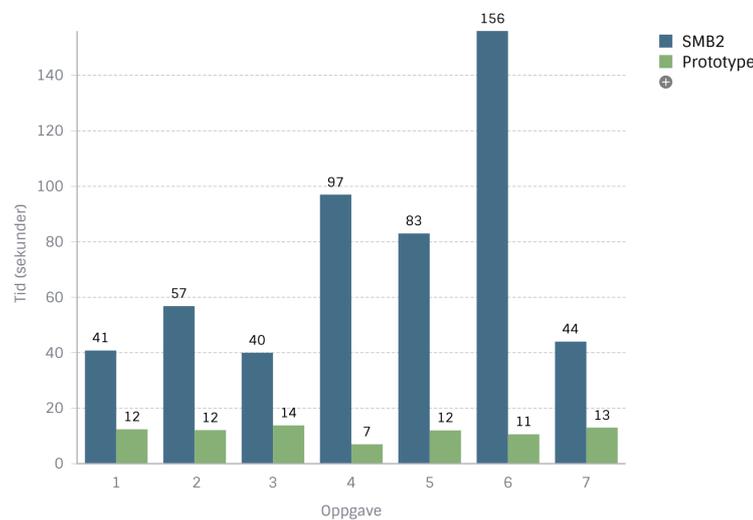
Figur 17: Forholdet mellom pragmatisk kvalitet og hedonisk kvalitet.

SMB2 ligger i venstre ytterkant av karakterregionen *neutral*, eller «nøytral». Dette tilsier at produktet har en mild overvekt av hedonisk kvalitet, og at vurderingen av både hedonisk kvalitet og pragmatisk kvalitet er under middels. Gjennomsnittsmålingen for PQ er -0,86, mens gjennomsnittsmålingen for HQ er -0,44, på en skala fra -3 til 3. Konfidensnivået er beregnet til henholdsvis 0,72 for PQ og 0,44 for HQ. Dette innebærer at verdien for PQ 95% sikkert ligger mellom -1,53 og -0,14. Dette konfidensintervallet er det største i målingen, noe som innebærer at testdeltagernes svar varierte mest i denne kategorien. Verdien for HQ ligger 95% sikkert mellom -0,88 og 0.

Rekkefølgen på testen av de to systemene ble byttet om for halvparten av deltagerne. Det vil si at halvparten prøvde forskningsprototypen først, og halvparten prøvde SMB2 først. Da oppgavene som skulle utføres i de to testene var de samme, var det forventet en viss læringseffekt når deltagerne gikk over på sitt andre system. Resultatene viste dog ingen indikasjoner på at det ene påvirker det andre. Årsaken til dette kan være at interaksjonsformene er fundamentalt forskjellige, og at de ikke nødvendigvis har så mye til felles som kan bidra til læring i en slik test. At man husker hvilke oppgaver som skal utføres, betyr ikke nødvendigvis at det blir enklere å vite hvordan oppgavene skal løses med et annet grensesnitt.

## 7.2 Loggdata

Resultatene fra beregningen av suksesscore er utelatt fra oppgaven. Å sette en karakter ved de ulike gjennomføringsgradene viste seg å være en utfordrende oppgave, som krevde mange subjektive evalueringer av hva som kan regnes som assistanse og ikke. Disse resultatene ville derfor ikke kunne brukt aktivt i evalueringen, da de baserer seg for mye på hva jeg som testleder anser som assistanse. Det var heller ikke satt noen øvre tidsgrense for hvor lenge en bruker fikk holde på før testleder grep inn og avbrøt oppgaven. Dette medførte at karakteren 0 ikke ble gitt i noen av oppgavene, med noen av de to systemene for noen av brukerne. Som en et bilde på hvor effektivt oppgavene ble utført, legges derfor tidsbruken til grunn, tatt i betraktning at alle oppgavene ble suksessfullt utført med begge systemene.



Figur 18: Gjennomsnittsmålinger av tidsbruk pr. oppgave for de to systemene.

Tid per oppgave var betydelig lavere for alle oppgavene med voice-first prototypen. Som illustrert i figur 18, bruker deltakerne betydelig mindre tid på å utføre oppgavene via tale enn via fjernkontrollen. Dette er i tråd med ideen om at fleksible inngangspunkter, som tale-input tillater, gjør navigasjon i interaktive systemer raskere. Disse resultatene bidrar til å validere den høye vurderingen av pragmatisk kvalitet i AttrakDiff undersøkelsen.

## 7.3 Intervjuer

Intervjuene som fant sted etter gjennomføring av de to testene, ga tydelige indikasjoner på at brukerne var svært positive til voice-first i TV-sammenheng. Samtlige deltakere ga uttrykk for at de foretrakk taleprototypen fremfor SMB2. Flere av brukerne lurte på om hvor lenge det er til produktet lanseres, og at de gjerne kunne tenkt seg å kjøpe et slikt

produkt. Dette indikerer to ting. For det første var det tydelig at de fleste deltagerne faktisk trodde de interagerer med et reelt talegrensesnitt, og at *Wizard of Oz* testen dermed fungerte slik den var tiltenkt. For det andre indikerer det at brukerne ikke ser på denne type produkt som noe som hører til fremtiden, men som noe de anser som gjennomførbart nå.

De aspektene ved prototypen som ble trukket frem som spesielt positive var friheten som taleinteraksjon tilbyr. Flere uttrykte en entusiasme når de snakker om hvordan man bare kan si det man er på jakt etter, og så får man det presentert, uten å gå gjennom en lang rekke operasjoner. En av deltagerne hevdet at en slik løsning ville kunne føre til at hun finner innhold hun ikke hadde klart å finne med TV-løsningen hun bruker i dag.

*«Alt blir jo mye lettere når man bare kan si det, og slipper å lete etter de rette knappene på fjernkontrollen» - Kvinne 56*

*«Det er deilig å føle at man ikke er låst til en sti, men at man kan velge hvordan man gjør ting selv» - Mann 30*

Mange av deltakerne pekte også på at det var gøy å interagere via tale, og at de så for seg at det kunne vært morsomt å ha en slik hjemme. Dette underbygger resultatene fra AttrakDiff undersøkelsen i kategorien hedonisk kvalitet stimulering. Ifølge Hassenzahl (2008) er det at et produkt kan gi positive følelsesresponser en essensiell del av en god brukeropplevelse.

En tilbakemelding som gikk igjen for flere av brukerne var at manglende erfaring med SMB2 gjorde at de brukte lengre tid enn de ville gjort dersom de hadde fått tid til å sette seg inn i det. Som en kommentar til dette, presiseres det, at deltagerne fikk like lite tid til å sette seg inn i de to systemene. På bakgrunn av dette kan det argumenteres for at prototypen fremstår som mer intuitiv for brukerne ved førstegangsbruk. Tilbakemeldingene bekrefter dog at testen ikke tar den temporale dimensjonen ved UX i betraktning. Som Hassenzahl presiserer, er UX noe som endrer seg over tid og avhenger av brukssituasjonen (2004). Utfordringene med at testen ikke fanger opp hvordan læringseffekt spiller inn på brukeropplevelse tas opp i diskusjonen.

Flere av deltakerne pekte på et potensielt problem ved å bruke en slik prototype sammen med andre. Dersom det er flere mennesker som sitter og snakker sammen samtidig som de ser på TV, trodde brukerne det kunne by på problemer. Denne utfordringen er definitivt reell, men har ikke blitt utforsket i dette prosjektet.

I løpet av de ti intervjuene, kom det ingen direkte negative tilbakemeldinger på prototypen. Det kan spekuleres i om dette kan ha noe med å gjøre at jeg personlig utførte både tester og intervjuer, og at deltagerne var ekstra positive fordi de ikke ville kritisere arbeidet jeg hadde utført. For å unngå dette oppfordret jeg deltagerne til å være helt ærlige i tilbakemeldingene, men det kan selvsagt hende at dette fortsatt kan ansees som en potensiell feilkilde. Det kan også tenkes at den frie og ustrukturerte stilen som intervjuene ble utført i kan ha medført en noe ukritisk vurdering. Dersom intervjuet hadde hatt en mer strukturert karakter, og inneholdt mer konkrete spørsmål, ville jeg kanskje ha fått et mer komplett bilde av de ulike deltagernes meninger.

# 8 Diskusjon

I denne oppgaven ønsket jeg å undersøke hvordan man kan utvikle og bruke en forskningsprototype til å gi innsikt i hvordan voice-first egner seg som interaksjonsform i TV-sammenheng. Som et svar på det overhengende forskningsspørsmålet, har jeg dokumentert og beskrevet min design-forskningsprosess, og forsøkt å gi leseren den svevende, kvalitative kunnskapen som har blitt generert under utviklingen av prototypen, for så å presentere en evaluering av brukernes opplevelse av prototypen i en komparativ laboriestudie. I dette kapitlet diskuterer jeg resultatene fra det komparative forsøket i lys av det teoretiske rammeverket og hvordan min spesifikke RtD-tilnærming har påvirket og muliggjort funnene.

## 8.1 Implikasjoner

### 8.1.1 Pragmatisk kvalitet

Resultatene fra alle de tre informasjonskildene gir indikasjoner på at voice-first, sett fra et pragmatisk perspektiv, egner seg svært godt til å interagere med TV. Både AttrakDiff dimensjonen som fanger opp de pragmatiske kvalitetsaspektene og ytelsesmålene som ble utført, i tillegg til tilbakemeldinger fra intervjuerne peker i retning av at voice-first, dersom det fungerer optimalt, er bedre egnet enn fjernkontrollen. Et viktig poeng å nevne her er at testen som er utført ikke tar stilling til tekniske utfordringer ved taleinteraksjon. Selv om dette ligger utenfor oppgavens fokus, sees det som nødvendig å vurdere resultatene i lys av de teknologisk utfordringene med taleinteraksjon i dag. For at et reelt system skal kunne fungere med like stor treffsikkerhet som prototypen, kreves en nærmest feilfri NLP. En sentral utfordring er hvordan potensielle feil som skjer kan rettes opp. Dette er utfordringer som det jobbes aktivt for å løse, og det forespeiles at teknologien vil bedre seg betraktelig fremover, i lys av den kommersielle suksessen som har funnet sted de siste årene. Dette vil kunne bidra til å gjøre det mulig å utvikle systemer som minner om prototypen som er utviklet i dette prosjektet.

Pragmatisk kvalitet dreier seg dog ikke bare om ytelsen til systemet, men også om i hvilken grad det er intuitivt og enkelt å bruke. Resultatene fra forsøket som er utført gir tydelige indikasjoner på at voice-first kan bidra til å bedre brukeropplevelsen på dette området. Den pragmatiske dimensjonen i AttrakDiff undersøkelsen, den som skilte mest mellom de to testede systemene. SMB2 fikk scoren -0.86, mens prototypen fikk scoren 2,13 (skalaen går fra -3 til 3). I tillegg viste loggdataen at deltagerene brukte betydelig kortere tid på å utføre oppgavene med taleprototypen. Intervjuene ga indikasjoner på at brukerne syntes det var

både raskt og effektivt å interagere via tale. Disse funnene bidrar til å bekrefte antakelsene, presentert i 4.2, om at kombinasjonen mellom tale og skjerm, og fleksibiliteten dette medfører er en effektiv interaksjonsform, også TV-sammenheng.

### **8.1.2 Hedonisk kvalitet**

Hedoniske egenskaper vektlegger brukerens psykologiske velvære (Hassenzahl, 2003). Ifølge Hassenzahl er de hedoniske egenskapene like viktige som de pragmatiske, og det er de hedoniske kvalitetsegenskapene som virkelig knytter brukeren til et produkt (2007). I følge Hassenzahl (2003) er det mye mindre sannsynlig at brukeren slutter å sette pris på produkter som gir hedonisk verdi, enn at de slutter å sette pris på produkter med pragmatisk verdi. Han begrunner dette med at brukeres emosjonelle preferanser er mye mer stabile enn brukeres atferdsmål. Resultatene fra forsøker peker i retning av at voice-first har et større potensiale for å frembringe hedoniske egenskaper, både i form av stimulering (HQ-S) og identifisering (HQ-I).

Stimulering dreier seg om brukerens iboende ønske om nye inntrykk, muligheter og innsikt (Hassenzahl, 2003). En klar faktor som kan bidra til å forklare hvorfor prototypen scorer høyt i denne kategorien er at taleinteraksjon fremstår som en ny og innovativ interaksjonsform. Resultatene fra AttrakDiff undersøkelsen og intervjuene sier at deltagerne opplevde det som gøy og spennende å snakke med TV-en. Det understrekes dog at måten prototypen ble presentert i studien, kan ha gitt et overdrevet godt inntrykk av taleinteraksjon. Resultatene kan ha blitt farget av det Desmet (2005) omtaler som en «WOW» effekt. Altså at den overlegene kapasiteten i systemet kan ha bidratt til å gjøre deltagerens vurdering kunstig høy.

Den andre dimensjonen innenfor hedonisk kvalitet, identifikasjon, dreier seg om menneskers behov for å bli anerkjent og kommunisere sosialt via sine fysiske eiendeler. At prototypen scorer høyere enn SMB2 i denne kategorien er ikke overraskende, da SMB2 representerer et svært tradisjonelt produkt som er standard i de fleste norske hjem, og dermed ikke kommuniserer hverken det ene eller det andre. Taleprototypen på sin side representerer ny teknologi og kan bidra til å kommunisere til andre at man henger med, og kanskje til og med er litt foran, på ny teknologi. Å fremstå som endringsvillig og oppdatert ansees av mange som fordelaktig.

### **8.1.3 APPEAL**

Attraktivitet, eller *APPEAL*, sammenfører følelser og opplevelser overfor et produkt i en bestemt situasjon til en målbar enhet. I AttrakDiff undersøkelsen, scorer taleprototypen betydelig høyere enn SMB2. Dette indikerer at den totale brukeropplevelsen ved bruk av prototypen er bedre enn i et tradisjonelt grensesnitt. Dette underbygges av deltageres tilbakemeldinger i intervjuene.

Hassenzahl hevder at vurderingen av *APPEAL* avhenger av hva slags modus brukerne befinner seg i. Er brukeren i *goal*-modus, er fokuset å gjennomføre en oppgave. Er brukeren i *action*-modus, kan fokuset være på selve bruken av produktet (Hassenzahl, 2004). Et sentralt spørsmål for denne oppgaven blir derfor: hva slags mål har brukerne når de ser på TV. TV-titting kan innebære både *be-goals*, i form av å slappe av og kose seg, bli underholdt, utforske innhold. Allikevel er det også en rekke *do-goals* involvert, for eksempel i å skulle navigere, ta opp programmer, finne innhold osv. I testsituasjonen var oppgavene som ble gitt til deltagerne tydelig definerte. Det sentrale målet for deltagerne i studien kan derfor antas å være fullføring av disse oppgavene. Dette medfører at modusen brukerne befant seg i under testen trolig var *goal-mode*.

Hassenzahl mener at brukernes vurdering av hedonisk kvalitet og pragmatisk kvalitet ikke avhenger av hvilken modus man befinner seg i, men brukernes vurdering av *APPEAL* og følelsesmessige reaksjoner avhenger av hvordan produktet passer inn i bruksmodusen (Hassenzahl, 2001). Dette kan tolkes som en svakhet, da brukernes vurdering av *APPEAL* kunne ha sett annerledes ut dersom brukerne hadde fått bruke prototypen i en annen bruksmodus. Dersom de for eksempel fikk bruke prototypen fritt over tid i en naturlig kontekst, kan det hende at de hadde flyttet fokuset over fra å kun gjennomføre konkrete oppgaver, til å bruke systemet på en friere, mer handlingsorientert måte. Dette kunne ha resultert i andre situasjoner og en annen vurdering av *APPEAL*.

## **8.2 Svakheter**

### **8.2.1 Manglende temporal dimensjon**

En svakhet med studien som er utført er at det ikke fanger opp den temporale dimensjonen i brukeropplevelse. Hassenzahl og Roto (2007) hevder at brukeropplevelse er en kontinuerlig begivenhet som endrer seg over tid. Dette kan for eksempel innebære at et produkt kan oppleves som svakt ved første møte, men etter hvert som man bruker det, bygger man seg opp erfaringer og blir fortrolig med systemet. På denne måten kan det være et stort skille mellom hvordan erfarne brukere av et produkt vurderer systemet og hvordan nybegynnere vurderer det. I det gjennomførte studien er interaksjonstiden målt til et

gjennomsnitt for to minutter for forskningsprototypen, og seks minutter med SMB2. Brukerne fikk ikke anledning til å gjøre seg fortrolige med systemene før testen startet. Dette begrunnes igjen med at oppløsningen på prototypen ikke var tilstrekkelig høy til å tillate en fri utforskning. For å gjøre det likt for de to systemene, ble det besluttet å ikke tillate tilvenningstid med noen av systemene.

Slik kan man si at testen kun fanger opp et øyeblikksbilde av systemet, og ikke gir noen indikasjon på hvordan systemene er å lære seg, eller hvordan brukerens holdninger til systemene endrer seg over tid. Dette gjelder trolig i særlig grad for SMB2. Flere av deltagerne kommenterte at dersom de hadde hatt muligheten til å sette seg inn i hvordan fjernkontrollen fungerte, trodde de de ville fått en mye bedre opplevelse av den. Dette kan trolig ha bidratt til å gjøre scoren til SMB2 kunstig lav.

Selv om et kriterium for utvalg av brukere var at de ikke skulle ha tidligere erfaring med SMB2, er det rimelig å anta at de fleste har brukt en form for TV-løsning. Dette kan ha medført en skjevhet i resultatene i begge retninger. På den ene siden vil brukernes tidligere bruk av liknende systemer ha kunne bidratt til å gjøre det enklere å interagere med SMB2, fordi grensesnittet kan minne om noe de har sett før. På den andre siden kan nettopp denne forutinntatte idéen om hvordan et system fungerer, bidra til frustrasjon hvis fjernkontrollen eller grensesnittet viser seg å samsvare med forventningene.

Forskningsprototypen på sin side er i stor grad fritatt fra slike skjevheter, da deltagerne i testen ikke hadde noen foregående erfaring med *voice-first* interaksjon i TV-sammenheng.

### **8.2.2 Predefinerte scenarioer**

En utfordring med laboratorieforsøk bestående av predefinerte scenarioer er at disse kan være forskjellige fra de oppgavene en bruker ville ha valgt i en fri, ikke pre-definert interaksjon. Bevan (1995) hevder at predefinerte scenarioer ikke gir innsikt i hva en bruker egentlig ville gjort dersom han var overlatt til seg selv, noe som medfører risikoen for at noe som fremstår som små utfordringer i en test, potensielt kan være store utfordringer i en reell brukssituasjon. Uten predefinerte scenarioer ville man trolig kunne etablert en mer naturtro og spontan interaksjon, og, ifølge McTear et al. (2016) vil en slik fri interaksjon kan vise mangfoldet i interaksjonen og gi resultater som er immune mot skjevheter og heterogenitet (McTear et al., 2016). På en annen side kan det hevdes at forsøk uten predefinerte scenarioer ville gjort det vanskelig å sammenlikne resultater fra ulike systemer, ettersom oppgavene som utføres i forsøkene i stor grad bestemmes av brukerne.

For å oppnå kunne la brukere interagere fritt med et produkt, er man avhengig av å ha nådd en tilstrekkelig høy oppløsning på produktet, som skal testes. I dette prosjektet var ikke ressursene til stede for å oppnå dette. For å teste prototypen vil det være nødvendig å lede brukeren gjennom et sett med oppgaver. For å forsøke å bevare validiteten i testresultatene, ble det forsøkt å gjøre testomgivelsene så realistiske som mulig, ved å plassere testen i deltagerens naturlige omgivelser for TV bruk.

### **8.2.3 Retrospektiv bedømmelse**

En annen potensiell feilkilde i studien er hukommelsesaspektet som spiller inn i en retrospektiv evaluering. At spørreundersøkelsen ble utført etter gjennomføring av samtlige oppgaver i ett system, medfører en risiko for at brukernes retrospektive bedømmelser av et system blir farget av enkelthendelser som ble spesielt godt lagt merke til, eller at oppgavene som ble løst mot slutten av testen blir en kunstig stor del av grunnlaget for vurdering. For eksempel kan brukeren oppdage en kritisk feil tidlig i studien, uten at dette nødvendigvis fører til en negativ bedømmelse av systemet i spørreundersøkelsen. Kanskje brukeren ikke ga systemet skylden for systemet, men sin egen mangel på erfaring. Kanskje ble feilen tidlig i prosessen overskygget av svært positive opplevelser mot slutten av testen. Et alternativ, som kunne bidratt til å utelukke denne risikoen, er å evaluere etter hver enkelt oppgave, fremfor å evaluere testen som helhet. Dette er dog svært tidkrevende og var ikke gjennomførbart i dette prosjektet.

## **8.3 Replikabilitet og validitet**

Utformingen av prototypen som ble brukt i den komparative studien erbasert på en kvalitativ utvikling i form av en brukersentrert designprosess, preget av undertegnedes egne erfaringer og tolkninger. Denne prosessen innebærer en rekke små valg og påvirkes av en rekke faktorer som er vanskelig å presisere og videreformidle. Overordnet kan man si at prototypen representerer min oppfatning av brukernes behov og preferanser, formidlet til meg gjennom brukersentrerte innsiktsmetoder (intervjuer, observasjoner, workshop). I tillegg har valg blitt gjort basert på mine erfaringer og antagelser om den aktuelle teknologien. Den svevende, kvalitative kunnskapen som har blitt generert under utviklingen av prototypen er vanskelig å videreformidle, og medfører dermed en utfordring når det kommer til replikabilitet. Det er svært sannsynlig at en annen person med andre erfaringer eller preferanser, ville produsert en annerledes løsning, dersom de samme metodene hadde blitt fulgt.

Metodene som ble brukt i evaluering er dog basert på en teoretisk modell, og består av kvalitative standardiserte målinger, som bidrar til å gjøre selve forsøket fullt ut replikabelt. Et viktig spørsmål med tanke på validiteten av funnene i eksperimentet er om de to evaluerte systemene gir en realistisk og rettferdig representasjon av de to interaksjonsformene som er sammenliknet. Når det gjelder SMB2, som representerer tradisjonell TV-interaksjon, i form av et grafisk grensesnitt som betjenes med en fysisk fjernkontroll, er dette én av mange ulike løsninger som eksisterer på dette området. Grunnlaget for valget av nettopp dette systemet er at det er RiksTV sin nåværende løsning. Det er ikke gjort en evaluering av de ulike tradisjonelle TV-løsningene på markedet, for å finne den som er best egnet for å representere interaksjonsformen. Det er en mulighet for at en annen løsning kunne gitt andre resultater.

Spørsmål prototypen som er utviklet og evaluert i dette prosjektet kan sees som en troverdig og rettferdig representant for taleinteraksjon er vanskelig å besvare. Prototypen er ikke forankret i eksisterende teknologi, og baserer seg i stor grad på menneskelig manipulasjon i form av Wizard of Oz. Som diskutert i 5.6.3 kan det være utfordrende for en menneskelig Wizard å begrense kapasiteten til systemet, da det vil være fristende å gjennomføre handlinger selv om de ikke står i manus. Det kan stilles spørsmålstegn ved hvor god jobb som ble gjort med tanke på denne problemstillingen i studien som er utført. Selv om det ble etterstrebet å holde seg til manus, kan min menneskelige språkforståelse ha bidratt til en interaksjon som var i overkant smidig og feilfri, og som i så måte ga et kunstig bilde av hvor godt taleinteraksjon egner seg i denne konteksten.

På en annen side kan også argumenteres for at prototypen som er utviklet ikke er basert på en grundig nok undersøkelse av brukerbehov og preferanser, og dermed gir en svak representasjon av mulighetene som kommer med introduksjon av taleinteraksjon i denne konteksten. På grunn av omfanget av prosjektet har tiden tilgjengelig til innsiktsarbeid og brukerstudier, vært begrenset. Dette kan ha påvirket flere aspekter ved prototypen, særlig knyttet til selve dialogen i systemet. En grundigere undersøkelse av aspekter som dialogflyt, personlighet, ordvalg og timing i taleinteraksjonen kunne trolig bidratt til gjøre interaksjonen bedre.

# 9 Konklusjon

## 9.1 Bidrag

Funnene fra denne oppgaven impliserer at voice-first som interaksjonsform er lovende for bruk i TV-sammenheng. Studien som er utført indikerer at taleinteraksjon kan bidra til å øke opplevd pragmatisk kvalitet i TV-produkter, i form av å gjøre det raskere og mer effektivt å navigere i systemet. I tillegg indikerer resultatene at en voice-first tilnærming kan bidra til en økt opplevelse av hedonisk kvalitet i form av å bringe frem en større grav av positive emosjonelle responser hos brukere.

Studien har gjennom en RtD-lab tilnærming, bestått i å utforme en prototype som gjenspeiler en ideell forespeilet fremtid, der taleinteraksjon fungerer optimalt. Denne metoden gjorde det mulig å evaluere hvordan brukeropplevelsen vil påvirkes av en alternativ interaksjonsform, uten faktisk utvikle den nødvendige teknologien. Evalueringen ble det utført i form av et kontrollert komparativt forsøk med prototypen og en tradisjonell tv-løsning. En AttrakDiff semantisk differensial, som måler deltageres subjektive responser til de to systemene, i tillegg til ytelsesmål og kvalitative intervjuer, gir positive resultater i favør voice-first prototypen. Resultatene indikerer at voice-first kan gjøre navigasjon enklere og mer effektiv og dermed bidra til en høyere grad av pragmatisk kvalitet. Når det gjelder hedonisk kvalitet, viste prototypen seg i stand til å frembringe mer positive følelser, noe som indikerer at voice-first interaksjon har en positiv innvirkning på den totale brukeropplevelsen i TV-produkter.

Det kan argumenteres for en manglende økologisk validitet i studien. En spesiell stimulus (en forskningsprototype), ble brukt for å vurdere HQ og PQ hos testdeltagerne. Resultatene som ble innhentet vil i stor grad avhenge av stimulusen de ble utsatt for, noe som begrenser generaliserbarheten i studien. Selv om utviklingsprosessen er basert på brukerinnsikt, er den også tungt forankret i forfatterens tolkning av denne innsikten og hvordan denne innsikten skulle gjenskapes i prototypen. Dette bidrar til å gjøre et slikt forsøk vanskelig å replikere, noe som svekker validiteten.

## 9.2 Fremtidig arbeid

For å utfylle denne studien, forslås det å gjennomføre et nytt studie som gjentar metodene så langt det lar seg gjøre. Det anbefales også at det utvikles en prototype med en høyere oppløsning, som gjør det mulig å teste den i reelle omgivelser over tid, uten predefinere

scenarier, slik at den temporale dimensjonen i brukeropplevelsen kan evalueres. En annen faktor som også bør undersøkes er hvordan brukeropplevelsen endres dersom man opplever produktet sammen med andre. Florizzi og Batterbe (2004) beskriver co-experience som en kollektiv brukeropplevelse som oppstår når mennesker bruker et produkt eller system sammen. I denne sammenhengen vil den sosiale dimensjonen som introduseres dersom mer enn en person interagerer med systemet medføre en rekke nye utfordringer. Ett av de opprinnelige målene i for denne oppgaven var å undersøke hvordan evalueringen av taleinteraksjon i TV-sammenheng varierer mellom målgrupper. På grunn av tidsmangel ble dette utelatt i dette prosjektet, men dette anbefales som et fokusområde ved videre arbeid.

# Referanser

Activate Analysis, (2018, 2. mai). *Smart Speaker Penetration, U.S., Years from Inception*. Hentet fra: <https://www.slideshare.net/ActivateInc/activate-tech-media-outlook-2018>

Alben, L., (1996). Quality of experience: defining the criteria for effective interaction design. *Interactions*, (3), 11 – 15, <https://doi.org/10.1145/235008.235010>

Amundsen, J.H. (2018). En utforskning av taleinteraksjon i TV-sammenheng. *Design 9, NTNU*. Upublisert

Amundsen, J.H. (2018). Conversational Interfaces: Are they changing us?. *Design Teori, NTNU*. Upublisert

Beauchaud, K & Kroemker, H. (2010). A framework to measure User eXperience of interactive online products. *ACM International Conference Proceeding Series*, 261-264. <https://doi.org/10.1145/1931344.1931358>

Bevan, N. (1995). Usability is quality of use. Proceedings of the International Conference on Human Computer Interaction (HCII), 349–354, [https://doi.org/10.1016/S0921-2647\(06\)80241-8](https://doi.org/10.1016/S0921-2647(06)80241-8).

Bevan, N. (2008). Classifying and selecting UX and usability measures. In *Proceedings of the International Workshop on Meaningful Measures: Valid User Experience Measurement*. 13–18. Hentet fra: <https://pdfs.semanticscholar.org/696f/f049aa613395edb4804a87425756af6762da.pdf>

Blythe, M., Hassenzahl, M., Vermeeren, A., Law, E. L-C. (2007). An Analys Framework for User Experience (UX) studies: A Green Paper. *Towards a UX manifesto - Proceedings of A Cost294-Affiliated Workshop on BHCI. 1-5*.

Boehner, K., DePaula, R., Dourish, P., Senger, P., (2007). How emotion is made and measured? *International Journal of Human–Computer Studies* 65(4), 275-291 <https://doi.org/10.1016/j.ijhcs.2006.11.016>

Bowden, M. A. (2006), *Mind As Machine: A History of Cognitive Science*, London: Oxford University Press.

- Brooke, J. (1996). SUS: A quick and dirty usability scale. I P. W. Jordan, B. Weerdmeester, A. Thomas, & I. L. McClelland *Usability evaluation in industry*, (s.189–194). London: Taylor and Francis.
- Desmet, P., Porcelijn, R & van Dijk, M. (2005). HOW to design WOW Introducing a layered-emotional approach. I S. Wensveen, *Proceedings of The International Conference on Designing Pleasurable Products and Interfaces*, (s.71-89).October 24-27, Eindhoven
- Durrant, A., & Price, J. (2015). RTD 2015 Provocation by Sir Christopher Frayling Part 1: Research Through Design Evolution. Hentet fra: <https://vimeo.com/129775325>
- Dybkjær, L., Bernsen, N. O., & Minker, W. (2004). Evaluation and usability of multimodal spoken language dialogue systems. *Speech Communication*, 43, 33–54.
- Fichter, D., & Wisniewski, J. (2017). Chatbots Introduce Conversational User Interfaces. *Online Searcher*, 41(1), 56-58.
- Forlizzi, J., & Battarbee, K. (2004). Understanding experience in interactive systems. I D. Benyon, P. Moody, D. Gruen & I. McAra-William (Eds.), *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS '04*, 261–268.
- Frayling, C. (1993). Research in art and design. *Royal College of Art Research Papers*, 1(1). 1-5.  
Hentet fra:  
[http://researchonline.rca.ac.uk/384/3/frayling\\_research\\_in\\_art\\_and\\_design\\_1993.pdf](http://researchonline.rca.ac.uk/384/3/frayling_research_in_art_and_design_1993.pdf)
- Gaver, W. (2012). What Should We Expect From Research Through Design?. *Conference on Human Factors in Computing Systems - Proceedings*.  
<https://doi.org/10.1145/2207676.2208538>.
- Gaver, W. (2014). Science and Design: The Implications of Different Forms of Accountability. I J. S. Olson & W. A. Kellogg (Eds.), *Ways of Knowing in HCI* (s. 143–165). New York, NY: Springer New York.
- Gibbon, D., Mertins, I., & Moore, R. (2000). Handbook of multimodal and spoken dialogue systems: Resources terminology and product evaluation. *Computational Linguistics*. 27(1). 149-150. <https://doi.org/10.1162/coli.2000.27.1.149b>. Norwell, MA, USA: Kluwer Academic.

- Hassenzahl, M. (2001). The Effect of Perceived Hedonic Quality on Product Appealingness. *International Journal of Human-Computer Interaction*, 13(4), 481-499.
- Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product. *Funology: From Usability to Enjoyment*. 31-42. [https://doi.org/10.1007/978-3-319-68213-6\\_19](https://doi.org/10.1007/978-3-319-68213-6_19)
- Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, 19(4), 319-349.
- Hassenzahl, M. (2007). The hedonic/pragmatic Model of User Experience. *Towards a UX manifesto - Proceedings of A Cost294-Affiliated Workshop on BHCI*. 10-14.
- Hassenzahl, M. (2008). User Experience (UX): Towards an Experiential Perspective on Product Quality. *ACM International Conference Proceeding Series*. 339. <https://doi.org/10.1145/1512714.1512717>.
- Hassenzahl, M., Platz, A., Burmester, M. & Lehner, K. (2000). Hedonic and ergonomic quality aspect determine a software's appeal. *Conference on Human Factors in Computing Systems – Proceedings*, 2(1), 201-208.
- Hassenzahl, M. & Roto, M. (2007). Being and Doing: A Perspective on User Experience and its Measurement. *Interfaces*. 72(1), 10-12.
- Hassenzahl, M. & Tractinsky, N. (2006). User Experience – A research Agenda, Behaviour & Information Technology, 25(2), 91-97.
- Hirschberg, J. & Manning, C. (2015). Advances in natural language processing. *Science*. 349(1), 261-266. Hentet fra: <https://cs224d.stanford.edu/papers/advances.pdf>.
- Houde, S. & Hill, C. (1997). What do prototypes prototype. I *Handbook of Human- Computer Interaction*. (s. 367–381).New York,NY: Elsevier,
- Hone, K. S., & Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering*,6(3–4), 287–303.
- Höök, K., Bardzell, J., Bowen, S., Dalsgaard, P., Reeves, S., & Waern, A. (2015). Framing IxD knowledge. *Interactions*, 22(6), 32–36. <https://doi.org/10.1145/2824892>

- Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, 37(1), 81-88. <https://doi.org/10.1080/02763869.2018.1404391>
- Igbaria, M., Schiffman, S. J., & Wieckowski, T. J. (1994). The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. *Behaviour & Information Technology*, 13(1), 349–361.
- International Organization for Standardization (2010). *Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: Guidance on usability* (ISO No. 9241). Geneva, Switzerland.
- International Organization for Standardization. (2010). *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*. (ISO 9241-210:2010) Geneva, Switzerland.
- Jordan, P.W. (1998). *An Introduction to Usability*. London: Taylor & Francis
- Kirakowski, J., & Corbett, M. (1993). SUMI: The software usability measurement inventory. *British Journal of Educational Technology*, 24(3), 210–212.
- Koskinen, I., Zimmerman, J., Binder, T., Redstrom, J., & Wensveen, S. (2011). *Design Research Through Practice: From the Lab, Field and Showroom*. Waltham, MA: Elsevier.
- Kühnel, C. (2012). *Quantifying Quality Aspects of Multimodal Interactive Systems*. Berlin: Springer-Verlag Berlin Heidelberg.
- Law, E. L-C., Vermeeren, A., Hassenzahl, M. & Blythe, M (2007). Towards a UX manifesto. *COST294-MAUSE affiliated workshop*, Lancaster UK.
- Law, E. L-C. & Hornbæk, K. (2007) Measures of Usability and User Experience (UX): Correlation and Confusion. *COST294-MAUSE affiliated workshop*, Lancaster UK.
- Law, E. L.-C., van Schaik, P., & Roto, V. (2014). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies*, 72(6), 526–541. <https://doi.org/10.1016/j.ijhcs.2013.09.006>
- Lim, Y-K., Stolterman, E. & Tenenber, J. (2008). The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Trans. Comput.-Hum. Interact.* 15(2), 1–27. <http://doi.org/10.1145/1375761.1375762>

- McTear, M., Callejas, Z., & Griol, D. (2016). *The Conversational Interface: Talking to Smart Devices*: Switzerland: Springer Publishing Company.
- Möller, S & Engelbrecht, K.-P & Kühnel, C & Naumann, Anja & Wechsung, I & Weiss, B. (2010). Evaluation of Multimodal Interfaces for Ambient Intelligence. *Human-Centric Interfaces for Ambient Intelligence*. 347-370. <http://doi.org/10.1016/B978-0-12-374708-2.00014-0>.
- Norman, D. (2013) *The Design of Evryday Things*. New York : Basic Books
- Norman, D. A., Miller, J., & Henderson, D. A. Jr. (1995). What you see, some of what's in the future, and how we go about doing it: HI at Apple Computer. In J. Miller, I. R. Katz, R. L. Mack & L. Marks (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI '95*. (s: 155). New York, NY: ACM.
- Obrist, M., Roto, V., & Väänänen-Vainio-Mattila, K. (2009). User experience evaluation: Do you know which method to use? I *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2009* (s. 2763–2766). New York, NY: ACM.
- Obrist, M., Roto, V., Law, E. L.-C., Väänänen-Vainio-Mattila, Vermeeren, A. P. O. S. & Buie, E. A. (2012). Theories behind UX research and how they are used in practice. I *Proceedings of the International Conference on Human Factors in Computing Systems, CHI '12* (s. 2751–2754). New York, NY: ACM.
- Odom, W., Wakkary, R., Lim, Y., Desjardins, A., Hengeveld, B., & Banks, R. (2016). From Research Prototype to Research Product. 2549–2561. <https://doi.org/10.1145/2858036.2858447>
- Ortony, A., Clore, G. & Collins, A. (1988). The Cognitive Structure of Emotion. *Contemporary Sociology* 18(6). 957-958. <https://doi.org/10.2307/2074241>.
- Prentice, D. A. (1987). Psychological correspondence of possessions, attitudes, and values. *Journal of Personality and Social Psychology*, 53(6), 993-1003.
- Roto, V., Law, E., Vermeeren A, Hoonhout, J., (2010). Demarcating User eXperience. *Dagstuhl Seminar*. Hentet fra: [http://drops.dagstuhl.de/opus/volltexte/2011/2949/pdf/10373\\_AbstractsCollection.2949.pdf](http://drops.dagstuhl.de/opus/volltexte/2011/2949/pdf/10373_AbstractsCollection.2949.pdf)
- Scherer, K. R. (2005). Appraisal Theory. I *Handbook of Cognition and Emotion*. <https://doi.org/10.1002/0470013494.ch30>

- Schulze, K. and Krömker, H. (2010). A framework to measure user experience of interactive online products. *Proc. Int. Conf. on Methods and Techniques in Behavioral Research*, 14(1). <https://doi.org/10.1145/1931344.1931358>
- Schwartz, S.H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, 53(3), 550-562.
- Seo, E. (2018). *The Voice Musings #1: Designing For Voice When Alexa Doesn't Really Understand*. UX Collective. Hentet fra: <https://uxdesign.cc/the-voice-musings-designing-for-voice-when-alexa-doesnt-re-ally-understand-469c32e9fabd>
- Simonsen, J. G. (2017). User Experience. I *The Wiley Handbook of Human Computer Interaction*. <https://doi.org/10.1002/9781118976005.ch10>
- Swallow, D., Blythe, M., Wright, P., (2005). Grounding experience: relating theory and method to evaluate the user experience of smartphones. I *Proceedings of the EACE'05*, 91–98.  
Hentet fra:  
[https://www.researchgate.net/publication/228577754\\_Grounding\\_experience\\_relating\\_theory\\_and\\_method\\_to\\_evaluate\\_the\\_user\\_experience\\_of\\_smartphones](https://www.researchgate.net/publication/228577754_Grounding_experience_relating_theory_and_method_to_evaluate_the_user_experience_of_smartphones)
- Tullis, T. & Albert, B. (2013) *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics* (2), Chapter 4 - Performance Metrics. s. 63 – 97.  
Burlington, MA: Morgan Kaufman
- Vermeeren, A., Law, E. L-C., Roto, V., Obrist, M., Hoonhout, J. & Väänänen, K. (2010). User experience evaluation methods: Current state and development needs. *NordiCHI 2010: Extending Boundaries - Proceedings of the 6th Nordic Conference on Human-Computer Interaction*. 521-530. <https://doi.org/10.1145/1868914.1868973>.
- Walker, M., Litman, D., Kamm, C. & Abella, A. (1997). PARADISE: A framework for evaluating spoken dialogue agents. *ACL '98/EACL '98 Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*. 271-280.  
<https://doi.org/10.3115/979617.979652>.
- Ware, C. (2012). *Information Visualization: Perception for Design (Interactive Technologies)*.(3). San Fransisco ,CA: Morgan Kaufmann Publishers.

- Wensveen, S. & Matthews, B. (2015). Prototypes and prototyping in design research. I P. A. Rodgers & Joyce Y. *Routledge Companion to Design Research* (s. 262-276) London: Routledge. .
- Whitenton, K. (2017). Voice First: The Future of Interaction? *Nielsen Norman Group*. Hentet fra <https://www.nngroup.com/articles/voice-first/>
- Whiteside, J., & Wixon, D. (1987). The dialectic of usability engineering. I G. Salvendy, S. L. Sauter, J. J. Hurrell Jr. *Proceedings of the Conference on Human-Computer Interaction, INTERACT '87* (s. 17–20). Amsterdam: Elsevier.
- Wright, P., McCarthy, J., & Meekison, L. (2004). Making sense of experience. I M. A. Blythe, K. Overbeeke, A. F. Monk & P. C. Wright (Eds.), *Funology: From usability to enjoyment* (s. 43–53). Dordrecht, Kluwer Academic Publishers.
- Zimmerman, J. & Forlizzi, J. & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. *Conference: Proceedings of the 2007 Conference on Human Factors in Computing Systems, CHI 2007*. 493-502.  
<https://doi.org/10.1145/1240624.1240704>.
- Zimmerman, J., & Forlizzi, J. (2014). Research Through Design in HCI. I J. S. Olson & W. A. Kellogg. *Ways of Knowing in HCI* (s. 167–189). New York, NY: Springer New York.  
[https://doi.org/10.1007/978-1-4939-0378-8\\_8](https://doi.org/10.1007/978-1-4939-0378-8_8)

# Appendix A: Intervjuguide

## Intervjuguide

- Kan du fortelle litt om dine TV vaner i dag?
  - Hva bruker du av hardware?
  - I hvilke situasjoner ser du på TV – beskriv en typisk setting?
- Dersom streaming: Hvordan går du frem når du skal velge noe å se på?
  - Bestemt på forhånd?
  - Leter etter noe som blir spennende ut?
- Kan du beskrive et typisk hendelsesforløp når du setter deg for å se på TV.
  - Streaming
  - Live
- Hvilke deler av tjenesten bruker du mest?
  - Streaming
  - Film
  - Serie
  - Tv-arkriv
  - Direkte TV
- Er det noe ved TV-opplevelsen du finner problematisk/frustrerende?
  - Manøvrere
  - Fjernkontrollen
  - Huske hvor kanalen er
  - Huske hvilken kanal programmet går på
  - Søke etter fil/serie
- Hva er det viktigste funksjonene for deg i et TV/streamingprodukt?
  - Holder styr på hvor langt du har kommet i en serie?
  - Lærer om deg, vet hva du vil se på.
  - Kommer med forslag basert på ting jeg har sett tidligere
  -
- Har du noen erfaring med bruk av VUI?
  - Hvilke tjenester
  - Hvordan har opplevelsen av dette vært?
- Har du noen tanker rundt det å interagere med TVen gjennom tale?
  - Eksempel på en situasjon det kunne vært nyttig?

# Appendix B: Workshopplan

## Workshopplan

**Tid:** 50 min.

**Deltagere:** 4 eller 6 personer.

Workshopen består av en enkel dramatiseringsøvelse der to og to personer går sammen og dramatiser tre gitte situasjoner. Den ene personen skal spille talestyrt-TV og den andre skal spille seer. Situasjonene som skal dramatiseres er valgt ut på bakgrunn av tidligere utførte brukerintervjuer og observasjoner. Disse tre situasjonene ble trukket frem av flere brukere som problematiske med dagens TV-løsning.

Gruppene får 15 minutter til å komme opp med de dialogene de mener er best for å løse de gitte situasjonene gjennom tale. Etter dette skal hver av gruppene presentere sine løsninger, ved å dramatisere for hverandre. Avslutningsvis vil det bli en liten diskusjon det vi snakker om de ulike løsningene og deler erfaringer.

### Spørsmål til diskusjonen:

- Hvordan opplevde dere denne øvelsen?
- Hva tror dere om talestyrt TV?

### Tidsplan:

13:00	Ønske velkommen. Deltagere setter seg til rette.
13:05	Introdusere prosjektet. Gi en liten beskrivelse av prosjektet. Og fortelle om hva som skal skje. Presentere øvelsen de skal gjennom.
13:10	Gruppene får utlevert et ark med tre situasjoner. De skal så sammen komme frem til en måte å løse dette på gjennom talestyring. Ikke la seg påvirke av teknologiske begrensninger. Gjør det slik dere skulle ønske at det kunne skje. Får utlevert TV-ramme for dramatisering.
13:30	Gruppene presenterer sine løsninger.
13:40	Oppsummering. Diskusjon.
13:50	Takk for i dag :D

## Hvordan snakke med en TV?

Dagens TV-løsninger inneholder som regel et underholdningstilbud utover direktesendt TV. I tillegg inkluderer TV-tilbudet streaming tjenester på nett og på casting enheter som Apple TV/Chromecast. I denne workshopen skal vi utforske hvordan vi kan bruke tale til å navigere i TV- og streamingtjenester. Ta utgangspunkt i følgende to situasjoner:

- Du setter deg ned for å finne noe å se på. Du vet ikke på forhånd hva du ønsker å se, men du vet at du kanskje kunne tenkt deg en komedie som minner litt om en film du har sett før.
- Du skal se en fotballkamp mellom Norge og Island, men aner ikke hvilken kanal den går på. Du har dårlig tid fordi du vet at kampen snart begynner.

Deres oppgave er å komme frem til et forslag til hvordan disse situasjonene kan løses gjennom tale. Dere skal ikke la dere begrense av de tekniske aspektene ved talestyring, tenk dere at alt er mulig! Det er allikevel viktig at dere fokuserer på å lage en entydig løsning med et spesifikt «manus». Dere velger selv hvor mye «TVn» skal svare/bekreft på seer-input, gjør det dere mener er det beste!

For å presentere løsningene deres skal dere dramatisere de to situasjonene. En skal spille seer og en skal spille talestyrt-TV.

# Appendix C: Forsøksguide

## Forsøksguide

### Introduksjon:

Introdusere meg selv. Gi deltageren en introduksjon i hva prosjektet går ut på. Forklare gangen i forsøket.

### Test 1:

Første testsystem settes opp. Brukeren informeres kort om hvordan systemet fungerer. Bruker gjennomfører følgende oppgaver på kommando fra testleder.

- Skru på TV
- Sett på bestemt program
- Finn ut når bestemt fremtidig program starter
- Ta opp program
- Finn film med bestemt sjanger
- Finn film med bestemt skuespiller
- Spill av arkivert opptak

Bruker fyller ut AttrakDiff undersøkelse for systemet som er testet.

### Test 2:

Prosedyre fra test 1 gjentas.

### Avsluttende intervju:

Spør bruker om hvordan opplevelsen av de to systemene var.

# Appendix D: Resultater ytelsesmål

SWB2															
Deltager	Oppgave 1 (fått all av program)	Oppgave 2 (fått all program starter)	Oppgave 3 (all programsløst)	Oppgave 4 (all program)	Oppgave 5 (fått alle deler)	Oppgave 6 (skulle)	Oppgave 7 (fått opp)								
Tid	Subscore	Tid	Subscore	Tid	Subscore	Tid	Subscore	Tid	Subscore	Tid	Subscore	Tid	Subscore		
1	30	128	40	81	124	259	20	1							
2	33	91	56	136	90	156	39	1							
3	35	109	59	137	90	156	39	1							
4	70	23	19	243	63	73	46	1							
5	85	58	7	219	45	79	31	1							
6	52	20	20	84	50	120	69	0,5							
7	40	45	28	143	28	143	68	0,5							
8	40	42	33	143	28	143	68	0,5							
9	22	14	9	32	29	79	20	1							
10	32	32	49	44	34	139	28	1							
Gjennomsnitt	40,8	100 %	49,2	85 %	26,7	95 %	78,1	0,75	75 %	66,7	0,75	75 %	40	0,9	90 %
Total tid: 421,5 sek															
Gjennomsnittstid pr. oppgave: 60,2 sek															
Gjennomsnittlig suksessrate: 0,84 p															
Taleprogramtype															
Oppgave 1	Oppgave 2	Oppgave 3	Oppgave 4	Oppgave 5	Oppgave 6	Oppgave 7									
Tid	Subscore	Tid	Subscore	Tid	Subscore	Tid	Subscore	Tid	Subscore	Tid	Subscore	Tid	Subscore		
1	9	14	11	6	10	21	15	0,5							
2	10	8	7	7	12	10	10	1							
3	18	12	6	6	12	10	18	1							
4	13	10	9	14	14	8	4	0,5							
5	13	15	10	9	14	15	11	1							
6	11	14	18	8	17	11	8	1							
7	11	12	17	5	9	7	8	1							
8	12	10	10	6	10	6	10	1							
9	12	10	6	6	12	11	11	1							
10	15	10	7	7	12	9	11	1							
Gjennomsnitt	12,4	100 %	12,1	100 %	12,2	100 %	11,6	0,9	100 %	10,6	1,008	11,6	0,9	90 %	
Total tid: 78,4 sek															
Gjennomsnittstid pr. oppgave: 11,2 sek															
Gjennomsnittlig suksessrate: 0,97 p															

