



Blekinge Tekniska Högskola

Campus Karlshamn

Datavetenskapliga institutionen för teknik och estetik

2018-05-18

# Röststyrning av mediauppspelning

ME1588 V18 Lp34 Kandidatarbete i Webbutveckling

Martin Fröberg

Patrik Svensson

Handledare: Peter Giger, Mattias Schertell, Siamak Khatibi

Examinator: Kåre Synnes

## Abstrakt

Med detta kandidatarbete vill vi undersöka funktionaliteten mellan röststyrning och andra styrenheter inom smarta hem, där fokuset för undersökningen är behov och intresse av tekniken. Med hjälp utav synsättet agentuell realism som förklaras av Karen Barads teorier finner vi en inblick om hur människan påverkas av dagens enheter och hur enheterna påverkas av oss.

Designprocessen uppvisar arbetet kring gestaltningen med hjälp utav ett sammankopplande av en röststyrt internet of things applikation, vilket under en kvalitativ intervju får ett resultat kring frågeställning som används till grund för undersökningen.

Resultatet för frågeställningen visar de intervjuade personernas subjektiva åsikt och tankar om röststyrning i dagens hem och om det är något som eftersträvas. Denna undersökning visar att röststyrningen är ett intresse och ett mindre behov i det vardagliga hemmet, men behöver mer utveckling för att fler människor skall vilja använda den.

**Nyckelord:** *Röststyrning, navigationssystem, intresse, behov, potential*

## Abstract

With this bachelor thesis, we want to examine the functionality of voice control and other smart home management units, where the focus of the study is the need and interest of the technology. With the help of the agential realism approach as explained by Karen Barad's theories, we find an insight into how people are affected by today's devices and how the devices are affected by us.

The design process shows the work on the project with the help of a combination of software for a voice-controlled internet of things application, which during a qualitative interview gets a result of the questioning used for the investigation.

The outcome of the question shows the subjective opinions and thoughts of voice control in today's home and if it is something that is sought. This study shows that voice control is an interest and a lesser need in the everyday home, but needs more development in order for more people to want to use it.

Keyword: Voice control, navigation system, interest, needs, potential

<b>1. Bakgrund</b>	<b>4</b>
1.1 Frågeställning	5
1.2 Syfte	6
1.3 Det gemensamma textskapandet	7
<b>2. Tidigare och aktuell forskning</b>	<b>8</b>
2.1 Internet of Things	8
2.2 Smarta hem	9
2.3 Röststyrning	9
2.4 Brusreducering	11
2.5 Behovet för röststyrning	12
<b>3. Metoder</b>	<b>13</b>
3.1 Agentiell Realism	13
3.2 Kvalitativ metod	15
3.3 Parprogrammering	16
<b>4. Designprocess</b>	<b>17</b>
4.1 Koncept	18
4.2 Val av gestaltningstekniker	19
4.3 Utförande	24
<b>5. Resultat av undersökningen</b>	<b>26</b>
<b>6. Diskussion</b>	<b>29</b>
<b>Separata bilagor</b>	<b>34</b>
Källförteckning	34
Ordlista	37
Övriga bilagor	38

# 1. Bakgrund

*I det kommande kapitlet förklarar vi hur vi kommit fram till ämnet vi valt, varför vi har valt det och vad vi hoppas kunna visa med arbetet.*

I takt med att vår teknik idag gör stora framsteg varje dag ser vi ett påbörjat beroende av att alltid vara uppkopplad till internet. Med hjälp av internet får vi idén om en effektivare värld där allt befinner sig inom ett räckhåll av ett enkelt knapptryck.

Utveckling inom våra hem börjar applicera dagens teknik och utvecklas till smarta hem som underlättar människors vardag på diverse sätt. Smarta hem ger dig åtkomst till att styra din teknik i hemmet som belysning, termostat, högtalare etc, med hjälp av olika webbapplikationer.

Sättet du kontrollerar ett smart hem är en viktig del, användaren måste både vara bekväm med tekniken och känna att den underlättar vardagen. Utöver röststyrda hem assistenter såsom Amazon Echo(Crist, 2018), Google Home(Gebhart, 2018) och Amazon Alexa(Crist, 2018) används ofta smarttelefoner för att styra funktioner inom ett smarthem. Huvudsakligen används touchskärmar från smart telefoner, men de har även inbyggda röststyrningssystem som går att koppla till ditt smarta hem, även då den här röststyrningen inte är gjord för att styra externa applikationer.

Även då det finns teknik idag för att kunna skapa ett röststyrt smart hem används det inte lika flitigt som vi hade hoppats, vi vill undersöka varför det inte används mer och vilka problem som tekniken har.

Röststyrning i sig är ett intressant ämne då det är naturligt för oss människor att använda rösten för att kommunicera, men också för att röststyrning kommer att fungera precis likadant för människor som har skriv och lässvårigheter, tex dyslexi. Därför kommer utveckling av den här tekniken att bidra till att lösa ett flertal problem med användning av digitala funktioner.

Översättning av språk har länge funnits digitalt och varit användbart, en potentiell förbättring hade varit just med implementationen av röststyrning, att endast använda sig utav en smarttelefon mellan parterna för att kontinuerligt översätta vad som sägs till den andra partens modersmål hade varit en större utveckling inom tolkning och samtidigt förbättra utvecklingen av röststyrning för fler språk.

Under en tidigare kurs har vi undersökt röststyrning inom webbläsare, där vi utvecklade en hemsida som enbart använde sig utav tekniken. Hemsidan var huvudsakligen en undersökning hur långt röststyrning har kommit under de senare åren, då det blivit ett allt populärare ämne via utvecklingen av hem assistenter. Röststyrningen på hemsidan fungerade bättre än förväntat, vilket fick oss intresserade att fortsätta undersökningen inom tekniken. Vi känner att röststyrningen har sin plats idag i samhället och vill därför undersöka både för- och nackdelar med att använda röststyrning över andra styrenheter, såsom mus och tangentbord, touchskärmar, handrörelser, sensorer och fjärrkontroller.

Februari 2018 i Sverige har det blivit olagligt att köra bil med en mobiltelefon i handen(Transportstyrelsen 2018), även den här diskussion har skapat scenarion där röststyrning som styrteknik har skapat ett intresse. Svara telefon, byta musik, skicka meddelanden är något som är fullt möjligt med röststyrning utan att behöva röra sin telefon. Vilket är ett av områdena vi känner röststyrningen kan underlätta över tid, då tekniken har sina ärvda problem att inte fungera när en extern ljudkälla skapar problem för röstigenkänningen.

## 1.1 Frågeställning

Det finns ett fåtal realistiska sätt att kontrollera funktionerna inom ett smart hem idag, vilket får användaren både känna att tekniken faktiskt underlättar deras liv, och även känner sig trygg med användningen.

Det vi vill undersöka i det här examensarbetet är interaktionen mellan människa och navigationssystem inom ett smart hem. Vi vill bland annat utveckla vårt perspektiv om vilken metod för att kontrollera funktion som är mest användbar idag och vad det potentiellt kan vara inom en framtid.

Vår frågeställning för att fortsätta arbetet lyder därför: ”Är röststyrning idag ett funktionellt och accepterbart alternativ till traditionella metoder?”

## 1.2 Syfte

Tekniken för röststyrningen har funnits länge men den har inte riktigt slagit igenom förens de senare åren, men även idag används inte till den grad vi känner att den borde. Vi försöker därför lyfta fram potentialen inom röststyrning i det här examensarbetet med avgränsningen inom användning för smarta hem. Samt att vi kommer jämföra röststyrningen med populära styrmetoder vi använder idag.

Under en tidigare kurs har vi arbetat med artificiell intelligens och interaktionen med en sådan, vi gestaltade en simpel hemsida som använde sig utav röststyrning som den enda form av styrning. Vi kände att användningen med den här hemsidan under utvecklingsprocessen underlättade vårt arbete på ett förvånansvärt sätt och såg därifrån potentialen kring att använda röststyrning i andra scenarion.

Röststyrning har potentialen att lösa förhinder med användning av alternativa styrmetoder för människor med diverse handikapp såsom former av förlamning, och läs- och skrivsvårigheter. Vår hypotes är att röststyrning kan visa sig vara en funktionell metod för att kontrollera digitala funktioner, eftersom att tala med rösten är ett naturligt sätt att kommunicera på för oss människor.

### 1.3 Det gemensamma textskapandet

Vi delade upp textdelen av vårt arbete till att fokusera på specifika forskningsområden och metoder, för att sedan arbeta tillsammans med en grund och genomföra resterande skrivdel. Google docs har använts i den gemensamma delen av skrivandet, där det huvudsakliga syftet var att kunna skriva separat inom samma dokument. Skrivandet har gått till genom diskussion av innehåll, för att sedan delas upp i hanterbara underrubriker, och sedan en utförlig renskrivning av texten.



## 2. Tidigare och aktuell forskning

*Forskningen för vår undersökning har huvudsakligen haft fokus på tekniken som används inom smarta hem samt behov och potential kring röststyrning.*

### 2.1 Internet of Things

Internet of Things är allt från våra telefoner, klockor, bilar, vitvaror etc, som är uppkopplade och använder internet till att kommunicera med andra elektroniska enheter.

1999 berättar Kevin Ashton om att den information och data vi använder och ser på webben är skapad av en människa. En människa är långsam och en informationskälla man inte alltid kan lita på. Därför är utvecklingen att försöka få alla elektroniska apparater att skicka information till varandra utan felaktig data, med hjälp av sensorer som skickar data direkt till internet utan någon mänsklig kontakt. Denna tanke var det som väckte visionen om Internet of Things. (Waher, 2015)

Idag innehåller majoriteten av våra elektroniska apparater diverse sensorer och användargränssnitt som ger användaren en möjlighet till att sätta sin personliga stil på inköpta elektroniska apparater. Man kan se på hur mikrovågsugnen har utvecklats, till en början gav den sin enkla funktion med att enbart värma upp den djupfrysta maten med en timer, till senare där man kunde bestämma hur snabbt den skulle värma maten. Redan i ett tidig stadié kan man kalla sin mikrovågsugn för en smart mikrovågsugn då den använder sig av inbyggda processorer för att bestämma värmen under uppvärmningen. En smart mikrovågsugn idag gör att du kan trycka på förinstallerade program som värmer maten till en perfekt temperatur, och utgör funktionalitet av även andra vitvaror, såsom ugn och spis.

Denna utveckling som sker på en mikrovågsugn är på grund utav internet of things, där möjligheterna blir fler ju fler elektroniska apparater som skapas.

## 2.2 Smarta hem

Smarta hem är uppbyggt av internet of things och är den nya trenden som växer varje dag som går. För att förstå smarta hem så bryter vi ner det i tre delar, första delen handlar om hur smarta apparater kan underlätta vardagslivet i ens hem, allt från att låsa dörren med ett knapptryck till att höja och sänka ljudnivån på sina högtalare. Andra delen handlar mer om hur man kan minska sin energiförbrukning i hushållet med, tex timers för att släcka lampor, miljövänligt läge på televisionen etc. Sista delen beskrivs mer som design fiktion där människor ser smarta hem som en ny våg av teknisk utveckling för mänskligheten och dess behov. (Wilson, Hargreaves & Hauxwell-Baldwin 2014)

## 2.3 Röststyrning

Med röststyrning i det här sammanhanget är tekniken eller metoden att kontrollera digitala funktioner med hjälp av sin röst. Tekniken bygger på att användaren talar normalt medans enheten översätter och plockar ut nyckelord ur frasen, sedan försöker matcha dem till en lista över ord och tillsist omvandla frasen till text som i sin tur kan användas för att styra förprogrammerade funktioner.

Efter en fras har blivit översatt till text så är det inga problem att använda sig fritt utav tekniken, problemet röststyrningen har är i själva översättningen av de talade språket. Svårigheten med att översätta mänskligt tal ligger i många års evolution som vår hjärna blivit anpassad till att förstå naturligt, som olika dialekter, språk, meningsstrukturer och även oklara ord som vi människor förstår genom kontexten det sägs i.

## Automatic Speech Recognition

Metoden för att översätta meningen som blivit talad blir kallad Automatic Speech Recognition (ASR). ASR är enligt Woodford ett utav de med komplexa områdena inom datavetenskap eftersom det involverar ett flertal ämnen som komplex språkvetenskap, matematik och datateknik(Woodford, 2018).

Huvudsakligen används fyra olika metoder för en dator att uppfatta mänskligt tal. För att snabbt summera de här metoderna:

- *Simple pattern matching* - Varje ord uppfattas i helhet av datorn utan en djupare analys försöker den sedan göra en tolkning vad som har sagts.
- *Pattern and feature analysis* - Talet analyseras vidare och bryts ner till olika kännetecken i former av vokaler, konsonanter och andra specifika normer inom språket.
- *Language modeling and statistical analysis* - Försöker förutse vilka ord som kan med störst sannolikhet följa ett annat ord, samt se om frasen är grammatisk korrekt.
- *Artificial neural networks / Deep neural network* - En simplificerad digital version av den mänskliga hjärnan som använder sig utav inlärningsmetoder för att kunna tränas upp att känna igen vanliga använda frasen och strukturer av meningar.

Utöver de här metoderna för att skapa en ASR, kommer även diskussionen om att ha röstigenkänningen på en server Online eller Offline i enheten. Både metoderna har sina för- och nackdelar och appliceras olika beroende på syftet med röststyrningen i specifika fall.

## Online Recognition

Med den här metoden menas att programmet för att översätta den talade meningen finns på en server som nås via ett nätverk från enheten, vanligtvis internet. Personen talar därför in i mikrofonen på sin enheten som senare skickas över till programmet på en server som översätter meningen och skickar tillbaka data till enheten som i sin tur kan använda det för att starta olika förinställda funktioner och applikationer.

Fördelarna med att använda den här metoden är för den arbetskraft en extern server kan ha emot t ex din mobiltelefon. Avancerade röstigenkänningsprogram idag kräver en större mängd kraft för att noggrant kunna översätta den talade meningen korrekt, vilket inte vanliga enheter som är menat för att använda sig utav röststyrning har. Genom att använda sig utav en extern server kan många fler enheter koppla till ASR systemet utan några större problem.

Den här metoden kräver dock att enheten konstant har ett bra nätverk förhållande till servern för att kunna användas.

## Offline Recognition

För de enheterna som möjligtvis inte kan hantera goda nätverks förhållanden implementeras en offline-version av ASR systemet. Huvudsakligen används den här metoden för att undvika att enheten har kontakt med internet, antingen från ett säkerhetsperspektiv eller användning i miljöer där kontakt med internet kan ha svårigheter.

Eftersom ASR oftast har ett krav på datorkraft både i prestanda och lagring måste till viss del funktionerna i programmet minimeras. Funktioner såsom maskininlärning, val av språk och större ordlista att matcha emot, kan bli svårt att implementera med tanke på vilken enhet som ska klara röststyrning offline.

## 2.4 Brusreducering

I studien om brusreducering som är skriven av Kevin Gulley (2014) förklaras det hur mikrofonen stänger ute de ljud som inte är lämpat att användas. Han förklarar att det finns tre olika nivåer av brusreducering (brusreducering, ultra brusreducering och mörkläggande brusreducering)

1. **Brusreducering:** Denna teknik är baserad på att inte vara känslig för ljud från alla håll. Med hjälp av två små mikrofoner känner den av vibrationer, är dessa vibrationer lika så släpps inget ljud igenom. Men när vibrationen från användaren blir större på ena mikrofonen så skickas dessa ljudvågor vidare.
2. **Ultra brusreducering:** För att få bättre brusreducering stänger man av ljud som går 270 grader runt mikrofonen och fokuserar på användarens mun. Placeringen av

mikrofonen är viktigare i detta stadie gentemot förra nivå för att inte mikrofonen ska stänga ute användarens röst.

3. **Mörkläggande brusreducering:** Konceptet kommer från den vanliga brusreduceringen där man använder sig av två mikrofoner. Skillnad till första steget av brusreducering är att mikrofonerna jämför informationen som plockas upp och efter det gör en slutsats på vad som är det viktiga ljudet att spela in.

## 2.5 Behovet för röststyrning

Enligt Simon Andersson och Simon Hagelins arbete *Röststyrning och text-till-tal i mobila enheter för personer med läs- och skrivsvårigheter*(2013) syftar de på att behovet med röststyrning för människor med läs- och skrivsvårigheter existerar. De skriver i resultatet av deras arbete så här:

*“Många av testpersonerna med läs och skrivsvårigheter föredrog röststyrning och text till tal före att läsa receptet själv. Således kan vi dra slutsatsen att röststyrning och TTS ökar användbarhet till en viss grad för personer med läs och skrivsvårigheter även om det går snabbare att läsa receptet på egen hand för samtliga testpersoner.”*

De nämner att ett utav problemen med röststyrning idag för människor med diverse handikapp har med igenkänningen av röststyrningen att göra. Där människor kan ha svårigheter med tekniken och ser den som mindre användbar när problem med mikrofon, dialekt och externa ljudkällor påverkar dess funktionalitet.

Enligt Brainline (Brainline 2010) är röststyrning inte bara en lösning för människor med läs och skrivsvårigheter. Under studien tas det upp hur människor med funktionsnedsättning ska kunna vara med på skollektioner utan att behöva förlita sig på sin lärare eller annan personal som hjälper till. Röststyrningen ska ge personerna fri tillgång till att skriva ner sina ideer och tankar utan att ha en rädsla för hur den andra parten kommer reagera.

## 3. Metoder

*Under det här kapitlet ger vi en förklaring till metoderna vi använder under arbetsgång, och ger exempel på varför vi har valt dem.*

### 3.1 Agentiell Realism

Agentiell Realism är en teori av Karen Barad (2007, 2003) där hon använder posthumanism som ett centralt begrepp i sin text. Posthumanism menas med att inget / ingen i vår värld kan leva eller vara självständigt, utan allt som händer och görs i dagens samhälle påverkar varandra på ett eller annat sätt. Barad skriver om intra-aktion som en utveckling av ordet interaktion där intra-aktion handlar om sammanhanget mellan människa och icke människa. till skillnad från interaktion som beskriver samverkan mellan människa och människa.

*The notion of intra-action (in contrast to the usual “interaction,” which presumes the prior existence of independent entities/relata) represents a profound conceptual shift. It is through specific agential intra-actions that the boundaries and properties of the “components” of phenomena become determinate and that particular embodied concepts become meaningful.*  
(Barad, 2003, s. 815)

För att få en bättre bild av vad Barad menar finns det ett exempel, när en människa springer är det inte bara ena foten framför den andra, utan luften, tiden, marken etc påverkas under samma tid som människan springer.

För att kunna förstå intra-aktion bättre så introducerar Barad (2007) onto-epistemologi som är en sammanflätning av epistemologi och ontologi. Epistemologi eller kunskapsteori som det även heter förklarar läran om kunskap och hur man kan få den kunskapen. Ontologi är läran om det varande, det som existerar och vad det betyder om något existerar. Barad vill att onto-epistemologi ska visa världen på ett mer icke traditionellt sätt

(människor/icke-människor, kunskapande/varande, objektivt/subjektivt etc), utan vi ska se hur vi uppfattar våra fenomen.

Under intra-aktioner uppstår det en agens eller *handlingskraft*. Agens är ingen attribut utan det är något som används av både mänskliga och icke-mänskliga aktörer under sin intra-aktion process där mänskligt medvetande ska lämnas utanför.

Även om diffraktion är en annan forskningsmetodologi av (Haraway, 1997) använder Barad (2007) det som ett begrepp i sin forskning. Exemplet som Barad använder sig av är när någon kastar en hand med grus ner i en lugn vattenpöl så skulle ett vågmönster uppstå där cirkelarna påverkas av varje gruskorn som når vattnets yta och sedan genomtränger varandra och sammanflätas. Barad knyter ihop detta exempel till hur kunskap och handling skapar nya betydelser för de mänskliga och icke-mänskliga tingen.

Anledning till varför vi har valt att använda agentuell realism som synsättet till det här arbetet är den direkta kopplingen hur människor har påverkat utvecklingen av teknik baserat på vårt naturliga sätt att kommunicera på, men även hur röststyrning påverkar människors sätt att interagera med den.

## 3.2 Kvalitativ metod

Kvalitativ metod är baserat på intervjuer med den grupp människor man har inriktat sin forskning över. Då datainsamlingen inom en kvalitativ intervju ger ett mer personligt resultat där man är ute efter informantens synpunkter gentemot den kvantitativa metoden som försöker få fler svar av en större folkmängd, tex enkäter och formulär. Under intervjuer som är baserat på den kvalitativa metoden ges det möjlighet att vara flexibel med sina frågor under tiden personen eller personerna intervjuas.

Djupintervju är en del av den kvalitativa metoden och det var den intervjuform som vi valde att använda under undersökningen. Djupintervju har som mål att skapa en relativt öppen diskussion under det ämne man valt att fråga om (Tjora, 2012). Med hjälp av den öppna diskussion som kan uppstå under svaren kan det även bildas nya frågor eller svar för intervjuaren som kan leda till ett bättre resultat och nya tankar för sin studie.

Tillvägagångssätt för en djupintervju inleder man intervjun med en allmän konversation om sin studie för att ge respondenten en mer avslappnad stämning. Under intervjun är det viktigt att informatören är avslappnad under hela samtalet för att få personen att känna att den kan svara så öppet som möjligt utan att känna att den förstör intervjun.

I grunden för en djupintervju så använder man sig av tre faser:

1. **Uppvärmningsfrågor** – enkla och konkreta frågor som inte kräver så mycket eftertanke.
2. **Reflektionsfrågor** – huvudfrågorna man vill få svar på.
3. **Avrundningsfrågor** – frågor som leder bort uppmärksamheten från reflektionsfrågorna med syfte att ge informatören tillfälle att ge fler svar eller ställa frågor till intervjuaren.

Valet att använda en kvalitativ metod istället för en kvantitativ är på grund av att vi ville få ett bredare svar genom att först visa upp röststyrning och få in synpunkter efter användning av gestaltning.



### 3.3 Parprogrammering

Eftersom vi endast var två stycken i vår grupp kände vi att det var lämpligt med parprogrammering då det gör att båda får en bra inblick i vad som händer i både kod och arbete.

Parprogrammering består av två personer som sitter och löser samma problem på samma skärm. En person kallas förare och den andre för navigatör, föraren är personen som sitter och styr kodningen. Navigatören sitter vid sidan om och har en översikt om hur koden ser ut och hur koden kommer byggas upp under tidens gång, det är viktigt att navigatören hänger med och ser problem som uppstår så att föraren snabbt kan rätta till problemen. (Ulf Rask, 2013)

Teorin bakom parprogrammering kommer från att hjärnan inte kan jobba med för många tankar samtidigt och därmed behöver en assistent som hjälper till att jobba vidare med delar av projektet på annat håll.

Parprogrammering valdes som metod för att få en mer omfattande konstruktiv kritik genom att ha en person som observerar och kommenterar under arbetet.

## 4. Designprocess

*I det här kapitlet kommer vi förklara processen med det huvudsakliga fokuset på utvecklingen av gestaltningen.*

S. Andersson, S. Hagelins (2013) studie visar behovet av röststyrning hos människor med inlärningssvårigheter som dyslexi, där röststyrningen kan vara ett hjälpmedel. Man kan läsa vidare i Brainline(2010) hur dom förklarar varför människor med funktionsnedsättning har det svårare med sin inlärning och hur dom känner sig låsta när en andra partner står vid sidan om och ska skriva ner deras ideer.

Efter att dessa texter blivit lästa tog vi en slutsats om att en till undersökning om behovet kring röststyrning inte behöver tillämpas då det redan finns mycket forskning kring det, och därför valt att undersöka varför utvecklingen av röststyrning inte blivit större.

Utöver den framgångsrika utvecklingen av *Deep Neural Networks* (DNN)(Woodford, 2013) där potentialen för röststyrning hänger samman med utvecklingen av artificiell intelligens för att på ett liknande sätt till en mänsklig hjärna uppfatta mänskligt tal, är potentialen endast spekulativ. Därför kommer gestaltningen och undersökningen inte lägga fokus till potential för utveckling, ämnet kommer därför utforskas mer under kapitlet analys och diskussion.

Vi har därför gjort valet att undersökningen med gestaltningen ska sätta fokus på intresset och vår hypotes om inlärningsförmågan av ett navigationssystem som är ett naturligt sätt för oss människor att kommunicera på.

### 4.1 Koncept

För att undersöka intresset och vår hypotes om undersökningen var det ursprungliga konceptet att utveckla en applikation för smarta hem, där röststyrning kan visas och jämföras mot andra navigationssystem.

Under ett handledningstillfälle fick vi förslaget att avgränsa applikationen till ett specifikt rum av det smarta hemmet, och utgå ifrån ett användningsområde. Baserat på det här förslaget och syftet om en undersökning av applikationer där ett flertal navigationssystem kan användas valde vi att arbeta utifrån tekniken som används i ett vardagsrum.

Vi valde att använda ett differaktivt synsätt av vilka användningsområden som existerar inom ett modernt vardagsrum. Definitionen av vardagsrum kan vara annorlunda från person till person beroende på fritidsaktiviteter, då vi valt att fokusera på funktionaliteten för röststyrning jämfört mot andra navigationssystem gav metoden oss tekniker för att visa media.

Media som används inom ett vardagsrum har fokuserat huvudsakligen på videor, musik och till viss del digitala spel.

Uppspelning inom vardagsrummet för videor och musik kommer huvudsakligen från två objekt, skärm och högtalare. För att en skärm och högtalare ska vara en applikation till ett smart hem där röststyrning används, måste applikationerna kommunicera med varandra eller med övriga applikationer i hemmet.

Diffractionen av ett vardagsrum gav därför oss ett par ord att utgå ifrån till undersökningen:

- Röststyrd
- Mediauppspelning
- Funktionell

Konceptet av undersökningen blev därefter att se om en röststyrd applikation för mediauppspelning intresserar population och vilket navigationssystem som väljs, i förhoppning om att undersöka vår hypotes om inlärningsförmågan av röststyrning.

## 4.2 Val av gestaltningstekniker

*I texten under kommer vi förklara vilka val vi gjort under gestaltningens gång och varför vi har valt att gjort dem.*

## Mediauppspelning

Eftersom att undersökningen inom röststyrning för digitala spel och dess funktioner inom ett vardagsrum inte visar syftet eller speglar vår frågeställning har valet blivit att fokusera på de andra två formerna av media, videor och musik.

## Svenska

I texten *Röststyrning och text-till-tal i mobila enheter för personer med läs- och skrivsvårigheter* (S. Andersson, S. Hagelin, 2013) tar de upp hur dialektskillnader av testpersoner med deras engelsk baserade röststyrning påverkade resultaten av studien. Därför har vi gjort valet att använda röstigenkänningen på svenska då detta kan påverka undersökningens resultat om testpersonerna upplever att röststyrningen har svårt att förstå dialektskillnader.

Då vi valt att använda oss utav svenska i röstigenkänningen har vi eliminerat användandet utav mer utvecklade mjuk- och hårdvaror inom engelsktalande röststyrning, ex Amazon Echo (Crist, 2018), Google Home (Gebhart, 2018) och Amazon Alexa (Crist, 2018).

## Operativsystem

Användningen av ett operativsystem för att kombinera mediauppspelning för både videor och musik har sina fördelar i huvudsakligen uppbyggnaden av gestaltning, men även för röststyrning då navigationssystemet kan fungera likadant för båda typerna av media.

Vanligtvis idag används ett flertal maskiner inom vardagsrummet för att hantera olika former av mediauppspelning, då ett operativsystem kan klara av att göra båda och samtidigt utvecklas som en IoT applikation för ett smart hem.

## Användarvänligt

För att undersöka vår hypotes om inlärningsförmågan har vi valt att skapa gestaltningen med användarvänlighet i fokus. Valen vi gjort för att illustrera detta är bland annat användning av svenska som huvudspråk för röstigenkänningen, och användargränssnittet till röststyrningen.

Detta är för att lyfta fram potentialen av röststyrning i nuläget genom att försöka minimera problemen med *Automatic Speech Recognition* (ASR). Vilket i vår forskning syftar på igenkänningen när externa ljudkällor tas upp av mikrofonen, kvalité på mikrofonen i sig, dialektskillnader, talhastigheten och det ej naturliga sättet att formulera meningar på i ett kommando baserat röststyrningssystem.

## Mjukvara

Då röststyrning är ett så väldigt komplext ämne efter utvecklingen av *Deep Neural Networks* (DNN)(Woodford, 2013) och just specifikt maskininlärning, blev planen att inte utveckla någon egen mjukvara för röststyrning utan undersöka potentiella program för användning inom vår gestaltning.

För att få IoT applikation att fungera med röststyrning och uppspelning av media, har vi gjort valet att använda oss utav ett program vid namnet Kodi(Kodi, 2018) för gränssnittet. Kodi används huvudsakligen i visning och organisation av lokal media, men utöver det även har ett tankesätt om öppen källkod. Öppen källkod inom kodi betyder att vem som helst kan skapa applikationer och externa kopplingar till programmet, vilket betyder för vår gestaltning kommer kunna använda sig utav ett externt program för röststyrning.

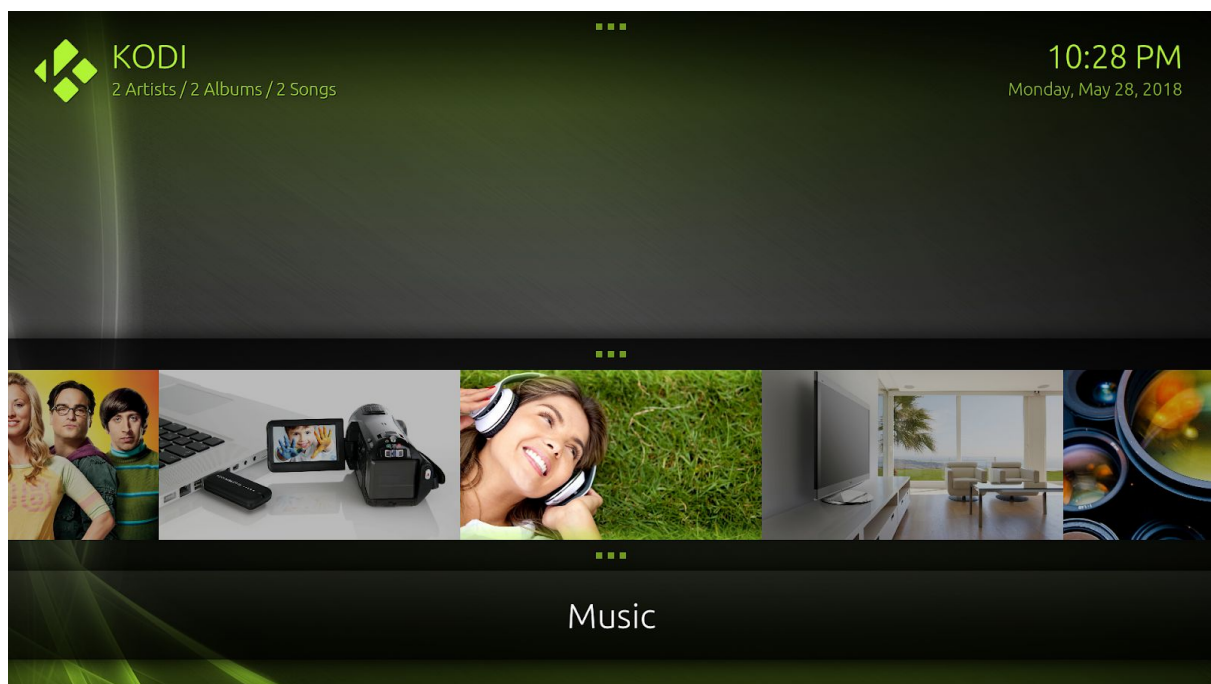


fig. 1. Bilden ovanför visar gränssnittet till Kodi vi använt för gestaltningen.

Programmet för röststyrning har vi valt att använda VoxCommando (VoxCommando, 2018), som är en mellanhand för microsofts ASR system inbyggt i Windows. VoxCommando i sig använder sig utav XML(Extensible Markup Language) för att programmera funktionerna för röststyrning. Kopplingen mellan VoxCommando sköts av att MSR(Microsoft Speech

Recognition) skickar översättningarna av din röst till VoxCommando som innehåller en XML-fil med vad som händer när ett specifikt kommando blir talat. När kommandot har hittats i listan startar VoxCommando en genväg för att utföra det vi vill ska hända, anledning varför programmen fungerar tillsammans är att gränssnittet till Kodi går att styra med externa anrop, som då skapar illusionen av röststyrning för mediauppspelning.

VoxCommando använder sig av en offline-recognition eftersom MSR är inbyggt i operativsystemet, för att tillåta röststyrningen fungera utan en internetuppkoppling. Vi hade först hoppats att använda ett online-recognition system inom röststyrningen då potentialen för maskininlärning och en mer komplex form av ASR hade kunnats användas. Då vi använder oss utav forskningen kring IoT applikationer kommer gestaltningen behöva en internetuppkoppling.

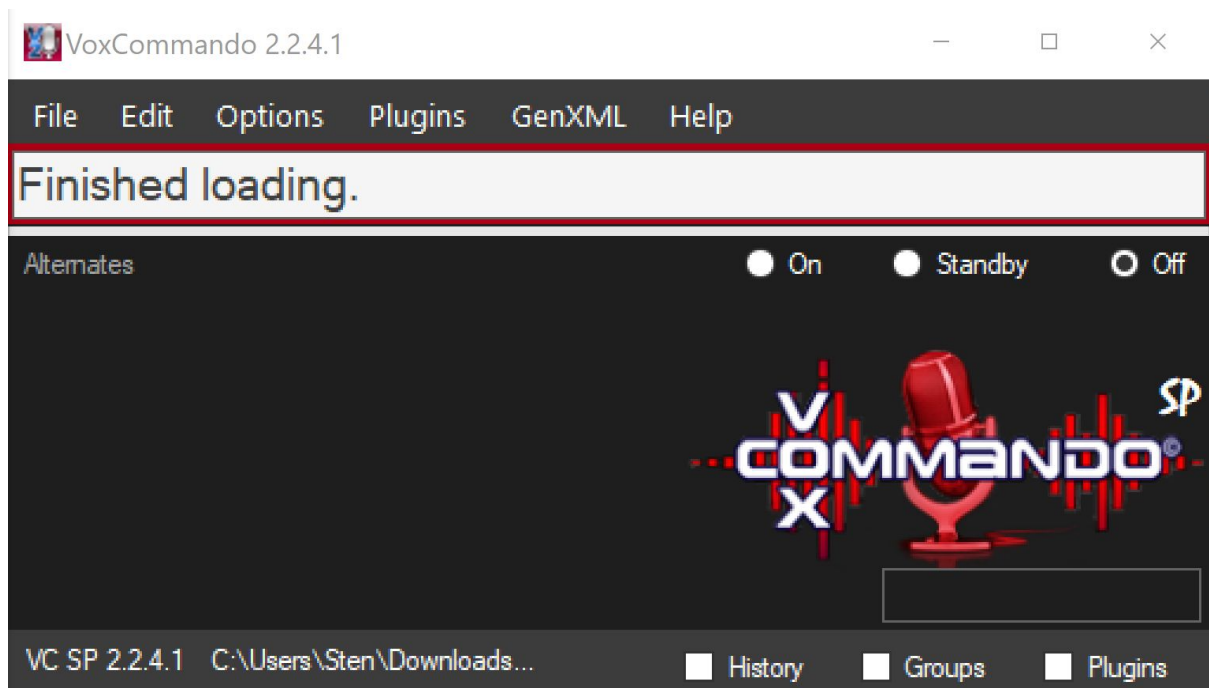


fig. 2. Bilden ovanför visar gränssnittet till VoxCommando vi använt för gestaltningen.

Kodi passar även till den jämförande delen av vår undersökning, då ett flertal navigationssystem går att använda från grunden, fjärr- och handkontroller, mus- och tangentbord, samt touchskärmar. Vilket gör att systemet i sig endast behöver ta hänsyn till röststyrningen i utvecklingen, då endast utseendet på applikation förändrar upplevelsen att använda alternativa navigationssystem.

## Mikrofon

Val av mikrofon är ett av problemen som har uppstått kontinuerligt i vår designprocess, då tekniken i mikrofonen speglar hur lättanvändligt röststyrning upplevs av användaren.

Tekniken för brusreducering och filtrering av andra typer av bakgrundsljud är den viktiga delen av en mikrofon, då vi till en början använde oss utav den inbyggda mikrofonen i en laptop som ej använder sig utav tekniken, vilket skapade ett flertal problem med röstigenkänningen. Därför har även mer forskning inom ämnet gjorts iterativt under gestaltningens gång för att ta hänsyn till detta.

Undersökande av mikrofoner i olika sammanhang, positionen ifrån användaren av röststyrning och hur lätt ASR kan uppfatta meningen som talas är problemen vi huvudsakligen undersökt med användning av mikrofoner.

Vi hade valet att använda en mobiltelefon som mikrofon till gestaltning, mobiltelefonen i sig kan använda sig utav lokala filer för ASR och sedan skicka informationen till applikationen. Problemet känner vi med användning av en telefon som mikrofon är bekvämligheten av användandet, då simpliciteten som gör potentialen kring röststyrning försvinner när en ny styrenhet måste användas.

## Agentiell Realism

Synsättet agentiell realism i gestaltningen har varit om hur röststyrningen ändrar sättet vi kommunicera med den, eftersom vi människor är vana att tala i fullständiga meningar där en dynamisk förståelse om vad som sägs händer kontinuerligt. I dagens iteration av röststyrning



påverkas detta, där ett antal specifika kommandon måste uttalas och formuleras på rätt sätt för att navigationssystemet ska fungera.

Sättet vi skriver dessa kommandona till människor påverkar därför också hur tekniken utvecklas i detta projekt, där vi måste ta hänsyn till hur vi människor använder specifika ord på ett logiskt sätt. Detta påverkar inlärningsförmågan av röststyrning, då en utvecklad artificiell intelligens med maskininlärning hade anpassats till hur specifika människor väljer att tala.

Tanken om ett informationsflöde från tekniken till användaren för att få system fungera mer naturligt, där både maskinen och människan kommunicera med varandra. Funderingarna över att ha ett text-till-tal system för att rikta användarna av röststyrning till rätt kommandon formulerades med hjälp av synsättet agentuell realism.

### 4.3 Utförande

Huvudsakligen har metoden parprogrammering används i den utförande delen av gestaltningen, anledningen till det här metodvalet var för att tillåta navigatören ge kontinuerlig konstruktiv kritik.

Då vi båda i arbetet har dialektskillnader har även parprogrammering resultera i en utveckling där kommandot fungerar till bådars dialekt, vilket i förhoppning kommer fungera till flera dialekter i uppvisandet.

Sättet vi arbetat med metoden är genom att turas om med programmeringen av kommandona och gränssnittet, där en fokuserar på funktionaliteten av arbetet just nu och den andra på helheten. Metoden skapade diskussioner om specifika användningsområden och användarvänlighet under arbetets gång, vilket skapade en iterativ process i utförandet av gestaltningen.

## Djupintervju

För att få ett svar på frågeställningen valdes det att få en djupare förståelse om kvantitativa- och kvalitativa metoder och hur des datainsamling kunde hjälpa oss i undersökningen. Den kvantitativa metoden fördes under diskussion men då den utformar sig till att ge statistiska svar som siffror, kön, ålder etc och inte ge möjlighet till en öppen diskussion antingen under eller efter intervju perioden. På grund utav detta valde vi att använda kvalitativa metoden i hopp om att få ett bredare och öppnare svar till vår frågeställning, vilket även ger oss ett tillfälle att observera hur röststyrningen används av testpersonerna.

Kvalitativa metoden ger oss en djupare förståelse för de val som respondenten gör under intervjuens gång genom att analysera kroppsspråk, åsikter och erfarenheter. Valet av struktur för djupintervjun kommer från den forskning som skett under tidigare metodkapitel där vi fann en indelning av frågor som gav grund till intervjun. Handlingen av dessa frågor ger respondenten en naturlig känsla av vad som skulle ske och vad syftet för undersökningen var.

1. **Uppvärmningsfrågor** - Hur gammal är du? Vad jobbar du med? Vad gör du på fritiden? Är du tekniskt intresserad? Har du använt röststyrning eller använder du röststyrning?
2. **Reflektionsfrågor** - Skulle du vilja använda röststyrningen istället för dina vardagliga navigations sätt? Varför skulle du vilja det eller varför skulle du inte vilja det? Tycker du det är ansträngande att ha fler än ett sätt att navigera runt sig på?
3. **Avrundningsfrågor** - Vad ska du göra nu? Kommer du följa dagens teknik mer?

För att ge personen som intervjuas en känsla för hur röststyrning kommer fungera i sitt egna hem, placerade vi ett bord och en stol där personen skulle få möjlighet att välja mellan röststyrning och de vanliga styrenheterna som används för samma ändamål. Stolen vi satt framför bordet kunde man på ett enkelt sätt få tag i kontrollerna som användes, sedan valde vi att ta bort stolen för att se om personen fick en obehaglig känsla till att använda samma kontroller som användes under tiden testpersonen satt ner. Valet av detta kom från att leda personerna in till att testa röststyrning på ett mer naturligt sätt.

## Valda personer

Vi valde att inte försöka intervjua personer som redan har bra erfarenhet av röststyrningen, då röststyrningsfenomenet är något som man har pratat om under längre tid. Att ha ett brett utbud av flera människor i olika åldrar tyckte vi även kunde ge ett bredare resultat.

## Intervjuprocessen

När respondenten anlände satte vi oss ner för att förklara det syfte vi har med undersökningen och varför det är viktigt att respondenten ger naturliga svar på de frågor som ställs. Svaren på dessa frågor som ställs kommer även behandlas konfidentiellt för att ytterligare ge respondenten en säkrare och tryggare miljö under intervjuens gång.

Respondenten har nu fått en förståelse för upplägget, därefter börjar vi med att ställa uppvärmningsfrågorna som uppkommit under planeringsfasen av intervjun. Nu känner sig respondenten bekvämare i miljön och vi ger den möjligheten att testa både den prototyp som utvecklats fram i tidigare process och de kontroller som används dagligen för samma program med och utan stol. Testfasen är sedan över och vi sätter oss i en bekvämare miljö för att ställa reflektionsfrågorna som ska ge oss svar på frågeställningen.

Reflektionsfrågor på intervjun:

- Är du tekniskt intresserad?
- Vad tyckte du om röststyrningen?
- Skulle du tänka dig att använda röststyrningen istället för dina vardagliga navigations sätt såsom tangentbord, mus, touch-screens etc?
- Varför skulle du vilja det eller varför skulle du inte vilja det?
- Vad anser du om att det kommer fler hjälpmedel att navigera sina vardagliga enheter med?

- Övriga tankar om systemet?

*Ovanför ser vi ett exempel på reflektionsfrågorna som ställdes efter testandet av prototyp.*

Valet av att ha få reflektionsfrågor var på grund utav att inte ta mycket tid från respondenten då vi även hade ett testande av prototyp inlagt under intervjuens tiden. Användningen av avrundningsfrågorna uppstod mot slutet av intervjun där vi gav respondenten en möjlighet att ställa övriga tankar om vårt arbete.

## Sammanfattning

Processen av arbetet var till ett försök att minimera de kända problemen kring röststyrning för att sedan visa människor med mindre erfarenhet av tekniken till försök att få ett svar till arbetets frågeställning.

## 5. Resultat av undersökningen

*I detta kapitel kommer resultatet och sammanfattningen av datainsamling, observationer och svar på reflektionsfrågorna uppvisas.*

Fem personer i olika åldrar blev intervjuade under samma plats där valet att all information förutom ålder och de svar de ger på reflektionsfrågorna ska behandlas konfidentiellt då syftet med studien inte har något med respondentens privata liv.

Under testfasen i intervjun där respondenten fick testa prototypen analyserade vi respondenterna för att se hur varje person reagerar med de navigationssystem som tillhandahölls. Respondenterna kommer förkortas med RD och därefter sin ålder då vi anser att åldern i undersökning är intressant för dess resultat och dessutom är ett anonymt sätt att skilja dem på.

## Test med stol

Först(se bilaga 1) analyserade vi hur respondenterna använde sig av de navigationssystem vi vald i en sittande position. I bilagan ser man att alla var öppna till att prova röststyrning och det fanns ett intresse hos majoriteten att vilja använda tekniken. RD16 och RD24 använde röststyrningen men valde att gå tillbaka till mus och tangentbord. RD48 och RD30 var mer nyfikna och var inte rädda för att säga fel, tested under dessa två respondenter hade en lekfullare jargong och ett större behov att testa hur det fungerade. Det var lättare för RD68 att säga till Kodi att göra något än att hitta själv och det uppskattades men efter det försvann intresset för att fortsätta experimentera.

## Test utan stol

Nästa del av intervjun(se bilaga 2) fick respondenterna testa navigation systemen en gång till utan att sitta ner, vilket gav användarna en lekfullare blick där majoriteten valde att försöka använda röststyrningen från olika positioner i rummet. En gemensam nämnare som uppstod var att mikrofonen till datorn hade svårt att fånga ljud när respondenten var i rörelse. RD30 började utforska olika tonlägen när kommandon inte utfördes och därmed fick för- och nackdelar med röststyrning.

## Reflektionsfrågor

Under reflektionsfrågorna till testpersonerna visade det sig att användandet av röststyrning var mer en intressant teknik att få testa använda. Där endast en av de fem testpersonerna var intresserad och hade gjort en egen undersökning av röstigenkännings tekniken. Intresset i de övriga respondenterna i inläringen speglade inte deras användbarhet av tekniken, där även två kände sig obekväma med användandet av röststyrning publikt.

Fyra av fem av testpersonerna var intresserade i utvecklingen av röststyrning och känner att de potentiellt används mer av tekniken om de funnits färre problem med igenkänningen, men väljer att använda mer populära navigationssystem då de har en större vana.

RD48 var den enda av testpersonerna som hade idén om att använda röststyrning som den enda formen av digital navigation, när tekniken blivit bra nog för att kunna jämföra med mer utvecklade tekniker för kontrollering av funktioner.

## Potential inom utveckling

För att summera observation och reflektionsfrågor från respondenterna kan vi se att resultatet av vår undersökning var att det är ett mindre intresse då de ser funktionaliteten som används idag, och endast kan spekulera om potentialen.

Många kände att tekniken i sig var något obekvämt att använda i sitt nuvarande stadiet och vill att den ska vara mer precis i igenkänningen för att få en mer användarvänlig upplevelse.

Vårt försök till att minimera problemen kring röststyrning med att använda svenska igenkänning, och utveckla ett lättanvänt gränssnitt påverkade inte respondenternas åsikt om bekvämligheten och användarvänligheten inom röststyrning.

Vår tidigare hypotes om att röststyrning har ett mindre krav på inläring speglas inte i resultatet som vi hoppats, även då testpersonerna förstod hur man använder röststyrning var inte en naturlig övergång då kommandon för användning var tvunget att visas. Där övrig teknik, invand eller ej hade en liknande lärande fas.

## Sammanfattning

Där många väljer att använda tekniken som skapar minst repetition i användning och är mer användarvänlig då röststyrning även kräver idag att du kan kommandona som kan användas.

Då vi inte upplevde att någon av testpersonerna hade svårigheter att använda tekniken, utan fick avsmaken av röststyrning genom dess problem, har vi tagit slutsatsen till att röststyrningen har potentialen inom ett smart hem att fungera som huvudsakligt navigationssystem därför anser vi att tekniken inom en framtid är funktionellt, men inte i dagsläget.

## 6. Diskussion

*I det här stycken kommer vi ta upp egna tankar och funderingar om röststyrnings utveckling och resultatet vi fick utav vårt arbete. Även reflektioner över metodval och dess påverkan på resultaten av undersökningen.*

### Metodval

Metodvalet för att undersöka med hjälp av en kvalitativ metod fick oss att få ett djupare perspektiv om intresset och problemen kring röststyrning. Fördelen med metoden var att testpersonerna fick använda gestaltningen först, där vi försökt minimera problemen med röststyrning, för att sedan säga sin åsikt. Hade vi istället valt en kvantitativ metod hade nog ett spritt resultat om varför röststyrning inte är en populär metod för att kontrollera funktioner, men även där varit spekulativt om varför det är så.

Frågeställningen i sig var svår att undersöka med vår gestaltning, där subjektiviteten av röststyrning är den drivande kraften. Att på ett objektiva sätt se om röststyrning kan fungera inom en framtid är endast spekulationer, intresset och behovet finns där delvis, men om

utvecklingen i sig kommer till punkten där den triumferar andra navigationssystem får framtiden visa.

## Gestaltning

Innan gestaltningen hade vi tre idéer för att undersöka vår frågeställning:

- Smart kylskåp: Tanken var ett samarbete med en annan grupp där vi med hjälp av en scanner, streckkoder och en databas försökt låta kylskåpet informera dig om vilka råvaror som finns tillgängliga, och låta användaren med hjälp av röststyrning kontrollera funktionerna.
- Röststyrning för bilister: Telefon lagen februari 2018 som förbjöds användandet av mobiltelefoner medans man kört skapade nästa idé där vi försökte använda röststyrning för att låta bilisterna svara på samtal, byta musik och läsa/skicka meddelanden.
- Röststyrning som säkerhet: Använda rösten som identifiering vid dörröppning, se om det på något sätt går att identifiera om rösten är inspelad eller i realtid i hoppet om att hitta fler användsområden för röststyrning.

Hitta en gestaltning som speglar vår syfte och frågeställning var en iterativ process där idéerna vi hade visade sig vara svåra att undersöka inom ett subjektivt ämne. Hade vi istället valt att avgränsa arbetet till en mer specifik fråga om problem, eller röststyrning inom artificiell intelligens hade vi troligen fått mer konkreta svar på vår frågeställning.

Eftersom vi gjort valet att använda svensk röststyrning, som idag inte använder sig utav maskininlärning för att anpassa röstigenkänning till talaren, har teknikens brister visats sig tydligare. Under den senare delen av 2018 har företaget Google visat nya framsteg i utvecklingen av röststyrning och har även tänkt lansera sitt egna svenska språkpaket, vilket i sin tur kan leda till att både intresset och behovet av röststyrning ökar.

Externa ljudkällor som förvirrar röstigenkänningen är ett ärvt problem av röststyrning där teknikens utveckling stått still. En av de teoretiska lösningarna till det här är användandet av



strupmikrofoner, som då istället för att använda sig utav ljudvågor rösten skapar, använder sig av vibrationerna från strupen. En strupmikrofon kräver att den är fast på strupen för att kunna ta upp vibrationerna, vilket gör att tekniken kan känna sig obekvämt att använda kring ett smart hem. Huvudsakligen är idén i vår åsikt om röststyrning att kontrollera funktionerna med endast rösten och inga enheter förutom en mikrofon i närheten.

## Behov

Människor med läs- och skrivsvårigheter, funktionsnedsatta och former av förlamning har ett behov för röststyrning, då många av navigationssystem som touchskärmar och tangentbord använder sig av händerna.

Behovet av röststyrning existerar för människor med svårigheter att använda alternativa tekniker, men behovet för vidareutveckla som navigationssystem för allmänheten har blivit ställt då problemen med tekniken ger avsmak.

Man kan även dra slutsatsen att när utvecklingen av artificiell intelligens tar fart tar även behovet av röststyrning det, baserat på teknikens lyft i utvecklingen med introduktionen av DNN emot den tidigare metoden av *Language modeling and statistical analysis*.

## Jämförelse

Resultatet av intervjun visade på att många av testpersonerna använde sig utav tekniken de kände sig bekväma med över röststyrning. Anledning till detta kan vi endast spekulera runt om eftersom undersökningen huvudsakligen fokuserar på användningen av röststyrning.

Förutom en vana till andra navigationssystem har i dagsläget röststyrning för många tekniska problem, en utveckling inom röststyrning kan ändra det här. Då att utveckla teknik anpassad

till människan är ett funktionellt sätt än att låta människan anpassa sig till tekniken, då det underlättar den lärande fasen och användarvänligheten.

I vår undersökning har vi funnit att förutom röststyrning används huvudsakligen touchskärmar och till en mindre grad handgester och olika typer av sensorer. Just sensorer och handgester är liknande till röststyrning då det normalt sätts upp med hjälp av interna enheter för navigationssystem. Medans en touchskärm kräver att du använder fingret eller handen emot en skärm med tillhörande gränssnitt.

Framtida maskininlärning har även potentialen att låta den utvecklade fasen av röststyrning endast kräva talande, där du sitter och förklarar till en maskin hur du pratar. Vilket vi anser är ett av sätten som får tekniken att sticka ut mot andra navigationssystem.

## Potential inom tekniken

Artificiell intelligens specifikt är vad vi tror kommer få röststyrning att lysa igenom som styrmetod, när du har väl utvecklade maskiner kapabla att hålla en naturlig människolik konversation med dig. Hem assistenter idag bygger mycket på den här utveckling, men artificiell intelligens är inte på en tillräckligt hög nivå att förstå komplexa och annorlunda formulerade meningar.

Utveckling av deep neural networks inom artificiell intelligens påskyndade utvecklingen av röstigenkänning, och vi tror i framtiden de två teknikerna kommer utvecklas i samma takt, då artificiell intelligens kommer vara den huvudsakliga delen av röststyrning. Vi tror även att ett informationsflöde i formen av en röst som du kontinuerligt har en konversation med är den delen av röststyrning som får människor att bli intresserade.

Röststyrning i sig har många användningsområden där du inte kan använda dig utav händerna, där som i en tidigare gestaltningsidé tanken var att inom köket kunna hämta information om vilka råvaror som existerar i hemmet medans du håller på med annat. Frågan

vi ställer oss är just om utvecklingen av röststyrning kommer vara bra nog för att sköta det här utan en komplex artificiell intelligens.

## Sammanfattning

Svaret vi fick ut utav vår frågeställning är att tekniken kring röststyrning baseras just nu på utvecklingen av artificiell intelligens, samt också att tekniken i sig idag är användbar vid behov när den specifikt är utvecklad för människan. Dvs att människor med intresse och lite kunskap inom ämnet kan skapa ett fungerande system för röststyrning när de ställer och väljer kommandon specifikt för sig själva. Att skapa fungerande röststyrningen för allmänhet är dock svårare när utvecklaren måste ta hänsyn till dialekt skillnaden och olika meningsuppbyggnader, det här har som sagt innan potentialen att förändras med utvecklingen för artificiell intelligens eller mer specifikt maskininlärning.

# Separata bilagor

## Källförteckning

Crist, R. (2018). *Everything you need to know about the Amazon Echo* hämtad 2018-06-12 från <https://www.cnet.com/how-to/amazon-echo-alexa-devices-everything-you-need-to-know>

Andersson, S. Haglin, S. (2013). *Röststyrning och text-till-tal för människor i mobila enheter för personer med läs- och skrivsvårigheter*.

Arkenson, C (2016). *Music Control in the Car – Designing voice interactions between user and a music service with a focus on in-car usage*.

Brainline (2010). *Speech Recognition for Learning*. Hämtad 2018-05-07 från <https://www.brainline.org/article/speech-recognition-learning>

Barad, K. M. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Durham, N.C;London;: Duke University Press

Barad, K. (2003) *Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter*, Journal of Women in Culture and Society, 28(3), University of Chicago

Creutz, J, Borgkvist, I (2015). *Smart Hem, smart för vem?: En kvalitativ studie om varför det Smarta Hemmet inte har fått sitt förväntade genomslag*.

C.Woodford (2018). *Speech Recognition software*. Hämtad 2018-04-19 från <http://www.explainthatstuff.com/voicerecognition.html>

Effektiv parprogrammering i utvecklingsorganisationer (2013) Hämtad 2018-04-21 från <https://www.citerus.se/effektiv-parprogrammering-i-utvecklingsorganisationer/>

Eklind. S (2017). *Samarbetet mellan människa och automation i det avhjälpande underhållet: behovsanalys och krav på framtida utveckling av automationsutrustning.*

Karimi. K, Atkinson. G (2013) *What the Internet of Things (IoT) Needs to Become a Reality.*  
Hämtad: 03012018 från <http://www.mouser.fr/pdfdocs/INTOTHNGSWP.PDF>

Gebhart, A. (2018) *Google Home review: A smart speaker that's its own worst enemy* hämtad  
2018-06-12 från <https://www.cnet.com/products/google-home/review/>

Gulley. K (2014) *Noise Cancelling Headsets: How the microphones work and what you need to know.* Hämtad 2018-04-25 från  
<http://www.theucbuyer.com/blog/noise-cancelling-headsets-how-the-microphones-work-and-what-you-need-to-know>

Kodi. (2018). *Kodi (Version 17.6)* hämtad från <https://kodi.tv/>

Liu, X. F., & Zhou, H. (2014). *A Chinese Small Vocabulary Offline Speech Recognition System Based on Pocketsphinx in Android Platform.*

Mircosoft Research (2017) *Automatic Speech Recognition - An Overview.* Hämtad  
2018-04-25 från <https://www.youtube.com/watch?v=q67z7PTGRi8>

Nilsson. J, Larsson. D (2015). *Dialekter och röstigenkänning: Ett röstigenkännings-API:s förmåga att uppfatta svenska dialekters kännetecken och röstkombinationer.*

Rehman. A, Arif. R, Khursheed. H (2014) *Voice Controlled Home Automation System for the Elderly or Disabled People.*

Song, Qunying, Shen, Hui (2012). *Intelligent Voice Assistant.*

Transportstyrelsen. (2018) *Sluta rattsurfa* hämtad 2018-04-25 från <https://www.transportstyrelsen.se/sv/sluta-rattsurfa/fragor-och-svar-om-mobilanvandning-i-fordon/#130876>.

Ulf, Rask (2013). *Effektiv parprogrammering i utvecklingsorganisationer* hämtad 2018-06-05 från <https://www.citerus.se/effektiv-parprogrammering-i-utvecklingsorganisationer>.

VoxCommando. (2018). *VoxCommando (Version 2.2.4.0)* hämtad från <http://voxcommando.com/home/>

Waher. P (2015). *Learning Internet Of Things*. Packt Publishing, UK.

Wilson, C. Hargreaves, T. & Hauxwell-Baldwin, R. (2014). *Smart homes and their users: a systematic analysis and key challenges*. *Personal and Ubiquitous Computing, Volume 19, Issue 2, Pages 463-476*. Hämtad 2018-03-26 från <https://link-springer-com.lib.costello.pub.hb.se/article/10.1007%2Fs00779-014-0813-0>

## Ordlista

**Amazon, Google, Microsoft** - Större företag med fokus på teknik

**Applikation** - Ett tillämpningsprogram för ett specifikt syfte

**Artificiell intelligens (AI)** - Intelligens som uppvisas av maskiner

**Automatic Speech Recognition (ASR)** - Igenkännings funktioner inom röststyrning

**Brusreducering** - Metod för att reducera störningar i ljudvågor kring en mikrofon

**Deep Neural Networks (DNN)** - Teknik inom artificiell intelligens för maskininlärning baserat på hur den mänskliga hjärnan arbetar.

**Gestaltning** - Ett arbete med syftet att illustrera en tanke, idé eller en teknik

**Hem Assistent** - Artificiell intelligens för hemmet, vilket i det här sammanhanget använder sig utav röststyrning

**Internet of Thing (IoT)** - Samlingsnamnet för enheter med en internetuppkoppling

**Kod** - Instruktioner till maskiner

**Maskininlärning** - En maskin som lär sig själv genom repetition

**Microsoft Speech Recognition(MSR)** - Microsofts eget utvecklade röstigenkänningsprogram inbyggt i Windows

**Mjukvara** - Kod, program och interna funktioner i en dator

**Navigationssystem** - Teknik för att kontrollera digitala funktioner, ex mus och tangentbord

**Operativsystem** - En samling datorprogram för att underlätta användandet

**Programmering** - Att skriva instruktioner till maskiner

**Röststyrning** - Kontrollera digitala funktioner med hjälp av rösten

**Sensor** - Enhet som känner av rörelse, och använder funktioner baserat på dessa

**Smart** - Sammanhanget i texten syftar på applikationer som kommunicerar varandra i ett nätverk

**Strupmikrofon** - Mikrofon som fästs vid strupen och lyssnar till vibrationer i din hals, istället för din röst

**Touchskärm** - En interaktiv skärm där du pekar och trycker på ett förprogrammerat gränssnit

**Windows** - Operativsystem utvecklad av Microsoft

**XML** - Extensible Markup Language, ett märkspråk använt i olika applikationer

**Öppen källkod** - *Open Source* projekt där alla kan läsa koden och lägga till egna funktioner

## Övriga bilagor

### Bilaga 1 - Val av styrenhet för testpersonerna

RD16: Första valet av navigations metod gick till mus och tangentbord för att på ett smidigt sätt starta musiken som valdes. Alla kontroller testades därmed röststyrning, röststyrningen fungera bra med den dialekt som användes men användaren valde att gå tillbaka till mus och tangentbord.

RD48: Valet av kontroll gick till datormusen och ingen användning av tangentbord, snabbt över till testande av de andra kontrollerna där röststyrningen blev ett vinnande koncept för att snabbt sätta på en låt. Dialekten var det inga problem med och allt fungerade utan problem.

RD30: Gav sig på röststyrningen direkt som sedan ledde till att personen fick ett starkt intresse för detta, började jämföra hur snabbt det gick att byta musik gentemot det kontroller som fanns tillgängliga.

RD24: Mus och tangentbord dominerade, även efter test av andra kontroller. Röststyrningen fungera men inget vidare intresse till att fortsätta och gick tillbaka till tangentbord och mus.

RD68: Val av datormus fanns men hade problem med förståelsen om hur Kodi fungerade, där vi bad personen att säga till programmet att starta en låt som då fungerade. Fanns uppskattning men inget intresse till att fortsätta.



## Bilaga 2 - Användning av röststyrningen för testpersonerna

RD16: Valde att styra allt med rösten denna gången, gick runt i rummet för att se om det funka. Personen pratade för tyst vissa gånger där inget hände, detta på grund utav olika ställen i rummet

RD48: Gick på röststyrningen där personen gick förbi bordet och sa något snabbt för att se om det reagera. Orden uppfattades men sades för tyst för att programmet skulle fortsätta kommandot.

RD30: Val av att gå runt i rummet samtidigt som att försöka säga kommandon. Olika tonlägen användes för att testa. Orden uppfattades men va för tyst för att komma vidare eller för höga toner så orden inte uppfattades.

RD24: Mus och tangentbord användes, men större intresse för röststyrning gentemot förra fas då man valde att gå runt lite i rummet och säga kommandon. Orden registrerades men kommandona utfördes inte alltid.

RD68: Personen valde att skippa detta moment.

## Bilaga 3 - Transkribering av reflektionsfrågor under intervju

RD16:

Är du tekniskt intresserad?

- Nej, jag har min mobil och det är allt.

Vad tyckte du om röststyrningen?

- Det var väl häftigt.

Skulle du tänka dig att använda röststyrningen istället för dina vardagliga navigations sätt såsom tangentbord, mus, touch-screens etc?

- Vet inte.

Varför skulle du vilja det eller varför skulle du inte vilja det?

- Skulle jag inte varit så van vid att använda tangentbord, mus och touch-screens så skulle jag säkert använda det.

Vad anser du om att det kommer fler hjälpmedel att navigera sina vardagliga enheter med?

- Vet inte riktigt, inget jag har tänkt på.

Övriga tankar om systemet?

- Ne, inget jag kan tänka på.

RD48:

Är du tekniskt intresserad?

- Nej, inte överhuvud taget. Eller jag det är intressant när det kommer till mig men det är inget jag följer.

Vad tyckte du om röststyrningen?

- Det var en upplevelse, jag har aldrig fått göra något sådant tidigare så det va kul.

Skulle du tänka dig att använda röststyrningen istället för dina vardagliga navigations sätt såsom tangentbord, mus, touch-screens etc?

- Absolut, skulle jag kunna få in det i mitt hus så skulle jag vilja ha det.

Varför skulle du vilja det eller varför skulle du inte vilja det?

- Ne men jag tycker det var en rolig sak man kan visa upp för sina släktingar och kompisar.

Vad anser du om att det kommer fler hjälpmedel att navigera sina vardagliga enheter med?

- Jag är helt för det, tycker det är roligt med nytt även om jag inte är så insatt i teknikens värld.

Övriga tankar om systemet?

- Inte mer än att jag tycker det är intressant och ska allt ha ögonen öppna för detta i framtiden.

RD30:

Är du tekniskt intresserad?

- Jodå, jag följer ändå med lite i vad som händer i dom stora företagen som google och amazon.

Vad tyckte du om röststyrningen?

- Jag har ju läst och kollat lite på videos om alexa och google home och överväger om jag ska ha en eller inte, men detta va intressant att man kunde ha röststyrning utan alexa och google home.

Skulle du tänka dig att använda röststyrningen istället för dina vardagliga navigations sätt såsom tangentbord, mus, touch-screens etc?

- Skulle nog inte byta ut det helt, men att ha det som en sekundär grej skulle jag tycka funkar bra.

Varför skulle du vilja det eller varför skulle du inte vilja det?

- Det känns som det inte riktigt är utvecklat för att bytas och jag vet inte hur det skulle funka i andra ljudmiljöer.

Vad anser du om att det kommer fler hjälpmedel att navigera sina vardagliga enheter med?

- A det tycker jag är skitkul. Att tekniken utvecklas och nya coola saker kommer som gör det lättare för oss.

Övriga tankar om systemet?

- Ne jag har inte så mycket mer att säga.

RD24:

Är du tekniskt intresserad?

- Lite är jag, men mest åt spel

Vad tyckte du om röststyrningen?

- Det är ju inget jag riktigt skulle använda idag, jag har testat det lite med telefonen men känner att det är så konstigt att springa runt och prata med sig själv.

Skulle du tänka dig att använda röststyrningen istället för dina vardagliga navigations sätt såsom tangentbord, mus, touch-screens etc?

- Kanske när alla andra börjar använda det så det inte känns så konstigt att prata.

Varför skulle du vilja det eller varför skulle du inte vilja det?

- (tog svar från föregående fråga)

Vad anser du om att det kommer fler hjälpmedel att navigera sina vardagliga enheter med?

- Jag tycker det är coolt men det är inget som man riktigt behöver känner jag.

Övriga tankar om systemet?

- Nej

RD68:

Är du tekniskt intresserad?

- Inte så insatt i det, det som kommer det kommer men det är inget jag letar efter

Vad tyckte du om röststyrningen?

- En konstig upplevelse

Skulle du tänka dig att använda röststyrningen istället för dina vardagliga navigations sätt såsom tangentbord, mus, touch-screens etc?

- Skulle jag kunna använda min röst till telefonen så skulle jag nog testat det.

Varför skulle du vilja det eller varför skulle du inte vilja det?

- Allt är så nytt för mig så om jag ändå kan använda rösten och prata med min telefon så skulle det vara bra tror jag.

Vad anser du om att det kommer fler hjälpmedel att navigera sina vardagliga enheter med?

- Desto mer utbud som kommer där man får välja fritt så är det jättebra

Övriga tankar om systemet?

- Vet inte, kommer inte på något nu

Fig. 1.

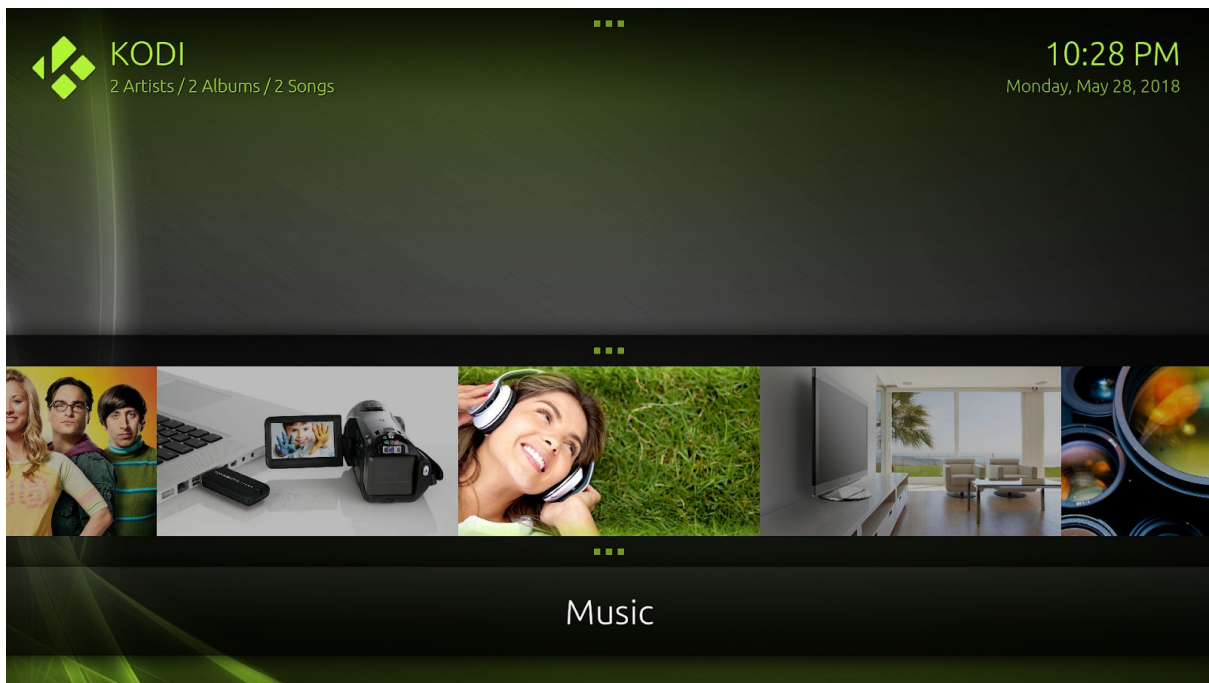


Fig. 2.

