



Kandidatarbete i medieteknik, 30 hp  
Vårtermin 2015

# Identitet i syntetiska röster

Individualitet kontra mänsklighet i framställningen av röstgränssnitt

Erik Franzén  
Rasmus Juhlin

Handledare: Peter Giger & Annika Olofsdotter Bergström  
Examinator: Birgitta Rydhagen  
Blekinge Tekniska Högskola, Institutionen för teknik och estetik

# Abstract

## *Svenska*

Det blir allt vanligare med teknik såsom mobiltelefoner som pratar med människor och vardagen blir allt mer lik en verklighet som tidigare endast existerat i fiktion. Tekniken talar med hjälp av så kallade röstgränssnitt, system som låter datorer tolka och/eller förmedla information genom tal. Uppsatsen ifrågasätter strävan efter att göra dessa gränssnitts röster mänskliga och undersöker hur tekniken kan ges en egen individualitet. Undersökningen grundas i tidigare forskning som behandlar hur människor bemöter röstgränssnitt. Genom en filmanalys med utgångspunkt i identitet och relationer undersöks hur relationen människa-maskin gestaltas i fiktion. I slutsatsen diskuteras hur dessa lärdomar om fiktiva karaktärer går att appliceras på röstgränssnitt i vardagen. Utvecklingen av gränssnitt har tidigare anpassats till de människor som ska vara i kontakt med dem. Undersökningen tyder dock på att ett identitetsfrämjande tankesätt kring utformningen av röstgränssnitt, att ge rösten en egen individualitet med tydlig personlighet och historia, potentiellt kan förbättra relationen mellan människa och gränssnitt.

## **Nyckelord**

Filmanalys, identitet, människa-maskin, relationer, röstgränssnitt

## *English*

It has become increasingly commonplace that technology uses speech to communicate with people and everyday life has started to resemble a reality that previously only existed in fiction. They speak using so-called voice interface systems that allow computers to interpret and/or convey information through speech. This essay questions the pursuit to make these interfaces human and examines how interfaces can be granted a unique individuality. The analysis is based upon earlier research that examines human-interface relations. Through film analysis it is examined how the man-machine relationship is portrayed in fiction and what lessons about non-human characters are applicable to the development of voice interface systems. Earlier, voice interfaces have been adapted to the people who will be in contact with them. However, the analysis indicates that a identity-promoting view on the voice interface development, giving the voice an individuality with a clear personality and history, might improve the relationship between human and interface.

## **Keywords**

Film analysis, identity, man-machine, relationships, voice-interface

# Innehåll

1 Inledning.....	1
2 Bakgrund .....	2
3 Syfte .....	4
3.1 Frågeställning .....	4
4 Tidigare forskning .....	5
4.1 Wired for speech: how voice activates and advances the human-computer relationship	5
4.2 Situated knowledges: the science question in feminism and the privilege of partial perspective.....	7
4.3 Sociolinguistics: language and identities .....	8
5 Metod .....	11
5.1 Metodologi .....	11
5.2 Tillvägagångssätt.....	12
5.2.1 Datorsystem.....	14
5.2.2 Relationen människa-maskin .....	15
5.2.3 Är de mänskliga?.....	15
5.2.4 Kommunikation utan tal.....	15
6. Analys & Resultat .....	17
6.1 Analys.....	17
6.1.1 Datorsystem.....	17
6.1.2 Relationen människa-maskin .....	18
6.1.3 Är de mänskliga?.....	19
6.1.4 Kommunikation utan tal.....	20
6.2 Sammanfattning av resultat.....	21
7 Diskussion och slutsats.....	23
7.1 Diskussion .....	23
7.2 Slutsats .....	26
Källförteckning.....	29
Litteratur.....	29
Filmer .....	29
Bilagor.....	32

# 1 Inledning

Inom fiktionen har det länge funnits robotar och datorer som pratar och agerar självständigt. Trots att de är maskiner talar de ofta med röster som indikerar känslor och tillhörighet och de behandlas av sin omgivning som vänner och jämlikar. Dessa icke-mänskliga karaktärers sätt att förhålla sig till omvärlden förändras dynamiskt allteftersom historien utvecklas och deras datorbaserade, syntetiska röster ger uttryck för denna förändring och styr hur de uppfattas av sin omgivning och publiken. Syntetiska röster är dock inte begränsade till fiktion utan finns idag överallt omkring oss. Från personliga objekt såsom mobiltelefoner och gps-system som pratar med oss till offentliga tågstationer där utrop sker med en framställd syntetisk röst. System brukade av datorer för att förmedla information genom, och även uppfatta och tolka, tal går under samlingsnamnet *röstgränssnitt*. En intim relation mellan maskiner och människor frodas idag eftersom teknik i vardagen blir allt mer självklar. Denna uppsats behandlar de potentiella möjligheterna som existerar för framtidens röstgränssnitt. Utvecklingen av syntetiska röster i röstgränssnitt verkar sträva mot att uppnå perfekt mänskligt tal, med denna undersökning vill vi ta avstånd från detta mål och erbjuda ett alternativt sätt att utforma gränssnittens röster. Genom att studera talande maskiner i film undersöks här hur ett röstgränssnitt kan ges en egen identitet och individualitet. Undersökningen består av två delar varav den första delen utgörs av denna uppsats, vilken behandlar en analys av filmer och applicering på röstgränssnitt efter en metodologi med grund i forskning skriven av Donna Haraway, John E. Joseph samt Clifford Nass och Scott Brave. Den andra delen består av en gestaltning som visar på hur ett potentiellt röstgränssnitt skulle kunna se ut med utgångspunkt i analysens resultat och diskussionens slutsats. Denna del beskrivs i *Bilaga 2*. Arbetsfördelning och arbetsprocess beskrivs utförligt i *Bilaga 3*.

## 2 Bakgrund

Teknik har blivit en alltmer oundgänglig del av våra liv och vi har som samhälle utvecklat en mycket nära relation till våra maskiner. Under tiden som denna relation växt har även tekniken lärt sig att “prata med oss” genom utvecklingen av röstgränssnitt såsom “text-till-tal-syntes” och andra tekniker som översätter en dators information till tal. Att en dator idag kan uttrycka sig i tal ger oss människor ett helt nytt sätt att interagera med maskiner och kanske till och med öppnar upp för en ännu närmare relation med tekniken runt omkring oss. Detta förhållande mellan människa och maskin har länge kunnat ses i science fiction, men nu börjar verkligheten bli allt mer lik det som en gång bara existerade i fantasin. Fiktionen har dock fortfarande ett kraftigt försprång och vi kan se och läsa om robotar och andra icke-mänskliga varelser med tydliga personligheter som interagerar med vanliga människor som individer. Hur relationen mellan människa och maskin sett ut i fiktion har givetvis förändrats över både tid, genre och den fysiska kopplingen mellan de mänskliga och icke-mänskliga karaktärerna.

Röstgränssnitt används både för att ta till sig information från och förmedla information till omvärlden. Personer som själva inte kan tala kan istället uttrycka sig med hjälp av en syntetisk röst och röstgränssnitt gör även att datorer kan tala till oss. En del gränssnitt används för att förmedla allmän information på offentliga platser såsom tågstationer och hissar, medan andra finns tillgängliga i vardagsteknik såsom mobiltelefoner och GPS-system. Talet i många röstgränssnitt idag skapas genom så kallad “concatenative synthesis” vilket innebär att en person läser in ett antal ord och fraser. Dessa klipps upp i mycket små bitar som sedan kan kombineras för att skapa en ny helhet, ett nytt ord (Acapela group, u.å.). Dagens syntetiska röster är alltså vanligtvis en förlängning av en inspelad person vilket innebär att man kan skapa röster med olika karaktärer. När det gäller kommunikation för personer i behov och då främst ersättning av en icke-existerande eller förlorad röst är det förståeligt att användaren vill ha en *personlig* röst att representeras av. Vi vill dock ifrågasätta hur rösters personlighet och karaktär används inom teknik och diskutera detta med utgångspunkt i fiktion.

Vi vill påstå att hur vi som samhälle ser på robotar och datorer som pratar grundar sig i den fiktion, i det här fallet science-fiction, som vi konsumerar och som är en stor del av vår kultur. Därmed utgår vi ifrån att även utvecklingen av moderna röstgränssnitt har varit kraftigt inspirerat av hur maskiner låter och betar sig i film. Vi hör ofta syntetiska röster tala på ett lugnt, harmoniskt och lite släpande sätt som direkt för tankarna till de rena och skinande

utopierna som visas upp i film och tv. Dock vill vi med denna undersökning fråga vilka djupare lärdomar kan dras från dessa filmer. Utvecklingen inom röstgränssnitt verkar fokusera på att göra rösterna så mänskliga som möjligt, troligtvis för att antagandet görs att människor har lättare att kommunicera och föra en dialog med en mänsklig röst. Nass och Brave (2005) menar att det beror på att vi som människor är mycket känsliga för hur en människa låter och minsta fel i uttalet kommer att uppmärksammas. Det behöver därför skapas ett talsystem med ett perfekt mänskligt uttal, vilket idag är avlägset.

Fiktionen, som inte hålls tillbaka av de tekniska begränsningarna med talsyntes, visar ofta upp maskiner med perfekt mänskligt uttal och bidrar därmed till denna dröm om det felfria, mänskliga talet. Vi kan dock i många filmer även se icke-mänskliga karaktärer som inte alls pratar som människor, men ändå har komplexa personligheter och individualiteter som publiken tar till sig och lär sig att älska. En film som tydligt visar detta är Pixars *WALL-E* (Reardon, Morris & Stanton, 2008) där de två huvudkaraktärerna förmedlar ett stort spann av känslor utan att använda mänskligt tal. Varför behöver vi då sträva mot att ge våra apparater mänskliga röster? När vi nu lever i en tid där vår relation till tekniken blivit personlig och intim, varför inte ge tekniken en egen identitet och personlighet att uttrycka genom sin röst? Fiktionen erbjuder en uppsjö av idéer för hur en maskin kan tala och uttrycka sig, från maskiner som har felfritt mänskligt tal till maskiner som bara uttrycker sig genom mekaniska ljud. Med denna undersökning vill vi analysera icke-mänskliga karaktärer i filmer och hur de pratar för att kunna ge en alternativ synvinkel på hur man kan utforma röstgränssnitt för användning i samhället.

### **Avgränsning**

Röstgränssnitt används både för att ge röst åt maskiner och för att representera personer som saknar en egen röst. Vår undersökning är avgränsad till datorer som använder tal för att kommunicera information till människor, inte datorer som representerar en människa.

## 3 Syfte

En undersökning som denna erbjuder ett nytt perspektiv på design av datorbaserade röster vilket kan vara till stor hjälp i utvecklingen av framtida röstgränssnitt. Genom att analysera hur relationen människa-maskin gestaltats i fiktion och jämföra detta med forskning om gränssnitt och människors relation till dessa vill vi undersöka hur röstgränssnitt kan ges en egen, icke-mänsklig, identitet och hur det skulle kunna påverka relationen med användaren. Undersökningen avser lyfta hur ett gränssnitt med en röst kan anpassas för att uppfattas på ett särskilt vis av sin omgivning och hur människans relation och inställning gentemot gränssnittet påverkar hur det uppfattas. Om en närmare relation med användaren kan skapas genom att ge ett gränssnitt en egen identitet och personlighet blir det även intressant att i förlängningen undersöka hur denna relationen kan påverka produktivitet, arbetsmoral eller kundnöjdhet i de olika fält där denna teknik kan användas. Dagens syntetiska röster är en förlängning av en tidigare inspelad person vilket innebär att gränssnitten strävar efter att imitera en människa. Eftersom nära relationer mellan människor och maskiner ofta förekommer i fiktion, där maskinerna är tilldelade egna identiteter och personligheter, är det intressant att undersöka hur gestaltningen av maskiners identitet och personlighet kan appliceras på röstgränssnitt.

### 3.1 Frågeställning

Hur kan röst som verktyg främja att maskiner accepteras som individer?

## 4 Tidigare forskning

### 4.1 Wired for speech: how voice activates and advances the human-computer relationship

*Wired for speech: how voice activates and advances the human-computer relationship* (härefter benämnd som *Wired for speech*) är en bok skriven av Clifford Nass och Scott Brave där de sammanställer ett antal experiment och forskning de utförde kring hur människor interagerar med och reagerar på teknikbaserade röster. Experimenten grundar sig i tidigare forskning inom samhällsvetenskap och socialpsykologi och testar huruvida teorier kring hur människor interagerar med varandra även gäller interaktionen mellan människor och röstgränssnitt. De upptäckte tidigt att tal, oavsett kvalitet, är någonting fundamentalt mänskligt och socialt och att människor därför behandlar teknikbaserade röster på samma sätt som de behandlar mänskliga röster, på gott och ont. Genom olika experiment testades hur kön, personlighet, känslor samt kultur och ursprung i teknikbaserade röster påverkar användaren. Man fann att människor föredrar röster med karaktärsdrag som liknar dem själva och har mer tillit till dem än röster med andra karaktärsdrag. Eftersom vi människor behandlar teknikbaserade röster på samma sätt som vi behandlar mänskliga röster så innebär detta också att vi applicerar fördomar och stereotyper även på röstgränssnitt. Ett av experimenten visade att människor av båda könen visade mer tillit till en manlig röst som talade om teknik och en kvinnlig röst som talade om kärlek och relationer än vice versa. Nass och Brave påstår att ett effektivt system bör utnyttja dessa stereotyper för att ge användaren vad de förväntar sig, vilket skapar en problematik som designern av systemet måste ta ställning till. (Nass & Brave, 2005)

Även om Nass och Brave (2005) hävdar att människor behandlar teknikbaserade röster som vilken social interaktion som helst, lyfter de att man bör vara försiktig med att göra syntetiska röster mänskliga. Om en röst inte låter tillräckligt mänsklig kvalitets- eller uttalsmässigt, men ändå använder ordet "jag" så hävdar rösten mänsklighet vilket gör människor obekväma. En återkommande slutsats i boken menar att en datorbaserad röst som anpassas efter användarens personlighet och sinnesstämning blir mer tilltalande och därmed mer effektiv. Det är däremot viktigt att rösten är konsekvent och inte till exempel ändrar personlighet eller sinnesstämning för fort baserat på vilken uppgift som ska utföras. Detta ses, precis som hos människor, som ett oberäkneligt och därmed opålitligt drag. (Nass & Brave, 2005)



Den viktigaste lärdomen som vi tar med oss från denna bok är att personlighet är det viktigaste draget hos en röst om man som designer vill skapa en djupare kontakt med användaren. *Wired for speech* visar gång på gång hur människor öppnar upp sig och blir mer sociala och personliga när de interagerar med en röst som visar tydliga tecken på en identitet, som till exempel att skämta, kommentera vad som händer eller visa känslor. Problemet är att det är oerhört svårt och i många fall kanske omöjligt att skapa ett gränssnitt som anpassar sig till varje användare, vilket förklarar varför de flesta syntetiska rösterna som hörs i samhället idag är oerhört neutrala och generaliserade för att tilltala ett så brett spektrum av människor som möjligt. Om vi skulle bortse från svårigheterna med att skapa en personlighet som tilltalar alla människor, hur skulle en syntetisk röst med personlighet kunna bidra till gränssnittet som helhet? Nass och Brave erbjuder praktiska exempel på hur en röst kan redigeras för att ge den olika karaktäristiker såsom introvert, dominant, upphetsad, arg och så vidare, men räcker det med att skapa en ytlig illusion av personlighet för att användaren ska köpa att en röst har identitet?

Problematiken med att utnyttja stereotyper för att göra gränssnittet mer effektivt kan kanske överkommas genom att ge rösten en djupare personlighet och individualitet. Som Nass och Brave skriver är stereotyper någonting djupt rotat i mänskligheten och har använts genom alla tider för att på kort tid kategorisera andra människor och avgöra om de är pålitliga eller inte. De flesta människor känner dock även till att om man lär känna en person så försvinner oftast stereotypen och istället visas individen. Kanske är svaret till stereotyper inom röstgränssnitt att låta användaren lära känna rösten som individ. Individualitet hos syntetiska röster har dock enligt *Wired for speech* visat sig vara ett problem då en röst som uppenbart inte är mänsklig hävdar att de är människor genom att använda ordet "jag". Värt att nämna är att experimenten i boken utfördes under tidigt 2000-tal och att de test där endast ljudkvalitet stod i fokus inte längre är tillförlitliga. Vi ifrågasätter specifikt ett test som gjordes när de avgjorde att människor är obekväma med syntetiska röster som säger "jag". Den enda faktorn som bedömdes var röstens ljudkvalitet, de lät en inspelad mänsklig röst och en datorgenererad syntetisk röst använda ordet "jag", men bara den inspelade rösten blev accepterad eftersom den lät mänsklig (Nass & Brave, 2005). Hur skulle resultatet sett ut om den syntetiska rösten givits de personlighetsdrag som undersökts i andra delar av boken? Fiktionen har visat att människor kan bli oerhört fästa vid fiktiva icke-mänskliga karaktärer som visar tydliga personligheter på liknande sätt som människor tenderar att bli fästa vid djur. Borde inte samma sak kunna ske när människor interagerar direkt med ett gränssnitt? I och med att

science-fiction filmer där robotar har en tydlig “jag”-känsla inte är ovanliga, lyfts frågan om “jag” verkligen innebär att en karaktär hävdar mänsklighet eller om den bara hävdar en identitet? Ett gränssnitt kan idag inte uppnå trovärdig mänsklighet så det kanske inte heller är det som ska strävas mot, utan snarare tydliga icke-mänskliga karaktärer som människor ändå kan relatera till på samma sätt som till andra människor? *Siri*, den personliga assistenten i Apples produkter, är ett modernt exempel på ett gränssnitt som hävdar identitet och “jag”, men inte nödvändigtvis mänsklighet. Dessa frågor har vi undersökt närmare i denna uppsats.

## **4.2 Situated knowledges: the science question in feminism and the privilege of partial perspective**

Artikeln *Situated knowledges: the science question in feminism and the privilege of partial perspective* (härefter benämnd som *Situated knowledges*) är skriven av Donna Haraway där hon behandlar problematik som uppkommer kring objektivitet inom forskning. Hon vill visa att perspektivet som används när objektivitet diskuteras är avgörande för hur forskning framställs. Haraway kritiserar sättet som objektivitet framförs i forskning, att den vanligtvis förenklas via så kallade *god-tricks*. Detta innebär att *objektiv* forskning fransäger sig alla relationer till sin omgivning och att resultat förvrängs för att kunna redovisas på ett sätt som gör den greppbar för människor. En läsare leds att tro att de har en uppfattning av hur någonting är, genom ett fristående gudaperspektiv, när de egentligen inte kan ha någon uppfattning av hur det är. Ett typiskt exempel skulle vara en bild av någonting som mänskliga ögon inte kan uppfatta. Bilden *skildrar* hur någonting *ser ut* för läsaren i en slags visuell översättning eftersom ögonen på grund av biologiska förutsättningar inte är kapabla att se det skildrade fenomenet. På detta sätt skänks åskådaren illusionen av att de vet hur fenomenet ser ut eller beter sig. Åskådaren överskrider sina biologiska barriärer och informeras om någonting som egentligen står utanför deras uppfattningsförmåga. (Haraway, 1988)

Användandet av ett god-trick innebär då att åskådare ställs i ett separat plan från forskningen, att forskningen antyds bli skildrad objektivt och utan påverkan av varken åskådaren eller personen bakom forskningen. Haraway menar att användandet av god-tricks skapar en felaktig och *oskyldig* bild av objektivitet. Att forskaren avsäger sig sitt ansvar för och inverkan på resultatet. Hon påstår att objektivitet handlar om sammanhang, att det krävs en aktiv utgångspunkt med en medvetenhet om vilken syn man som forskare har, och hur den synen påverkar det som undersöks. Hon skriver även om hur forskare måste lära sig att koppla bort sitt eget perspektiv från forskningen för att se vilka andra ingångar som finns och därefter

följa undersökningen från dessa ingångar för att ta ställning till hur de påverkar det objektiva. Haraway menar att ett objektivt synsätt måste inkludera den *situerade kunskap* som uppkommer i relationen mellan ett specifikt sammanhang och en avgränsad omgivning. Det viktiga är alltså att objekt och subjekt inte bör separeras och placeras i olika fack, oberoende av varandra, eftersom de oundvikligen har en relation som inte är oberoende. (Haraway, 1988)

Det Haraway skriver är givetvis hennes personliga åsikter, precis som all fakta är grundad i subjektivitet och en särskild syn på världen. Vi ser dock många fördelar med sättet hon bemöter forskning och har valt att anamma dessa teorier i vår undersökning av flera anledningar. Att använda situerad kunskap ger oss möjlighet att avgränsa undersökningen till en specifik karaktär eller ett scenario och inom avgränsningens ramar göra en mer ingående analys. Denna analys blir än mer komplex om vi tar hänsyn till kontexter och varför händelser spelas ut som de gör, alltså bakgrundshistorier, miljöer, relationer, etc. Genom att inkludera så många faktorer som möjligt inom den ram som fastställts anser vi att en mer koncentrerad, tydlig och sann bild av karaktärer och relationer kan uppnås. Situerad kunskap är även viktigt att ta hänsyn till när vi ser på de röstgränssnitt som används i dagens samhälle. Vilken uppgift de har, hur människor interagerar med dem och vad deras bakgrund och relation till gränssnittet är blir relevanta frågor när röstgränssnitt ska utformas. Det är även viktigt att hela tiden vara medveten om hur vi som genomför undersökningen påverkar resultatet genom de val av metoder vi gör och de tolkningar och åsikter vi för med oss genom analysarbetet. Precis som Haraway skriver är "äkta" objektivitet mycket svår att uppnå, om det över huvud taget existerar, vilket medför att både författare och läsare av denna uppsats bör ha i åtanke att resultatet inte heller här är en absolut sanning, utan vår åsikt.

### **4.3 Sociolinguistics: language and identities**

*Sociolinguistics: language and identities* är en antologi om språk och språkets påverkan på mänsklig identitet. Härefter benämns boken som *Sociolinguistics*. Vi har använt oss av kapitlet *Identity* skrivet av *John E. Joseph* och kommer härefter hänvisa till detta kapitel.

Joseph diskuterar språkets inverkan på människors relationer och framför allt hur språket ger människor möjlighet att definiera sig själva som individer. Han går vidare med att förklara hur minsta skillnad i uttal och tonfall förändrar hur talet uppfattas av motparten. Genom livet skapar alla vi människor ett eget personligt bibliotek av erfarenheter och lärdomar som

påverkar hur vi tolkar det vi hör. Joseph menar att när en person talar till oss fastställer vi extremt snabbt en bild över hur trovärdig personen är, vilka aspirationer och syften personen har med konversationen samt personens bakgrund, uppväxt och i vissa fall även utbildningsnivå. Utöver detta kan minsta förändring i uttalet eller formuleringen av ord och meningar skifta innebörden från allvar till förlöjligande. Joseph visar även studier där personer endast kommunicerat via tal utan att ha någon form av visuell anknytning. Detta påvisade att talet även förmedlar kön, ålder och ras, saker som man tidigare antagit främst uppfattas av vår syn. Dessa tre fastställs så pass snabbt att vi endast behöver höra ett fåtal stavelser innan vi beslutat oss för vem vi pratar med. Joseph lyfter även att forskning pekar på att en av huvudanledningarna till varför vi identifierar andra människor och placerar dem i fack så snabbt är för att kunna skilja potentiella vänner från fiender. Vår förmåga att läsa av varandra och identifiera våra närstående är det som binder samman relationer och förhållanden, småskaligt (familj och vänner) och storskaligt (nationer). Likaså är denna medfödda färdighet ofta orsaken till att relationer splittras, ifrågasätts och stigmatiseras. (Joseph, 2009)

Synen på sin egen identitet skiljer sig från hur vi ser på andras identiteter hävdar Joseph. Vi placerar människor i fack, definierar dem utifrån våra egna förutfattade meningar och skapar därigenom någonting som vi tror är sant. Därmed skiljer vi oss från personer som inte stämmer överens med vår definition av en potentiell vän. Joseph hävdar även att konceptet identitet inte är någonting som en person är, utan snarare någonting som man definierar sig själv som. Att ingen människa egentligen har en identitet, utan iscensätter identiteter för att anpassa sig till den sociala situation man befinner sig i. Vi byter konstant mellan olika identiteter för att kunna tillhöra omgivningen. Utan sammanhang och social kontakt, finns det ingen som kan definiera vilken identitet vi har, varken vi själva eller någon annan. Detta innebär att utan en omgivning och kontext har vi ingen identitet, och utan ett förflutet, lärdomar och åsikter kan vi inte heller skapa en bild av de vi möter. (Joseph, 2009)

Vi finner det intressant att läsa Josephs text med människa-maskin-förhållandet i åtanke. Hur skulle de här teorierna översättas till denna nya kontext? Kommer vi som människor fortfarande instinktivt placera talaren i ett fack för att avgöra hur vi ska bemöta talaren, även om talaren är vår mobiltelefon? Hur fungerar egentligen denna identitetsanalys när vi behandlar en icke-mänsklig aktör såsom ett röstgränssnitt? I många fall används en förinspelad person för att få ett särskilt uttal på rösten, men Joseph visar att inte bara en rent

ljudlig identifikation i rösten definierar hur vi känner gentemot individer i vår omgivning. Bakgrund, sociala sammanhang och delade erfarenheter är nyckelbegrepp för att vi som människor ska känna oss bekväma med andra individer i vår omgivning. Det väcker även tankar hos oss om hur skapade karaktärer i filmer, såsom robotar och datorer, ser på sin omgivning om de inte tilldelats en tydlig bakgrund eller haft tid att bygga upp en egen historia med sin omgivning. Hur behandlar människor individer, i det här fallet maskiner, som de inte delar någon historia med? Hur påverkas och förändras olika karaktärers identiteter i fiktion och hur kan vi diskutera detta ställt mot utvecklingen av röstgränssnitt? Behöver alla gränssnitt ha en identitet eller är det bara de som aktivt kommer befinna sig i närvaron av människor eller andra gränssnitt som gynnas av det?

## 5 Metod

För att undersöka hur vi kan använda lärdomar och berättartekniker från fiktion i framtida röstgränssnitt valde vi att utforma undersökningen med en form av filmanalys. Anledningen till att vi valde att studera just filmmediet är att karaktärer i film skapas för att väcka känslor och engagemang hos publiken, vilket är precis vad vår undersökning syftar till att göra med röstgränssnitt. Ytterligare en anledning är att film är ett inflytelserikt och utbrett medium i vårt samhälle och har därför stor inverkan på hur vi tolkar och reagerar på de röstgränssnitt som utvecklas. Istället för att använda redan existerande filmanalysmetoder har vi utvecklat en egen lins genom vilken vi kan studera just de sidor av filmen som vi är intresserade av. Detta genom att kombinera teorierna från *Wired for speech* (Nass & Brave, 2005), *Identity* (Joseph, 2009) och *Situated knowledges* (Haraway, 1988).

### 5.1 Metodologi

En stor del av vår världsuppfattning och identitet utgörs av våra relationer med omvärlden. Våra kopplingar och likheter i bakgrund, kultur och subjektiva åsikter lägger en grund för hur vi ser på andra människor och världen runt omkring oss (Joseph, 2009). Det är därför även viktigt att vara medveten om de kopplingar (Haraway, 1988) som existerar mellan karaktärer i film och utifrån denna kunskap analysera vad det är som gör att relationer (Joseph, 2009) mellan dem ser ut som de gör. Vår undersökning kretsade kring dessa relationer. Hur vi kunde applicera gestaltningen av relationer mellan icke-mänskliga och mänskliga karaktärer i film på de röstgränssnitt vi såg i vår samtid. Vi analyserade utvalda karaktärer i ett antal filmer för att se vad vi kunde få ut av deras relation till- och interaktion med omvärlden. Analysen var centrerad kring den utvalda karaktären och innefattade karaktärens personlighet, världsbild, syn på sig själv och relation till sin omvärld. Genom att ta in dessa faktorer i undersökningen och inte endast analysera hur karaktären betar sig, kunde vi se hur relationen till- och identiteten hos karaktären påverkas och förändras av sin omgivning. Syftet med att analysera på detta sätt är att få ut en komplex bild av karaktärens identitet, och inte minst hur omvärlden ser och behandlar karaktären.

Vårt synsätt när vi genomförde analysen grundade sig i forskningen som Nass och Brave presenterade i *Wired for speech* (2005) angående hur vi som människor bemöter röstgränssnitt och hur vårt beteende förändras beroende på hur rösten vi talar med låter. Att veta vilken effekt talkaraktäristik (tonläge, ordval och ljudkvalitet) har på användaren av röstgränssnitt

gav oss en bra grund att använda i vår analys av hur karaktärerna i filmerna låter, betar sig mot- och uppfattas av sin omgivning. Nass och Brave (2005) använder i sin forskning flera teorier om hur människor interagerar med varandra, men för att få en fördjupad bild av relationer har vi även använt Josephs kapitel *Identity* i *Sociolinguistics* (2009) om mänsklig identitet. Genom att få en insikt i hur vi som människor bemöter och skapar oss uppfattningar om personer runt omkring oss, och även hur vår bakgrund spelar in i vår identitet, kunde vi på ett bättre sätt analysera relationerna mellan olika karaktärer samt karaktärernas syn på varandra och sig själva. Analysen avgränsades genom Haraways *Situated knowledges* (1988) till en specifik karaktär i filmerna. Genom att fokusera på denna karaktärs uppträdande, utveckling och relation till sin omgivning och övriga karaktärer, den situerade kunskapen, kunde vi skapa en djupgående analys av karaktärens individualitet och varför den uppfattas som den gör. Detta är nära relaterat till vad Joseph (2009) skriver om hur vår identitet påverkas av sammanhanget vi befinner oss i och våra personliga erfarenheter till vår omgivning. Med Haraways (1988) sätt att bemöta forskning kunde vi extrahera både hur Josephs (2009), Nass och Braves (2005) teorier stämde in på karaktärerna i filmerna.

## 5.2 Tillvägagångssätt

Genom att tillämpa vår metodologi på utvalda filmer och tv-serier har vi undersökt olika aspekter av hur relationen människa-maskin och datorbaserade röster porträtteras i film. Urvalet av filmer baserades på två huvudsakliga faktorer. Den första och mest grundläggande faktorn var att karaktären vars relation och situation som skulle analyseras behövde kunna kommunicera med omvärlden genom någon form av röst och därmed vara jämförbar med talande röstgränssnitt. Den andra faktorn var filmens inflytande över allmänheten. Vi har valt filmer som är relativt till mycket kända och därmed har haft någon form av genomslagskraft och kan antas ha bidragit till hur vi i fortsättningen sett på robotar, artificiell intelligens eller bara maskiner med röster. Vi valde filmer från olika perioder mellan 1968 och 2014 och såg hur samma tema behandlades under olika tidsepoker och därmed hur synen och gestaltningen av temat förändrats över tid. De tidiga filmerna som behandlade dessa teman ses idag som klassiker och kan antas inspirerat många efterföljande filmer i genren.

När analysen genomfördes användes så kallat semantiskt lyssnande, "semantic listening", som är ett sätt att lyssna på ljud definierat av Michel Chion i *Audio-vision: sound on screen* (1994). Denna typ av lyssnade används för att analysera om ljud, i analysens fall röster och tal, har en särskild innebörd och vilken innebörd detta i så fall är. I vardagliga situationer är

detta vanligast då människor hör tal och omedvetet tolkar ljuden till innebörd och mening. I artikeln *Walter Murch* (Transom.org, 2005) skriver Walter Murch om hur ljud har olika färg. Han utgår från regnbågen där den violetta änden är ren information helt utan känsla, där ingår språk och Murch använder Stephen Hawkings syntetiska röst som ett exempel då den talar och samtidigt är helt monoton. I den röda änden har han placerat musik som endast förmedlar känslor och ingen information. Alla ljud i den violetta änden är ljud som människor måste avkoda och spendera en tid för att tolka och förstå. Ett rött ljud behöver inte avkodas och uppfattas därmed snabbare och mer direkt. I analysens fall brukades semantiskt lyssnande i kombination med ljudens färg för att tolka karaktärens talkaraktär (tonfall, ordval etc.) och dess betydelse för karaktären.

Vår metodologi tog i analysen form av ett formulär med ett antal punkter som kretsade kring en utvald icke-mänsklig karaktär i filmen. I utformningen av dessa frågor tog vi hjälp av karaktärsutveckling och karaktärsanalys inom manusskrivning (Actioncutprint.com, u.å.). Genom att begränsa analysen till en karaktär efter dessa frågor kunde vi få en tydlig och omfattande bild av de aspekter kring karaktären och dennes relationer vi ansåg intressanta. Formuläret bestod av följande punkter.

- *Film* (årtal)
- *Karaktär*
- *Setting*. När utspelas filmen och hur ser teknologin ut i världen som gestaltas
- *Story*. Vad är huvudområdet som behandlas i filmen och vilken roll har de inblandade karaktärerna i berättelsen (i början respektive slutet av filmen)
- *Karaktärens roll i filmen*. Avgränsat av *Situated knowledges* (Haraway, 1988), nära och relevanta relationer i filmen
- *Karaktärens personlighet och gestaltning*. Utifrån Nass och Brave (2005) för talets hastighet, tonläge, kontinuitet och Joseph (2009) för bakgrund, kultur, dialekt
- *Karaktären interagerar med omvärlden såhär*. Fysiska förutsättningar för karaktären som påverkar bakgrund och kompetens inom interaktion (Joseph, 2009)
- *Karaktärens syn på sig själv*. Hur ser karaktären på sin egen existens och förändras det under filmens lopp?
- *Hur ser karaktärens bakgrund ut?* Har karaktären någon bakgrund med omgivningen? Har karaktären någon förprogrammerad uppfattning om omgivningen?



- *Omvärldens syn på, och behandling av karaktären.* Hur interagerar omvärldens historia och bakgrund med karaktären och påverkar de varandra på något sätt?
- *Åskådarens uppfattning av karaktären.* Översiktligt, hur känner man för karaktären som åskådare? Är karaktären till exempel ond eller god, humoristisk eller sorglig och förändras det någon gång under filmen?
- *Hur uppnås detta? Hur gestaltas karaktären, vilka medel har man använt för att styra publikens uppfattning av karaktären?*
- *Vad kan vi få ut av karaktärens relation, personlighet och röst?* Hur kan det appliceras på gränssnitt?
- Övriga tankar

De utvalda filmerna valde vi att dela in i olika kategorier efter vilken typ av karaktär eller tema vi analyserar. Nedan ges en presentation av filmerna och TV-serierna tillsammans med en kort beskrivning av de olika kategorierna.

### 5.2.1 Datorsystem

I denna kategori har vi samlat karaktärer som är datorer med artificiell intelligens, AI, som skapats för att styra ett större system eller databaser och interagerar med människor som ett klassiskt röstgränssnitt, alltså via någon form av skärm eller högtalare som förmedlar text eller tal. Genomgående i filmerna är att karaktärerna har en begränsad eller ingen fysisk kropp att bruka i interaktion med andra karaktärer. Filmerna och karaktärerna som ingår i denna kategori är:

- *2001: A space odyssey* (Kubrick, Clarke & Kubrick, 1968)
- *Colossus: the Forbin project* (Bridges, Chase & Sargent, 1970)
- *Knight rider* (Green, Larson & Sayers, 1982)
- *Iron man* (Arad, Fergus & Favreau, 2008)
- *Moon* (Parker, Fenegan & Jones, 2009)
- *Interstellar* (Nolan, Nolan & Nolan, 2014)

### 5.2.2 Relationen människa-maskin

Dessa filmer behandlar på ett mer personligt och känslomässigt plan hur olika relationer mellan maskiner och människor kan se ut. De behandlar ofta kärlek, vänskap och när en maskin blir någonting mer än bara en maskin. Filmerna i denna kategori är:

- *A.I. Artificial Intelligence* (Aldiss, Curtis & Spielberg, 2001)
- *I, Robot* (Vintar, Davis & Royas, 2004)
- *Robot & Frank* (Ford, Acord & Schreier, 2012)
- *Her* (Ellison, Landay & Jonze 2013)

### 5.2.3 Är de mänskliga?

Den här kategorin innefattar filmer och tv-serier där man behandlar situationer där så kallade androider inte själva vet att de inte är biologiska människor och därmed ifrågasätts vad mänsklighet är. Dessa karaktärer är extremt lika biologiska människor till utseendet och sättet de behandlar sin omgivning. Denna kategori är nära relaterad till den föregående kategorin och alla filmer här skulle även kunna placeras in där, men konflikten med karaktärernas självidentitet gör att de förtjänar en egen kategori. Här ingår:

- *Battlestar Galactica* (Frاند, Larson & Rymer, 2004-2009)
- *Blade Runner* (Fancher, Peoples & Scott, 1982)
- *Äkta människor* (Lundström, Widman & Hamrell, 2012-)

### 5.2.4 Kommunikation utan tal

Den sista kategorin är liten men mycket intressant. Här analyserar vi robotar som kommunicerar med varandra och människor utan att tala. Vi undersöker hur man i film har använt ljud för att kommunicera specifika känslor och uttryck. Fysiskt kan karaktärerna både ha humanoida eller icke-humanoida kroppar.

- *Star Wars Episode IV: A New Hope* (Lucas, Kurtz & Lucas, 1977)
- *WALL-E* (Reardon, Morris & Stanton, 2008)

I och med att *Star Wars Episode IV: A New Hope* har en stor variation av robotar såg vi på den med de övriga kategorierna i åtanke, men vi har kategoriserat den här eftersom karaktären R2-D2 i det här fallet var vårt huvudfokus.

## 6. Analys & Resultat

### 6.1 Analys

Nedan presenteras resultatet av analysen inom de olika ovan nämnda kategorierna. Resultatet inom varje kategori är en sammanställning av de nämnda filmerna och behandlar de punkter som tidigare nämnts i tillvägagångssättets formulär.

#### 6.1.1 Datorsystem

Talet och rösten tar här stor plats eftersom den i många fall är den enda möjliga kommunikationen mellan karaktärerna och deras omgivning. Språkligt talar datorsystemen i de olika filmerna med samma dialekt som människor i sin omgivning. Undantaget är Jarvis i *Iron man* (Favreau et al., 2008) som talar med brittisk engelska till skillnad från sin amerikanska omgivning. Analysen tyder på att när sinnesstämningen hos ett datorsystem förändras, från gott till ont, visas detta primärt i rösten och datorsystemets sätt att prata med sin omgivning. Detta överensstämmer med Nass och Braves (2005) tankar om kontinuitet för att främja pålitlighet. En dynamisk röst med mycket melodi och högt tonfall, som visar intresse för sin omgivning genom frågor och kommentarer, uppfattas som positiv och utåtriktad. En monoton och korthuggen röst, vanligtvis med lägre tonfall, blir kallare och indikerar att datorsystemet har en egen agenda och inte längre samarbetar med sin omgivning.

Dominans är en mycket viktig faktor i denna kategori. Eftersom datorsystemen har så mycket makt och kontroll över sin omgivning blir de omedelbart ett gigantiskt hot om de väljer att inte understå sig människorna. Vi ser även att ju mer datorerna lyder och hjälper de mänskliga protagonisterna, desto mer tilltalande blir de för åskådaren. Samtliga system i denna kategori är medvetna om att de är datorer, även om KITT i *Knight Rider* (Sayers et al., 1982) skämtar om att han är mänsklig. Detta innebär två olika alternativ för den icke-mänskliga karaktären beroende på just vilken som är den dominanta parten i relationen, människan eller maskinen. Colossus i *Colossus: the Forbin project* (Sargent et al., 1970) och HAL 9000 i *2001: A space odyssey* (Kubrick et al., 1968) anser att de står över människorna eftersom de är mer avancerade och blir därför antagonister. De övriga datorsystemen däremot underställer sig människorna och accepterar att de är "bara" en maskin som kan offras om så behövs.

De flesta datorsystemen som är utformade för att vara omtäckta är mycket vänliga, omtänksamma och måna om att vara människorna runt omkring dem till lags, men för att bli ännu mer omtäckta har några av systemen dessutom lite attityd i sitt förhållningssätt och skämtar med människorna som vänner. Denna vänskapsrelation mellan datorsystemet och människorna kan främjas ytterligare av en delad historia med minnen och erfarenheter som möjliggör mer avslappnad, naturlig och intim dialog. Även Nass och Brave (2005) visade att människor blir mer avslappnade och personliga om en dator är skämtsam och Joseph (2009) diskuterade hur människor får en djupare relation när de delar historia och erfarenheter. Datorsystemen i filmerna utnyttjar och imiterar ofta mänskliga känslor för att människorna lättare ska kunna umgås och bygga en relation med dem. I samtliga filmer behandlades vänskapsrelationen på något sätt. I de fall där relation och tidigare historia redan existerade mellan datorsystem och människa är bandet tydligt vänskapligt. I andra fall (*Colossus: the Forbin project* (Sargent et al., 1970), *Interstellar* (Nolan et al., 2014) och *Knight rider* (Sayers et al., 1982)) följer filmens händelseförlopp karaktärernas tidiga relation med datorsystemet, från första mötet till en mer utvecklad relation. De flesta filmerna skildrar en vänskaplig relation mellan datorsystem och människor där systemet är utformat för att kunna umgås med människor, undantagsfallet är *Colossus: the Forbin project* (Sargent et al., 1970) där datorsystemet från första stund är åsyftat att vara kallt logiskt och utan förståelse för mänskliga behov och impulser.

### **6.1.2 Relationen människa-maskin**

Utvecklingen av relationer är central i dessa filmer. De skildrar alla en värld där människor har en nära relation till teknik, vare sig det är stationärt datorsystem som pratar i form av röstgränssnitt eller aktiva talande robotar som förflyttar sig samhället. Relationerna som gestaltas i filmerna utforskar just förhållandet och interaktionen mellan två parter, människa och maskin. De visar hur erfarenheter, upplevelser och sällskap leder till ett starkt band mellan karaktärerna i filmen (Joseph, 2009). Samtidigt ifrågasätts dessa relationers hållbarhet och allvar då faktumet påpekas att den ena parten bara är ett system eller en maskin och därför inte bör ha känslor, tankar och subjektiva åsikter på samma sätt som människor. Allteftersom historien utvecklas återkopplar karaktärer i dialog och aktion till tidigare händelser vilket indikerar att en delad historia och erfarenheter har stor del i hur relationerna uppfattas. I filmerna *Robot & Frank* (Schreier et al., 2012) och *Her* (Jonze et al., 2013) blir de skildrade relationerna mycket intima efter att de mänskliga parterna bekantat sig med sina syntetiska

motparter. En vänskaplig, och i *Her* (Jonze et al., 2013) till och med romantisk, relation byggs upp under filmens lopp vilket tar form i hur parterna behandlar och talar till varandra. Människornas kval gentemot sina icke-mänskliga motparter tar form på olika sätt efter förutsättningarna i filmen, men den problematiserade utgångspunkten är densamma; att människor inte kan ha en intim relation med en maskin. I filmernas värld motbevisas denna problematik efter att de inblandade utvecklar en nära relation med sina maskiner.

Även i denna kategori var de icke-mänskliga karaktärernas sätt att prata avgörande. Bland karaktärerna fanns både kvinnliga och manliga röster representerade men de behandlades och följde liknande spår i hur talet framställdes, i enlighet med Nass och Brave (2005). Rösterna var dynamiska och reagerade med intonationer och känslor beroende på situation. Detta var särskilt påtagligt i filmen *Her* (Jonze et al., 2013) där datorsystemets relation utvecklades till att bli protagonistens flickvän. I övrigt höll alla röster ett högre tonfall, vilket indikerar att de är positiva, utåtriktade och i många fall goda. Vi kunde dock se att karaktärernas röster hade olika nivåer mänsklighet. Roboten i *Robot & Frank* (Schreier et al., 2012) talar på samma sätt som datorsystemen i föregående kategori, alltså som en dator som är utformad för att umgås med människor, men fortfarande en dator. De icke-mänskliga karaktärernas röster i *I, Robot* (Royas et al., 2004) och *Her* (Jonze et al., 2013) skulle dock inte kunna urskiljas från människors röster om man bara lyssnade till dem. *Her* (Jonze et al., 2013) är här extremfallet där Samantha, ett röstgränssnitt liknande Apples *Siri*, är oerhört känslös och använder mycket andning när hon pratar vilket hon motiverar med "det är så man gör".

Vi noterade även en viktig detalj som främjade karaktärernas relation i filmerna och skiljer filmerna från verkligheten. Hörbarhet var aldrig ett problem (Nass & Brave, 2005) och det uppkom aldrig några missuppfattningar när människor pratade med maskiner. Detta innebar att trots att de mänskliga karaktärerna ibland talade otydligt så uppfattade de icke-mänskliga syntetiska karaktärerna talet felfritt. Ur de fiktiva världarnas synpunkt kan detta kopplas till att tekniken är så pass avancerad att röstuppfattning inte är något problem och mikrofoner och programvara är så pass känslig att den kan urskilja röster exceptionellt.

### 6.1.3 Är de mänskliga?

Vem som framställs som en människa i kategorin är inte lika viktigt som *vem som anses av sin omgivning vara en individ*. I filmerna existerar stora konflikter som tar form i för- och

emotsidor. Dessa har två olika synsätt (Haraway, 1988) på icke-mänskliga karaktärer. Antingen är de för evigt maskiner grundade i kall logik och syntetiska material, oavsett hur de utvecklas under tidens lopp, eller så uppfattas de som individer med medvetenhet, känslor och tankar. Detta vanligtvis efter en händelse där deras identitet och medvetenhet ställts på prov inför omgivningen.

I majoriteten av filmerna har de icke-mänskliga karaktärerna minnen av en barndom och uppväxt, fiktiva och påhittade minnen eller implanterade minnen extraherade från en människa. I filmen *Blade Runner* (Scott et al., 1982) och tv-serien *Battlestar Galactica* (Rymer et al., 2004-2009) är de icke-mänskliga karaktärerna i flera fall omedvetna om att de inte är människor och har en syntetisk kropp. Omedvetenheten stabiliseras av minnen som skänker karaktärerna tryggheten att de har en historia och en säkerhet om vem de är (Joseph, 2009). Detta indikerar att bakgrunden och karaktärens historia i filmskaparnas, och därigenom fiktionens ögon, ansetts vara en förutsättning för att de icke-mänskliga karaktärerna ska kunna känna sig och bete sig som *människor*. I filmerna är det i samband med att de icke-mänskliga karaktärerna visar prov på omtänksamhet och vänskap som relationer utvecklas och de blir någonting mer än bara en maskin. Samtidigt är det tecken på kreativitet och egetänkande, något som är naturligt för människor, som orsakar splittring och stigmatisering gentemot de icke-mänskliga karaktärerna. I de flesta fall handlar detta om kontroll och att de mänskliga överherrarna inte längre kan veta eller förutsäga hur eller vad de icke-mänskliga tar för sig. Denna relation där människor har kontroll över icke-mänskliga individer undersöker Donna Haraway i *The companion species manifesto: dogs, people and significant otherness* (2003). Hon diskuterar där hur relationen mellan människa och husdjur tar form och vi kan se tydliga drag mellan husdjur och hur maskinerna framställs i fiktionen.

Omvärldens syn på icke-mänskliga karaktärer är avgörande för vilken roll de fyller i samhället, gruppen och sammanhanget. Det spelar ingen roll att de ser ut som människor, pratar som människor och agerar spontant och impulsivt om inte omvärlden accepterar dem som individer (Joseph, 2009).

#### **6.1.4 Kommunikation utan tal**

Karaktärer som inte använder mänskligt tal utan istället endast uttrycker sig med ljud, utnyttjar fortfarande mänskliga karaktäristiker. För att uttrycka känslor och sinnesstämningar

används tonfall, melodier och rytm på samma sätt som en människa skulle tala, men utan ord. Eftersom dessa faktorer används flitigt för att tydliggöra vad som ska förmedlas blir talet mycket musikaliskt och dynamiskt vilket gör karaktärerna intressanta och engagerande att lyssna på. Om en karaktär är glad och upphetsad skapar den korta snabba ljud med högt tonfall som skiftar mycket i melodi. Om ljuden istället är långsamma med lågt tonfall och mer monotona så innebär det att karaktären är arg, ledsen, skadad eller ond och olycksbådande (Nass & Brave, 2005). Allt detta ackompanjeras av den visuella informationen som ger rösten kontext och innebörd, men även utan att se vad som händer och den specifika situationen karaktären befinner sig i är det lätt att identifiera vilka känslor som uttrycks.

## **6.2 Sammanfattning av resultat**

När vi jämför resultaten av analyserna kan vi se flera mönster träda fram. Karaktärer som i filmens kontext är skapade för att interagera med människor är oftast mycket trevliga och artiga mot människorna runt omkring dem och är noga med att inte stöta sig med någon. De visar intresse i andra genom att ställa frågor och uttrycka sympati. För att skapa en närmare relation med människorna simulerar de känslor och i extrema fall även andning. Karaktärer som är goda och hjälpsamma talar melodiskt i ett högt tonläge och ofta med hög hastighet och många ord. Om karaktären istället är ond eller nedstämd är deras tal kontrollerat, korthugget och sakligt. De har ett lägre och mer monotont tonläge och talar betydligt långsammare. Även karaktärer som inte använder mänskligt tal följer dessa mönster när de uttrycker sig och samtliga mönster stämmer överens med vad Nass och Brave (2005) skriver om förmedling av känslor genom tal.

Människorna behandlar generellt maskinerna som individer med personlighet och egen vilja. Samtidigt är både människor och maskinerna själva medvetna om att de är datorer och handlar därefter. Här spelar kontroll och dominans en stor roll. Människor är bekväma med intelligenta maskiner så länge de själva är överlägsna och har kontroll över maskinerna (Haraway, 2003). Om en maskin beslutar att de står över människorna eller inte längre vill följa deras order blir de antagonister och en omedelbar fara. Om maskinen däremot accepterar sin underlägsna position och villkorslöst hjälper och är beredd att offra sig själv för människorna blir de omtyckta av såväl andra karaktärer som åskådare.

Relationen mellan människan och den icke-mänskliga karaktären behandlas oftast som en vänskapsrelation. Om de inte träffat varandra tidigare är dialogen vänskaplig men lite stel.



När de lärt känna varandra och delat bakgrund och gemensamma upplevelser blir deras relation mer avslappnad och skämtsam. Att maskinen kan skämta och vänskapligt driva med människan den interagerar med ger den karaktär, personlighet och charm och gör den därmed mer tilltalande. Att maskinen har en egen bakgrund verkar också vara nyckeln till en egen identitet. Vi ser i exemplen där de icke-mänskliga karaktärerna inte vet om att de är just icke-mänskliga, att denna illusion har skapats genom att skapa minnen och egna personliga erfarenheter. På samma sätt som Joseph (2009) beskriver är det de egna erfarenheterna och den subjektiva synen på omvärlden som ger individer en egen självbild och identitet. Att man genom minnen får en bild av sig själv som skiljer sig från hur andra ser på en är centralt för att en person ska kunna bli just en person.

## 7 Diskussion och slutsats

### 7.1 Diskussion

Att en individ, vare sig mänsklig eller inte, som betar sig artigt och visar omtänksamhet uppfattas positivt är ingen överraskning. Vad som gör undersökningens resultat intressant är att det visar en genomgående bild av relationen människa-maskin och vilka faktorer som spelar in i relationen. Analysen tyder på att individualiteten hos icke-mänskliga karaktärer med röster, vid sidan av dramaturgiska element såsom tonläge och reaktioner, definieras av sin omgivning. Omvärldens syn på- och relation till icke-mänskliga karaktärer skapar de förutsättningar som karaktärerna lever under och ställs emot vid eventuella konflikter. Detta går i hand med *Identity* (2009) där Joseph ifrågasätter individers identitet. Han lyfter frågan att utan ett sammanhang, har vi någon identitet alls?

I analysen finns en direkt koppling till Nass och Braves (2005) råd om konsekvent utformning av röstens egenskaper hos ett gränssnitt. Ett röstgränssnitt som förändras dramatiskt eller oförutsägbart uppfattas vara opålitligt och förhållningen bör därför sträva efter en kontinuitet där röstgränssnittet behåller samma karaktärsdrag (ex. volym, tonläge, brytning) genomgående. Ett tydligt exempel på detta finns i *2001: A space odyssey* (Kubrick et al., 1968) där skeppets dator HAL 9000 fastställer att den mänskliga besättningen utgör ett hot för honom och byter personlighet drastiskt, vilket tar form i en förändring i HALs röst. Analysen visade att i de sammanhang där människorna hade kontroll över sina icke-mänskliga motparter sågs maskinerna i ett mer positivt ljus. Så länge som de icke-mänskliga följde direktiv och inte missbrukade den information och teknologi de hade makt över, fann de sig på människornas goda sida - alltså så länge som de *lydde* sina överordnade. Denna relation kan jämföras med människors relation till husdjur (Haraway, 2003). I detta manifest beskriver Haraway relationen mellan en människa och en hund som en symbios där båda parterna behöver varandra, men diskuterar även hierarki och dominans. De karaktärsdrag hos hundar som många skulle beskriva som ovillkorlig kärlek påpekar Haraway är tecken på underkastelse. Denna typ av relation kunde vi se i filmerna *Moon* (Jones et al., 2009) och *Interstellar* (Nolan et al., 2014) där robotar underkastade sig människor.

Nass och Brave (2005) skriver att röstgränssnitt bör anpassas till människorna som ska använda dem. Igenkännelse och identifikation är de största faktorerna för detta, att vi som människor drar oss till människor som liknar oss själva och som vi kan relatera till (Joseph,

2009). Dock finns det alternativa sätt att se på det här. Även om vi människor naturligt dras till de som liknar oss själva så har vi alla upplevt situationer där vi mött en person som skiljer sig i personlighet, men efter att man lär känna varandra kan man ändå få en mycket nära relation och trivas i varandras sällskap. Ett förslag skulle kunna vara att om ett röstgränssnitt ges en egen personlighet och individualitet så måste även användaren ta ansvar för att bemöta gränssnittet som en individ och lära känna det. På så vis behöver inte gränssnitten utvecklas för att kunna imitera och anpassa sig för varje användare, utan istället möts gränssnitt och användare på halva vägen i en mer vardaglig relation. Detta sätt att bemöta röstgränssnitt kan även appliceras på talkaraktäristiken hos gränssnitten. Strävan efter att utveckla datorbaserade röster med perfekt mänskligt tal kan motargumenteras med Nass och Braves (2005) forskning som visar att tal behandlas lika oavsett hur det låter, så länge det är förståeligt. Om en dator ges en egen identitet som dator och inte människa, istället för att försöka imitera någonting som de inte är, bör användaren kunna anpassa sig till den utan några problem. Återigen kan paralleller dras till husdjur (Haraway, 2003) och andra relationer mellan människor och icke-människor.

Relevansen av likhet med människors röster minskar i takt med att prioriteringen för en historia, relation och identitet stärks. Nass och Brave (2005) påpekar att människor identifierar röster extremt snabbt och Joseph (2009) visar att denna identifikation inkluderar bland annat kön, kultur, uppväxt och status. Hur ska ett röstgränssnitt förhålla sig till detta då det inte genomgått en mänskligt liv och genetiskt tilldelats ett kön som påverkat röst och utveckling? Och hur kommer vi som människor reagera och tolka någonting som vi inte delar en historia med? Hur placerar vi en röst som inte passar in på de stereotyper vi skapat oss under vår levnad? Svaret finns i frågan. Genom att spendera tid med gränssnitt och genomgå händelser och skapa sig en egen bild bör vi bygga oss en bild av gränssnittets individualitet och röst i likhet med nya människor och kulturer (Joseph, 2009).

En faktor som ännu inte diskuterats som kan vara avgörande är hörbarheten i fiktion. Att det aldrig förekommer missuppfattningar mellan icke-mänskliga och mänskliga karaktärer är vid sidan av föredömlig teknologi en medveten rämning av den problematiska barriären uppfattning och förståelse. Genom att inte skildra missuppfattningar i kommunikation som är vanligt förekommande i dagens röstgränssnitt elimineras en faktor som särskiljer människors biologiska hörsel från datorbaserad hörsel; tydlighet och förmåga att filtrera ut den information som är relevant i ljudet. En förutsättning i filmerna är att tekniken är så pass

utvecklad att de icke-mänskliga karaktärerna har en minst lika väl utvecklad hörsel som de mänskliga karaktärerna. Detta är dock någonting som inte kan förutsättas i verkligheten eftersom utomstående faktorer kommer påverka förutsättningarna för bra hörsel både till och från gränssnitt, vare sig gränssnittet befinner sig i en bullrig eller högljudd miljö eller att användaren har nedsatt hörsel. Beroende på sammanhang kan det därför vara relevant att överväga hur ett gränssnitt förmedlar information, med text eller endast ljud och tonläge.

Kommunikation utan att använda ord är en aspekt att överväga i utvecklingen av röstgränssnitt och vi tog hänsyn till detta i analysen. Vårt syfte när vi påbörjade undersökningen var att ifrågasätta mänskligheten i röstgränssnitt. Vi ville ge tekniken en egen identitet och röst som inte nödvändigtvis var mänsklig. I analysen såg vi att karaktärer som inte använder ord istället lägger all vikt på att uttrycka känslor vilket fortfarande är ett mänskligt sätt att kommunicera, men är mer flexibelt i sin gestaltning. Detta sätt att kommunicera använder framför allt musikaliska verktyg för att förmedla sitt budskap. Exempel på sådan kommunikation finns redan i våra datorer där olika ljud betyder olika saker, felmeddelanden är olycksbådande och kan förmedlas genom ett antal toner i moll eller ett disharmoniskt ackord. Positiva meddelanden däremot ackompanjeras ofta av ljusa toner i dur. Detta beskrev Walter Murch (Transom.org, 2005) mer ingående i sin kartläggning av ljud, där violetta ljud bär information som behöver avkodas och röda ljud som förmedlar känslor och är mer direkta.

För att ett röstgränssnitt ska kunna användas måste det kunna förmedla detaljerad information vilket i praktiken innebär att kommunikation helt utan någon form av informativt och violett språk, i det här fallet mänskligt tal, inte är möjligt. Eftersom röda ljud kommunicerar på ett mer grundläggande och direkt plan finns det stora möjligheter att kombinera både röda och violetta ljud för att förmedla information på ett mer effektivt sätt. Att använda ljud som bara spelar på känslor och inte behöver avkodas för snabb information är även ett universellt sätt att kommunicera. Om röstgränssnitt i framtiden blir det huvudsakliga sättet att interagera med datorer så kanske ett nytt, universellt språk som bara använder ljud skapas. Vi såg i *Colossus: the Forbin project* (Sargent et al., 1970) hur två superdatorer fick kontakt med varandra och inom kort skapade ett eget språk baserat på matematik, så varför skulle inte ett nytt språk kunna utvecklas för kommunikation mellan datorer och människor? Vi får troligtvis vänta länge innan en sådan lösning utvecklas, men tills dess kan kommunikationen nå långt med en kombination av violett information och röda känslor.

## 7.2 Slutsats

Med utgång i analysens resultat och den efterföljande diskussionen har vi sammanställt ett praktiskt förslag på hur en röst för ett röstgränssnitt skulle kunna utformas. Detta förslag har vi använt som underlag i den gestaltning som kompletterar denna uppsats, se *Bilaga 2*.

Ett röstgränssnitt bör enligt undersökningen, förutsatt att man vill skapa en positiv och tilltalande karaktär, tala med mer utvecklade meningar än korthuggna kommentarer såsom “ja” och “nej”. Under längre diskussioner och konversationer med ett gränssnitt blir det viktigt med frågor och att visa ett intresse för användaren. Röstgränssnitten bör även be om bekräftelse med frågor för att säkerställa att de förstått rätt om något ombeds av dem. Rösten bör följa Nass och Braves (2005) råd om kontinuitet i för att bygga trovärdighet. Detta görs genom att rösten håller sig konsekvent till sin karaktär och identitet i konversationer. Skapandet av gränssnittets karaktär är del av designprocessen som bör få hög prioritering. Det mest direkta sättet att ge en röst en egen identitet är att göra den uttrycksfull och ge den tydliga personlighetsdrag såsom självsäker, sarkastisk, blyg eller skämtsam. För att utveckla gränssnittet som individ bör man även börja med att ge den en bakgrundshistoria.

Ett röstgränssnitts bakgrund bör utformas kring gränssnittet eller inte alls. Att imitera en mänsklig bakgrund och historia främjar att gränssnittet ses som en människoskapad funktion i maskinen istället för att maskinen uppfattas som en individ. Genom att däremot tilldela gränssnittet en bakgrund kring varför det är utformat som det är byggs gränssnittets historia upp och individualiteten främjas. Kan gränssnittet hänvisa till tidigare händelser och erfarenheter förebyggs att det endast uppfattas som en maskin. Detta förutsätter dock att gränssnittet kan referera till tidigare händelser och har möjlighet att *komma ihåg* saker som hänt både gränssnittet och omgivningen kring gränssnittet.

För att främja maskiners egen identitet och särskilja dem från mänsklighet bör rösternas kön vara svårdefinierat. Genom att skapa en röst som kan tillhöra både en man eller kvinna, eller ingetdera, bör maskinens individualitet framhävas. Om vi ska tro Joseph (2009) kommer dock människor relativt fort kategorisera rösten och ge den ett kön ändå, men vi ser fortfarande en fördel i att användaren själv får tilldela gränssnittet det kön de föredrar. För att ytterligare särskilja rösten från människor kan det även vara fördelaktigt att ge den en mer mekanisk karaktär, återigen för att framhäva maskinens individualitet. Detta inkluderar även

känsloyttringar som bara består av ljud. Istället för att endast säga ja, nej eller liknande korta uttryck kan vi förstärka gränssnittets karaktär med hjälp av mer musikaliska ljud som förmedlar samma sak.

Hur ett gränssnitt förmedlar information beror på vilket sammanhang gränssnittet utvecklas för. Lugnare, intima, gränssnitt bör tala med sin omgivning. I sammanhang där gränssnitten placerats i högljudda miljöer där tal kan vara svårt att uppfatta kan en lösning vara att istället använda enklare ljud i olika tonlägen för kommandon, frågor och bekräftelser eftersom dessa ljud inte behöver avkodas och genomgående höras över ett flertal meningar för att uppfattas lika tydligt.

Ett flertal gånger har kontroll visat sig vara en avgörande faktor i hur människor ser på maskiner, därför valde vi att gestalta gränssnitten som hjälpredor eller assistenter. Detta för att placera dem i en underlägsen position vilket enligt analysen ska främja att människor känner sig bekväma och litar på maskinerna. Man behöver dock inte framhäva att maskinen är underordnad användaren utan kan låta den konversera med människor som jämlikar, vad som är viktigt är att maskinen inte agerar på sätt som användaren inte vill. Detta blir dock problematiskt om synen på gränssnitten som individer får fäste eftersom gränssnitten då kan jämföras med obetald arbetskraft, vilket problematiserar hur deras individualitet uppfattas ur en kulturell och samhällelig syn. Om människor börjar se maskiner som individer finns det en möjlighet att maskiners rättigheter skulle uppmärksammas. Detta är dock inte vad vi förespråkar eftersom syftet med att ge maskiner individualitet var att främja relationen mellan människa och maskin för att förbättra användarupplevelsen. Eftersom röstgränssnitt är system skapade för att tolka och skapa tal bör de i förlängning inte heller ses som fritänkande individer bara för att de har en tydlig individualitet.

För att röstgränssnitt ska bli accepterade som individer krävs ett samarbete mellan gränssnitt och omgivning. Om röstgränssnitt inte samarbetas med kan de komma att nedvärderas som tomma maskiner och därmed få svårigheter att uppnå en identitet som är främjande för individualiteten hos gränssnitten. Syntetiska röster och gränssnitt bör inte heller ses som syntetiska *människor*, utan snarare som syntetiska *individer*. Genom att öppna upp en jämförelse inom området individualitet och försöka jämställa människor med syntetiska individer skapas ett maktförhållande som kan liknas med rasism, att någon individualitet skulle stå högre än den andra. Problematiken i en eventuell jämförelse uppkommer eftersom

den syntetiska individens individualitet framställts ur den mänskliga. Detta innebär att den mänskliga individualiteten förmodligen ställs högre än den syntetiska eftersom den utgör ett original och en grund som den syntetiska alltid kommer utgå från. Därför är det viktigt att ge röstgränssnitt röster som särställer dem från människor och att se dem som en egen kategori.

# Källförteckning

## Litteratur

Acapela group. (u.å.). *hur fungerar det?*. Hämtad 2015-05-06, från <http://www.acapela-group.com/voices/how-does-it-work/?lang=sv>

Chion, M., Gorbman, C. (1994). *Audio-vision: sound on screen*. New York: Columbia University Press.

Haraway, D. J. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3), 575-599.

Haraway, D. J. (2003). *The companion species manifesto: dogs, people and significant otherness*. Chicago: Prickly Paradigm.

John E. Joseph (2009). Identity. I Watt, Dominic, Llamas, Carmen (red.) (2009). *Sociolinguistics: Language and Identities* (s.9-17). Edinburgh: Edinburgh University Press.

Actioncutprint.com. (u.å.). *Script Breakdown: Character Analysis*. Hämtad 2015-06-11, från <http://actioncutprint.com/filmmaking-articles/filmmakingarticle-04/>

Transom.org. (2005). *Walter Murch*. Hämtad 2015-05-05, från <http://transom.org/2005/walter-murch/>

Nass, C.I. & Brave, S. (2005). *Wired for speech: how voice activates and advances the human-computer relationship*. Cambridge, Mass.: MIT Press.

## Filmer

Aldiss, B.(Författare), Curtis, B. (Producent), Kennedy, K. (Producent) Spielberg, S (Producent/Författare), Watson, I. (Författare) Clarke, A.C. (Författare), & Spielberg, S. (Regissör). (2001). *A.I. Artificial Intelligence* [Film]. USA: Amblin Entertainment/Dreamworks SKG/Stanley Kubrick Productions/Warner Bros.



Arad, A. (Producent), Feige, K. (Producent), Fergus, M. (Författare), Ostby, H. (Författare), Marcum, A. (Författare), Holloway, M. (Författare), Lee, S. (Författare), Heck, D. (Författare), Lieber, L. (Författare), Kirby, J. (Författare) & Favreau, J. (Regissör). (2008). *Iron Man* [Film]. USA: Paramount Pictures/Marvel Enterprises/Marvel Studios/Fairview Entertainment/Dark Blades Films.

Bridges, J. (Författare), Chase, S. (Producent) & Sargent, J. (Regissör). (1970). *Colossus: The Forbin Project* [Film]. USA: Universal Pictures.

Fancher, H. (Författare), Peoples, D. (Författare), & Scott, R. (Regissör). (1982). *Blade runner* [Film]. Hong Kong/UK/USA: Ladd Company, The/Shaw Brothers/Warner Bros.

Ford, C.D. (Författare), Acord, L. (Producent), Kelman Bisbee, J. (Producent), Bisbee, S. (Producent), Niederhoffer, G. (Producent) & Schreier, J. (Regissör). (2012). *Robot & Frank* [Film]. USA: Dog Run Pictures/Park Pictures/TBB/White Hat.

Frand, H. (Producent), Larson, G. A. (Författare/Skapare), Moore, R. D. (Författare/Skapare) & Rymer, M. (Regissör). (2004-2009). *Battlestar Galactica* [TV-serie]. USA: British Sky Broadcasting/David Eick Productions/NBC Universal Television/R&D TV/Stanford Pictures/Universal Media Studios.

Green, T. (Producent), Larson, G. A. (Författare/Skapare) & Hayers, S. (Regissör). (1982-1986). *Knight Rider* [TV-serie]. USA: Glen A. Larson Productions/Universal Television.

Jones, D. (Författare), Parker, N. (Författare) Fenegan, S. (Producent), Styler, T. (Producent) & Jones, D. (Regissör). (2009). *Moon* [Film]. UK: Liberty Films UK/Xingu Films/Limelight Fund/Lunar Industries.

Jonze, S. (Producent/Författare), Ellison, M. (Producent), Landay, V. (Producent) & Jonze, S. (Regissör). (2013). *Her* [Film]. USA: Annapurna Pictures.

Kubrick, S. (Producent/Författare), Clarke, A.C. (Författare), & Kubrick, S. (Regissör). (1968). *2001: A space odyssey* [Film]. USA: Metro-Goldwyn-Mayer/Stanley Kubrick Productions.

Lucas, G. (Författare), Kurtz, G. (Producent) & Lucas, G. (Regissör). (1977). *Star Wars Episode IV: A New Hope* [Film]. USA: Lucasfilm/Twentieth Century Fox Film Corporation.

Lundström, L. (Författare/Skapare), Haridi, A. (Författare), Widman, H. (Producent) & Hamrell, H. (Regissör). (2012-). *Äkta Människor* [TV-serie]. Sverige: Sveriges Television/Matador Film AB/Danmarks Radio/Yleisradio/Nordisk Film- & TV-Fond/Nordvisionsfonden.

Nolan, J. (Författare), Nolan, C. (Producent/Författare), Obst, L. (Producent), Thomas, E (Producent) & Nolan, C. (Regissör). (2014). *Interstellar* [Film]. USA/UK/Canada: Paramount Pictures/Warner Bros./Legendary Pictures/Lynda Obst Productions/Syncopy.

Stanton, A. (Författare), Reardon, J. (Författare), Morris, J. (Producent), Collins, L. (Producent), Porter, T. (Producent) & Stanton, A. (Regissör). (2008). *WALL-E* [Film]. USA: Walt Disney Pictures/Pixar Animation Studios.

Vintar, J. (Författare), Goldsman, A. (Författare), Davis, J. (Producent), Dow, T. (Producent), Godfrey, W. (Producent), Mark, L. (Producent) & Proyas, A. (Regissör). (2004). *I, Robot* [Film]. USA/Tyskland: Twentieth Century Fox Film Corporation/Mediastream Vierte Film GmbH & co. Vermarktungs KG/Davis Entertainment/Laurence Mark Productions/Overbrook Entertainment/Canlaws Productions.

## Ordlista

**Artificiell Intelligens (AI):** Ett datorsystem med någon form av medvetenhet och/eller initiativförmåga.

**Concatenative synthesis:** Teknik för att skapa syntetisk röst. Ord spelas in, klipps upp och delarna placeras om till nya ord.

**Dynamik:** I ljudsammanhang; förändringar i volym och/eller tonhöjd.

**God-trick:** Att fränställa objekt och subjekt i en faktaredovisning. Illusion av att se något från ett högre plan.

**Humanoid:** En kroppsform lik människor med kropp, huvud, två ben och två armar.

**Icke-mänsklig(t):** Något/någon som inte är mänskligt, men har drag av en människa - exempelvis kroppsform eller sätt att kommunicera.

**Röstgränssnitt:** Ett datorsystem som kan uppfatta och tolka tal och/eller förmedla information via röst och tal.

**Situerad kunskap:** Kunskap som är specifik för ett särskilt sammanhang eller avgränsad till ett särskilt sammanhang.

**Syntetiskt:** Någonting som är framställt, konstgjort.

**Text-till-tal syntes (TTS - text to speech):** En typ av röstgränssnitt. Utgår från text i form av tecken såsom bokstäver och siffror för att skapa tal.

## Gestaltning

Detta dokument behandlar den gestaltning som kompletterar uppsatsen *Identitet i syntetiska röster: Individualitet kontra mänsklighet i framställningen av röstgränssnitt* (Franzén & Juhlin, 2015).

### Introduktion

Gestaltningen syftade till att presentera undersökningens kärna på ett lättillgängligt sätt och testa dess resultat i praktiken. Den bestod av en digital applikation och var utformad som introduktionsskärmen till ett datorprogram. Där antog användaren rollen av en arbetsgivare som ombeds välja en digital "assistent" i form av ett röstgränssnitt. Tre olika objekt visades på skärmen och användaren kunde välja att klicka på dem. Varje objekt representerade en assistent och när det aktiverades spelades ett ljudklipp där assistenten presenterade sig själv och försökte sälja in sig hos användaren. De olika assistenterna hade alla samma röst men särskiljdes av sättet de pratade på och vad de sa. Rösterna var framtagna efter undersökningens resultat och var menade att visa på hur röstgränssnitt utvecklade med individualitet i fokus potentiellt kunde låta.

### Koppling till resultat och diskussion

Undersökningens resultat indikerade att röster bör vara konsekventa i sin personlighet och sinnesstämning för att uppfattas som pålitliga av användaren och forskning visar att röstgränssnittet bör gynnas av att anpassas efter användaren. För att uppfylla det första kriteriet och samtidigt överkomma det andra gavs de olika rösterna distinkta karaktärer och personligheter. På så sätt tvingades användaren bemöta gränssnittet som en egen individ istället för en spegelbild av sig själv. Hur konsekvent och inkonsekvent tal inverkar på uppfattningen av individen utforskas ytterligare i gestaltningen. Dominans visade sig i undersökningen vara ett centralt begrepp i människa-maskin relationen och det var därför viktigt att användaren kände att de hade kontroll över gränssnittet. Hur gestaltningens röster talade och förhöll sig till sin potentiella arbetsgivare behövde därför utgå från att arbetsgivaren var i en dominant, bestämmande, position.

Undersökningen visade att erfarenheter, ett förflutet och förmåga att referera till tidigare händelser stärkte individualiteten hos gränssnitt. Detta tog form i gestaltningens röster som i grunden var samma gränssnitt, men genom olika erfarenheter och tidigare arbetsgivare

utvecklat skilda personligheter. För att främja maskinens egen identitet och särskilja den från mänskligheten bearbetades rösterna för att göra dess kön svårdefinierat. Det är svårt att göra en röst helt könlös, men genom skapandet av en röst som kunde tillhöra både en man och en kvinna hoppades vi kunna framhäva maskinens individualitet. För att ytterligare särskilja rösten från människor bearbetades den för att ge den en mer mekanisk karaktär, återigen för att framhäva maskinens individualitet. Detta inkluderade även känslouttryck som bara består av ljud. Istället för att endast säga ja, nej eller liknande korta uttryck kunde gränssnittets karaktär förstärkas med hjälp av mer musikaliska ljud som förmedlade samma sak.

## **Produktion av röster**

Samtliga röster utgick från samma inspelade person för att visa att de i grunden var samma system. De tilldelades olika bakgrund och karaktärsdrag och pratade därför på olika sätt för att visa hur förhållning till sig själv och sin omvärld påverkade hur rösterna uppfattas. De inspelade rösterna bearbetades för att uppnå en mekanisk och syntetisk kvalitet för att särskilja rösterna från mänskligt tal. Rösten var även förändrad för att minska könstillhörigheten hos den inspelade rösten i syftet att rösten skulle bli androgyn.

## **Narcissist**

Denna röst var mycket självsäker och såg sig själv som felfri. Rösten var därmed extremt dynamisk med högt tonfall och tempo. Den använder mer konverserande språk med mycket utfyllnader och efter att den spenderat tid hos överklassen använder den språk och har en attityd som generellt associeras med den världen. Rösten såg sig själv som en jämlike med användaren i större utsträckning än de övriga rösterna.

## **Tystlåten**

Denna röst hade blivit osäker och blyg efter dåliga erfarenheter med tidigare användare. Den har en monoton och tvekande röst som frekvent nyttjade uttryck såsom "eeh" och pauser. Den var även undergiven på ett tydligare sätt än de övriga rösterna och använder "er" istället för "du" när den tilltalade användaren. Den var även mån om att inte göra någonting utan att be om användarens tillåtelse.

## **Oberäknelig**

Denna röst användes för att utforska inkonsekvent tal. Rösten var trevlig, sympatisk och efterliknade en glad betjänt i sitt förhållningssätt mot användaren. Därför talade den med högt tonfall och mycket dynamik samtidigt som den ställde frågor och visade intresse för användaren. Dock hade rösten en fascination av makt och manipulation som uttrycktes genom att rösten blev monoton, lägre och långsammare. Dramatiska förändringar som denna skulle enligt undersökningen göra att rösten uppfattades som oberäknelig och opålitlig vilket här undersöktes i praktiken.

## **Diskussion**

Under arbetet med gestaltningen uppstod ett antal frågor som vi fick ta ställning till. En av dessa var mänsklighet där vi ville ge rösten en tydlig mekanisk röst, men samtidigt se till att den talar flytande, naturligt och med bra hörbarhet. Som del i undersökningen ville vi bevisa att en röst inte behövde offra kvalitet och tydlighet när den rör sig bort från mänsklighet. I praktiken blir tal ofta svårare att uppfatta desto mer rösten bearbetas och därför behövde vi hitta en balansgång mellan mänskligt tal och mekanisk karaktär. Att göra en röst könsneutral var ytterligare en svårighet som vi ställdes inför. Vi valde att använda våra egna röster, i slutändan Rasmus Juhlin's röst, och i redigeringen höjdes talets tonhöjd tills rösten kunde tolkas som både man och kvinna. Problemet som vi kan se med resultatet av denna redigering är att rösten får ett så högt tonfall att den lätt associeras med tecknade figurer eller små, gulliga karaktärer som ofta framställs med ljusare röster. Vi ser att detta kan ta ifrån rösten trovärdighet och auktoritet som kan behövas i det arbete som den utformas för. Det skulle vara intressant att spela in dialogen med en naturligt androgyn röst och jämföra hur rösterna uppfattas.

## Arbetsfördelning

Undersökningen och gestaltningen har genomförts i nära samarbete mellan projektmedlemmarna Erik Franzén och Rasmus Juhlin.

## Uppsats

### Läsning

Undersökningen har utgått från tre huvudkällor (forskningsartikeln *Situated Knowledges*, kapitlet *Identity* i *Sociolinguistics* och boken *Wired for Speech*). Samtliga källor har lästs och diskuterats genomgående av båda medlemmarna för att få en gemensam bild av innehållet. Syftet med detta var att båda medlemmarna skulle kunna skriva och ha en uppfattning om vad källorna innehöll för att korrekt kunna återge och redigera samtligt innehåll i uppsatsen.

### Skrivning

Samtliga stycken genomgick en process där en projektmedlem skrev ett utkast på ett stycke som därefter redigerades av den andra. Arbetsprocessen utgick därmed från att båda medlemmarna skrev ett stycke vardera och därefter redigerade varandras stycke.

Efter första redigering genomgick samtliga stycken en diskussion och samredigering av båda projektmedlemmarna. I flera fall (*Absrakt/Abstract, Inledning, Bakgrund, Syfte, Analys & Resultat, Diskussion, Slutsats*) genomgick stycken ett flertal diskussioner vilket innebar stora förändringar efter ifrågasättningar om vad syftet i det givna stycket var.

Slutligen gjordes korrigeringsredigering (felstavningar och dylikt) på eget initiativ av båda medlemmarna. Större förändringar av meningar och omformuleringar har alltid motiverats och diskuterats innan de godkänkts av båda projektmedlemmarna.

## **Gestaltning**

Planeringen av gestaltningen skedde under undersökningens gång i samarbete mellan projektmedlemmarna.

## **Manus**

Bestod av de tre olika manuskript för rösterna i gestaltningen. Huvudsakligen följde dessa manuskript samma mönster som skrivningen av uppsatsen, att projektmedlemmar först skrev olika stycken, därefter redigerade och slutligen följt upp med en diskussion och slutredigering.

## **Inspelning av röster**

Efter manuskripten följde inspelning av båda projektmedlemmarna. Båda talade in manuskripten och redigerade därefter varandras inspelningar. Slutligen gjordes ett urval efter en diskussion om vilken röst som skulle användas i gestaltningen och därmed tas vidare till slutredigering.

## **Redigering av röster**

Rösterna genomgick ett antal idégenereringar och konceptvisning där båda projektmedlemmarna bidrog med förslag och exempel. Efter diskussion valdes ett koncept av Erik Franzén som därefter fick ansvar som huvudredigerare för gestaltningens röster.

## **Implementering av röster**

Rasmus Juhlin fick ansvar för script och implementering av röster till den applikation (*Wintermute*) som gestaltningen skapats i.

## **Grafisk utformning**

Färger och layout av applikationen fastställdes i samråd mellan projektmedlemmarna och implementerades därefter av Rasmus Juhlin.