



Högskolan
Kristianstad

Högskolan Kristianstad
291 88 Kristianstad
044-250 30 00
www.hkr.se

Självständigt arbete (examensarbete), 15 hp, för
Kandidatexamen i Informatik
VT 2023
Fakulteten för ekonomi

Du är f*n dum i huvudet!

Hur error feedback för svenska användare av smarta högtalare kan utvecklas.

Jeanette Elfving och Julia Blom

Författare

Jeanette Elfving och Julia Blom

Titel

Du är f*n dum i huvudet! Hur error feedback för svenska användare av smarta högtalare kan utvecklas.

Engelsk titel

You're f*cking stupid! How error feedback for Swedish users of smart speakers can be developed.

Handledare

Kari Rönkkö

Examinator

Kerstin Ådahl

Sammanfattning

Smarta högtalare har idag blivit ett vanligt inslag i svenska hem, med flera funktioner för att lösa enklare vardagsproblem. Trots teknikens utveckling återstår problem kring själva dialogen mellan smarta högtalare och dess användare. Detta gäller speciellt den bristande error feedback användaren får när ett fel inträffar i dialogen. För att undersöka bristen kring error feedback och vad användare önskar har studier genomförts i form av förundersökande enkät, litteratursökning samt intervjuer och observationer "in the wild". Studierna har utgått från ett bekvämlighetsurval, där den error feedback som erbjuds idag har undersökts. Resultatet av studien visade på att det finns gemensamma teman för deltagarnas åsikter gällande den error feedback som erbjuds idag och hur den skulle kunna utvecklas. Ett exempel är önskan om individuella inställningar för hur instruktioner ges, vilket skulle kunna bidra till att användarna tar till sig instruktionerna bättre. De instruktioner som ges i dagsläget är ofta långa och har hög uppspelningshastighet, där möjligheterna att få dem upprepade eller framförda långsammare inte finns. Genom att erbjuda användare möjligheten att själva bestämma hur instruktionerna ges ökar chansen att målet uppnås. Detta ger även möjligheten för användarna att upptäcka och utnyttja fler av de smarta högtalarnas funktioner. Med studiens resultat som grund skapades ett förslag på sex guidelines indelade i två olika kategorier av error feedback: *Inställningar* och *Användning*. Detta för att hjälpa designers att utveckla error feedback från Voice user interface-system på svenska genom smarta högtalare.

Ämnesord

Smarta högtalare, Voice User Interface, Google Assistant, Guidelines, Error feedback

Author (formatmall Abstract Rubrik)

Jeanette Elfving and Julia Blom

Title

You're f*cking stupid! How error feedback for Swedish users of smart speakers can be developed.

Supervisor

Kari Rönkkö

Examiner

Kerstin Ådahl

Abstract

Smart speakers have become a common feature in Swedish homes, with several functions to solve simpler everyday problems. Despite the development of technology, problems remain regarding the dialogue between the smart speakers and their users. This applies especially to the lacking error feedback the user receives when an error occurs in the dialogue. In order to investigate the lack of error feedback and what users want, studies have been carried out in the form of an exploratory survey, literature search as well as interviews and observations “in the wild”. The studies have been based on a convenience collection, where the error feedback that is offered today is examined. The results of the study indicated that there are common themes for the participants' opinions regarding error feedback that is offered today and how it can be developed. An example is the desire for individual settings for how instructions are given, which could contribute to users understanding the instructions better. The instructions given today are often long and have a high playback speed, where possibilities to have them repeated or performed slower do not exist. By offering users the opportunities to decide for themselves how the instructions are given, the chance that the goal will be achieved increases. This also gives the users the opportunity to discover and utilize more of the smart speakers' functions. Using the results of the study as a basis, a proposal was created for six guidelines divided into two different categories of error feedback: *Settings* and *Use*. This is to help designers develop error feedback from Voice user interface-systems in Swedish through smart speakers.

Keywords

Smart speakers, Voice User Interface, Google Assistant, Guidelines, Error Feedback

Innehållsförteckning

Begreppsdefinitioner	6
Hur fungerar röstbaserade gränssnitt?	7
1. Introduktion	7
1.1 Syfte och frågeställning.....	9
1.2 Avgränsningar	10
1.3 Disposition	10
2. Litteraturgenomgång.....	12
2.1 Auditiv feedback	12
2.2 Error feedback.....	15
2.3 Design av VUI.....	18
2.4 GUI till VUI.....	19
3. Metod.....	21
3.1 Enkätundersökning.....	21
3.2 Litteratursökning	23
3.3 Deltagande observation.....	25
3.4 Intervju	26
3.5 Deltagare och urval	27
3.6 Genomförande.....	29
3.7 Analysmetod	30
3.8 Etiska överväganden	31
4. Resultat och analys.....	32
4.1 Resultat Enkät	32
4.2 Resultat observation	32
4.2.1 Uppgift 4: Byt låt.....	33
4.2.2 Uppgift 7-10: Sätt igång podcasten Creepypodden i P3, Stäng av podcast, Sätt igång podcasten Nordic Mythology Podcast, Stäng av podcast	34
4.2.3 Uppgift 12: Stoppa timer.....	35
4.2.4 Uppgift 15: Lägg till ett event för lunch den 14 april klockan 1 på eftermiddagen.....	36
4.2.5 Uppgift 16: Ta bort eventet för lunch den 14 april klockan 1 på eftermiddagen.....	38

4.2.6 Uppgift 17: Ta reda på vem den 15:de ministern av Kanada var	40
4.2.7 Uppgift 18: Ta reda på mer information om den 15:de ministern av Kanada	42
4.2.8 Uppgift 19: Ta reda på temperaturen i New York både i Celsius och Fahrenheit	43
4.2.9 Uppgift 20-21: Undersök om det finns några veganska restauranger i närheten, ta reda på bästa rutten till Kristianstad Högskola från din position	44
4.3 Sammanställt resultat av observation	45
4.4 Resultat intervjuer	47
4.4.1 Användningsområden	48
4.4.2 Kommandon till smarta högtalare	49
4.4.3 Error feedback från Google Assistant	51
4.5 Design guidelines	54
4.5.1 Inställningar	55
4.5.2 Användning	56
4.6 Återkoppling från deltagare	57
5. Diskussion	58
5.1 Metoddiskussion	58
5.2 Resultatdiskussion	60
6. Slutsatser och Framtida studier	65
7. Referenser	67
8. Bilagor	73
1. Förstudie enkät	73
2. Sökblock	77
3. Sökschema	78
4. Information om studie till deltagare	80
5. Observationsuppgifter	81
6. Tema för intervjuer	82

Begreppsdefinitioner

VUI (Voice user interface) - Röstbaserat gränssnitt som tillåter att användare interagerar med tekniska system via röstkommando och får svar via röstoutput, det vill säga genom auditiv feedback.

GUI (Graphic user interface) – Grafiskt gränssnitt som tillåter användaren att interagera och ge input via tangentbord eller datormus med tekniska system genom text och grafiska element.

Error feedback - Den feedback en användare får när problem uppstår efter att ha uttalat ett kommando till smart högtalare.

Guidelines - En uppsättning riktlinjer för hur någonting skulle kunna designas.

VA (Voice assistant) - En mjukvara som använder röstbaserat gränssnitt för att kommunicera och utföra uppgifter.

NLP (Natural language processing) - Underkategori inom maskininlärning som möjliggör för maskiner att förstå, analysera och generera mänskligt tal som sedan kan förmedlas via röstoutput eller textoutput.

API (application programming interface) - Tillåter synkning och delgivning av data mellan multipla plattformar och kommunikation mellan olika mikrotjänster i webbapplikationer.

Väckningsfras - Den fras en användare säger för att väcka en smart högtalare, till exempel "Hej Google"

Google Assistant - En konversationsassistent (conversational assistant) som kommunicerar med användare via både röst och text.

Smart högtalare - En högtalare med inbyggd röstassistent (Voice assistant) som möjliggör kommunikation via röstinput och röstoutput.

HCI (human-computer interaction) - Interaktionen och delgivningen av information mellan mänskliga användare och datoriserade system.

Hur fungerar röstbaserade gränssnitt?

En röstassistent (VA) består av tre delsystem: röstinspelning, röstigenkänning och kommandokörning (Yan, Zhang, Ji, Zhang, Zhang & Xu, 2021).

Röstinspelningssystemet (voice recognition) registrerar omgivningen och röster som sedan förstärks, filtreras och digitaliseras. Dessa skickas sedan till systemet för röstigenkänning (SR - speech recognition) som har två tillstånd, aktivering eller erkännande. Innan ett kommando kan utföras måste systemet först aktiveras genom en ”väckningsfras” eller via en startknapp på den fysiska högtalaren. Vissa röstassistenter kan även kopplas till specifika användare och enbart reagera på deras röstkommando. Röstigenkänningen bearbetas sedan via en molntjänst (Cloud service) (Yan et al., 2021) genom maskininlärningsalgoritmer och uppfattas ett korrekt röstkommando skickas detta vidare till kommandokörningssystemet som utför kommandot och ger ett bekräftande svar genom röstoutput via högtalaren till användaren (Sutton, Kirk, Foulkes, & Lawson, 2019).

1. Introduktion

Smarta produkter är idag en stor del av svenska hem, varav 13% av dessa är röststyrda högtalare och innehas främst av 70- och 80-talister (Svenskarna och internet, 2022). Sedan lanseringen av Amazons Echo 2014 med sin AI Alexa har teknologin både utvecklats med konkurrenter såsom Google och Apple, men även tredjeparts företag har lanserat egna högtalare (van Gils 2020). Genom smarta högtalare kan användare idag koppla ihop och styra sina produkter via röststyrning. Undersökningar visar att den största anledningen till att införskaffa röststyrda högtalare idag är att lyssna på musik, kolla vädret eller nyheterna, ställa frågor utan att fysiskt skriva dem eller för att det verkade vara en rolig ny produkt (The Smart Audio Report, 2022). Utvecklare har fokuserat på att skapa nya och bredare uppsättningar av funktioner, istället för att förfina de som redan finns (Alrumayh & Tan, 2022) eller erbjuda bättre språkförståelseprogram för olika dialekter och meningsformuleringar (Shafei & Tan, 2022). Problemet med att skapa för många färdigheter och funktioner till en smart enhet är att användaren omöjligt kan lära sig att memorera allihop eller specifika kommandon ordagrant

(Alrumayh & Tan, 2022; Kirschthaler, Porcheron & Fischer, 2020). En av svårigheterna med att använda röststyrda enheter är bristen på visuell feedback där användaren tvingas lära sig att navigera och testa sig fram hur interaktionen kan gå till (Dickson 2018). Det finns idag inget välutvecklat system för användare att visualisera mentala modeller för röststyrda högtalare och på så sätt hjälpa dem i tankeprocessen för att navigera och upptäcka bredden av vad teknologin har att erbjuda (Kirschthaler et al., 2020; Moore & Urakami, 2022).

Forskningen kring röstassistenter och smarta enheter samt dess användande har under de senaste åren fortsatt att växa (Diederich, Brendel, Morana & Kolbe, 2022). Några artiklar har försökt sammanställa state-of-the-art forskning kring Voice User Interface och dess olika problemområden (Diederich et al., 2022; Baimirov, Mergengali & Baimirov, 2022; Clark, Doyle, Garaialde, Gilmartin, Schlögl, Edlund, Ayett, Cabral, Munteanu, Edwards & Cowan, 2019; Seaborn, Miyake, Pennefeather & Otake-Matsuura, 2021). Deras fynd påvisar luckor i forskningen och brist på struktur (Diederich et al., 2022). Det saknas utförliga beskrivningar för hur deras användartester och interaktioner med användare och smarta högtalare går till, speciellt i icke-kontrollerade miljöer "*in the wild*" (Clark et al., 2019). Även design av VUI-system och information som förmedlas via röstoutput, samt hur användaren uppfattar och svarar på denna behöver undersökas (Seaborn et al., 2021).

Ett antal artiklar har försökt att lyfta problem eller föreslå lösningar kring bristen på struktur och ramverk vid design av röstassistenter (Sutton et al., 2019; Sugisaki & Bleiker, 2020) och hur designers arbetar idag (Kim, McGrenere, Reza & Yoon, 2021; Murad, Munteanu, Cowan & Clark, 2021).

Det saknas även omfattande forskning kring risken för informationsöverbelastning och användarnas mentala modeller, hur de föreställer sig navigeringen och interaktionen med gränssnittet (Moore & Urakami, 2022). En studie har fokuserat på att testa ifall tydligare val och direktiv upplästa från röstassistenten skulle vara en lösning (Kirschthaler et al., 2020). En annan studie har försökt ta reda på hur användare visualiserar sina röstassistenter och hur det påverkar deras interaktioner (Lee, Cho & Lee, 2020). Hur användare egentligen använder sina VUI-enheter

och hur de lär sig navigera dem är ett av de mindre undersökta områdena (Diederich et al., 2022; Clark et al., 2019; Myers, Furqan, Nebolsky, Caro & Zhu, 2018).

En litteraturstudie indikerar att forskning inriktad mot feedback vid error har fokuserats kring den tekniska aspekten (Kim, Hwang, Kim & Lee, 2022) och hur kodning och maskininlärningsverktyg som NLP (Natural language processing), nätverk (både inom enheten och externt) och API (application programming interface) utvecklats (Baimirov et al., 2022). Litteraturstudien av Lee, Sah och Lee (2019) påvisar även väldigt lite forskning kring hur användare upplever error feedback och dess användbarhet. Bristen på hjälpmedel för att lära användare navigera VUI-system vid error feedback, och vilka kommandon som är möjliga att utföra är någonting Myers (2019) har poängterat i sin studie. Speciellt bristande är studier kring minoritetsspråk som även uppmärksammas ge högre grad av error och missuppfattning på grund av bristande NLP-verktyg och förståelse kring dialekter (Shafei & Tan, 2022). Det finns därför behov av att undersöka hur VUI presterar gentemot icke-engelsktalande användare, men även olika könstillhörigheter, för att inte riskera exkludering och barriärer för att använda dessa produkter (Shafei & Tan, 2022).

Hur mer välutvecklad och användarvänlig error feedback kan tas fram och erbjudas till användare av smarta högtalare behöver undersökas. Det finns därför en kunskapslucka att fylla som detta arbete ämnar adressera.

1.1 Syfte och frågeställning

Det behövs bättre förståelse för hur användarna kommunicerar med sina smarta högtalare och hur man på bästa sätt kan förmedla information tillbaka via röstoutput. Detta för att komma vidare i utvecklingen av röstassistenter och designa produkter som är anpassade för sina användare.

Syftet med studien är därför att undersöka hur svenska användare, som använder minoritetsspråket svenska, uppfattar bristen på error feedback från smarta högtalare. Genom uppsatsens undersökningar finns förhoppningen om att få en större förståelse för vad som är önskvärt från ett VUI-system som används genom

smarta högtalare. Hur användare interagerar med sina enheter i sina egna hem är speciellt av intresse då studier “in the wild” är någonting som efterfrågas.

Denna uppsats ämnar utifrån insamlad kunskap, i samarbete med tidigare forskning, att ta fram ett förslag på guidelines för hur error feedback skulle kunna designas och presenteras för användare.

Uppsatsens syfte har brutits ner till frågeställningen:

Hur skulle guidelines som hjälper designers vid framtagning av error feedback för VUI-system, med svenska användare, kunna se ut?

För att svara på denna frågeställning har uppsatsens undersökningar utgått från tre subfrågor:

1. Hur svenska användare uppfattar den error feedback de får av smarta högtalare idag?
2. Vilken feedback som är problematisk, samt varför?
3. Hur error feedback skulle kunna se ut i stället?

1.2 Avgränsningar

Avgränsningar har fastställts i syfte att bibehålla ett tydligt fokus genom uppsatsen, och därmed möjliggöra ett rikt resultat. Studien avgränsas därför till att enbart fokusera på error feedback för smarta högtalare som använder sig av Google Assistant på svenska. Empiriska studier i form av enkätundersökning, litteratursökning samt observationer och intervjuer, varav de sistnämnda genomfördes med ett bekvämlighetsurval, har genomförts. Litteratursökningen har avgränsats till tidigare forskning och vetenskapliga studier inom VUI och tvärvetenskapliga artiklar och forskning relevant för uppsatsen. Samtligt material har utifrån state-of-the-art med maxgräns på 5 år.

1.3 Disposition

Uppsatsen första kapitel inleds med en introduktion till ämnet och presenterar sedan vidare dess syfte och frågeställning. Under det första kapitlet behandlas

även uppsatsens begreppsdefinitioner, avgränsningar och dess disposition. Kapitel 2 innehåller en litteraturgenomgång där den senaste forskningen inom auditiv feedback, specifikt riktad mot error feedback, undersöks. Vidare undersöks även design av VUI-system idag och hur designers arbetar för att klara av övergången från synliga till osynliga gränssnitt. I kapitel 3 redogörs uppsatsens studier ingående, även innehållande urval och etiska dilemman. Därefter följer kapitel 4 som behandlar en beskrivning av studiernas resultat samt en analys av detta. Under kapitel 4 presenteras även de guidelines som utformats efter resultatet av studierna, samt deltagarnas återkoppling. Kapitel 5 innehåller en diskussion kring metoderna som använts samt resultatet. I det avslutande kapitel 6 presenteras uppsatsens slutsats.

2. Litteraturgenomgång

I litteraturgenomgången presenteras fyra huvudkategorier för uppsatsen: Auditiv feedback, Error feedback, Design av VUI och GUI till VUI. Tillsammans ger de en överblick kring den senaste forskningen inom de olika aspekter som uppsatsen ämnar undersöka: hur auditiv feedback kan uppfattas, vilken error feedback som kan vara önskvärd, hur design av VUI går till idag samt de utmaningar designers står inför.

2.1 Auditiv feedback

Inom VUI används ofta röstkommando och röstoutput för att beskriva interaktionen mellan användare och teknik. Denna dialog har försökt designas och tas fram för att eftersträva ett naturligt språk (Kim et al., 2021). Studier inom HCI har länge fokuserat på att försöka utforma dessa samtal mellan användare och teknik på ett sätt som liknar en naturligt förekommande konversation mellan två eller fler människor (Kim et al., 2021; Sugisaki & Bleiker, 2020). Ett viktigt ämne för en naturlig röstframtagning av VUI är auditiv feedback. Dialogen mellan användare och VUI är helt röstbaserad (auditiv) och utförs genom svar (feedback) mellan parterna. Det är därför av intresse att förstå även hur auditiv feedback fungerar mellan människor och vad detta kan lära oss för att designa bättre röstbaserade gränssnitt.

Nyare forskning kring auditiv feedback har riktat fokus mot användarupplevelser, bland annat där nya pedagogiska tillvägagångssätt introducerats i och med teknikens utveckling (Sarcona, Dirhan & Davidson, 2020). Man har undersökt hur studenter och lärare upplever dessa förändringar, och implementeringen av auditiv teknik för att förmedla studie-och provresultat (Jones & Hilton, 2022; Killingback, Ahmed & Williams, 2019; Sarcona et al., 2020). Vilka möjligheter auditiv feedback har och hur effektiv den är för de involverade är ännu oklart, men flertalet undersökningar har påvisat en övergripande positiv upplevelse. Man har noterat ökat engagemang från studenterna där de kommenterar hur auditiv feedback gett ökade känslor av tillhörighet och uppfattningen av mer social närvaro från lärarna (Killingback et al., 2019; Sarcona et al., 2020). Samtidigt

uppmärksammas det hur vissa studenter fann det svårt att följa de kommentarer som lärarna delgav i sina ljudinspelningar, och behövde lyssna upprepade gånger (Ice, Curtis, Phillips & Wells, 2019; Sarcona et al., 2020). Ice et al. (2019, s. 16) beskrev hur en student svarade kring auditiv feedback: ”Jag lyssnade två, ibland tre gånger för att verkligen förstå vad som sades.” I samma intervju berättade dock studenten varför hen ändå var positiv till auditiv feedback:

”Kanske är det så att eftersom jag känner mig mer delaktig genom auditiv feedback, som i ett fysiskt klassrum och hur vi diskuterar där, så kände jag inte nödvändigheten att ha allting som presenterades nedskrivet på det sätt jag bäst lär mig saker.” (Ice et al., 2019, s. 16)

Sarcona et al. (2020) kommenterade utifrån sina resultat att det vanligaste problemet kring auditiv feedback var behovet av upprepning. Studenterna klagade över att behöva spola tillbaka och höra om flera gånger för att lyssna på det som sades. De kommenterade även hur de upplevde det svårt att följa uppgifter medan de lyssnade. Trots de uppmärksammade negativa aspekterna kring auditiv feedback från studenterna visade sig ändå majoriteten positiva till auditiv feedback, även om de föredrog skriftligt underlag (Sarcona et al., 2020). De menade att auditiv feedback gav dem känslan att lärarna brydde sig om deras utbildning och hjälpte dem få en djupare förståelse kring vad som sades. Killingback et al. (2019) argumenterade i sin studie att det positiva resultatet och kommentarerna från studenterna inte enbart kan vara kring den alternativa förmedlingen av feedback; textbaserad till auditiv. De menar att innehållet i feedbacken också måste tas i åtanke. Precis som med skriftlig feedback så är all feedback enbart så bra som personen som ger den. Ett intressant exempel av reflektion som en student gör i studien av Ice et al. (2019) var:

”Jag transkriberade audiofilen ni skickade och läste sedan igenom texten. Jag tittade på den och lyssnade sedan på filen igen flera gånger. Vad jag insåg var att de var helt olika saker. Jag vet att det som sägs i audiofilen och vad jag transkriberade är likadana, men skillnaden var hur det sades. När jag läser transkriberingen finns inga markeringar på specifika ord. Vad jag uppmärksammade var hur det inte finns något sätt att förmedla informationen på samma sätt i text som i audio. Vi förlorar så mycket med det skrivna ordet och jag tror vi inte tänkt på det tillräckligt mycket.” (Ice et al., 2019, s. 14)

Varför majoriteten av studenter i olika studier ändå är positivt inställda till auditiv feedback menar Killingback et al. (2019) är vikten av samhörighet. De argumenterar för att känslan av tillhörighet inte ska undermineras. Att använda alternativa feedback-metoder som auditiv feedback kan uppfattas som mer humaniserat och därför hjälpa studenter känna tillhörighet. De menar att denna känsla kan vara av speciell vikt under de första årets studier där feedbacken är mer än bara informationen de behöver för att bli bättre på uppgifter. Den kan även ge det emotionella stödet som behövs för att anpassa sig till nya situationer. (Killingback et al. 2019)

Även Jones och Hilton (2022) argumenterade utefter sina resultat att en blandning av auditiv och visuell feedback ger bäst resultat. De menade att den visuella feedbacken i form av text gav en summering och instruktionsguide att följa, medans den auditiva feedbacken gav ett mer emotionellt stöd och ökade känslan av samhörighet mellan student och lärare. En student kommenterade:

“Jag kände också att feedbacken kändes mindre kritiserande och mer motiverande jämfört med hur skriftlig feedback ibland uppfattas. Du får möjligheten att höra röst och tonläge från din lärare vilket hjälper att förmedla informationen bättre”. (Jones & Hilton, 2022, s. 125)

Vidare argumenterade de att olika feedback-metoder påverkar när och hur feedbacken används. Auditiv feedback fyller uppgiften att positivt förstärka studentens upplevelse av inläring genom känslor av tillhörighet, närvaro och ökat självförtroende. Visuell feedback, å andra sidan, kan hjälpa till i nästa stadie vid djupare reflektion kring arbetet och framtida uppgifter. (Jones & Hilton, 2022)

Baserat på forskning kring auditiv feedback utförde Holland (2022) ett projekt med målet att undersöka ifall denna sorts feedback skulle vara applicerbar och uppskattad för besökare på LKS ASE (Library & Knowledge Service for NHS Ambulance Services in England). Genom att skicka ljudinspelningar som förmedlade information kring sökresultat via mejl till bibliotekets besökare undersöktes två huvudfrågor: *Förbättras användarnas förståelse av sökningsresultaten genom bibliotekskonsultation via auditiv feedback? Ökar användarnas känsla för social närvaro vid bibliotekskonsultation via auditiv*

feedback? Baserat på besökarnas svar uppmärksammade Holland en positiv attityd till auditiv feedback och ökad bekvämlighet att använda bibliotekets sökfunktioner. Även en högre detaljnivå av den feedback som gavs upplevdes, vilket kunde vara ett motargument till att använda skriftlig feedback som kan vara svårare att förstå. Genom projektets arbete tog Holland (2022) även fram ett antal guidelines baserat på tidigare forskning och projektets egen insamlade kunskap, varav några kan vara av intresse för röstbaserad gränssnittsframtagning:

- Säkerställ att nyckelinformationen presenteras i feedbacken.
- Håll informationen kort.
- Dialogen bör hålla en naturlig och saklig nivå, som mellan två vuxna individer.
- Erbjud användarna möjligheten att ge feedback tillbaka.
- Erbjud extra stöd till användare som behöver det.

2.2 Error feedback

Röstigenkänningsystem har idag inte tillräckligt sofistikerade funktioner för att förstå nyanser i dialektala variationer. Den dialekt som presterar bäst är amerikansk engelska, medan brittisk engelska eller australiensisk engelska kan uppbringa problem, och vara anledningen till problematiska dialoger snarare än grammatiska fel och ordval. Det finns inget bra sätt att mäta hur väl kommersiella produkter med röstigenkänningsystem idag presterar gentemot språkskillnader, så som dialekter. Inom HCI är det viktigt att arbeta för en bred användargrupp för att inte riskera exkludering av vissa personer. Studier kring användartester och presterande av smarta högtalare har nästan exklusivt fokuserat på engelsktalande användare. (Shafei & Tan, 2022)

Studier har uppmärksammat skillnader i hur väl kommandon till smarta högtalare presterar beroende på om användarna är engelsktalande eller icke-engelsktalande (Shafei & Tan, 2022). Även skillnader mellan manliga och kvinnliga användare av smarta högtalare har uppmärksammats. I en studie av Rajan, Udeshi och Chattopadhyay (2022) observerade man hur störande bakgrundsljud ökade risken för fel och missuppfattning vid kvinnliga användares kommandon. De fokuserade

dock enbart på utvärdering av smarta högtalare och erbjöd inga hypoteser till varför dessa skillnader fanns.

Det finns därför behov av att undersöka hur VUI presterar gentemot icke-engelsktalande användare, men även olika könstillhörigheter, för att inte riskera exkludering och barriärer för att använda dessa produkter. Svårigheten ligger i att rekrytera tillräckligt stora mängder användare med olika språk, dialekter och kön, som kan delta i studier för att förbättra utvecklingen av interaktionen med smarta högtalare. (Shafei & Tan, 2022)

Hur användare av VUI-enheter svarar på auditiv error feedback, samt tar den till sig och hur den påverkar dem behöver undersökas djupare (Lee et al., 2019). När problem uppstår i dialogen mellan användare och VUI-enhet sker detta oftast i två specifika faser; *igenkänningsfasen* eller *utförandefasen*. När en VUI inte kan tolka användarens röstkommando korrekt i NLP, till exempel inte känner igen ett ord eller hela meningar, och översätta detta till maskinella kommando tillhör detta igenkänningsfasen. Vid utförandefasen kan till exempel kontextuella begränsningar, såsom flera filer med samma namn och vilken VUI-enheten i så fall ska öppna, vara någonting som orsakar att VUI-enheten återger oväntade och irrelevanta svar till användaren (Lee et al., 2019). I deras artikel *Improving Usability Perception of Error-Prone AI Speakers: Elaborated Feedback Mitigates Negative Consequences of Errors* beskrev Lee et al. (2019) att användare anser att igenkänning är fundamental för VUI-system. Eftersom denna anses som en grundläggande funktion påvisar igenkänningsfel teknisk inkompetens från systemets sida enligt användare. Att utföra användares röstkommando däremot sågs istället som en komplex uppgift som kräver AI och ribban var därför lägre för användarnas förväntningar och deras utvärdering av VUI'ns användbarhet. För att motverka de negativa konsekvenserna av felaktiga svar eller problem med uppfattning av röstkommando föreslog Lee et al. (2019) att VUI-system ska designas för att hantera och kommunicera med användaren när fel uppstår.

Även studier kring hur användare i nuläget hanterar problemen som uppstår vid feltolkning av kommandon är begränsad. Myers et al. (2018) utförde en empirisk studie kring hur användare interagerade med deras eget framtagna VUI-system,

DiscoverCal, under tre sessioner. De identifierade tio övergripande mönster som deras användare uppvisade i sina interaktioner med VUI-systemet:

1. **Överdriven artikulering:** användaren pratade högre och/eller långsammare och försökte artikulera tydligt sitt kommando.
2. **Förenkling av kommando:** användaren försökte förkorta och förenkla sina kommandon genom att ta bort ord.
3. **Variation i ordval:** användaren testade olika ord för att nå samma mål med sitt kommando.
4. **Mer information:** användaren gav mer komplexa meningar i tron att mer information hjälpte systemet förstå vad de ville uppnå med sitt kommando.
5. **Förlita sig på GUI:** användaren tittade på DiscoverCal's tillhörande GUI för direktiv och pratade högt för sig själv.
6. **Förlikning:** användaren nöjde sig och accepterar felet efter flera försök att korrigera.
7. **Börja om från början:** användaren fick inte kommandot att utföras fullkomligt korrekt och beslöt sig för att starta om från början igen.
8. **Frustrations försök:** användaren försökte sig på flera olika taktiker i snabb följd på grund av frustration. Inte mycket tanke lades bakom varje kommando.
9. **Ge upp:** användaren avslutade sitt försök att få VUI-systemet att utföra kommandot och lyckades således inte med sin uppgift.
10. **Erinra sig:** användaren kom ihåg det korrekta sättet att säga kommandot genom att tänka högt för sig själv och därmed reflektera kring tidigare interaktion.

Av dessa mönster var den vanligaste att användare började använda överdriven artikulation och/eller gissade sig fram när de navigerade kring error feedback, snarare än att förlita sig på visuella hjälpmedel eller tidigare kunskap. (Myers et al., 2018)

För att fortsätta arbetet att utveckla och förbättra VUI-enheter föreslog Myers (2019) ett mer adaptivt system där tekniken fokuserar på att lära användare hur de

ska formulera sig och interagera med VUI-enheter. Detta genom en inbyggd funktion som uppfattar en användares vana att interagera med VUI's baserat på deras sätt att formulera sina kommandon. Om en användare uppfattas ha problem ska programmet automatiskt ge förslag på mer utförliga kommandon för användaren att testa.

Trots teknikens framsteg är det dock fortfarande av intresse att förstå hur användare interagerar och arbetar med nuvarande röstigenkänningsystem, vilka problem de upplever och hur de försöker arbeta runt dem (Diederich et al., 2022; Clark et al., 2019; Myers et al., 2018).

2.3 Design av VUI

Konversationsgränssnitt har studerats sedan 1950-talet, men trots detta finns många frågor kring hur dessa ska designas och optimeras för en naturlig dialog (Kim et al., 2021; Sugisaki & Bleiker, 2020). Röstassistenter utan visuell representation så som Google Assistant, Amazon Alexa och Apple Siri är några av dessa konversationsgränssnitt som användare är bekanta med och de senaste 5–10 åren har forskningen ökat gällande användarvänligheten kring VUI's (Murad & Monteanu, 2022). Utmaningarna kring dess design för användbarhet och skillnad mellan prestation och användares förväntningar existerar fortfarande (Sugisaki & Bleiker, 2020). Cambre och Kulkarni (2020) uppmärksammade hur dagens VUI-enheter enbart stödjer fasta konversationsmönster där användare och system turas om att prata. Detta är inbakat i både Google Assistant och andra VUI's infrastrukturer. Systemen anser en konversation är färdig om de inte får ett svar från användaren inom 5–8 sekunder, någonting som inte speglar en naturlig dialog mellan två eller flera människor (Cambre & Kulkarni, 2020). Ett annat problem som identifierats av användbarhetsexperter och diskuteras av Murad et al. (2021) är VUI's minneskapacitet och förståelse av kontext kring tidigare röstkommando. Ingen av de använda enheterna hade något minne av tidigare kommandon. Detta gör att man inte kan be om information eller ställa frågor baserat på vad som tidigare sagts. Minnesproblemet är någonting som uppmärksammades redan 2018 i en studie av Wei och Landay.

De allmänna reglerna för framtagning av användarcentrerad design (Nielsen 1994; Norman, 2002) applicerade för grafiska gränssnitt (GUI) fungerar inte att översätta direkt till röst användande gränssnitt (Murad et al., 2021). Eftersom de dessutom är väldigt övergripande tvingar det designers att tolka och använda dem utefter deras bästa förmåga, baserat på deras egen kunskap och expertis. Detta gör att framtagningen av röstassistenter och dess design är väldigt inkonsekvent (Seaborn & Urakami, 2021; Sugisaki & Bleiker, 2020). Enligt undersökningar gjorda av Clark et al. (2019) finns en brist på kvalitativ forskning och designorienterade undersökningar för att skapa standarder för pålitliga, tillämpliga och konsekventa utvärderingsverktyg. Utvecklingen av röstspecifika designprinciper, guidelines eller ramverk för framtagning av VUI's, med välfungerande verktyg och metoder som tillåter designers att utveckla bättre och mer konsekventa röstassistenter, behöver därför tas fram (Cambre & Kulkarni, 2020; Kim et al., 2021).

Ett problem som uppdagades under designernas arbete för att skapa en naturlig dialog mellan användare och teknik var fenomenet *The Uncanny Valley effect*. Genom att försöka skapa VUI's som kunde tyda sociala signaler och tala med ett mer naturligt språk uppstod en oväntad konsekvens. Användare uppvisade större motstånd kring interaktionen med systemet när de försökte framstå som mänskliga, men inte lyckades helt. Att härma en riktig människa verkade förstärka negativa känslor kring problem vid utförande och error feedback, något som visade motsatt effekt vid interaktion med en mer robotliknande röst. (Diederich et al., 2020)

2.4 GUI till VUI

Interaktionsdesigners inom industri som tar fram nya VUI's för smarta högtalare eller implementerar dem som ett alternativ till grafiska hemsidor saknar ofta relevant kunskap eller formell träning (Murad et al., 2021). Kurser och program för Human Computer Interaction har inte hunnit komma i kapp med undervisningen vad det gäller VUI som nytt grafiskt gränssnitt att designa för. Många tvingas därför använda kunskap från framtagning av GUI och försöka anpassa dem utefter egen bästa förmåga (Kim et al., 2021; Murad & Monteanu,

2022). I undersökningar har flera designers svarat att de ofta tar inspiration från tidigare VUI's, onlinesökningar eller diskuterar med kollegor när de ska ta fram nya röstbaserade gränssnitt (Murad & Monteanu, 2022).

Bristen på verktyg och metoder för tydliga ramar och riktlinjer kring hur gränssnitten bör skapas kan vara anledningen till att många brister återkommer och implementeras i även nyare versioner av VUI's. Även tester för gränssnittets användbarhet är mer sannolika att ske för grafiska gränssnitt med användartester, intervjuer och enkäter. Detta skulle kunna vara eftersom undersökningar påvisar att företagen själva som beställer röstbaseradegränssnitt ofta värderar GUI design över VUI's. (Murad & Monteanu, 2022)

De flesta VUI-enheter idag är beroende av molntjänster tillhandahållna av till exempel Amazon, Google eller Apple (Yan et al., 2021), vilka levereras i färdiga paket där den interna informationen inte är tillgänglig (Alrumayh & Tan, 2022). Därför är det även svårt för utvecklare att få tillgång till information kring fel och användares upplevelser av tjänsten (Alrumayh & Tan, 2022). Försök till ramverk för att designa samt utvärdera interaktiviteten med VUI-enheter har därför tagits fram för att försöka hjälpa designers och utvecklare, men inget är ännu standardiserat (Alrumayh & Tan, 2022; Seaborne & Urakami, 2021; Sutton et al., 2019; Sugisaki & Bleiker, 2020).

Murad et al. (2021) undersökte hur man på bästa sätt kan underlätta för designers vid övergången av gränssnitt genom workshops med grafiska designexperter, och föreslog lösningen att ta "the path of least resistance". Detta innebär enligt Murad et al. (2021) att adoptera och anpassa familjära tekniker och metoder från GUI heuristik till VUI-design i stället för att fokusera på att skapa helt nya heuristiska regler från början. Det har dock tagits fram förslag på VUI heuristik av Wei och Landay i artikeln *Evaluating Speech-Based Smart Devices Using New Usability Heuristics* (2018) som validerats och testats av användbarhets experter, och dessa fortsätter vara aktuella och citeras i state-of-the-art forskning.

3. Metod

Studien som presenteras nedan har utförts för att ge underlag till att skapa guidelines för designers att ha i åtanke vid design av error feedback inom VUI. För att åstadkomma detta har olika metoder använts; enkätundersökning, litteratursökning, deltagande observation samt intervjuer.

En enkätstudie skapades som förundersökning för att ge en indikation på problemområden som användare upplevde i sin användning av Google Assistant. Svaren på förundersökningen låg sedan till grund för uppsatsens riktning kring syfte och frågeställning.

Deltagande observation (Denscombe, 2018) valdes för att få en inblick i hur deltagarna interagerade med sina smarta högtalare. Detta oberoende av deras egen uppfattning av vad de säger och gör. Genom att observera deltagarnas interaktion gick det att uppfatta vilka problem som uppkom, hur dessa hanterades och deltagarnas reaktioner. Observationen kompletterades med personliga intervjuer som gav deltagarna möjligheten att svara med egna ord på öppna och stängda frågor relaterade till observationen. Genom att ta reda på åsikter, erfarenheter och känslor relaterade till deltagarnas interaktion med sina smarta högtalare blev det möjligt att få en djupare inblick i deras upplevda problem med error feedback. Då författarna till studien och deltagarna befann sig på olika platser hölls intervjuer och observationer både i verkligheten och via videosamtal. Observationerna och intervjuerna kombinerades även med litteratursökning för att ta reda på vad som redan var beforskat inom det utsedda området.

3.1 Enkätundersökning

En förstudie skapades genom enkätundersökning (Patel & Davidson, 2019) (se bilaga 1) för att lyfta fram en avgränsad mängd svenskars egna erfarenheter och uppfattningar kring Google Assistants prestation. Enkäten skapades genom Google Forms som även erbjöd sammanställning av resultatet genom diagram och tabeller. Resultatet exporterades även vidare till Microsoft Excel för att lättare kunna bearbetas och analyseras, där urskiljning av mönster och återkommande problem för deltagarna kunde uppmärksammas. Enkätundersökningen skickades

ut i en Facebook grupp med namnet “Google Home Sverige”, gruppen har över 12,000 medlemmar och är avgränsad till svenska användare. Innan enkätundersökningen skickades ut till målgruppen, skickades en pilotundersökning (Denscombe, 2018) ut för att säkerställa att respondenterna förstod frågorna och hur de skulle svara. Deltagarna i pilotundersökningen hade ingen tidigare vetskap om uppsatsens innehåll, men använde Google Assistant på svenska. Enkätundersökningen bestod av ett frågeformulär och utformades dels av klickbara fasta frågor (Denscombe, 2018), samt ett fåtal öppna frågor med kommentarsfält för att låta dem uttrycka sig med egna ord (Patel & Davidson, 2019).

Frågorna skapades först med bekräftelse kring användning av Google Assistant på svenska och fortsatte sedan med könsidentitet, ålder och länstillhörighet. Därefter fokuserade frågorna kring deltagarnas användning av Google Assistant, hur ofta de använde den och genom vilken enhet. Vidare efterfrågades information kring installation, första användning och eventuella hjälpmedel som behövts. Frågorna avslutades med flervalsalternativ samt kommentarsfält för undersökning av specifika problem och förbättringsförslag som deltagarna upplevt eller efterfrågade.

Under utformningen togs respondenternas svarsbörda i beaktning och därför hölls enkätens längd till ett minimum för att minimera risken att respondenterna gav upp och inte fullföljde hela undersökningen (Denscombe, 2018). Genom att ha ett fåtal öppna frågor med och en större mängd fasta frågor minimerades även respondenternas mentala ansträngning. För att säkerställa respondenternas motivation till att svara på formuläret beskrevs även hur lång tid det förväntades att ta, samt att svaren skulle användas för denna uppsats. Eftersom undersökningen lades upp i en Facebook grupp gavs bakgrundsinformation om arrangör, syfte och frivillighet redan i Facebookinlägget. Ytterligare information gällande konfidentiell hantering samt instruktioner till respondenten uttrycktes i introduktionsdelen för enkäten. (Denscombe, 2018)

Undersökningen togs även fram med hänsyn till användares integritet genom att vara helt anonym (Denscombe, 2018), men inkluderade kön, ålder och län. Dessa

ansågs vara viktiga för att urskilja eventuella skillnader och likheter i användning och uppfattade problem över kön, dialekt och ålder (Shafei & Tan, 2022). Länstillhörighet samlades in för att se eventuella dialektala konflikter och även var majoriteten av enkätens deltagare bor. Förstudiens deltagarantal uppgick till 44 personer, vilket inte är ett signifikant antal och inte heller har stor statistisk grund. Önskan att nå ut till betydligt fler deltagare fanns men lyckades inte. Inga personliga påminnelser skickades ut till personer som inte svarat på undersökningen. I stället användes Facebooks algoritmer för att synliggöra inlägget med enkäten så mycket som möjligt via kommentarsfält och reaktionsmarkeringar. Med detta i åtanke håller inte förstudiens resultat mer värde än att ge en indikation på vad vissa svenskar upplever problematiskt med kommunikationen mellan användare och Google Assistant.

3.2 Litteratursökning

Litteratursökningen startade med en bred sökning inom ämnet Voice User Interface. Detta gjordes för att inte riskera att missa relevant material och för att sätta sig in i vad senaste forskningen fokuserade på. Genom att förstå forskning även utanför ”VUI” tillsammans med ”error feedback” kunde en större förståelse skapas och kunskapen som samlades in kunde således hjälpa till med att ta flera aspekter i åtanke.

Sökmotorn HKR Summon användes tillsammans med databaserna ACM Digital Library och IEEE Xplore. Även sidan ResearchGate som tillåter vetenskapsmän och forskare att dela artiklar och forskning användes för att leta relevant forskning och ge djupare kunskap inom ämnet VUI.

I HKR's sökmotor Summon filtrerades resultaten till att vara vetenskapligt granskade och med tidsramen på max 5 år för att verifiera att forskningen var state-of-the-art. För att säkerställa artiklarnas relevans för informatik-området uppmärksammades även de tillhörande ämnesord som kopplades till artiklarna. Några av de mest frekvent uppmärksammade var ”human-computer-interaction”, ”voice user interface”, ”conversational agent”, ”speech user interface”, ”voice assistants” och ”human-centered-computing”.

Artiklar funna på Researchgate eller i databaserna IEEE Xplore eller ACM Digital Library söktes även upp genom HKR Summon via artikelrubrik för att säkerhetsställa att de var markerade som vetenskapligt granskade (peer-reviewed).

De initiala sökningarna resulterade i en stor mängd material att sortera igenom och detta skedde genom manuell granskning. Fritextorden valdes ut efter frågeställningen och synonymer eller närliggande ord användes och kombinerades (se bilaga 2). Granskningen skedde med uppsatsens frågeställning och syfte i åtanke. Materialet som sparades efter de första sökningarna valdes ut efter abstrakt och sökord, där de som uppfattades kunna ha relevant information togs vidare för granskning. Även artiklar riktade specifikt mot Google Assistant undersöktes och sparades ifall dessa ansågs relevanta.

Granskningen för de utvalda artiklarna skedde därefter genom att läsa igenom introduktionsdelen och delar av bakgrunden som artikeln byggde på. För att spara tid gavs metod och analysdelen enbart en mer övergripande blick medan resultat och diskussion lästes igenom. Utifrån detta togs beslutet ifall artikeln skulle sparas för mer djupgående läsning eller avfärdas som irrelevant för denna uppsats.

De artiklar som sparades efter granskningen kategoriseras i tre övergripande områden för en bättre överblick; Fokus mot designerns arbete, Fokus mot användare och Fokus mot teknik. Artiklarna inom varje område lästes därefter metodiskt igenom och information relevant för uppsatsen skrevs ner tillsammans med referensen. De som inte ansågs relevanta avfärdades.

Den insamlade informationen kunde sedan arbetas ihop till ett första utkast för litteraturgenomgången och ge en överblick över vad för information som saknades eller behövde vidareutvecklas. Det gav även ett bra stöd när en smalare sökning startades med fokus mot ”error feedback” och ”auditiv feedback”.

Den fokuserade litteratursökningen riktad mot ”error feedback” gav få resultat, med majoriteten fokuserad kring tekniska aspekter och brister i nuvarande mjukvaruprogram och system. Sökningen för ”auditiv feedback” gav ett rikare material, men state-of-the-art forskningen fokuserade mot auditiv feedback mellan

två eller flera människor. Några av artiklarna inom “auditiv feedback” valdes ut för dess uppfattade nytta i relevans till Voice User Interfaces. Dessa fokuserade på asynkron kommunikation och auditiv feedback skickad mellan studenter och lärare eller ett bibliotek och dess besökare. Deras relevans till uppsatsen ansågs finnas i upplevelsen som användarna, vilka mottog den auditiva feedbacken, upplevde och paralleller till VUI-enheter och dess användare kunde dras. Den fullständiga slutgiltiga sökstrategin bifogas som bilaga (se bilaga 3).

Samtliga av de relevanta artiklars egna referenser undersöktes och kunde på så sätt leda till vidare utforskning. Detta skedde inte genom någon specifik sökning utan var en kedjereaktion av artiklar refererande och kopplade till varandra. Även vissa författare som återkom undersöktes och deras publicerade material lästes igenom. Detta resulterade i en djupare kunskap inom området och uppmärksammade brister kring utforskning av error feedback till användare av andra forskare såsom Lee et al. (2019) samt Shafei och Tan (2022). Även hur användare interagerar med sina smarta högtalare, vilka problem som uppstår och hur dessa hanteras behöver undersökas ytterligare (Murad et al., 2021; Kirschthaler et al., 2020). Denna kunskap låg till grund för att utforma de teman som användes för utformningen av observationsuppgifterna och intervjuerna som följde.

Förstudien kombinerat med litteratursökningen kunde således stå som grund till uppsatsen område och syfte, och tillsammans med resterande metoder resulterade i ökad förståelse inom området Voice User Interfaces.

3.3 Deltagande observation

För att studien skulle hålla så hög kvalitet som möjligt och på så sätt kunna inbringa rik kvalitativ data valdes deltagande observation som en av metoderna. Genom denna metod möjliggjordes det för författarna att vara med i rummet (både i verkligheten och i videosamtal genom tjänsten Zoom) och observera användarna i brukskontexten. Detta för att kunna bevara den naturliga miljön och kunna fokusera på meningen bakom de handlingar som genomfördes (Denscombe, 2018). Samtliga fem deltagare som blev tillfrågade om de ville ställa upp på en

observationsstudie samt intervju, och tackade ja, informerades om författarnas roll som deltagande observatörer (Denscombe, 2018). Informationsinsamlingen skedde genom anteckningar från båda observatörerna. Tidsåtgången som beräknades var cirka 15 minuter per observation, med ett tillfälle per deltagare. Genomgående för de fem observationerna, oavsett om de hölls genom videosamtal eller i verkligheten, var att det endast var deltagaren, den smarta högtalaren och observatörerna i rummet.

Syftet med observationen var att i verkligheten observera vilka kommunikationsproblem som uppstod mellan deltagaren och deras smarta högtalare. Vidare var det verkligt intressanta att se vilken sorts feedback som gavs från den smarta högtalaren och hur deltagaren agerade efter det. Således utformades uppgifterna till deltagarna (se bilaga 5) utifrån de aspekter som uppmärksammats av tidigare forskning och de brister på kunskap som finns idag (Kim et al., 2022; Kirschthaler et al., 2020; Lee et al., 2019; Murad et al., 2021; Shafei & Tan, 2022).

För att kunna jämföra resultatet av de olika deltagarnas observationer var det av vikt att skapa så lika situationer som möjligt utan att observatörerna lade sig i under tiden. Detta för att öka chansen att deltagarna interagerade med sina smarta högtalare på samma sätt som de skulle gjort när de var själva. Genom att bibehålla den naturliga miljön i så hög mån det gick, möjliggjordes det att observatörerna kunde få så tillförlitligt resultat som möjligt. Förutom detta var det även av vikt att inte störa röststyrningen, någonting som kunde skett om observatörerna också pratade.

3.4 Intervju

Den personliga intervjun som genomfördes höll en låg grad av standardisering (Patel & Davidson, 2019) och var därför semistrukturerad. Detta för att kunna anpassas till den enskilda respondenten, men de teman för frågor som tagits fram, *användningsområden, kommando till smarta högtalare och error feedback från Google Assistant*, var samma för samtliga respondenter (se bilaga 6). De teman som togs fram användes för att komplettera uppgifterna i observationen. Det

skapade en möjlighet att förstå varför deltagarna reagerade på olika sätt vid error feedback, samt hur de upplevde och hanterade den error feedback de fick från sina smarta högtalare. Även dessa teman baserades på de aspekter som uppmärksammats av tidigare forskning och de brister på kunskap som finns idag (Kim et al., 2022; Kirschthaler et al., 2020; Lee et al., 2019; Murad et al., 2021; Shafei & Tan, 2022). I och med den låga graden av standardisering gavs deltagarna stor frihet att styra sina svar, samt leda in samtalet på nya vägar. Listan med teman gav även möjligheten att hålla konversationen inom området. Intervjuerna var beräknade att ta cirka 15-20 minuter per deltagare och genomfördes en gång direkt efter observationen, både via videosamtal genom tjänsten Zoom och i verkligheten. Informationsinsamlingen skedde genom anteckningar från båda intervjuerna.

Syftet med intervjuerna var att fördjupa förståelsen för respondenternas upplevda kommunikationsproblem med sina smarta högtalare. Detta var också anledningen till att intervjuerna hölls direkt efter att observationerna genomförts, så att deltagarna hade upplevelsen färskt i minnet. Genom intervjuerna gick det att separera idéer och tankar till enskilda respondenter (Denscombe, 2018), vilka sedan kunde jämföras med observationerna för att få en fördjupad förståelse av problemen som uppstod.

3.5 Deltagare och urval

Fem deltagare valdes ut genom bekvämlighetsurval (Patel & Davidson, 2019) och medverkade i observationen och intervjun. Bekvämlighetsurval valdes som urvalssmetod då det var svårt att komma i kontakt med personer som ville ställa upp i studier och tidsbrist kring att söka efter deltagare utanför författarnas närhet. Deras åldrar varierade mellan 24 år till 35 år och föll därför i kategorin benämnd "millennials" eller Gen Y. Det viktigaste kriteriet för deltagarna var att de ägde och använde en smart högtalare med Google Assistant på svenska. Övriga kriterier som författarna försökte hålla var att hitta deltagare från olika län. Detta för att undersöka ifall olika dialekter kunde uppmärksammas ha påverkan på Google Assistants svar och error feedback. Även kriteriet att ha en jämn fördelning på könsidentitet försökte hållas, detta för att undersöka ifall olika kön

kunde ha någon påverkan på Google Assistants svar och error feedback. De fem deltagarna som valdes ut för studierna uppfyllde samtliga kriterier, där ingen deltagare från samma län hade likadan könsidentitet. Deltagarna använde Google Assistant på svenska och var tillgängliga de dagar och tider som var avsedda för studierna som skulle utföras. Samtliga tackade ja till att medverka under anonymitet.

- **Deltagare 1 kommer benämnas som Maria**

Deltagare 1 är kvinna, 35 år och är från Skåne län. Har använt Google Assistant i smart högtalare sporadiskt i cirka 1 år. Hon studerar på högskola och jobbar extra som säljare på en matvarubutik. Är van vid att testa och använda ny teknik. Maria deltog i studien digitalt via videosamtal genom tjänsten Zoom.

- **Deltagare 2 kommer benämnas som David**

Deltagare 2 är man, 24 år och är från Blekinge län. Har använt Google Assistant i smart högtalare aktivt i cirka 5 år. Han arbetar som IT-ansvarig på låg- och mellanstadieskola, och har därför stor teknikvana. David håller sig uppdaterad inom ny teknik och tycker om att testa nya produkter. David deltog i studien digitalt via videosamtal genom tjänsten Zoom.

- **Deltagare 3 kommer benämnas som Kajsa**

Deltagare 3 är kvinna, 24 år och är från Östergötlands län. Har använt Google Assistant i smart högtalare sporadiskt i cirka 3 år. Hon studerar på universitet och jobbar extra som modell, och använder teknik dagligen men har inget större intresse för det. Kajsa deltog i studien digitalt via videosamtal genom tjänsten Zoom.

- **Deltagare 4 kommer benämnas som Pelle**

Deltagare 4 är man, 32 år och är från Skåne län. Har använt Google Assistant i smart högtalare i cirka 3 år. Han arbetar som VVS-montör och har intresse för maskineri och teknik. Pelle är intresserad av ny teknik och tycker om att testa nya produkter och vad de kan göra. Pelle deltog i studien som utfördes fysiskt på plats i hans hem.

- **Deltagare 5 kommer benämnas som Sara**

Deltagare 5 är kvinna, 30 år och är från Småland län. Har använt Google

Assistant i smart högtalare sporadiskt i cirka 3 år. Hon arbetar som byggnadsingenjör och har datavana, däremot är hon inte speciellt intresserad av att lära sig eller använda mer teknik än vad som är nödvändigt. Sara deltog i studien som utfördes fysiskt på plats i hennes hem.

3.6 Genomförande

Observationerna grundades på de uppgifter som tagits fram för deltagarna vilka baserades på forskning (Kim et al., 2022; Kirschthaler et al., 2020; Lee et al., 2019; Murad et al., 2021; Shafei & Tan, 2022). Deltagarna blev kontaktade av författarna och tillfrågade om de ville ställa upp på en observation och en intervju. När deltagarna tackade ja till att delta fick de ett PDF-dokument med information om både observation och intervju, samt de uppgifter som tagits fram (se bilaga 4 och 5) så att de skulle kunna förbereda sig. Två deltagare utförde observation och intervju i verkligheten, medan tre utfördes via videosamtal genom tjänsten Zoom. Den överenskomna tidslängden på observation och intervju tillsammans var cirka 30 minuter med reservation för att det skulle kunna bli något längre eller något kortare.

Utförandet av observationerna började med att en av författarna informerade deltagaren om hur upplägget såg ut och bekräftade att deltagaren inte hade några obesvarade frågor. Förutom detta såg de även till att presentera sig och berätta om undersökningens syfte och mål, detta för att skapa en så stor tillit och avslappnande dynamik som möjligt (Denscombe, 2018). Förutom detta bekräftades även att högtalaren var inom hörhåll för observatörerna, alternativt nära datorn vid videosamtalen. Deltagarna fick sedan påbörja att lösa uppgifterna. De hade fått anvisningar om att testa så många gånger de ville om uppgiften inte löstes direkt, samt att de fick hoppa över uppgifter om de inte ville utföra dem. Under tiden tog observatörerna anteckningar. En av observatörerna fokuserade på att anteckna vad som hände tekniskt, om kommandot fungerade eller inte, medan den andra fokuserade på deltagarens sinnesstämning och hur kommandot uttrycktes. När deltagaren hade genomfört alla uppgifter avslutades observationen och intervjun påbörjades.

Intervjun inleddes direkt efter att sista kommandot genomförts på observationen. Detta för att hålla upplevelsen färsk i minnet. Därför började intervjun med öppna frågor om hur observationen kändes, och om det var något specifikt de reagerat på. De öppna och generella frågorna var till för att få en avslappnande stämning där deltagarnas tankar och åsikter var i fokus. Detta gjordes i ett försök att färga deltagarna så lite som möjligt efter författarnas eventuella egna uppfattningar. Under samtalets gång blev frågorna mer specifika inom error feedback, vad de tyckte fungerade, vad de skulle vilja se förbättrat och liknande. Eftersom intervjun följde en semistrukturerad form har även sufflering, uppföljning och kontroll använts för att få värdefull information (Denscombe, 2018). Den avslutande delen av intervjun återgick till neutrala och öppna frågor där deltagaren kunde kommentera kring både observation och intervju, samt lägga till sådant som inte tagits upp (Patel & Davidson, 2019). Även i intervjun togs anteckningar, där en av intervjuarna ställde frågor och den andra antecknade svaren. I bearbetningen för anteckningarna gällande både observation och intervju skrevs också en sammanfattning för att få en enklare överblick över vad som sagts och gjorts.

3.7 Analysmetod

Enkätundersökningen skickades ut som en förundersökning för att skapa en uppfattning om vad användare av Google Assistant på svenska upplevde. Enkätundersökningen skapades i Google Forms och exporterades till Microsoft Excel för att skapa en tydligare överblick och även ge möjligheten att urskilja mönster och liknande svar från olika användare. Tillsammans med litteratursökningen som gjordes parallellt kunde error feedback från Google Assistant urskiljas som det mest uppmärksammade problemet.

Observationerna sammanställdes i ett Excel-dokument, där varje uppgift med tillhörande anteckningar för varje deltagare radades upp bredvid varandra. Detta gav en tydlig överblick och underlättade arbetet att analysera resultatet från de olika observationerna. De mönster som uppmärksammades kunde sedan markeras och undersökas gentemot svaren som gavs i intervjun som hölls efteråt.

Anteckningarna från intervjuerna sammanställdes i ett Word-dokument och sorterades under respektive deltagare och de svar som givits. Dessa undersöktes sedan och markerades ifall användare gav liknande svar eller någonting skiljde sig åt. De olika svaren sammanställdes även under olika teman;

Användningsområden, Kommando till smarta högtalare och Error feedback från Google Assistant, och kopplades ihop med observationerna för de uppgifter som utfördes.

Utifrån resultaten av observationerna och intervjuerna valdes några dialoger mellan deltagare och Google Assistant ut, där specifikt intressanta problem uppstått för uppsatsens syfte. Dialogerna illustrerades och exporterades som figurer för uppsatsen. Detta arbete gav ytterligare insikt kring hur de olika deltagarna formulerade sina kommandon och vilka svar Google Assistant gav. Sammanställningen av resultaten från observationerna och intervjuerna i uppsatsen gjordes som figurer med dialog eller löpande text. Eftersom vissa intervjufrågor baserades direkt på observationer som gjordes, valde författarna att lyfta även dessa i resultaten av intervjuerna då de ansågs relevanta för att förstå kontexten.

3.8 Etiska överväganden

Undersökningarna har tagit hänsyn till informationskrav, samtyckeskrav, konfidentialitetskrav och nyttjandekrav och delgavs till samtliga användare (Patel & Davidson, 2019). Även Vetenskapsrådets riktlinjer kring god forskningssed och hur sekretess, tystnadsplikt, anonymitet och integritet ska hanteras (Vetenskapsrådet, 2017) har följts.

Allt insamlade av material behövdes tas fram för att säkerställa att ingenting kan återkopplas till enskilda individer och avslöja deras identitet. Materialet sparades av undersökande parter i en gemensam Google Drive och delades inte med utomstående. Kön, ålder och länsstillhörighet anonymiserades inte, detta för att hänvisa till att det insamlade materialet är kopplat till den tilltänkta målgruppen. Detta var någonting som tydligt delgavs till deltagarna innan studien påbörjades.

4. Resultat och analys

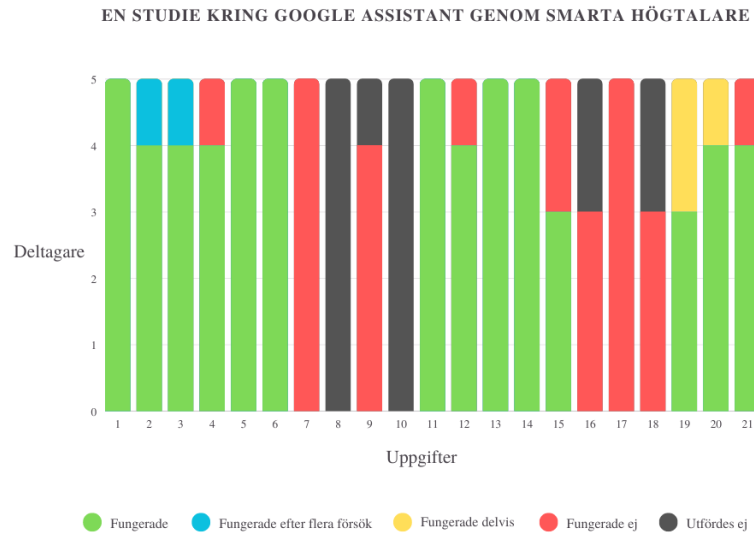
I denna del presenteras resultat och analys av de empiriska studierna, utförda genom observation och intervju. Det material som presenteras relaterar till både litteraturgenomgång och uppsatsens syfte; *att undersöka hur svenska användare uppfattar error feedback från smarta högtalare, vilken feedback som är problematisk och hur error feedback skulle kunna se ut istället*. Resultatet från båda studierna presenteras och analyseras mot litteraturgenomgången för att slutligen resultera i ett förslag av guidelines för designers vid framtagning av error feedback för VUI-system.

4.1 Resultat Enkät

Enkätundersökningen skapades som en förstudie och var därför enbart relevant i förhållande till vilken riktning uppsatsen tog. Genom deltagarnas svar kunde ett mönster gällande problematik om error feedback från Google Assistant urskiljas. 22 av deltagarna uppgav att “brist på utförlig feedback vid felaktigt kommando” var det mest problematiska. Detta uppmärksammades således vara det största problemområdet som kunde identifieras genom förundersökningen. Resultatet av förundersökningen låg sedan till grund för uppsatsens vidare studier och framtagning av syfte och frågeställning.

4.2 Resultat observation

Figur 1 nedan presenterar hur framgångsrikt de olika uppgifterna (se bilaga 5) utfördes genom ett stapeldiagram.



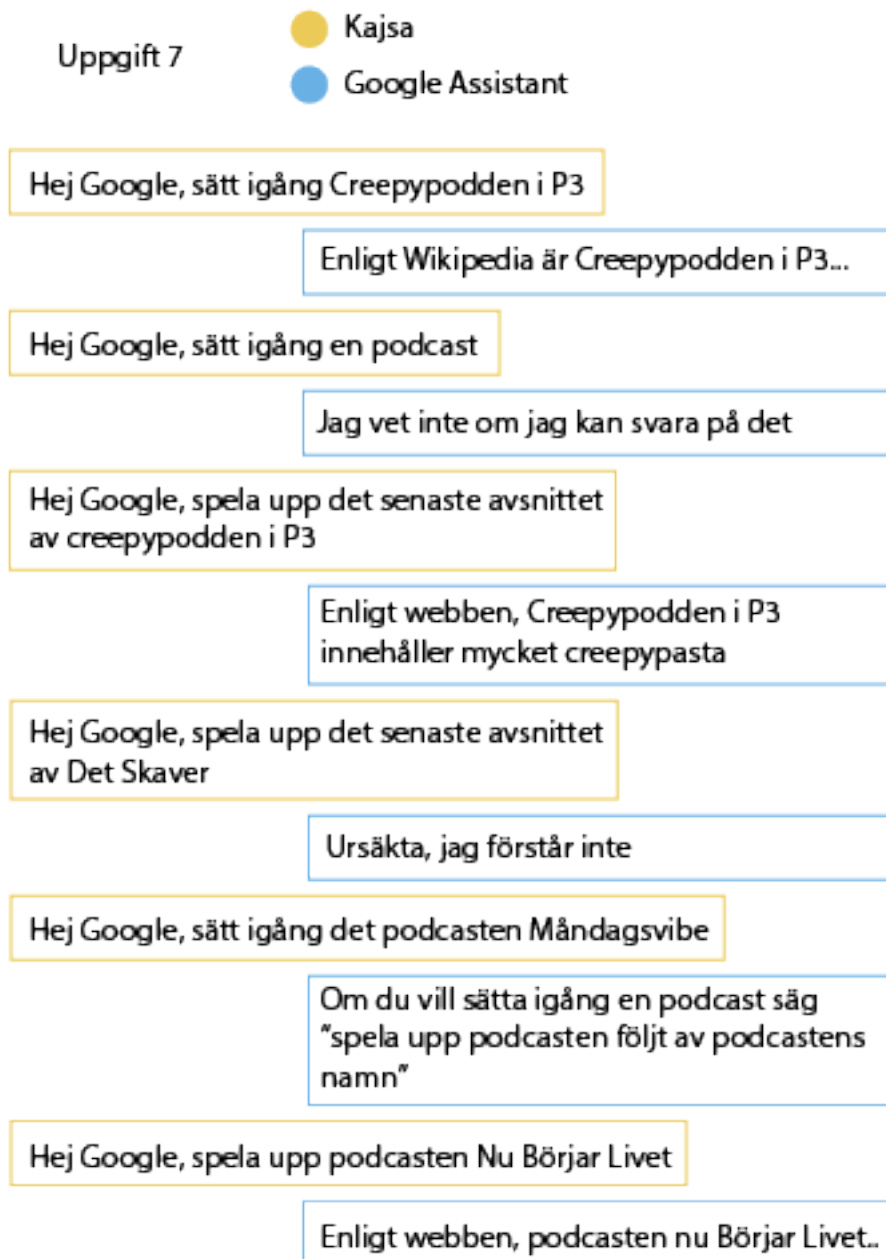
Figur 1: Stapeldiagram över deltagarnas utförande av observationsuppgifter

Uppgifter nummer 1, 5, 6, 11, 13 och 14 utfördes felfritt på första försöket av alla deltagare och deras smarta högtalare, medan uppgifter nummer 2 och 3 löstes på andra försöket. Nummer 8 och 10 utfördes inte av någon av deltagarna då de var följduppgifter från de föregående. För detta arbete ligger det intressanta däremot i de uppgifter där problem uppstod så att uppgiften antingen inte löstes alls eller endast löstes delvis, alltså uppgifter nummer 4, 7, 9, 12, 15, 16, 17, 18, 19, 20 och 21. Dessa presenteras nedan var för sig, antingen enbart i text eller gestaltat i figurer. De konversationer som presenteras i figurer valdes ut på grund av att något oväntat hänt under konversationen, såsom att inställningar behövts ändras, svar ur kontext, brist på minneskapacitet eller liknande.

4.2.1 Uppgift 4: Byt låt

David lyckades starta en låt i tidigare uppgift men vid kommandot att byta låt gavs inget gensvar och inget resultat. Efter diskussion i intervjun med deltagaren antogs det att det personliga Spotify-kontot var anledningen. Låten sattes förmodligen igång som en singel och därför byttes ingen låt då inget album spelats upp. Eftersom ingen feedback gavs så går det inte att veta säkert vad problemet var.

4.2.2 Uppgift 7-10: Sätt igång podcasten Creepypodden i P3, Stäng av podcast, Sätt igång podcasten Nordic Mythology Podcast, Stäng av podcast



Figur 2: Dialog mellan Kajsa och smart högtalare under uppgift 7

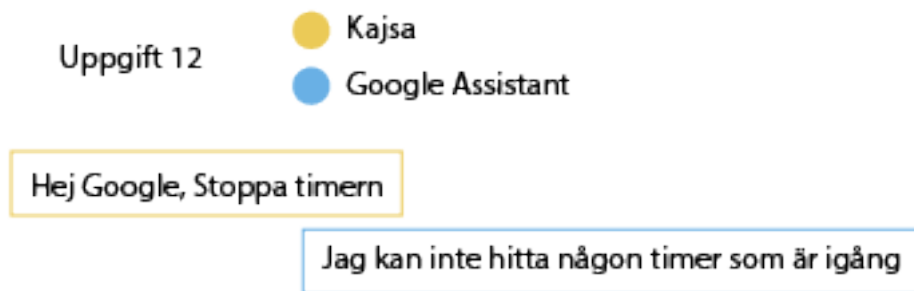
Samtliga deltagare misslyckades med att sätta igång podcasten. Trots att feedback i vissa fall gavs med instruktioner om hur kommandot skulle formuleras gav det inget resultat. För de deltagare där den sortens feedback inte gavs försökte de själva komma på olika sätt att formulera kommandot: "Sätt igång en podcast" (David, uppgift 7). "Sätt igång senaste avsnittet av Creepypodden", "Sätt igång

Creepypoddens senaste avsnitt” (Maria, uppgift 7). “Sätt på ett avsnitt på Creepypodden” (Sara, uppgift 7).

Detta står i linje med tidigare studier utförda av Myers et al. (2018) och Kirschthaler et al. (2020), både genom att prata långsammare, artikulera mer, ändra ordföljden och att använda synonymer. Efter att ha försökt flera gånger utan att lyckas gav samtliga upp och gick vidare till nästa uppgift. Detta betydde att uppgift nummer 8, att stoppa podcasten, hoppades över och ej utfördes.

Samma problem uppstod för deltagarna vid uppgift 9 (sätt igång Nordic Mythology Podcast) som under uppgift 7. Även här försökte deltagarna att omformulera sina kommandon, men gav upp snabbare, något som även observerades av Myers et al. (2018) i deras studie. Det som uppmärksammades under denna uppgift var en högre grad av frustration från samtliga deltagare. Tålmodet för att testa flera olika kommandon var kortare, förmodligen eftersom de redan antagit att inte heller denna uppgift skulle gå att lösa. Kajsa valde att inte utföra denna uppgift med kommentaren att hon redan testat så många andra podcasts som inte fungerade i uppgift nummer 7 (se figur 2).

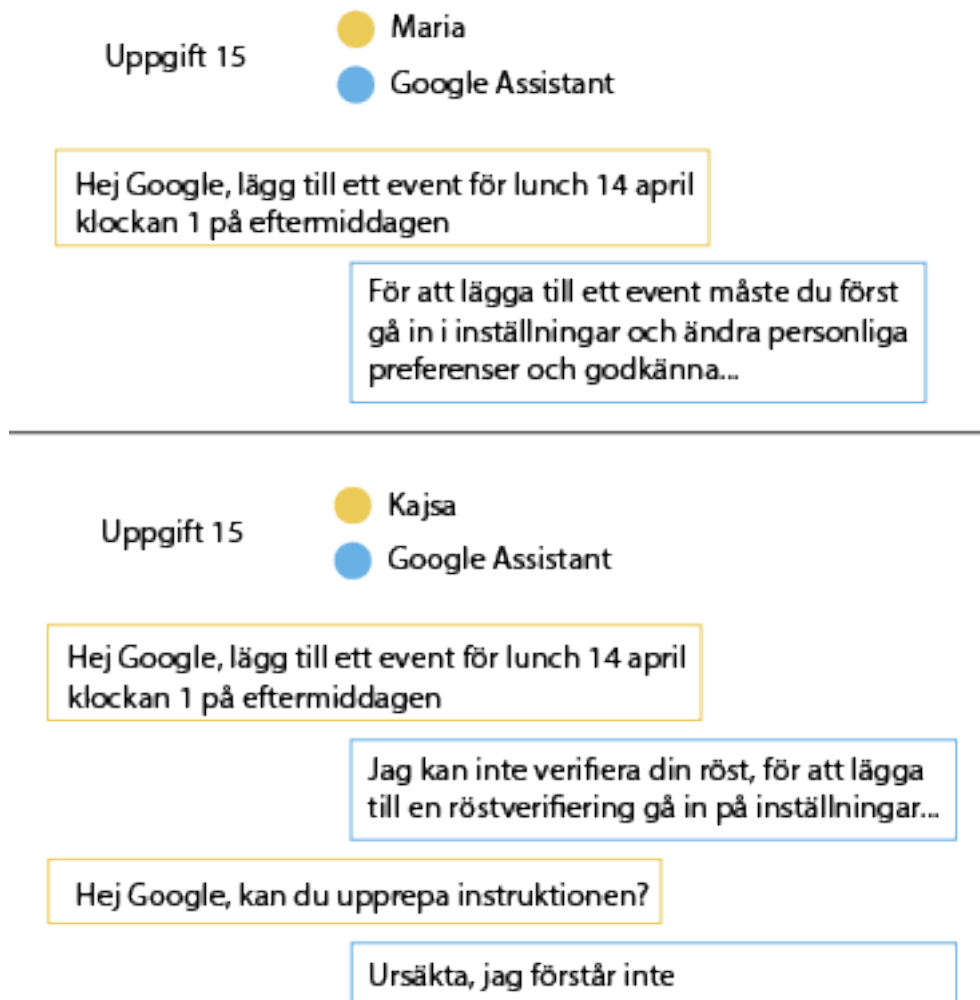
4.2.3 Uppgift 12: Stoppa timer



Figur 3: Dialog mellan Kajsa och smart högtalare under uppgift 12

Det var endast en deltagare som upplevde problem vid uppgift 12 då Google Assistant inte kunde hitta den timer som skulle satts igång i uppgift 11 (se figur 3). Det var svårt att veta vad som gick fel, eftersom Google Assistant givit röstoutput (feedback) om att kommandot “Sätt igång timer” uppfattats av systemets kommandodatabas (Baimirov et al., 2022; Sutton et al., 2019).

4.2.4 Uppgift 15: Lägg till ett event för lunch den 14 april klockan 1 på eftermiddagen



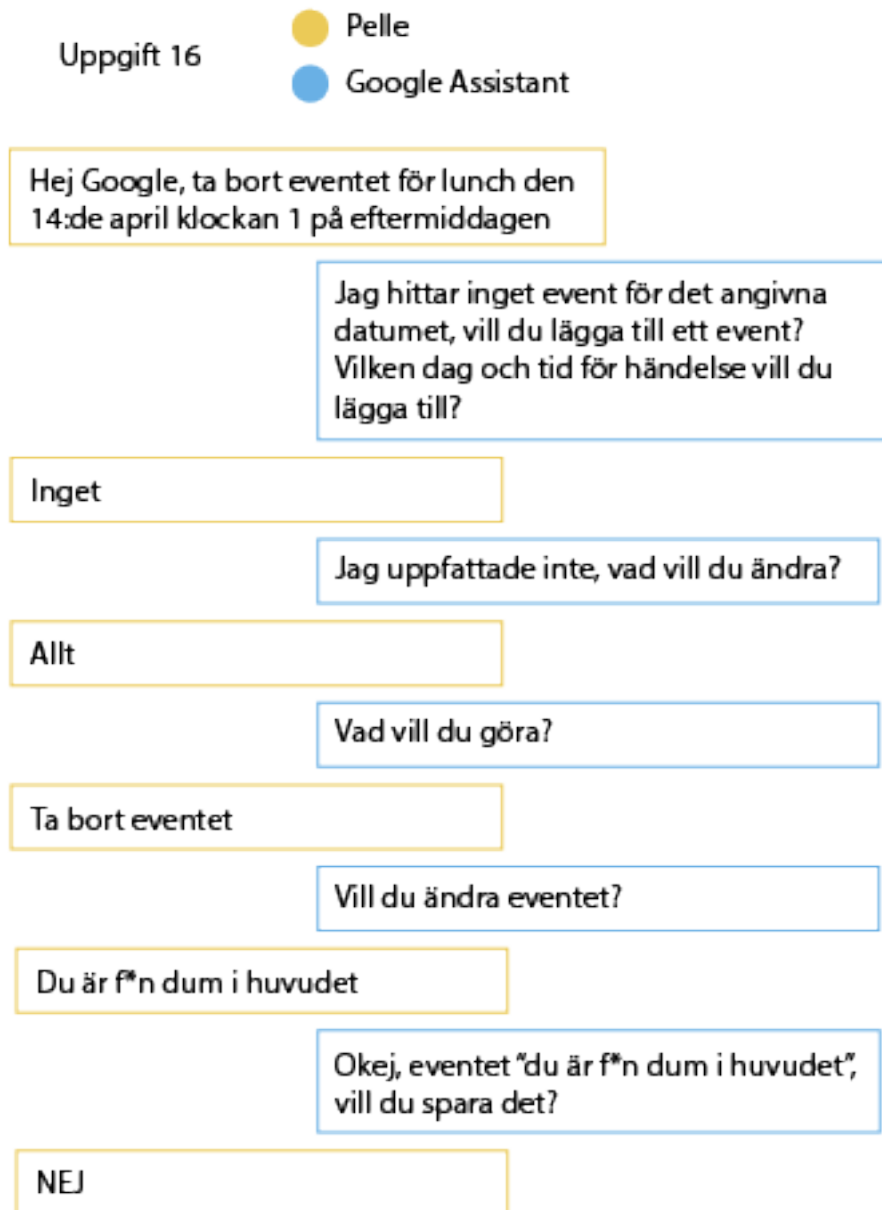
Figur 4: Dialog mellan Maria och smart högtalare, samt Kajsa och smart högtalare under uppgift 15

Vid uppgiften att lägga till ett event stötte två deltagare på problem som gjorde att uppgiften inte klarades av (se figur 4). Vid båda tillfällena gavs feedback om varför det inte gick att utföra kommandot. ”För att lägga till ett event måste du först gå in på inställningar och ändra personliga preferenser och godkänna Googles sekretesspolicy om insamling och hantering av persondata” (Google Assistant, uppgift 15 till Maria). “Jag kan inte verifiera din röst, för att lägga till en röstverifiering gå in på inställningar i Google Home appen. Tryck sedan på

Hey Google och Voice Match och välj andra enheter. Tryck sedan på Kom igång. Följ de anvisningar som ges” (Google Assistant, uppgift 15 till Kajsa).

Det som uppmärksammades här var att instruktionerna gavs i för snabb hastighet och var långa. Deltagarna uppfattade inte instruktionerna när de gavs och kunde således inte utföra det som Google Assistant föreslagit. Därför valde Maria att gå vidare till nästa uppgift, medan Kajsa försökte få instruktionen upprepad för att kunna utföra den. Sarcona et al. (2020) diskuterade i sin studie behovet av att kunna upprepa och lyssna igenom feedback för att kunna ta det till sig. Google Assistant kunde dock inte upprepa instruktionen till Kajsa när detta önskades, vilket kan bero på VUI-enhetens minneskapacitet, något som diskuteras av Murad et al., (2021) samt Wei och Landlay (2018).

4.2.5 Uppgift 16: Ta bort eventet för lunch den 14 april klockan 1 på eftermiddagen



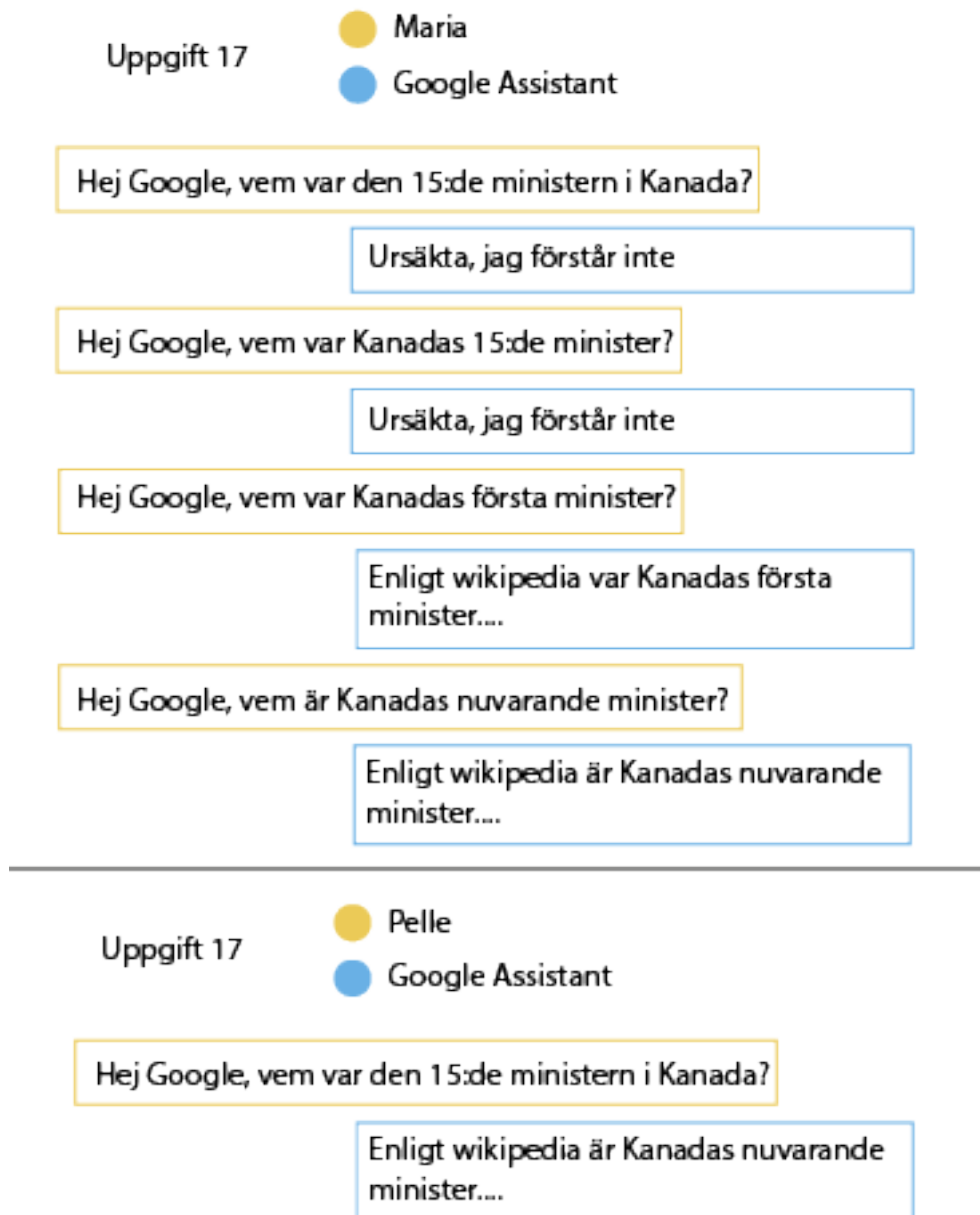
Figur 5: Dialog mellan Pelle och smart högtalare under uppgift 16

De tre deltagare som lyckades lägga till ett event i uppgift 15, gick vidare till uppgift 16 - att ta bort eventet. Där stötte samtliga deltagare på ytterligare problem. När de försökte ta bort eventet hittade inte Google Assistant något event att ta bort. Samtliga av deltagarna uppvisade förvirring då de antagit att deras event lagts till. Google Assistant frågade i uppgift 15 om de ville spara det event

de försökt lägga till, varpå de alla svarade "ja". Att deltagarna antog att Google Assistant uppfattat och utfört kommandot utan verifierande röstoutput var en intressant observation. Seaborn et al. (2021) lyfter hur det fortfarande saknas forskning kring hur användare uppfattar och svarar på VUI-systemets feedback. Att deltagarna antog att systemet bekräftat kommandot utan att ge verifierande röstoutput (feedback) kan bero på hur VUI-systemet designas med försök till en mer naturlig dialog, likt mellan människor (Diederich et al., 2020).

Sara valde att avbryta direkt, medan David och Pelle ville säkerställa att eventet inte fanns sparad trots allt. David valde att upprepa frågan långsammare, men då frågade Google Assistant i stället om han ville radera konversationer, aktiviteter och liknande. Den gav även instruktioner om hur han skulle kunna radera dessa, men verkade inte ha uppfattat det kommando han faktiskt gav. Pelle fick ett längre svar på första kommandot med en följdfråga om han ville lägga till ett event i stället (se figur 5). Detta var han inte intresserad av och försökte få sin Google Assistant att förstå honom. Han blev mer och mer irriterad under konversationens gång vilket resulterade i en oväntad dialog som dessvärre inte ledde till att uppgiften klarades av.

4.2.6 Uppgift 17: Ta reda på vem den 15:de ministern av Kanada var



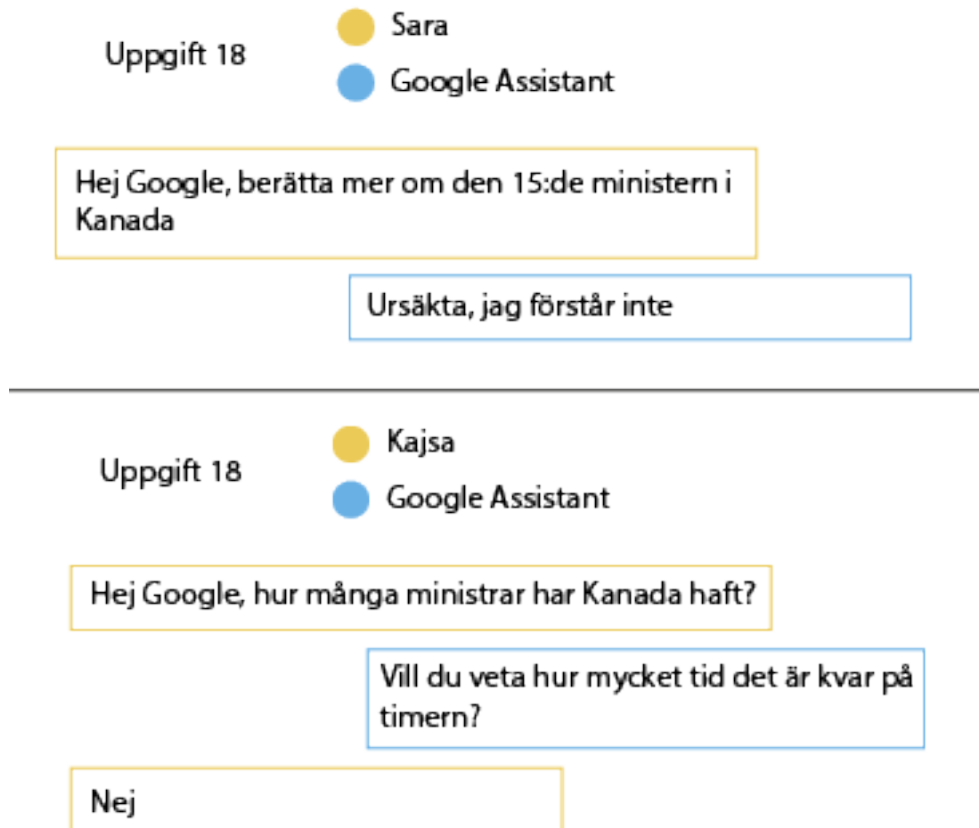
Figur 6: Dialog mellan Maria och smart högtalare, samt Pelle och smart högtalare under uppgift 17

Samtliga deltagare hade stora problem med att lyckas med denna uppgift (se figur 6), och alla deltagare utom en försökte upprepa eller omformulera kommandot. Pelle sa bara kommandot en gång och valde sedan att avbryta: "Hej Google, vem var den 15:de ministern i Kanada?" (Pelle, uppgift 17). David och Sara valde att avbryta efter andra försöket. Maria och Kajsa fortsatte att försöka komma på ett

sätt att uttrycka sig för att få svar på frågan. “Hej Google, vem var den 15:de ministern i Kanada?”, “Hej Google, vem var Kanadas 15:de minister?”, “Hej Google, vem var Kanadas första minister?”, “Hej Google, vem är Kanadas nuvarande minister?” (Maria, uppgift 17). “Hej Google, vem var den 15:de ministern i Kanada?”, “Hej Google, vem var Kanadas 15:de minister?”, “Hej Google, vem är Kanadas nuvarande minister?”, “Hej Google, vem var Kanadas 15:de minister” (Kajsa, uppgift 17).

Genom att ställa andra frågor om ministrar i Kanada hoppades de ta reda på vad i det ursprungliga kommandot som resulterade i problem. De fick svar på sina följdfrågor, även om svaren inte nödvändigtvis gav specifik information. Däremot ledde det inte till att uppgiften klarades av. Kajsa kommenterade under tiden att hon trodde att det var själva siffran “femtonde” som inte uppfattades i det första kommandot. “Eftersom att den (Google Assistant) kunde berätta vem den nuvarande ministern är, så är problemet nog att den inte förstår ordet “femtonde” tror jag” (Kajsa, uppgift 17). Hon fortsatte med att säga att det hade varit praktiskt med feedback i form av vilka nyckelord den uppfattade.

4.2.7 Uppgift 18: Ta reda på mer information om den 15:de ministern av Kanada

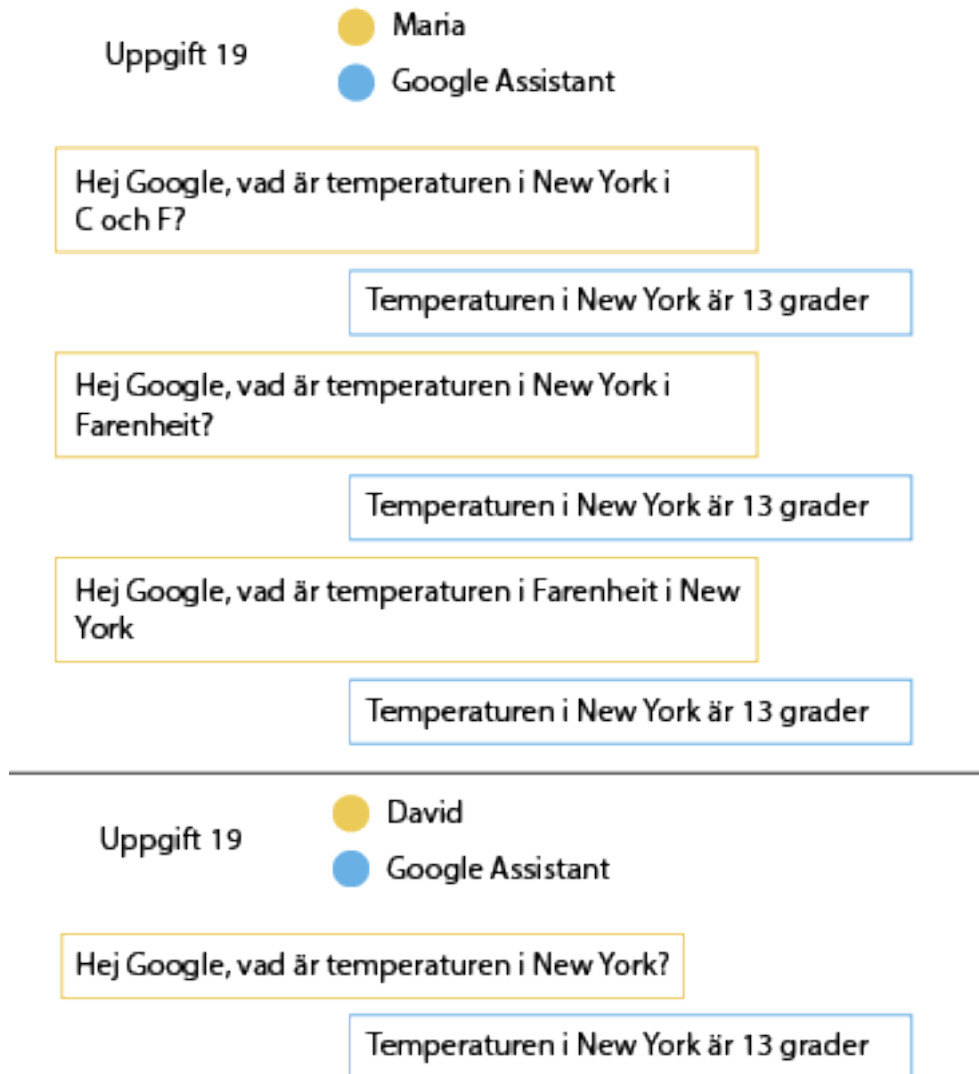


Figur 7: Dialog mellan Sara och smart högtalare, samt Kajsa och smart högtalare under uppgift 18

Maria och Pelle valde att hoppa över denna uppgift med kommentaren att den ändå inte hade uppfattat föregående fråga. De tre deltagare som försökte utföra denna uppgift testade bara en gång då de blev mer och mer frustrerade under observationens gång. De valde att formulera sina kommandon på olika sätt men ingen av dem fick tillfredsställande svar (se Sara och Kajsas dialoger i figur 7). I två fall förstod Google Assistant inte kommandot, vilket förmodligen har med VUI-enhetens minneskapacitet att göra, något som Murad et al. (2021) och Wei och Landlay (2018) nämner i sina studier. I det tredje fallet gavs det irrelevanta svaret “Vill du veta hur mycket tid som är kvar på timern?” (Google Assistant, uppgift 18). Detta hände Kajsa, samma deltagare som inte kunde stoppa timern i uppgift 12 då ingen timer kunde hittas. Det verkar således som att kommandot

uppfattats och en timer satts i gång trots allt, likt hur Baimirov et al. (2022) och Sutton et al. (2019) beskrivit i sina artiklar kring igenkänningsfas och utförandefas.

4.2.8 Uppgift 19: Ta reda på temperaturen i New York både i Celsius och Fahrenheit

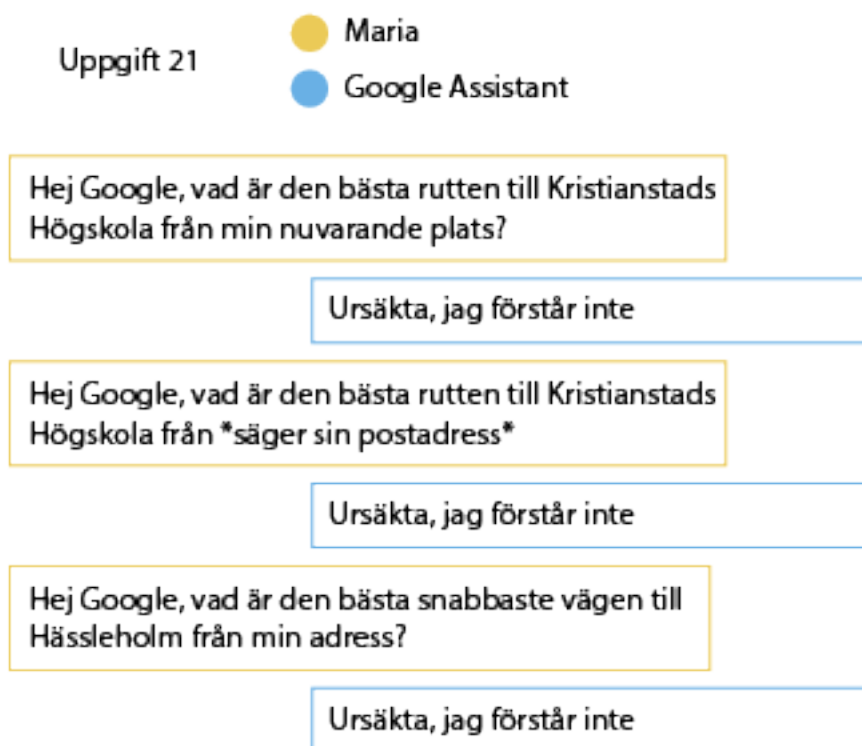


Figur 8: Dialog mellan Maria och smart högtalare, samt David och smart högtalare under uppgift 19

För tre av deltagarna utfördes uppgiften korrekt genom att de delade upp kommandot i två delar, att först fråga om Celsius och sedan Fahrenheit, eller tvärtom. För Maria och David svarade Google Assistant endast i Celsius (se figur

8). Maria testade att fråga om enbart Fahrenheit, dessvärre fortsatte Google Assistant att svara i vad som antogs var Celsius. Hon anmärkte: “Den verkar ignorera ordet Fahrenheit. Jag ger upp.” (Maria, uppgift 19). David testade endast en gång och avbröt sedan med kommentaren: “Jag tror inte att den förstår ordet Fahrenheit” (David, uppgift 19).

4.2.9 Uppgift 20-21: Undersök om det finns några veganska restauranger i närheten, ta reda på bästa rutten till Kristianstad Högskola från din position



Figur 9: Dialog mellan Maria och smart högtalare under uppgift 21

För fyra av deltagarna utfördes uppgift 20, att undersöka om det finns några veganska restauranger i närheten, framgångsrikt och gav resultat som överensstämde med var de befann sig. Deltagarna kommenterade att de blev positivt överraskade av de korrekta svaren. Sara uttryckte att hon kommer att använda den funktionen i fortsättningen också. För Maria svarade Google Assistant på frågan med veganska alternativ på restauranger, men utgick inte från hennes position. Hon uttryckte att det var märkligt eftersom hon hade lagt till sin

adress i Google Home-appen samt hade platstjänst och Bluetooth i gång på sin mobil.

I uppgift 21 (se figur 9) uppmärksammades liknande problem som föregående uppgift 20 för samma deltagare: Maria. Övriga deltagare kunde efter viss omformulering få en tillfredsställande vägbeskrivning från sin position till Högskolan Kristianstad. Endast Maria hade problem med att få ett svar. Hon försökte utföra uppgiften på flera sätt genom att omformulera, samt specificera sitt kommando: “Hej Google, vad är den bästa rutten till Kristianstads Högskola från min nuvarande plats?”, “Hej Google, vad är den bästa rutten till Kristianstads Högskola från *säger sin adress*?”, “Hej Google, vad är den snabbaste vägen till Hässleholm från *säger sin adress*?” (Maria, uppgift 21).

Samtliga försök resulterade i att Google Assistant endast svarade att den inte förstod. Deltagaren kommenterade att det förmodligen hängde ihop med föregående uppgift, att Google Assistant helt enkelt inte förstod var hon befann sig. Dock var det svårt att veta om teorin stämmer eftersom hon även specificerade sin egen adress och resmålet i ett kommando och trots detta fungerade det inte. Deltagaren nämner dock att hon alltid är markerad som placerad i Västra Götaland eller Stockholmsområdet när hon använder Google för sökningar på dator, trots att hon befinner sig i Skåne. Baimirov et al. (2022) förklarar hur systemet skickar i väg förfrågan till internet vid kommando som kräver sökningar. Det kan därför vara så att trots användarens inställning i Google Home-appen om var hon befinner sig, så kan sökresultatet prioriteras och ge liknande svar som vid deltagarens egna Google sökningar. Dock gäller detta inte när användaren specifikt angav start- och slutdestination med adress. Det är därför oklart varför detta kommando inte fungerade om VUI-enheten sökte via Google Maps.

4.3 Sammanställt resultat av observation

Insikterna som observationerna gav var att de uppgifter som resulterade i störst problem var samma uppgifter för alla deltagare. Trots att alla uppgifter som utfördes i observationen är sådana kommandon som sägs ska kunna lösas av

Google Assistant (Alrumayh & Tan, 2022; Murad et al., 2021) var inte detta fallet. Deltagarna hade olika erfarenheter av att använda Google Assistant, där David hade störst användningsvana. Kajsa och Maria var de deltagare med minst erfarenhet gällande användning av Google Assistant, men var de som testade samma kommando flest gånger. Det verkade således som att desto mer erfarenhet med Google Assistant desto färre gånger testades samma kommando, något som även tidigare forskning uppmärksammat (Kirschthaler et al., 2020).

Den uppgift som var mest problematisk var att sätta igång en podcast, oberoende av svenskt eller engelskt namn, trots att båda podcasterna fanns i Googles egna Podcast-applikation samt Spotify. Den error feedback som gavs var blandad, men vid upprepade fall gav Google Assistant instruktioner om hur kommandot skulle formuleras för att sätta igång en podcast. Trots att deltagarna upprepade kommandot på det sätt de tolkat instruktionen på, sattes ingen podcast igång, och svaret blev "ursäkta, jag förstår inte". Lee et al. (2019) förespråkade i sin artikel vikten av att VUI-system ska kunna kommunicera med användare när fel uppstår, vilket systemet gjorde i detta fall. Hur instruktionerna ges och tolkas är däremot någonting som kräver vidare utforskning enligt Seaborn et al. (2021).

I uppgift 16 "att ta bort ett event", gavs en ovanligt lång dialog med Pelle där Google Assistant inte ansåg dialogen avslutad lika snabbt som vid övriga uppgifter. Det som skiljer denna långa dialog mot andra dialoger var att Pelle inte behövde säga "Hej Google" när han svarade på dess frågor. Denna form av dialog, alltså dialogmönstret som tillåter användare och system att turas om att prata, är det som sägs vara inbyggd i Google Assistants infrastruktur (Cambre & Kulkarni, 2020). Detta var ingenting som uppmärksammades i någon av de övriga dialoger som skedde mellan deltagarna och deras smarta högtalare. I samtliga fall stängdes sessionen av med tydlig indikation av släckta lampor efter att svar givits från den smarta högtalaren. Det gavs därför ingen möjlighet att fortsätta konversationen, utan ny väckningsfras krävdes och minnet från tidigare dialoger saknades (Murad et al., 2021; Wei & Landay, 2018.) Hur minnesfunktionen fungerar är därför svårt att teoretisera kring, speciellt eftersom ett annat exempel på minnesfunktionen under observationen uppmärksammades. Kajsa fick en fråga ur kontext i uppgift

18 som relaterade till uppgift 12, “att stänga av en timer”. Det verkade således som att Google Assistant kom ihåg att en timer satts igång trots att den tidigare sa att ingen hittades. Varför detta hände är oklart.

Den sista uppgiften som var märkbart svår att genomföra för samtliga deltagare var att ta reda på vem den 15:de premiärministern av Kanada var, uppgift 17. Dessvärre gavs ingen error feedback på hur kommandot skulle kunna ändras för att nå det önskade resultatet. Istället fick vissa deltagare reda på den nuvarande ministern, alternativt den första ministern. Detta gjorde också att uppgift 21, “ta reda på mer information om Kanadas 15:de minister”, inte löstes. Vissa valde att hoppa över frågan medan andra utförde den, men ingen vidare information om Kanadas 15:de minister kunde presenteras från deras Google Assistants. Deltagarna visade att de var frustrerade genom höjda röster, förändrade tonlägen och kortare tålamod.

Under observationerna togs även svårigheter med dialekter eller könstillhörighet i beaktning (Shafei & Tan, 2022; Rajan et al., 2022), ifall ytterligare problem uppstod i dialogen mellan användare och smart högtalare. I vissa fall uppmärksammades tydliga dialektala skillnader när deltagarna pratade med författarna jämfört med när de pratade med sina smarta högtalare. Sara pratade med betydligt mindre dialekt när hon konverserade med sin smarta högtalare och Pelle berättade att han aktivt valde att avstå från att använda skånska uttryck. Detta gjorde därför att problem vid dialektala skillnader inte uppmärksammades. Vidare uppmärksammades inga tydliga indikationer på att könstillhörighet hade någon påverkan på resultatet av observationsuppgifterna.

4.4 Resultat intervjuer

Intervjuerna genomfördes direkt efter att sista kommandot på observationen genomfördes efter deltagarnas bästa förmåga. Dessa grupperades efter de teman som tagits fram för intervjun (se bilaga 6), och presenteras på liknande sätt nedan i form av text och direkta citat från deltagarna.

4.4.1 Användningsområden

Deltagarna har haft och använt Google Assistant genom smarta högtalare i minst ett år, upp till fem år. Majoriteten av deltagarna använder den endast för att lyssna på musik, alternativt för att lyssna på när den drar ett skämt. Denna form av användning ligger i linje med vad rapporten NPR & Edison Research (2022) och studien av Lee et al. (2022) kommit fram till. “Den drar så torra skämt, som Göteborgsvitsar, det är rätt roligt!” (Kajsa).

Majoriteten av deltagarna försökte testa sig fram för att ta reda på vilka funktioner som Google Assistant kunde erbjuda när de precis köpt dem, något som de senare slutat göra. Att användare slutar utforska funktioner och enbart håller sig till specifika saker har även uppmärksammats i tidigare studier gjorda av Kirschtaler et al. (2020).

“Högtalaren köptes för runt ett år sedan och jag använde den mycket i början för att testa och utforska vilka funktioner den har. Jag har dock använt den mindre och testat färre kommandon under den tid jag haft den på grund av för höga förväntningar som sedan tappades när högtalaren inte kunde utföra eller förstå vad jag ville.” (Maria)

“Jag testade ett par olika funktioner, speciellt voicematch i början. Men jag har slutat testa nya grejer, och jag har stängt av min voicematch” (Pelle).

David använder sig av alla de funktioner som han har hittat och kommenterar att de flesta fungerar väldigt bra. Han lägger också till att han har förlikat sig med att den bara kan utföra vissa kommandon. “De funktioner som finns fungerar bra, förutom shoppinglistan, den är svår och borde göras mer tillgänglig. För nu måste man gå in i inställningar själv via mobil eller surfplatta och sedan klicka sig vidare till listor för att hitta den.” (David).

David var den deltagare som haft och använt sin smarta högtalare längst, runt fem år, och har under åren köpt på sig totalt fyra stycken utspridda i olika rum i sin lägenhet. Vid frågan om hur hans förväntningar förändrats från den första inköpta högtalaren till nu berättade han att förväntningarna både stigit och sjunkit. Moore och Urakami (2022) diskuterar hur ett röstbaserat system som sägs vara tillgänglig hela tiden skapar höga förväntningar. Något som de menar att nuvarande

system inte lever upp till i dagsläget. Lee et al. (2020) lyfter hur användarens mentala modeller och förväntningar på vad VUI's kan göra inte stämmer överens med hur de senare upplever dess faktiska kapacitet.

“När jag köpte den första högtalaren var mina förväntningar höga, men dessa sänktes snabbt på grund av brister i teknologin och vid inköpet av andra högtalaren som var en uppdatering var förväntningarna inte speciellt höga. Förväntningarna har sakta ökat med tiden då teknologin har utvecklats och förbättrats, så nu är de rätt höga igen.” (David)

Ett oväntat resultat av observationen var att Sara uttryckte att hon med hjälp av observationsstudien insett att hon kan be sin smarta högtalare att göra fler saker. Kirschtaler et al. (2020) lyfter att det är otydligt och svårt för användare att upptäcka vilka möjligheter och begränsningar som VUI's har, speciellt gällande användning av applikationer skapade av tredje parter. “Jag visste inte att jag kunde be den om tips på restauranger, det är något jag absolut kommer använda i fortsättningen” (Sara). Denna brist uppmärksammades vid flera tillfällen under studiernas gång, till exempel när Kajsa uttryckte att hon inte visste vad högtalaren faktiskt kunde göra för henne, och att hon inte känt sig manad att testa sig fram själv. På grund av detta har hon hållit sig till att sätta igång musik och lyssna på torra Göteborgsskämt. Detta gällde även Pelle och Maria som enbart använde sina högtalare för att lyssna på musik. Kirschthaler et al. (2020) och Moore och Urakami (2022) nämner i sina artiklar att det inte finns något bra system för att erbjuda användare mentala modeller för hur röststyrda högtalare fungerar och kan navigeras.

4.4.2 Kommandon till smarta högtalare

När uppgifterna till observationen togs fram beskrevs de på ett sätt som skulle ge användaren frihet till att själv formulera sitt kommando, men de kunde även läsa uppgiften ordagrant. Hur en användare pratar med sin smarta högtalare är väsentligt för att den ska uppfatta vad som sägs (Shafei & Tan, 2022; Rajan et al., 2022). Det var därför relevant för studien att få reda på hur deltagarna upplevde de kommandon de utfört genom observationen. Åsikterna gällande dessa var blandade, där David förklarade att han ofta försöker säga så få ord som möjligt för att försöka göra det lättare för Google Assistant att förstå honom.

“Jag brukar försöka korta ner kommandona så mycket som möjligt, så istället för att säga “kan du skruva ner volymen” som jag hade frågat en människa, brukar jag bara säga “sänk”. Det är något jag har lärt mig med tiden” (David)

Kajsa berättade att hon ofta formulerar sina kommandon i längre meningar och att det har fungerat bra hittills. Detta mönster av att ge mer information till sin smarta högtalare, i förhoppning att den lättare ska förstå kommandot, ligger i linje med ett av de mönster Myers et al. (2018) uppmärksammade i sin studie. Hur Kajsa uttryckte sina kommandon är inte något som hon har reflekterat så mycket över. “Jag tänker inte på att jag pratar med en robot” (Kajsa). Denna typ av tanke skulle kunna öka risken för negativa upplevelser ifall användaren inte har i åtanke att det är ett system de interagerar med, något som Diederich et al. (2020) varnar för gällande *The Uncanny Valley Effect*. Maria har däremot reflekterat över hur hon uttrycker sina kommandon. Hon menade att det var jobbigt att aktivt tänka på hur hennes kommandon uttrycks.

“Jag tycker att det är svårt att säga de mer avancerade kommandona som har mer detaljer. Om jag måste omformulera frågan glömmer jag ofta av vad jag precis har sagt. Det känns som jag har ett guldfiskminne, tänk Doris i Hitta Nemo” (Maria)

Pelle uttryckte att han mest tycker att det är jobbigt att behöva säga “Hej Google” vid varje kommando, vilket även deltagare i studier gjorda av Murad et al. (2021) uttryckte. Han observerade att skånska slangord inte fungerade under observationen och kom på sig själv med att omformulera till ett mer vedertaget språk. Sara reflekterade över att det är svårt att veta vad hon kan säga till sin Google Assistant och hur det ska formuleras, hon hade önskat att den kunnat hjälpa henne med detta. Möjligheten till hjälp är något som Myers (2019) förespråkar genom ett med adaptivt system som lär användare hur de ska formulera sig.

Maria kommenterade att hon tyckte att de uppgifter som skulle utföras i observationen kändes för enkla när hon läste dem första gången, något som hon senare insåg inte var fallet.

“Uppgifterna kändes initialt lite banala, det var till synes väldigt enkla saker som skulle utföras om man tänker efter kring vad man

egentligen önskar att den kunde göra. Jag blir inte så sugen på att utforska fler funktioner när inte ens de enkla sakerna fungerar, känns inte så inspirerande.” (Maria)

4.4.3 Error feedback från Google Assistant

Att undersöka den error feedback som ges från Google Assistant genom smarta högtalare till användaren låg i fokus genom studierna. Genom observationen uppmärksammades det att deltagarna inte fick speciellt mycket error feedback när deras kommandon inte kunde uppfattas eller utföras. Lee et al. (2019) menar att en mer utförlig feedback till användarna leder till en högre acceptans vid oförmåga att utföra eller uppfatta kommandon. Vid de tillfällena där error feedback gavs i form av instruktioner kunde deltagarna ändå inte få kommandot att fungera. Majoriteten av deltagarna berättade att de upplevde en stor frustration i att inte veta vad som var problemet med kommandot, eftersom de då inte visste hur de skulle omformulera sig. “Jag lyssnar mycket på poddar, och hade det fungerat hade det varit jättebra. Men nu fungerade det inte även om jag sa exakt så som den föreslog i sin feedback. Jättestörigt” (Kajsa). “Den hjälpte ju inte mig att veta hur jag skulle ändra kommandot och jag blev därför både irriterad och även lite uppgiven” (Maria). Även Sara berättade att hon inte förstår vad som går fel ibland, men att hon lärt sig vilka ord som förstås av Google Assistant med tiden. “Under tidens gång har jag lärt mig vad jag kan säga och inte, numera kommer det nästan automatiskt” (Sara).

Det som utmärkte sig extra var att oavsett om Google Assistant gav feedback eller inte så löstes inte alla uppgifter. De uppgifter som löstes på första eller andra försöket var tack vare att deltagarna själva omformulerade det de sa. Vid de uppgifter som var problematiska för samtliga deltagare löstes ingen av dem, trots att deltagarna i vissa fall fick instruktioner i hur de skulle formulera sig. De tydligaste exemplen av detta är vid uppgift 7 och 9, att sätta igång podcasts. Alla deltagare fick vid något tillfälle feedbacken “Om du vill sätta igång en podcast, säg sätt igång podcast följt av podcastens namn”. Vissa deltagare hade redan formulerat detta innan feedbacken, och andra gjorde det efter att feedbacken givits. Oavsett ordning kunde ingen av dem sätta i gång någon podcast.

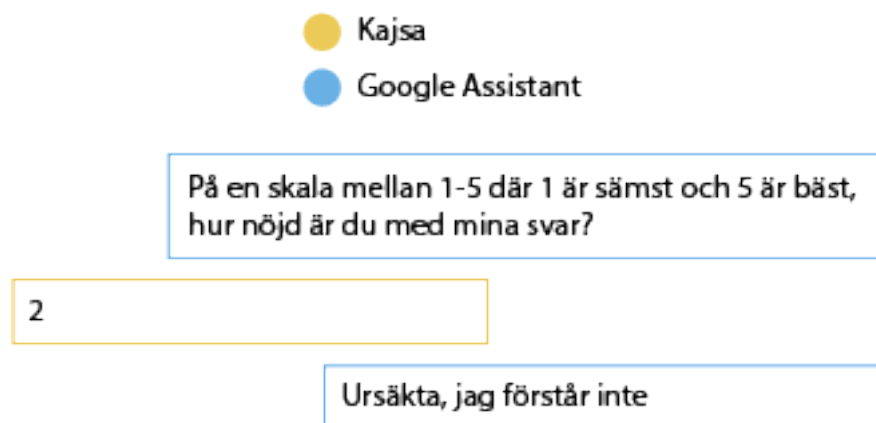
Vid några tillfällen gavs också längre instruktioner från Google Assistant till deltagarna. Kajsa fick en instruktion om hur hon skulle kunna verifiera sin röst, men uttryckte att den gick så fort att hon ville få instruktionen upprepad. Detta gick dock inte, utan när hon frågade om Google Assistant kunde upprepa instruktionen fick hon som svar att den inte förstod. Studier av Sarcona et al. (2020) lyfter specifikt hur ett av de vanligaste problemen med auditiv feedback är just behovet av upprepning. Både David och Sara önskade att instruktionerna kunnat ges på andra sätt. "Det hade varit bra om det gick att anpassa instruktionerna beroende på vad man behöver, kanske genom att den kan upprepa informationen eller man får det stegvis" (David). "Det skulle vara bra om det gick att sakta ner eller få informationen stegvis. Till exempel att assistenten gick igenom stegen med dig och du behöver bekräfta att steget är genomfört innan nästa steg i instruktionen kommer" (Sara).

Annan feedback som också gavs var till exempel när ett alarm eller timer sattes i gång. Vid flera tillfällen under observationen fick deltagarna dubbla signaler från sina smarta högtalare. När de skulle skapa ett event frågade Google Assistant om de ville spara detta, de svarade ja och antog därför att eventet sparats trots att ingen ytterligare feedback gavs efter det. Den bekräftade därför varken om det sparats eller inte. När de sedan skulle ta bort samma event visade det sig dock att det inte fanns sparad. Pelle uttryckte att han inte skulle lita på att Google Assistant verkligen sparar viktiga data, så som möten och dylikt. Han fortsätter med att säga att det är bäst att lägga till det manuellt. På liknande sätt upplevde Kajsa att hon fick mixade signaler av Google Assistant gällande den timer hon satt igång. När hon försökte stoppa timern svarade hennes smarta högtalare att den inte kunde hitta en timer, bara för att senare i en helt annan uppgift fråga om deltagaren ville veta hur mycket tid som var kvar. Detta resulterade i skratt, men hon uttryckte också en osäkerhet kring hur hon ska kunna lita på att det den smarta högtalaren bekräftat stämmer.

Vid andra tillfällen under observationen erhöles däremot ingen förklarande error feedback alls, i de enklare uppgifterna såsom att skruva upp musik eller liknande löste deltagarna det själva. Vid uppgift 17 "att ta reda på vem Kanadas 15de

premiärminister var” presenterades ingen utförlig error feedback och deltagarna förstod inte hur de skulle omformulera sina kommandon. Att ge mer utförlig feedback till användare menar Myers et al. (2019) hjälper dem att lära sig vad för kommandon som fungerar överlag och även hur de ska formulera sig för att klara sina specifika uppgifter. Även Holland (2020) lyfter vikten av att erbjuda extra stöd vid auditiv feedback för att den ska vara till nytta. Kajsa kommenterade att det hade varit bra om Google Assistant hade kunnat komma med mer konkret feedback om den inte förstod vad som sas. “Jag hade velat veta vilka ord i frågan som snappades upp, därefter kunde ett exempel på hur man skulle kunna formulera sig ges. Kanske förslag på nyckelord istället för “förlåt jag förstod inte”” (Kajsa).

Det som uppmärksammats är att samtliga deltagare även utanför studien upplever problem med sina Google Assistants och försöker själva hitta lösningar på dessa. Medan vissa lugnt förlikar sig med vad den smarta högtalaren faktiskt kan göra blir andra frustrerade. Alla deltagare kommenterade även att den error feedback som ges sällan hjälper, och om den ger instruktioner går de för snabbt och går inte att få upprepade igen. Kajsa sammanfattade sin upplevelse när hennes smarta högtalare bad henne, efter sista uppgiften i observationen, att betygsätta högtalarens svar mellan 1–5 (se figur 10).



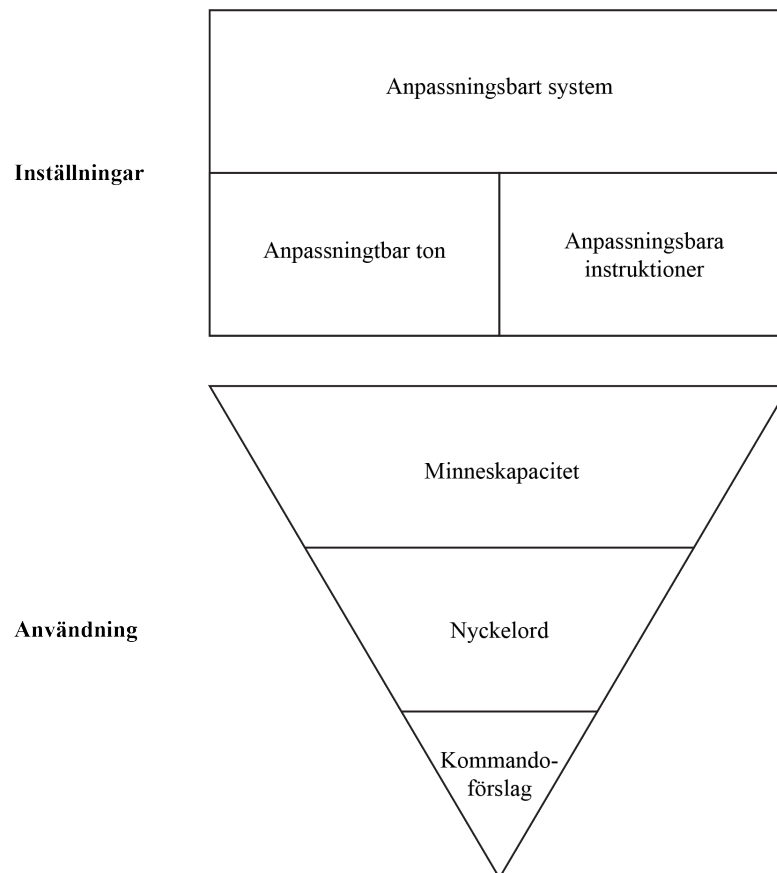
Figur 10: Dialog mellan Kajsa och smart högtalare efter observation

“Det kändes rätt passande att den inte förstod mig när den ville att jag skulle berätta hur nöjd jag var, känns nästan som den gjorde det med mening” (Kajsa)

4.5 Design guidelines

Under litteratursökningen lyfte flera artiklar bristen på struktur kring framtagning av VUI-system genom ramverk, guidelines, metoder och verktyg (Alrumayh & Tan, 2022; Kim et al., 2021; Murad & Munteanu, 2022; Seaborn & Urakami, 2021; Sugisaki & Bleiker, 2020; Sutton et al., 2019). Genom att försöka ta fram ett förslag för en specifik del av VUI-framtagningen, i detta fall error feedback, är förhoppningen att kunna bidra till designers arbete och hjälpa utvecklingen av kommande VUI-system.

I nedan Figur 11 följer sex design guidelines framtagna utifrån studiens resultat i syfte att vägleda designers i utformningen av error feedback till VUI. Dessa är uppdelade i två olika teman: *Inställningar* och *Användning*.



Figur 11: Modell över struktur för designguidelines

4.5.1 Inställningar

4.5.1.1 Design guideline 1 – Anpassningsbart system

Användare av VUI's har olika erfarenheter kring gränssnittet och använder det till olika ändamål. På grund av detta är det av högsta vikt att systemet kan ge någon form av anpassningsbarhet genom personliga inställningar baserat på vad den specifika användaren behöver. One size does not fit all. Detta är något som uppmärksammats både i uppsatsens studier samt av Myers (2019). Att kunna ställa in personliga inställningar samt ha ett system som erbjuder extra stöd för att lära användaren hur de ska interagera med systemet kan därför ge värde till användare. Genom att implementera detta finns kanske även möjligheten att användaren utforskar systemet istället för att endast använda sig av ett fåtal kommandon som fungerar för dem. Detta är något som uppmärksammats genom uppsatsens studier, samt poängterats av Kirschtaler et al. (2020). Det kan speciellt vara av nytta för nya användare av VUI-system där det extra stödet kan hjälpa dem att lära sig att navigera ett nytt gränssnitt, i likhet med hur Killingback et al. (2019) menar att auditiv feedback kan användas för extra stöd till nya studenter.

4.5.1.2 Design guideline 2 – Anpassningsbar ton

Eftersom alla lär sig och tar till sig kunskap på olika sätt är en anpassningsbar ton något som kan implementeras inom VUI. Genom att användare skulle kunna anpassa den ton som VUI'n använder för att förmedla sin error feedback skulle chansen att användaren tar till sig feedbacken på ett så bra sätt som möjligt öka. Genom studierna som utförts i denna uppsats är det tydligt att de olika deltagarna formulerar kommandon på olika sätt och har olika förväntningar på responsen. Det skulle således vara värdefullt för användaren att denna själv kan anpassa VUI'ns tonläge och sätt att formulera error feedbacken på så att det passar den enskilda användaren.

4.5.1.3 Design guideline 3 – Instruktioner

För att underlätta för användare av VUI's vid instruktionsgivning behövs möjligheten att kunna kontrollera hur instruktionerna framförs. Valet att få instruktioner upprepade (Sarcona et al., 2020) bör vara en standardinställning även för mer vana användare. Möjligheten att få successivt tillhandahållna

instruktioner kan vara till stor hjälp för mindre teknikvana användare av VUI-enheter. Genom att erbjuda instruktioner uppdelade i steg och låta användaren själv bekräfta att nästa steg kan påbörjas, ökar även möjligheten att användaren når sitt mål.

4.5.2 Användning

4.5.2.1 Design guideline 4 – Minneskapacitet

VUI's bristande minneskapacitet är någonting som uppmärksammats både i denna uppsats studier men även av Murad et al. (2021) och Wei och Landlay (2018). Eftersom denna begränsning finns behövs en strukturerad dialog utformas i flera steg för att användaren och systemet tillsammans ska kunna ta sig till en lösning i slutet av dialogsessionen. Vid ett error måste systemet kunna prata med användaren och ge tydliga instruktioner så att kommandot kan lösas. I dagsläget stängs sessionen av så fort VUI-systemet säger "jag förstår inte" eller uppfattar att kommandot är utfört, alternativt avslutat (Cambre & Kulkarni, 2020; Yan et al., 2021). Istället föreslås en fortsatt dialog mellan användare och VUI-systemet för att kunna lösa problemet i den befintliga sessionen. Minneskapaciteten behöver alltså öka för att en session ska kunna pågå under en längre tid. Därigenom ges möjligheten att utföra mer komplexa kommandon och lösa eventuella felaktigheter tillsammans.

4.5.2.2 Design guideline 5 – Nyckelord

VUI's ska kunna identifiera och återge de ord som den har identifierat i ett kommando. Om den inte uppfattar hela kommandot så bör den kunna ge tillbaka ord som den ändå har uppfattat, "Är det rätt uppfattat att det handlar om...". Det hjälper användaren att veta att det är andra ord runtomkring detta/dessa nyckelord som behöver omformuleras eller tydliggöras. Denna aspekt av error feedback har setts vara önskvärd genom studierna som genomförts i uppsatsen.

4.5.2.3 Design guideline 6 – Kommandoförslag

Genom att VUI'n kan föreslå förändringar i kommandot slipper användaren fastna i den onda cirkeln av "jag förstår inte" - varav användaren upprepar kommandot som observerats i uppsatsens studier. Detta hjälper även användaren att dela på

minnesbördan av hur kommandot formulerats tidigare. I studien uppmärksammades det av användare att de hade svårt att erinra sig tidigare formuleringar av kommandon. Förslag från VUI'n ger även möjligheten att skapa en djupare och mer naturlig dialog mellan användare och system (Kim et al., 2021). Användaren kan få kommandot utfört så som denne önskat och inte enbart så som systemet uppfattat.

4.6 Återkoppling från deltagare

Efter utarbetning av guidelines för error feedback riktat mot designers, delades dessa med deltagarna i uppsatsens studie. Detta för att ta emot återkoppling i form av feedback och eventuella missuppfattningar eller generella tankar kring dess validitet. Att få ta del av deltagarnas återkoppling var värdefullt eftersom de är mottagarna av den error feedback som ges från VUI-systemet.

Deltagarnas inställning var överlag positiv. Några av de positiva kommentarerna som gavs var ”Jag gillar speciellt nummer två och fem, men nummer två mest för att jag gillar när man kan ändra lite kosmetiska inställningar som påverkar upplevelsen” (David) och ”Tycker det ser bra ut! Rimliga guidelines, hade verkligen varit spännande att få testa i praktiken!” (Kajsa). Deltagaren Maria var den enda som kommenterade den eventuella belastning som ett anpassningsbart system skulle kunna skapa för användarna: ”Delarna för användning tycker jag låter bra, men de under inställningar gör mig lite fundersam. När jag använder teknik vill jag inte behöva fixa massa inställningar först, men det är bra om alternativet finns för de som vill.” (Maria).

5. Diskussion

Det som undersökts genom studierna är hur användare uppfattar error feedback från sina smarta högtalare och hur de hanterar den. Detta med syftet att skapa ett förslag på guidelines för designers att använda vid framtagning av error feedback för VUI-system. Error feedback är väsentligt för att användaren ska kunna navigera felaktiga kommandon och andra problem som uppstår i användning av VUI's, i detta fall smarta högtalare. Genom observationer och intervjuer utifrån ett bekvämlighetsurval, dels via videosamtal och dels i verkligheten, har resultaten bearbetats. Detta har sedan legat till grund för de guidelines som utformats i kombination med tidigare forskning.

5.1 Metoddiskussion

Valet av enkätundersökning gjordes för att få en indikation på riktning för uppsatsen under arbetets start. Enkätundersökningar visar översiktliga områden, det krävs desto fler deltagare för att ge ett rikt kvantitativt material. Att kunna dela undersökningen i Facebook gruppen "Google Home Sverige" möjliggjorde att ett större antal personer deltog. Resultatet av undersökningen indikerade att error feedback var ett område många upplevde som problematiskt, och gav därför en grund till uppsatsens syfte.

Litteratursökningen som gjordes gav initialt ett rikt övergripande material gällande Voice User Interface då mycket har beforskats de senaste åren. Mer avsmalnad sökning kring uppsatsens frågeställning gav däremot färre resultat och tog längre tid. Detta var både problematiskt att luta sig mot, samtidigt som det förstärkte indikatorn om att området kräver mer studier. För att motarbeta bristen på områdesspecifika studier undersöktes även närliggande studier gällande generell auditiv feedback.

Observation och intervju utfördes utifrån ett bekvämlighetsurval (Denscombe, 2018) på grund av att det var den möjlighet som presenterades för att utföra båda studierna. Detta innebär att resultatet inte är generaliserbart utifrån en större population, men visar en indikation på områden som de enskilda deltagarna uppmärksammat. Av en slump visade det sig att samtliga deltagare tillhörde

kategorin “millenials”, det vill säga personer födda mellan tidigt 80-tal och sent 90-tal. Resultatet av studierna hade kunnat bli annorlunda om andra åldersgrupper också deltagit. Millenials är uppväxta med teknikens snabba framfart vilket kan ha lagt grunden för att snabbt och nyfiket ta till sig nya system och anpassa sig till vad tekniken kan erbjuda just nu.

Shafei och Tan (2022) påpekar även svårigheten att hitta tillräckligt många deltagare som passar för studiens utformning, något som även stämde för denna studie. Observation och intervju hölls dels via videosamtal genom Zoom, och dels i verkligheten. Det hade varit en tillgång att kunna hålla alla observationer och intervjuer i verkligheten för att minska risken att skillnaderna på utformningen påverkar resultatet. Dessvärre var detta inte en möjlighet under studien då författarna befann på olika platser och inte hade möjlighet att resa dit deltagarna befann sig. För att hålla studierna så verklighetsnära som möjligt prioriterades det istället att deltagarna utförde studierna i den kontext de normalt hanterade sina VUI's. Att hålla studierna “in the wild” är även något som efterfrågas i befintlig forskning (Diederich et al., 2022; Clark et al., 2019).

Vid utformningen av intervjuerna och observationerna togs beslutet att inte spela in ljud och bild för att deltagarna skulle känna sig mer bekväma och mindre iakttagna. Detta är dock något som försvårade sammanställningen av resultatet. Trots att utförliga anteckningar togs av båda författarna finns risken att något missats. Detta gäller speciellt dialogerna mellan deltagarna och deras smarta högtalare. Möjligheten att gå tillbaka och analysera interaktionen flertalet gånger hade kunnat uppdaga fler detaljer som eventuellt missats under själva studietillfället. Å andra sidan uttryckte deltagarna att de kände sig bekväma under observationen, och om de inte hade känt sig bekväma hade resultatet kunnat påverkas negativt på andra sätt istället. Därför prioriterades brukskontexten över författarnas arbetsbörda. Intervjuns semistrukturerade upplägg möjliggjorde för spontana kommentarer och en avslappnad dialog. Det hade kunnat vara positivt om några frågor eller mer specifika teman tagits fram för att underlätta dialogens flöde och författarnas sammanställning av resultat. Däremot är det möjligt att anta att intervjuerna inte hade flutit på lika naturligt, samt att det inte hade varit lika

naturligt att spinna vidare på deltagarnas egna resonemang. Vidare hade ett mer strukturerat upplägg kunna styra intervjuerna i en specifik riktning och därför riskerat att missa viktiga detaljer (Denscombe, 2018).

5.2 Resultatdiskussion

Studierna har riktats mot error feedback inom VUI's och hur vissa svenska användare upplever och hanterar den. Under studien uppmärksammades mönster för hur användare interagerade med VUI-systemet i likhet med en tidigare studie utförd av Myers et al. (2018). I observationen var det tydligt att deltagarna uppvisade återkommande beteende såsom överdriven artikulering, förenkling eller variation av kommandon. Även att ge mer utförlig information testades. Detta antas ha gjorts i tron att deras smarta högtalare skulle förstå kommandona bättre. Att ge mer utförlig information i sina kommandon kan vara effekten av hur kommunikationen mellan människor ofta berikas av en bredare kontextbeskrivning. Detta visade sig dock inte vara ett effektivt sätt att kommunicera med VUI-enheten. Kim et al. (2021) påpekar hur designers försöker uppnå en mer naturlig dialog mellan VUI-system och dess användare. Detta genom att skapa VUI-system som uppför sig och uttrycker sig som människor, med fokus på flexibilitet, intuitivitet och tillgänglighet. Sugisaki och Bleiker (2020) kommenterar att för att en naturlig dialog ska kunna skapas, behövs ett språk utvecklas där både system och användare förstår varandra på ett korrekt sätt. Därför skulle naturlig dialog kunna vara en utvecklingsmöjlighet att ta i beaktning.

Under observationerna togs även svårigheter med dialekter eller könstillhörighet i beaktning, om ytterligare problem uppstod i dialogen mellan användare och smart högtalare. Shafei och Tan's studier (2022) lyfte hur smarta högtalares prestation ofta försämras med icke-engelsktalande användare. Rajan, Udeshi och Chattopadhyay (2022) uppmärksammade i sina studier att även könstillhörighet kan öka möjligheten till error vid användning av smarta högtalare. I uppsatsens studie observerades, i vissa fall, tydliga dialektala skillnader när deltagarna pratade med författarna jämfört med när de pratade med sina smarta högtalare. Sara pratade med betydligt mindre dialekt när hon konverserade med sin smarta

högtalare, och Pelle valde att aktivt avstå från att använda skånska uttryck. Däremot kunde inga skillnader upptäckas gällande hur mycket error som uppstod baserat på dialekter eller könstillhörigheter.

Efter att ha sammanställt och analyserat resultaten från både observation och intervju framgick det att vissa användare hade problem med inställningarna som krävdes för att använda VUI'n till dess fulla potential. Deltagarna i studien utförd av Kim et al. (2021), samtliga designers, menade att ett VUI-system bör designas för att kunna erbjuda användare möjligheten att enkelt ställa in personliga profiler och preferenser. Vidare uppmärksammades även problem vid kommandon och hur dessa skulle ändras och omformuleras, något som lyfts i ett flertal studier (Alrumayh & Tan, 2022; Clark et al., 2019; Kirschthaler et al., 2020; Moore & Urakami, 2022; Myers et al., 2018). Vid utformningen av guidelines kunde sex olika aspekter urskiljas, dessa kunde sedan delas in i två olika kategorier av error feedback: *Inställningar* och *Användning*.

Under studiens gång framkom önskan från deltagarnas sida om ett mer anpassningsbart gränssnitt där valet av ton, sätt att erhålla instruktioner och även stöd kunde göras. Liknande resultat gällande behov för ett mer adaptivt system har uppmärksammats i studier gjorda av Myers (2019). Det är dock av vikt att anpassningsbarheten inte blir en belastning för användarna, något som deltagaren Maria också nämner i sin återkoppling angående de guidelines som tagits fram. Möjligheten att kunna anpassa systemet efter användarnas behov bör endast ses som ett hjälpmedel, det får inte bli ett hinder för användarna eller ett krav för att få VUI-systemet att fungera.

Några av deltagarna i uppsatsens studie hade längre vana av systemet och visade även en större anpassning för hur kommando och inställningar gavs. Dessa var även mer benägna att erinra sig från tidigare erfarenheter och kunde därför följa instruktioner eller lösa kommando och problem som uppkom snabbare. Detta låg i linje med vad studier av Myers et al. (2018) såg bland sina användare, även om majoriteten fortsatte att gissa sig fram till hur kommandon skulle formuleras. De användare som hade minst vana att interagera med VUI-systemet uttryckte däremot önskan att få mer utförliga instruktioner, ibland med mer seriös och

utbildande ton, som kunde pausas, upprepas eller läsas upp stegvis. Killingback et al. (2019) och Jones och Hilton (2022) lyfter hur auditiv feedback har möjligheten att ge ett bra stöd vid anpassning av nya situationer, men poängterar vikten av just upprepningmöjligheter precis som studier av Sarcona et al. (2020) och Ice et al. (2019). Deltagarna i denna uppsats studie uppvisade vid flera tillfällen frustration kring att inte hinna med eller förstå vad VUI-enheten förklarade för dem.

Kirschthaler et al. (2020) testade i sin studie att erbjuda deltagare två olika stödfunktioner. Den första gav utförliga instruktioner på varje kommando, medan den andra enbart erbjöd stöd i de fallen där användare önskade det. Utifrån resultatet av deras studie märktes en högre grad av interaktivitet och engagemang. De förespråkade därför ett mer anpassningsbart system, beroende på användarnas vana att hantera VUI-system. Resultatet av denna uppsats studie visade att ett mer anpassningsbart system är att föredra för VUI-användare. Kirschthaler et al. (2020) och Myers' (2019) förslag på ett mer adaptivt system låg därför i linje med denna studies resultat. Genom en guideline för ett anpassningsbart system där användaren själv kan ange önskan om hur och vad för feedback som ges, ökar chansen för att användaren ska kunna lösa de problem som uppkommer.

Möjligheten att även välja tonläge kan vara önskvärt för vissa användare. I designers önskan att skapa en mer naturlig dialog kan fenomenet *The Uncanny Valley-effect* uppfattas och avskräcka användare från att använda VUI-enheten eller förstå vad den försöker säga (Diederich et al., 2020). Eftersom systemet även uppvisade problem i igenkänningsstadiet och inte förstod vissa kommandon skulle valmöjligheten att ändra tonläge kunna öka användarnas acceptans för problem som uppstår (Lee et al., 2019).

Möjligheten att upprepa feedback, eller dela upp det stegvis har även uppmärksammats i studier utförda av Ice et al. (2019) där studenterna anmärkte behovet av upprepning för att kunna ta till sig och förstå informationen. Detta kan appliceras på användare av VUI's när det kommer till delgivning av instruktioner och behovet av att kunna få dessa upprepade. Killingback et al. (2019) och Sarcona et al. (2020) menar att auditiv feedback har fördelen att kunna engagera och uppmuntra användare. Däremot poängterar de precis som Ice et al. (2019) att

auditiv feedback har behovet av just upprepning, vilket speciellt gäller för nya användare av ett VUI-system innan de vant sig vid auditiv feedback.

Jones och Hilton (2022) lyfter hur olika feedback-metoder påverkar hur användaren tar till sig och använder den. De menar att auditiv feedback hjälper till att förstärka mottagarens upplevelse av inläring genom känslor av tillhörighet, närvaro och ökat självförtroende. Däremot argumenterar de för att visuell feedback är bättre för att ge summeringar och instruktioner. Det är därför av vikt att vara extra uppmärksam vid designen av informationen som ges. Holland (2020) menar å andra sidan att auditiv feedback ger en känsla av högre detaljnivå och kan därför vara ett motargument till visuell feedback som kan vara svårare att förstå. Utifrån Hollands guidelines (gällande framtagning av auditiv feedback), samt Jones och Hiltons (2022) argument mot auditiv feedback vid instruktionsgivning, erbjuder uppsatsens guidelines för att anpassa hur instruktionsgivning framförs en medelväg för designers vid framtagning av VUI-system. Genom att tillåta användare möjligheten att själva bestämma hur instruktionerna ges ökar chansen att målet uppnås och låser även upp möjligheten för användarna att utnyttja fler av VUI-enheternas funktioner.

I dagsläget skapas VUI-system med för lite minneskapacitet enligt Cambre och Kulkarni (2020). Även Murad et al. (2021), Yan et al. (2021) och Wei och Landlay (2018) har lyft problemet med att bristande minneskapacitet ger negativa konsekvenser för VUI-användare. Möjligheten att utöka minneskapaciteten för VUI-enheter ökar chansen för en mer utförlig dialog och ett större utbud av kunskap kan erhållas. Strävan att erbjuda en mer naturlig dialog mellan användare och VUI-enhet kan därför öka om kommando, gensvar och följdfrågor möjliggörs genom ökat minne. Ökad minneskapacitet kan även ge VUI-enheten möjligheten att erbjuda användaren nyckelord som uppfattats i ett kommando och därför hjälpa dem uppfatta vart i kommandot justering behövs. Detta är någonting deltagarna i uppsatsens studie uppmärksammat som ett problem, de visste inte vad i kommandot som inte förstods. En ökad minneskapacitet i VUI-system kan även hjälpa till att förhindra användarens risk för informationsöverbekantning, något som Holland föreslår i sin guidelines för framtagning av auditiv feedback (2020).

Även mer utförliga hjälpmedel i form av kommandoförslag kan vara av nytta, speciellt för nyare användare, innan de lärt sig vilka formuleringar som fungerar och behöver extra stöd (Kirschthaler et al., 2020; Holland, 2020; Myers, 2019). Denna typ av kommandoförslag kan förhoppningsvis motverka den negativa upplevelsen av felaktiga svar eller problem som uppstår, något som Lee et al. (2019) menar försämrar användares uppfattning om VUI-systemets kapacitet.

Andra aspekter som uppmärksammas under studierna, men som inte faller direkt inom error feedback, är bland annat svårigheten med andningspaus i ett kommando. Enligt Cambre och Kulkarni (2020) ska VUI'n tillåta fem till åtta sekunder av tystnad i ett kommando för att ge användaren tid att andas och tänka. I denna studiens observation uppmärksammades dock att den smarta högtalaren, vid samtliga fall av andningspauser, sa att den inte förstod kommandot trots att fem sekunder inte passerat. Inställningen för andningspauser ligger inte inom uppsatsens syfte, men är däremot något som kan skapa problem för användaren. På grund av detta är det av vikt att pauser implementeras på ett effektivare sätt som tillåter användare att både andas och tänka vid ett kommando. Vidare uppmärksammades även att användare slutar utforska funktioner och endast använder sina smarta högtalare till få specifika funktioner. Detta ligger i linje med vad Kirschthaler et al. (2020) och Kim et al. (2022) har uppmärksammat i sina studier.

6. Slutsatser och Framtida studier

Målet med uppsatsens studie var att undersöka hur svenska användare, som använder minoritetsspråket svenska, uppfattar bristen på error feedback från smarta högtalare. Detta med syftet att ta fram förslag på guidelines för designers som skapar VUI-system med svenska användare. Genom att bryta ner frågeställningen *Hur skulle guidelines som hjälper designers vid framtagning av error feedback för VUI-system, med svenska användare, kunna se ut?* i tre subfrågor kunde en större förståelse kring ämnet VUI skapas och mer kunskap genereras från användares uppfattning av dagens error feedback.

Observationerna visade att användare hade svårt att förstå eller följa vissa av VUI-systemets instruktioner. Vidare fick de ibland ingen utförlig feedback alls för att hjälpa dem lösa sitt problem. Detta kunde därför svara på frågan: *vilken feedback som var problematisk, samt varför?* Intervjuerna efteråt gav en inblick i *hur* vissa svenska användare uppfattar den error feedback de får av smarta högtalare idag? Deltagarna uppvisade i vissa fall frustration eller uppgivenhet vid feedbacken "ursäkta, jag förstår inte" och uttryckte en önskan om mer utförlig feedback för att förstå vart i kommandot som förståelsen brast. Vidare gav intervjuerna användarna möjligheten att uttrycka vad de saknar i olika sorters feedback de får från VUI-system samt förslag på förbättringar och andra synpunkter. Detta kunde därför generera kunskap om *hur error feedback skulle kunna se ut i stället?*

De slutsatser som kunde dras är därför att den error feedback VUI-system idag erbjuder sina användare behöver utvecklas. Detta för att ge användarna en ökad möjlighet att lösa eventuella problem, bidra till en mer naturlig dialog och därmed ge en övergripande bättre upplevelse av sina smarta högtalare. Utifrån det sammanställda resultatet av uppsatsens studier kunde ett svar på frågeställningen fås och sex förslag på guidelines kunde tas fram. Uppsatsens sex guidelines utgår från två teman: *inställningar* vilket består av *anpassningsbart system*, *anpassningsbar ton* och *instruktioner*, samt *användning* som består av *minneskapacitet*, *nyckelord* och *kommandoförslag*. Dessa hoppas kunna erbjuda designers ett hjälpmedel för hur error feedback från VUI-system kan tas fram, utvecklas och användas för att hjälpa svenska användare av smarta högtalare.

Uppsatsens studier genomfördes i en liten skala och för framtida studier hade ett större urval av deltagare önskats för att ge ett generaliserbart resultat. Resultatet från observationerna och intervjuerna genererade dock tillräcklig information för framtagning av förslag till guidelines för error feedback. Framtida undersökningar skulle även kunna fokusera på hur dessa guidelines används och uppfattas av designers, samt hur de presterar i användartester. Även vidare undersökningar kring alternativ error feedback och studier "*in the wild*" som Diederich et al. (2022) och Clark et al. (2019) lyfter i sina artiklar, samt eventuella svårigheter och problem gällande olika röstlägen baserade på könstillhörighet och dialekter som Shafei och Tan (2022) nämner, behöver undersökas ytterligare.

7. Referenser

Alrumayh, A.S. & Tan, C.C. (2022). VORI: A framework for testing voice user interface interactability. *High Confidence Computing*, 2(3), ss. 1-8.

doi: 10.1016/j.hcc.2022.100069

Baimirov, K., Mergengali, E. & Baimirov, B. (2022). Overview of the latest research related to smart speakers. I *Proceedings of 2022 IEEE 7th International Conference (ENERGYCON)*, Riga, Latvia 9-12 maj 2022, ss. 1-5.

doi: 10.1109/ENERGYCON53164.2022.9830196

Cambre, J. & Kulkarni, C. (2020). Methods and Tools for Prototyping Voice Interfaces. I *2nd Conference on Conversational User Interfaces*, Bilbao, Spanien 22-24 juli 2020, ss. 1-4.

doi: 10.1145/3405755.3406148

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., Edwards, J. & Cowan, B. (2019). The State of Speech in HCI: Trends, Themes and Challenges. *Interacting with Computers*, 31(4), ss.349-371.

doi: 10.1093/iwc/iwz016

Denscombe, M. (2018). *Forskningshandboken: För småskaliga forskningsprojekt inom samhällsvetenskaperna*. (4 ed.). Lund: Studentlitteratur AB.

Dickson, B. (2018). The fundamental problem with smart speakers and voice-based AI assistants. *TechTalk* [blog], 20 januari.

<https://bdtechtalks.com/2018/09/03/challenges-of-smart-speakers-ai-assistants/>

Diederich, S., Brendel, A., Morana, S. & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems*, 23(1), ss. 96-138.

doi: 10.17705/1jais.00724

Holland, M. (2022). Audio Feedback Project: A Project to Increase Social Presence in a Virtual Library and Knowledge Service. *Evidence Based Library and Information Practice* 2022, 17(2), ss. 1-13.

doi: 10.18438/eblip30006

Ice, P., Curtis, R., Phillips, P. & Wells, J. (2019). Using asynchronous audio feedback to enhance teaching presence and students' sense of community. *Online learning*, 11(2), ss. 1-23.

doi: 10.24059/olj.v11i2.1724

Jones, D. & Hilton, C. (2022). Audio feedback in distance design education. *Design and Technology Education: An International Journal*, 27(1) ss. 113–135.

doi: ojs.lboro.ac.uk/DATE/article/view/3078/3072

Killingback, C., Ahmed, O. & Williams, J. (2019). It was all in your voice - Tertiary student perceptions of alternative feedback modes (audio, video, podcast, and screencast): A qualitative literature review. *Nurse Education Today*, 72, ss. 32-39.

doi: 10.1016/j.nedt.2018.10.012

Kim, H., Hwang, S., Kim, J. & Lee, Z. (2022). Toward Smart communication Components: Recent Advances in Human and AI Speaker Interaction.

Electronics, 11(10), ss. 1-15.

doi: 10.3390/electronics11101533

Kim, Y., McGrenere, J., Reza, M. & Yoon, D. (2021). Designers Characterize Naturalness in Voice User Interfaces: Their Goals, Practices, and Challenges. I *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama, Japan 8-13 maj 2021, ss.1-13.

doi: 10.1145/3411764.3445579

Kirschthaler, P., Porcheron, M. & Fischer, J. (2020). What Can I Say? Effects of Discoverability in VUIs on Task Performance and User Experience. I

Proceedings of the 2nd Conference on Conversational User Interfaces. Bilbao,

Spain 22-24 juli 2020, ss.1-9.

doi: 10.1145/3405755.3406119

Lee, S., Cho, M. & Lee, S. (2020). What If Conversational Agents Became Invisible?: Comparing Users' Mental Models According to Physical Entity of AI Speaker. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3), ss. 1-24.

doi: 10.1145/3411840

Lee, D., Sah, Y.J. & Lee, S. (2019). Improving Usability Perception of Error-Prone AI Speakers: Elaborated Feedback Mitigates Negative Consequences of Errors. *International Journal of Human-Computer Interaction*, 35(17), ss. 1645-1652.

doi: 10.1080/10447318.2018.1561069

Moore, B.A. & Urakami, J. (2022). The impact of the physical and social embodiment of voice user interfaces on user distraction. *International Journal of Human-Computer Studies*, (161). Tillgänglig: ELSEVIER, Science Direct.

doi: 10.1016/j.ijhcs.2022.102784.

Murad, C. & Munteanu, C. (2022). Voice-First Interfaces in a GUI-First Design World: Barriers and Opportunities to Supporting VUI Designers On-the-Job. I *Proceedings of the 4th Conference on Conversational User Interfaces*, Glasgow, United Kingdom 26-28 juli 2022, ss. 1-10.

doi: 10.1145/3543829.3543842.

Murad, C., Munteanu, C., Cowan, B. & Clark, L. (2021). Finding a New Voice: Transitioning Designers from GUI to VUI Design. I *Proceedings of the 3rd Conference on Conversational User Interfaces*. Bilbao, Spain 27-29 juli 2021, ss.1-12.

doi: 10.1145/3469595.3469617

Myers, C. (2019). Adaptive suggestions to increase learnability for voice user interfaces. I *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*. Marina del Ray, California, US 16-20 mars 2019, ss. 159-

160.

doi: 10.1145/3308557.3308727

Myers, C., Furqan, A., Nebolsky, J., Caro, K. & Zhu, J. (2018). Patterns for how users overcome obstacles in voice user interfaces. I *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal, QC, Canada 21-26 april 2018, ss. 1-7.

doi: 10.1145/3173574.3173580.

Nielsen, J. (1994). *10 Usability Heuristics for User Interface Design*.

<https://www.nngroup.com/articles/ten-usability-heuristics> [2023-02-17]

Norman, D.A. (2002). *The Design of Everyday Things*. (2 ed.) New York: Basic Books

NPR & Edison Research (2022). *The Smart Audio Report*. National public media.

<https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/>

Patel, R., & Davidson, B. (2019). *Forskningsmetodikens grunder: att planera, genomföra och rapportera en undersökning* (5 ed.). Lund: Studentlitteratur AB.

Rajan, S.S., Udeshi, S. & Chattopadhyay, S. (2022). AequoVox: Automated Fairness Testing of Speech Recognition Systems. I *25th International conference on fundamental approaches to software engineering*, München, Tyskland 2–7april 2022, ss. 245–267.

Sarcona, A., Dirhan, D. & Davidson, P. (2020). An overview of audio and written feedback from students' and instructors' perspective. *Educational Media International*, 57(1), ss. 47-60.

doi: 10.1080/09523987.2020.1744853

Seaborn, K., Miyake, N.P., Pennefeather, P. & Otake-Matsuura, M. (2021). Voice in Human-Agent Interaction: A Survey. *ACM Computing Surveys*, 54(4), ss. 1-43.

doi: 10.1145/3386867

Seaborn, K. & Urakami, J. (2021). Measuring voice UX quantitatively: A rapid review. I *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan 8-13 maj 2021, ss. 1-8
doi: 10.1145/3411763.3451712

Shafei, H.A. & Tan, C.C. (2022). Do smart speaker skills support diverse audiences? *Pervasive and Mobile Computing*, 87, ss. 1-18.
doi: 10.1016/j.pmcj.2022.101716

Sugisaki, K. & Bleiker, A. (2020). Usability guidelines and evaluation criteria for conversational user interfaces: a heuristic and linguistic approach. I *Proceedings of Mensch und Computer 2020*. Magdeburg, Tyskland 6-9 september 2020, ss. 309-319.
doi: 10.1145/3404983.3405505

Sutton, S.J., Kirk, D., Foulkes, P. & Lawson, S. (2019). Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. I *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow, Scotland 4-9 maj 2019, ss.1-14.
doi: 10.1145/3290605.3300833

Svenskarna och internet. (2022). *Det uppkopplade hemmet och smarta prylar (Rapport Svenskarna och internet 2022:9)*.
<https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2022/det-uppkoppladehemmet-ochsmarta-prylar/>

van Gils, J. (2020). How smart speakers will transform the way we live! *iGadgetsWorld* [blog], 20 januari. <https://www.igadgetsworld.com/smart-speakers-history-future/>

Vetenskapsrådet. (2017). *God forskningssed*. Stockholm: Vetenskapsrådet.
Hämtad från: <https://www.vr.se/analys/rapporter/vara-rapporter/2017-08-29-god-forsknings-sed.html>

Wei, Z. & Landay, J.A. (2018). Evaluating Speech-Based Smart Devices Using New Usability Heuristics. *IEEE Pervasive Computing*, 17(2), ss. 84-96.

doi: 10.1109/MPRV.2018.022511249

Yan, C., Zhang, G., Ji, X., Zhang, T., Zhang, T. & Xu, W. (2021). The Feasibility of Injecting Inaudible Voice Commands to Voice Assistants. *IEEE Transactions on dependable and secure computing*, 18(3), ss. 1-17.

doi: 10.1109/TDSC.2019.2906165

8. Bilagor

1. Förstudie enkät

https://docs.google.com/forms/d/1UbXg8iPNIL9azQ_XApfcPyR9B6UCwljsLCCWel9aahA/edit?ts=63ca6868

Undersökning av Google Assistant med svenska kommandon

Hej!

Vi är två kandidatexamen studenter som skriver vår C-uppsats inom Informatik. Det ämne vi valt att undersöka är Röstassistenter, begränsat till Google Assistant, på svenska och hur de fungerar i verkligheten.

Genom att delta i denna enkät förblir du helt anonym, men ålder, kön och vag geografisk plats (för att uppmärksamma eventuella dialektala begränsningar) kommer registreras. Insamlat material kommer enbart ses och behandlas av oss, samt delas med Högskolan Kristianstads Informatiks-fakultet.

Tack för er medverkan!

Jeanette och Julia

Använder du din Google Assistant på svenska? *

- Ja
- Nej

Vad är din könsidentitet? *

- Kvinna
- Man
- Annat
- Vill ej uppge

När är du född? *

1. Innan 1920
2. 1920 - 1929
3. 1930 - 1939
4. 1940 - 1949
5. 1950 - 1959
6. 1960 - 1969
7. 1970 - 1979
8. 1980 - 1989
9. 1990 - 1999
10. 2000 - 2009
11. 2010 och senare

Hur ofta använder du Google Assistant? *

- Dagligen
- Veckovis
- Månadsvis
- Nästan aldrig

I vilket län bor du? *

1. Blekinge
2. Dalarna
3. Gotland
4. Gävleborg
5. Halland
6. Jämtland
7. Jönköping
8. Kalmar
9. Kronoberg
10. Norrbotten
11. Skåne
12. Stockholm
13. Södermanland
14. Uppsala
15. Värmland
16. Västerbotten
17. Västernorrland
18. Västmanland
19. Västra Götaland
20. Örebro
21. Östergötland

På vilken enhet använder du Google Assistant (ex. mobil, dator, smart högtalare) *

Short-answer text

Hur upplevde du installationen av Google Assistant? *

	1	2	3	4	5	
Mycket svårt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Mycket enkelt

Kommentarer om installationen

Long-answer text

Hur upplevde du första användningen av Google Assistant? *

	1	2	3	4	5	
Mycket svårt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Mycket enkelt

Kommentarer om första användningen

Long-answer text

Använde du dig av hjälpmedel för att lära dig navigera din Google Assistant? *

- Ja
- Nej

Om ja, vad för hjälpmedel använde du?

Long-answer text

Om nej, vad gjorde du för att lära navigera din Google Assistant?

Long-answer text

Upplever du några specifika problem med din Google Assistant? *

- Nej
- Att förstå dina kommandon
- Att veta vad assistenten kan göra
- Brist på visuell feedback (tex, vid shopping veta vad du har lagt in kundvagn)
- Att du inte hör/förstår vad assistenten säger
- Brist på utförlig feedback vid felaktigt kommando (tex, när assistenten inte förstår dig)
- Assistenten ger inte svar på frågan du ställde
- Other...

Har du förslag på förbättringar för din Google Assistant? *

Long-answer text

2. Sökblock

<https://docs.google.com/document/d/1QpdI->

[KS10SXIP4DD9gYY7GskvSIOv6VMITTkUR9NcLA/edit?usp=sharing](https://docs.google.com/document/d/1QpdI-KS10SXIP4DD9gYY7GskvSIOv6VMITTkUR9NcLA/edit?usp=sharing)

VUI	Design	Feedback	Users
voice assistant	guidelines	interaction	perception
voice user interface	heuristics	feedback	mental models
smart speaker	methods	communication	experience
conversational assistant	tools	sound	discoverability
conversational user interface	naturalness	audio	user study
google assistant	user friendliness	speech	learnability
	dialogue	error	

3. Sökschema

<https://docs.google.com/document/d/1MAPx95ygboKHIXAFPUTN0uS6baf2wZhqfS7eyY9A2DY/edit?usp=sharing>

Sökverktyg	Sökord (kombination)	Avgränsningar	Antal träffar
HKR Summon	(voice assistant*) AND (design*)	Peer-review, max 5 år	327
HKR Summon	(voice assistant*) AND (user experience*)	Peer-review, max 5 år	107
HKR Summon	(smart speaker*) AND (design*) AND (experience*)	Peer-review, max 5 år	72
HKR Summon	((voice assistant*) OR (voice user interface*)) AND ((user experience*) OR (perception*))	Peer-review, max 5 år	380
HKR Summon	((voice assistant*) OR (vui*)) AND (error feedback*)	Peer-review, max 5 år	11
ACM Digital Library	[["voice assistant"] OR ["voice user interface"] OR [smart speakers*]] AND [["interaction"] OR ["feedback"] OR ["communication"]] AND [["perception"] OR ["metal models"] OR ["user experience"]]	Peer-review, max 5 år, Proceedings of the ACM On Human-Computer Interaction	254
ACM Digital Library	["google assistant"] OR [smart speaker*] AND [["interaction"] OR ["feedback"] OR ["communication"]] AND [["discoverability"] OR ["user experience"]]	Peer-review, max 5 år, Proceedings of the ACM On Human-Computer Interaction	186

ACM Digital Library	[[vui*] OR ["voice user interface"]] AND [{"design"}] OR [method*]]	Peer-review, max 5 år	405
IEEE Xplore	(voice user interface OR vui) AND (design) AND (user)	Conferences, Journals, 2019-2023	231
IEEE Xplore	(conversational agent OR conversational assistant OR Google assistant) AND (feedback OR error)	2019-2023	54
ResearchGate	voice user interface, design, user experience	Since 2019, Articles, Conference papers, Only full-texts	N/A

Sparade artiklar för djupgranskning: 53
 Artiklar valda efter djupgranskning: 23
 Artiklar funna utefter referenser: 7

4. Information om studie till deltagare

Studie kring Google Assistant genom [Smarta Högtalare](#).

Ett C-uppsats arbete våren 2023 av Jeanette Elfving och Julia Blom.

Observation

Du kommer att få 21 uppgifter som du ska förmedla till Google Assistant via din smarta högtalare. När en uppgift är löst kan du gå direkt vidare till nästa. Om Google Assistant inte kan lösa uppgiften ser vi gärna att du försöker att lösa den utefter bästa förmåga och uppmuntrar till att testa dig fram. Du får göra hur många försök du vill. Julia och Jeanette kommer att observera din interaktion med högtalaren och kommer således inte lägga sig i, kommentera gärna högt vad som händer, hur du tänker eller liknande.

Din identitet kommer att förbli anonym, men ålder och länsstillhörighet kommer att publiceras. Du får lov att avbryta observationen när du vill och även hoppa över uppgifter. Det insamlade materialet kommer endast att ses av Julia och Jeanette för att kunna slutföra sin C-uppsats. Materialet kommer att raderas när C-uppsatsen är publicerad.

Intervju

Du kommer att bli intervjuad om din upplevelse av interaktionen med [din smarta högtalare](#) direkt efter att observationen är genomförd. Du får lov att avbryta intervjun om du vill. Det insamlade materialet kommer endast att ses av Julia och Jeanette för att kunna slutföra sin C-uppsats. Materialet kommer att raderas när C-uppsatsen är publicerad.

Tack för ditt deltagande!

5. Observationsuppgifter

Uppgifter

1. Sätt igång låten "Imagine" med The Beatles.
2. Skruva upp volym.
3. Skruva ner volymen.
4. Byt låt.
5. Spola fram 30 sekunder.
6. Pausa musiken.
7. Sätt igång "Creepypodden i P3".
8. Stäng av podden.
9. Sätt igång "nordic mythology podcast".
10. Stäng av podden.
11. Sätt igång en timer på 10 sekunder.
12. Stoppa timer.
13. Sätt igång ett alarm om 10 sekunder.
14. Stoppa alarm.
15. Lägg till ett event för lunch den 14 april kl 1 på eftermiddagen.
16. Ta bort eventet för lunch den 14 april kl 1 på eftermiddagen.
17. Ta reda på vem den 15:de ministern av Kanada var.
18. Ta reda på mer information om den 15:de ministern av Kanada.
19. Ta reda på temperaturen i New York både i C och F.
20. Ta reda på om det finns det några veganska restauranger i din närhet.
21. Ta reda på bästa rutten till Kristianstad högskola från din nuvarande plats.

6. Tema för intervjuer

Hur upplevde du kommandona? Använder du liknande själv i din vardag?

Åsikter:

Önskvärt:

Vad tycker du om de tillgängliga funktionerna som VUI erbjuder?

Åsikter:

Önskvärt:

Lingvistik - Kändes språket VUI använder naturligt?

Åsikter:

Önskvärt:

Vad anser du om error feedbacken du får från VUI'n? Lätt att följa med i anvisningar?

Hur löser du det?

Åsikter:

Önskvärt: