GÖTEBORGS UNIVERSITET

Bör intelligenta assistenter tala naturligt?

En undersökning om betydelsen av intelligenta assistenters tal för användarupplevelsen

Is natural speech desirable in intelligent assistants?

A study of the relevance of intelligent assistants' speech for the user experience

PAULINE LORIN
LINN THORSAGER

Kandidatuppsats i kognitionsvetenskap

Rapport nr. 2017:125

Göteborgs universitet
Institutionen för tillämpad informationsteknologi
Göteborg, Sverige, maj 2017

Sammanfattning

Intelligenta personliga assistenters tal är idag begripligt. Trots detta låter det fortfarande konstgjort och talets prosodi tycks vara en av de stora svårigheterna. Både företag och forskare inom ämnet strävar efter en naturlighet i syntetiskt tal, men om detta är något som är eftersträvansvärt har dock varken motbevisats eller bekräftats. Syftet med denna studie var att undersöka betydelsen av intelligenta personliga assistenters tal för användarupplevelsen. Apples intelligenta personliga assistent Siri valdes för att utföra ett experiment med inomgruppsdesign bestående av tre nivåer: Siris tal, naturligt tal som efterliknar Siris prosodi samt naturligt tal. Resultatet visade att ingen statistiskt signifikant skillnad förelåg mellan de tre nivåerna och därvid kunde ingen skillnad påvisas mellan naturligt tal och Siris tal gällande användarupplevelsen. Inte heller huruvida prosodin har en inverkan på användarupplevelsen kunde påvisas. Ytterligare statistisk analys konstaterade att en statistiskt signifikant skillnad förelåg mellan de naturliga talen och Siris tal angående hur enkla de var att förstå, där Siris tal var signifikant svårare att förstå. Att en statistiskt signifikant skillnad går att påvisa vad gäller förståelsen av intelligenta assistenters tal men inte för användarupplevelsen i sin helhet kan vara en intressant upptäckt, men kräver vidare undersökning.

Nyckelord: Intelligent personlig assistent, användarupplevelse, talsyntes, Siri, prosodi, människa-datorinteraktion, konkateneringssyntes.

Modern intelligent personal assistants are comprehensible. Despite this, their speech still sounds artificial, where the prosody of the speech seems to be one of the major difficulties. Both companies and researchers in the subject strive for naturalness in synthetic speech, but if this is something that is desirable has neither been demented nor confirmed. The purpose of this study was to investigate the importance of intelligent personal assistants' speech for the user experience and if it increases with more naturalness. The intelligent personal assistant Siri was chosen to perform an experiment with repeated measures design consisting of three levels: Siri's speech, natural speech mimicking Siri's prosody and natural speech. The results revealed no statistically significant difference between the three levels. Hence, no difference was found between natural speech and Siri's speech regarding the user experience. The study was also not able to confirm whether prosody has an impact on the user experience. Further statistical analysis found that there was a statistically significant difference between the natural speeches and Siri's speech as to how easy they were to understand, where Siri's speech was significantly more difficult to understand. That a statistically significant difference can be found in terms of understanding the intelligent assistants but not for the user experience could be an interesting discovery, but requires further investigation.

Keywords: Intelligent personal assistant, user experience, speech synthesis, Siri, prosody, human-computer interaction, concatenative speech synthesis

Förord

Uppsatsen är ett resultat av en genomgående gemensam process av Pauline Lorin och Linn Thorsager. Därmed står båda ansvariga för dess utformning, innehåll och bakomliggande arbete.

Ett stort tack till Gustaf Lindblad för hans engagemang och handledning. Vi vill även tacka personen bakom ‘de naturliga talen’, Hannah Sommargren, för att vi fick låna hennes röst.

Innehållsförteckning

1 Inledning

“Siri, vilken är din favoritfärg?”. “Min favoritfärg är… tja, den är lite grönaktig, men med fler dimensioner” är svaret. Rösten är hackig, talmelodin underlig och betoningen på ‘tja’ låter snarare som hälsningsfrasen än som en fundering. Siri kan hjälpa till med ett flertal saker då händerna är upptagna, ’hon’ har oftast inga problem med att höra vad som sägs och inte heller med att göra sig förstådd (Savage, 2017, mars). Men trots en ömsesidig förståelse mellan Siri och användaren innebär det inte att Siri är färdigutvecklad. Talet är inte alltid behagligt att lyssna på och vissa uttal kan komma att kräva särskild ansträngning för full förståelse.

Siri är en intelligent personlig assistent utvecklad av Apple och programvaran är tillgänglig i flertalet av deras produkter, såsom iPhone och Mac (Apple, 2017). En intelligent personlig assistent kan förklaras som en mjukvaruagent vars syfte är att hjälpa en användare med diverse problem eller uppgifter (Ponciano, Pais & Casal, 2015). Hur interaktionen med en intelligent assistent sker skiljer sig från assistent till assistent, men kommunikationen med Siri sker främst genom röststyrning (Preece, Rogers & Sharp, 2015). Det är möjligt att bland annat be Siri att boka in möten, skicka ett meddelande eller svara på frågor (Apple, 2013).

Siri presenterades för första gången i samband med smartmobilen iPhone 4S i oktober 2011 (Apple, 2011). Det blev då möjligt att tala med sin enhet, både för att styra den och för att ställa frågor. Dessutom producerades ett talat svar, något som Apple tillsammans med Siri var först med att introducera till allmänheten (Campbell & Li, 2015). Det talade svaret var skapat med hjälp av talsyntes, eller text-till-tal, som är den teknik som används vid skapandet av syntetiskt tal. Apple och Siri följdes dock av flertalet konkurrenter såsom Google, Microsoft och Intel. För utvecklingen inom talsyntes innebar det ett nytt syfte, att ha ett konverserande tal framför att endast ”läsa högt” som tidigare varit syftet. Detta innebar inte bara ett annat sätt att tala på utan även en mycket mer komplicerad process för att få det att låta naturligt.

Dagens intelligenta assistenter är begripliga, syntetisk meningsuppbyggnad och uttal är under ständig utveckling, men begriplig behöver inte innebära naturlig (Tatham & Morton, 2005). Språkets komplexitet sätter idag begränsningar på känslan av naturlighet i det syntetiska talet och Tatham och Morton (2005) menar på att ett av de större hindren är talets prosodi. Prosodi är en del av naturligt tal som inte är faktiska ord eller meningar, utan snarare förmågan att kunna variera talet med till exempel rytm, ton och melodi. I stora drag är det prosodin som gör talet levande, och utan naturlig prosodi kan det syntetiska talet uppfattas som hackigt, konstlat eller monotont. Huruvida syntetiskt tal bör låta naturligt tycks dock endast tas för givet, Lee (2010) menar att området inte är tillräckligt studerat för att dra en sådan slutsats. Ett tal som låter näst intill naturligt men som egentligen är syntetiskt skulle kunna framkalla fenomenet *uncanny valley*, en effekt som ursprungligen presenterades av Masahiro Mori (1970) och innebär en känsla av obehag då en artefakt är mycket människolik. Romportl (2014) har testat huruvida detta fenomen framkallas vid interaktion med syntetiskt tal. Två syntetiska tal med olika grad av naturlighet användes i studien. Resultatet kunde inte styrka huruvida *uncanny valley* framkallades, dock med viss diskussion kring huruvida respondenternas tekniska bakgrund kan ha påverkat utfallet. Huruvida fenomenet existerar då talet är desto mer naturligt kräver dock vidare undersökning.

Om det är önskvärt att ha ett människolikt tal tycks även tas för givet av de företag som jobbar med att ta fram bättre algoritmer för att inkludera en mer naturlig prosodi, exempelvis Deepminds (2016) Wavenet. Vad som borde vara av relevans i

sammanhanget är huruvida detta innebär en förhöjd positiv användarupplevelse, något som enligt författarnas egen observation ännu inte är studerat.

1.1 Syfte

Syftet med denna studie är att undersöka betydelsen av intelligenta personliga assistenters tal för användarupplevelsen.

1.2 Frågeställningar

Syntetiskt tal kommer att jämföras med naturligt tal och naturligt tal som efterliknar syntetisk prosodi. Bidrar ett mer naturligt tal hos intelligenta personliga assistenter till en förhöjd användarupplevelse? Påverkar talets prosodi användarupplevelsen av intelligenta personliga assistenter?

1.3 Avgränsningar

Studien kommer endast att använda sig av en implementering av syntetiskt tal, Apples Siri (i macOS Sierra version 10.12.2). Valet föll på Siri då hon är en av de bättre intelligenta personliga assistenterna på marknaden, men också för att hon talar svenska (Dunn, 2016, 4 november). Det finns flertalet metoder för att skapa syntetiskt tal. Då företaget Nuance tycks ligga bakom Siris tal kommer studien främst fokusera på den typ av talsyntes som de använder sig av, så kallad konkateneringssyntes (Anderson, 2013, 17 september). Vidare undersöks naturligt tals betydelse för användarupplevelsen och lämnar det till framtida studier att undersöka eventuella andra parametrar. Även termen naturligt tal är bred och kommer således avgränsas genom att fokusera på prosodi, den faktor som framförallt tycks vara problematisk i syntetiskt tal (Tatham & Morton, 2005).

2 Bakgrund

Bakgrunden behandlar intelligenta personliga assistenter, framförallt Siri, hur de styrs och hur de fungerar med syfte att skapa kunskap om dess funktion samt förståelse för vilken roll talet har. Vidare förklaras produktionen av naturligt tal följt av prosodins syfte, för att förbereda läsaren för nästkommande avsnitt som behandlar syntetiskt tal och problematiken med prosodi. Syntetiskt tal följs av de två avsnitten människa-teknikinteraktion samt användarupplevelse.

2.1 Intelligenta personliga assistenter

En intelligent personlig assistent är en mjukvaruagent vars syfte är att hjälpa en användare med diverse problem eller uppgifter (Ponciano et al., 2015). De röststyrda assistenterna i mobila enheter, såsom Siri, Cortana och S voice, är mycket snarlika, dels vad gäller deras funktionalitet och dels i hur interaktionen utspelas (Apple, 2017; Microsoft, 2017; Samsung, u.å.). Nedan används Siri som exempel för röststyrda intelligenta personliga assistenter.

2.1.1 Interaktion

Siri styrs genom att antingen manuellt starta programmet ‘Siri’ eller genom att säga ”Hej Siri” till enheten. Interaktionen som följer är att Siri lyssnar på vad som efterfrågas av användaren och ger ett svar alternativt ställer en motfråga. Ställs en motfråga behöver inte Siri aktiveras för att återigen tala till henne, däremot krävs det om Siri ger ett svar. För att tala till henne igen finns det en knapp i programmet att trycka på. Programmet använder sig av en ljudsignal för att signalera när det är möjligt att tala till Siri.

2.1.2 Funktion

För att interaktionen med Siri ska fungera krävs att ett flertal komponenter samarbetar. Först och främst behöver Siri uppfatta vad användaren säger. Detta sker genom röstigenkänning som i stora drag konverterar tal till text (Anderson, 2013, 17 september). Siris röstigenkänning fungerar bra, och hon har inga större problem med att förstå ett flertal svenska dialekter (Westerholm, 2015, 27 februari).

Utöver detta behöver Siri även förstå innebörden av texten. Under Apples lanseringsevent av Siri förklaras det att Siri utgår ifrån nyckelord och deras associationer istället för att bearbeta hela meningar som användarna tillhandahåller (Apple, 2013). Detta leder till att Siri kan förstå den bakomliggande meningen, och har möjlighet att ge liknande svar till frågor av olika variation. Siri kan till exempel visa väderleksrapporten för den aktuella dagen och platsen oavsett om frågan är “Hur är vädret idag?”, “Hur ser prognosen ut timme för timme?” eller “Behöver jag en regnjacka idag?” (Apple, 2013. Författarnas översättning).

För att producera dessa svar kommunicerar Siri med molnbaserade servrar som i sin tur är kopplade till tredjepartstjänster som tillhandahåller exempelvis bordsbokning, filmrecensioner samt söktjänster (How stuff works, 2013). Slutligen behöver svaret kommuniceras till användaren i talformat, och det är denna konvertering som kallas text-till-tal (Anderson, 2013, 17 september).

2.2 Naturligt tal

Människans sätt att tala är unikt och något som tycks ha evolverat fram (Traxler, 2012). Det krävs flertalet system samt kontroll över en mängd muskler för att skapa det naturliga talet. Exakt hur produktionen går till och hur människan lär sig att tala är ännu inte helt kartlagt och det är således en utmaning att efterlikna det.

För att förstå syntetiskt tal och dess problematik och utmaningar följer nedan en kort förklaring av det naturliga talet; hur det produceras och vidare prosodins roll för talet samt dess komplexitet.

2.2.1 Talproduktion

Traxler (2012) beskriver grunden till talproduktion som följande: en idé uppstår, ord väljs för att uttrycka idén, orden uttalas verbalt. Griffith och Ferreira (citerad i Traxler, 2012) menar på att talproduktion är mycket komplext och kräver minst tre olika mentala processer. Först krävs tanken på vad som ska sägas, vilket innefattar processen konceptualisering. Vidare krävs ett bra sätt att uttrycka den idén givet de verktyg som språket ger, en process som kallas formulering. Avslutningsvis krävs rörelse av de muskler som behövs för att skapa ljudvågor som lyssnaren kan varsebli. Dessa processer

kallas artikulering. Artikulation är en komplex handling som kräver mycket god kontroll över fler än 100 muskler som rör sig samtidigt.

2.2.2 Prosodi

Bruce (1998) går närmare in på språkets ljudegenskaper och förklarar att en vanlig indelning av dessa är följande tre huvudklasser: konsonanter, vokaler och prosodi. Konsonanter och vokalers huvudsakliga syfte är att verka distinktivt, alltså som betydelseskiljande enheter. De genererar olika fonem, vilka i sig är betydelseskiljande enheter som kan skapa olika mening men som inte bär på en egen betydelse. Det distinktiva kan påvisas då exempelvis *mor* och *bor* har olika mening endast på grund av konsonanterna *m* och *b*, där *m* och *b* bildar olika fonem.

Även prosodi agerar distinktivt, men har fler egenskaper (Bruce, 1998). *Vit* och *vitt* är exempel på prosodins distinktiva egenskaper i talat språk där *vit* har lång vokal och kort konsonant och *vitt* det motsatta. Bruce (1998) definierar prosodi som "det vi preliminärt kan karakterisera som talets rytmiska, dynamiska och melodiska egenskaper." (s. 9). Dess mest påtagliga roll för det talade språket är att ge struktur, utan prosodin finns risk för en ström av endast språkljud.

Vidare delar Bruce (1998) upp prosodins huvudsakliga funktion i tre delar (bortsett från den distinktiva): prominens, gruppering och diskurs. Inom prominens menar han att framhävning och undanhållande är typiska funktioner. Framförallt är betoning, för framhävning, och obetoning, för undanhållande, nyckelord. Inom gruppering är istället signalering av samhörighet samt gräns typiska funktioner. Kort beskrivs det som hur grupper av ord bildar fraser som i sin tur blir till yttranden, detta vilket lyssnaren kan uppfatta. Avslutningsvis handlar diskurs framförallt om prosodins praktiska roll vid exempelvis dialog. Exempel är signalering av attityd, återkoppling, engagemang samt reglering av tur.

Jämfört med vokaler och konsonanter hamnar prosodi ofta i skymundan, men dess roll i talad kommunikation är minst lika central (Bruce, 1998). De prosodiska egenskaperna är knapphändigt kodifierade på grund av att de är svåra att komma åt och begripa. En konsekvens av detta är att skriftspråket i stor grad utesluter dessa och framförallt innehåller vokaler och konsonanter. Prosodiska egenskaper kan ibland presenteras i form av accenttecken eller interpunktionstecken, dock mycket begränsat i jämförelse med prosodins faktiska betydelse.

Intresset för studiet av prosodi ökar markant, mycket på grund av språk- och talteknologin (Bruce, 1998). För att syntetiskt tal ska låta naturligt och vara begripligt krävs reglering och kontroll av prosodin, något som var uppenbart tidigt i forskningen. Med detta introducerades även ett intresse för en fullständig kartläggning av prosodin. Det har dock visats vara en mycket komplex uppgift där forskarna inte är eniga. Problemet ligger delvis i att det inte finns en överenskommen definition av prosodi, något som påvisar prosodins komplexa innebörd. Även Bruce (1998) själv menar på att hans egen definition inte är heltäckande.

2.3 Syntetiskt tal

Talsyntes, eller text-till-tal, är en teknik som konverterar text till vågformer som datorer vidare kan generera till tal (Khan & Chitode, 2016). De moderna syntetiska talen är baserade på mänskliga röster, till skillnad från tidigare versioner som skapades genom att

programmera datorer utifrån akustiska parametrar (Anderson, 2013, 17 september). Vilket företag som har skapat just Siris tal är inte känt, det råder hög konkurrens inom text-till-tal-industrin och därmed även hög sekretess. Däremot verkar företaget Nuance, som är en samarbetspartner till Apple, vara inblandade även om de inte explicit vill förtälja huruvida de ligger bakom skapandet av Siris tal eller ej.

2.3.1 Konkateneringssyntes

Vid skapandet av syntetiskt tal använder sig Nuance av tekniken konkateneringssyntes, som i enlighet med den moderna tekniken är baserad på en mänsklig röst (Anderson, 2013, 17 september). En röstskådespelare läser in förutbestämda meningar som är rika på variationer av språkljud. Inspelningarna delas sedan upp i olika ljudenheter, såsom fonem, difoner, stavelser, ord eller meningar (Khan & Chitode, 2016). Dessa binds sedan samman, konkateneras, för att skapa ett syntetiskt tal. Med andra ord behöver inte alla befintliga ord och meningar läsas in av röstskådespelaren, de bör däremot kunna skapas med hjälp av talsyntesen, vilket är grundprincipen bakom tekniken (Anderson, 2013, 17 september). Tekniken bakom konkateneringssyntes består av system som har till uppgift att välja lämpliga delar från en databas bestående av de olika språkljuden (Khan & Chitode, 2016). Vidare består systemen av algoritmer som sätter samman de valda språkljuden för att sedan släta till gränsen mellan dem för en så mjuk övergång som möjligt.

2.3.2 Naturlig talsyntes

Att målet med talsyntes är att få talet att låta så naturligt som möjligt tycks tas för givet både av företag och av forskare (Deepmind, 2016; Reeves & Nass, 1996). Ett av de centrala problemen för att uppnå naturlighet vid användning av konkateneringssyntes är att de uppdelade ljudenheterna behåller sina ursprungliga prosodiska egenheter (Tatham & Morton, 2005). När segmenten sedan kombineras kan en inte helt perfekt matchning sinsemellan upplevas som ett “hack” i det syntetiska talet. Det syntetiska talet i intelligenta personliga assistenter kräver dessutom inte endast att talet ska vara begripligt samt utan hackighet, utan även att det ska vara uttrycksfullt och socialt anpassat (Campbell & Li, 2015). För att uppnå en sådan naturlighet i det syntetiska talet krävs framförallt naturlig prosodi (Bruce, 1998). Om detta ska åstadkommas krävs en kartläggning, eller uppmärkning, av prosodin för att kunna sortera och konkatenera ljudenheterna rätt. Prosodins komplexitet och brist på heltäckande definition sätter begränsningar för detta. Tillsammans med talets komplexitet i sig är det en svårlöst uppgift, inte bara för företagen som jobbar med tekniken, utan även för språkvetare.

2.4 Människa- teknikinteraktion

Fenomenet *computers are social actors* (förkortat CASA) är väl studerat och har fått stöd av flertalet empiriska undersökningar (Lee, 2010; Nass & Lee, 2001; Nass, Steuer & Tauber, 1994; Nass & Moon, 2000). CASA innebär att människor behandlar datorer som om de vore mänskliga, bland annat tillskrivs sociala regler som vanligtvis gäller vid mänsklig interaktion även vid människa-datorinteraktion. Trots medvetenhet kring att datorer inte kräver det sociala samspel vid interaktion som människor gör tillämpas ändå liknande sociala förväntningar och reaktioner på datorers ageranden som mänskliga. Fenomenet är även studerat inom syntetiskt tal där det bland annat har konstaterats att människor kan utröna personligheter samt identitet i syntetiska röster (Nass & Lee, 2001).

Med tanke på detta har det påstått att en utveckling mot mer mänskliga gränssnitt verkar önskvärt för att minska behovet av anpassning (Reeves & Nass, 1996). Dock menar Lee (2010) att det saknas empirisk forskning som stödjer en sådan slutsats. Även Ilves och Surakka (2013) menar på att det inte finns kongruent forskning inom detta ämne och att det ännu inte går att dra någon sådan slutsats. Schaumburg (2001) menar på att CASA inte behöver innebära en strävan mot människolik teknik, trots att det kan tyckas gå emot den intuitiva känslan. Det förklaras att det istället skulle kunna innebära en tendens till att överskatta teknikens förmågor. Det finns även risk för att något som är väldigt människolikt även blir dömt likt en människa, att användaren kan känna en känsla av motvilja mot tekniken. Det är svårt att reagera ogillande gentemot vad som kan tyckas vara en gullig robot, däremot är det problematiskt att skapa något väldigt människolikt med samma resultat. Ett mänskligt gränssnitt kan även framkalla det tidigare nämnda fenomenet *uncanny valley*, som innebär ett obehag när en icke-mänsklig enhet uppträder människolikt (Mori, 1970). Att människor tillskriver mänskliga attribut åt datorer behöver med andra ord inte innebära att mänskliga gränssnitt är önskvärda.

2.5 Användarupplevelse

Tullis och Albert (2013) menar att det i princip finns lika många definitioner av användarupplevelse som det finns personer som jobbar med det. Men de sammanfattar det med att säga att följande tre egenskaper utgör kärnan: användaren, en artefakt som användare kan interagera med samt användarens upplevelse av interaktionen. Artefakten kräver egentligen inga specifika egenskaper då det vidare förklaras att alla artefakter skapar en användarupplevelse vid interaktion med en mänsklig användare.

Termen användarupplevelse i sig refererar till upplevelsen hos användaren vid en interaktion med en produkt och kan innefatta många möjliga känslor och tillstånd (Preece et al., 2015). Exempel på sådana följer i nästa stycke. Syftet med att tala om en bra användarupplevelse är att avsaknaden av den kan innebära en rad komplikationer och konsekvenser som både kan vara frustrerande och kostsamma. I företagsvärlden kan en produkt som bidrar till en dålig användarupplevelse innebära en stor kostnad, bland annat på grund av att populariteten minskar vilket i sin tur innebär dålig försäljning. Det finns många uttalade upplevelser då en användare interagerar med en artefakt. Preece et al. (2015) har sammanfattat dessa i ett antal önskvärda och icke-önskvärda emotioner och upplevelser vilka presenteras i tabell 1 och tabell 2.

Tabell 1. Önskvärda emotioner och upplevelser vid en användarupplevelse. Författarnas översättning, original i parentes. Anpassad från *Interaction design: Beyond human- computer interaction* (s. 22), av J. Preece et al., 2015, West Sussex: John Wiley & Sons.

Tillfredsställande (*satisfying*)	Hjälpsam (*helpful*)	Rolig (*fun*)
Angenäm (*enjoyable*)	Motiverande (*motivating*)	Provokativ (*provocative*)
Engagerande (*engaging*)	Utmanande (*challenging*)	Överraskande (*surprising*)
Njutbar (*pleasurable*)	Förhöja sällskaplighet (*sociability*)	Belönande (*rewarding*)
Spännande (*exciting*)	Stödjer kreativitet (*supporting creativity*)	Emotionellt givande (*emotionally fulfilling*)
Underhållande (*entertaining*)	Kognitivt stimulerande (*cognitively stimulating*)	

Tabell 2. Icke-önskvärda emotioner och upplevelser vid en användarupplevelse. Författarnas översättning, original i parentes. Anpassad från *Interaction design: Beyond human- computer interaction* (s. 22), av J. Preece et al., 2015, West Sussex: John Wiley & Sons.

Tråkig (*boring*)	Barnslig (*childish*)	Gulligt tillgjort (*cutesy*)
Frustrerande (*frustrating*)	Obehaglig (*unpleasant*)	Plojartad (*gimmicky*)
Skuldbeläggande (*making one feel guilty*)	Nedlåtande (*patronizing*)	
Irriterande (*annoying*)	Förlöjligande (*making one feel stupid*)	

Preece et al. (2015) förklarar vidare att de flesta av dessa är subjektiva kvalitéer och handlar om hur en artefakt upplevs för användaren. Många av dem överlappar dessutom vilket har ett syfte. Det blir då möjligt att upptäcka små förändringar i upplevelse av samma artefakt över tid. Att få svar på hur användarna upplever artefakten med hjälp av termer som har som syfte att till exempel spegla känslor, tillstånd och sensationer kan öka förståelsen för att en användarupplevelse har många sidor och kan förändras över tid. Vidare förklarar Preece et al. (2015) att alla artefakter inte nödvändigtvis är kompatibla med alla dessa önskvärda och icke-önskvärda upplevelser samt att det även kan krävas fler utöver dessa. Utöver detta går det även att se en skillnad i antal termer vid önskvärda och icke-önskvärda. Anledningen är att målet med denna typ av studier är att komma åt de positiva upplevelserna, de negativa är snarare en indikation på vad som kan vara fel.

3 Metod

Studien ämnade undersöka eventuella skillnader i användarupplevelse mellan Siris tal, naturligt tal och naturligt tal som efterliknar Siris prosodi. Studien använde sig av en så kallad *Wizard of Oz technique* (Martin & Hanington, 2012). Tekniken innebär att respondenten tror att de interagerar med ett fungerande system när det i själva verket är experimentledaren som styr systemet. I denna studie ombads respondenterna att utvärdera tre demoversioner av intelligenta assistenter. Respondenterna ställde frågor till den intelligenta assistenten och den intelligenta assistenten svarade. I själva verket styrdes de intelligenta assistenterna av en experimentledare, som spelade upp förinspelade svar till respondenten. För att förstärka illusionen av att respondenterna interagerade med en faktisk intelligent assistent genomfördes ett flertal åtgärder. Respondenterna ombads att trycka på en knapp på en smartmobil som genererade ett plingande ljud innan de kunde kommunicera med den intelligenta assistenten, vilket överensstämmer med interaktion med Siri. Smartmobilen var även sammankopplad med en högtalare med en synlig AUX-kabel. En experimentledare spelade upp de förinspelade svaren via en iPad som var trådlöst kopplad till samma högtalare som smartmobilen.

3.1 Nollhypotes

Studien ämnade pröva följande nollhypotes: det finns ingen statistiskt signifikant skillnad i användarupplevelsen mellan Siris tal, naturligt tal och naturligt tal som efterliknar Siris prosodi.

3.2 Design

Studien utfördes med hjälp av ett experiment med inkomplett inomgruppsdesign. En oberoende variabel bestående av tre nivåer användes där den oberoende variabeln var olika typer av tal och nivåerna följande: Siris tal, naturligt tal och naturligt tal som efterliknar Siris prosodi. Den beroende variabeln var respondenternas uttalade upplevelse av interaktionen och mättes med en enkät efter varje avslutad nivå (se bilaga A & B).

3.2.1 Experimentdesign

Inomgruppsdesign valdes för att undvika individuella skillnader som kan komma att påverka vid en så pass subjektiv bedömning som användarupplevelse (Shaughnessy, Zechmeister & Zechmeister, 2012). Vid användandet av inomgruppsdesign hade dock en jämförande effekt kunnat uppstå mellan de olika talen. Enligt Hardman (2016) fattar människor olika beslut beroende på om de olika alternativen presenteras samtidigt i tid och rum alternativt separat i tid och rum. Om de olika alternativen presenteras samtidigt i tid och rum sker en jämförande effekt som annars inte hade uppstått eller lagts märke till. Då en sådan effekt inte var önskvärd i denna studie skedde en kortare fördröjning mellan utförandet av de olika nivåerna.

3.2.2 Oberoende variabeln

Genom att använda tre nivåer kunde prosodins eventuella inverkan på användarupplevelsen ringas in bättre än om endast nivåerna Siris tal och naturligt tal användes. Tabell 3 visar hur de tre nivåerna samspelade.

Tabell 3. Likheter mellan de olika nivåerna. Siris tal och naturligt tal som efterliknar Siris prosodi har likvärdig prosodi. De två naturliga talen har lika röster. Allt som skiljer mellan två tal förutom röst och prosodi, ’övrigt’, har även de naturliga talen gemensamt.

Nivå	Prosodi	Röst	Övrigt
1. Siris tal	x		
2. Naturligt tal		x	x
3. Naturligt tal som efterliknar Siris prosodi	x	x	x

I kategorin prosodi var nivåerna Siris tal och naturligt tal som efterliknar Siris prosodi likvärdiga. Det naturliga talet hade en naturlig prosodi medan prosodin hos Siris tal och naturligt tal som efterliknar Siris prosodi var skilda från den naturliga. Med röst menades den unika röst som användes. Då samma person spelade in det naturliga talet och naturligt tal som efterliknar Siris prosodi skiljde sig de nivåerna från Siris tal vad gällde röst. Det är inte enbart prosodi eller röst som skiljer sig mellan olika tal, och det var detta som kategorin övrigt återspeglade. Kategorin innefattade alltså alla skillnader mellan nivåerna som inte inkluderades i skillnader i prosodi eller röst. Det enda som skiljde mellan nivåerna naturligt tal och naturligt tal som efterliknar Siris prosodi var själva prosodin, därför ansågs dessa nivåer likvärdiga i denna kategori.

Om Siris tal föredras över naturligt tal och naturligt tal som efterliknar Siris prosodi innebär det att varken prosodi eller övriga delar av naturligt tal föredras över syntetiskt. Om naturligt tal föredras över de andra nivåerna innebär det att naturligt tal föredras över syntetiskt samt att naturlig prosodi är en bidragande faktor, annars hade naturligt tal som

efterliknar Siris prosodi föredragits tillsammans med naturligt tal. Om naturligt tal som efterliknar Siris prosodi föredras innebär det att naturligt tal föredras över syntetiskt, men inte specifikt naturlig prosodi.

3.2.3 Beroende variabeln

Efter varje utförd interaktion med en intelligent assistent ombads respondenterna att fylla i en enkät om hur de upplevde interaktionen (se bilaga A & B). En respondent interagerade med tre intelligenta assistenter och besvarade tre enkäter. Enkätens utformning grundades i linje med Tullis och Albert (2013) rekommendationer gällande likertskalor. Den bestod av nio påståenden med en för varje påstående följande femgradig skala för respondentens svar. De nio påståendena utgjorde tillsammans ett mått på användarens upplevelse av interaktionen.

Enkäten efter den tredje interaktionen innehöll förutom de nio påståendena även frågor om respondenten samt en fråga om vilken av de tre intelligenta assistenterna som föredrogs och varför denna föredrogs (se bilaga B). Frågan om vilken av assistenterna som föredrogs samt varför baserades på Tullis och Alberts (2013) rekommendationer för användarstudier vid jämförandet av alternativa designer. Påstående ett till åtta baserades på ett urval av de tidigare nämnda önskvärda och icke-önskvärda emotioner och upplevelser vid skapandet av en bra användarupplevelse (Preece et al., 2015). Påstående nummer sju (“Jag tyckte det var obehagligt att interagera med den intelligenta assistenten”) användes specifikt för att kunna upptäcka en eventuell effekt av *uncanny valley* (Mori, 1970).

Preece et al. (2015) påpekade att deras listade önskvärda och icke-önskvärda emotionerna och upplevelserna endast var ett urval, och att fler kan komma att behövas beroende på artefakt. Med tanke på detta användes även påståendet “Jag tyckte den intelligenta assistenten var enkel att förstå”, för att undersöka hur Siris begriplighet stod sig gentemot det naturliga talet och naturligt tal som efterliknar Siris prosodi. Tre av nio påståenden var negativt laddade medan de resterande sex var positivt laddade.

3.2.4 Frasernas inspelning, variation och presentation

Vid skapandet av de tre nivåerna spelades Siris röst och en mänsklig röst in. Den mänskliga rösten spelades även in när denna efterliknade Siris prosodi. Siris svar spelades in med hjälp av programmet Apowersoft Mac Audio Recorder. Den mänskliga rösten spelades in i ett *medialab* med hjälp av en *large diaphragm condenser microphone*.

Då respondenterna interagerade med alla tre nivåer (Siris tal, naturligt tal och naturligt tal som efterliknar Siris prosodi) en gång vardera varierades fraserna. En respondent hörde tre uppsättningar av skilda, men likvärdiga, svar från de intelligenta assistenterna (se bilaga C). Då tre olika uppsättningar användes spelades således nio uppsättningar av svar in. Tre uppsättningar av Siris tal, tre för naturligt tal och tre för naturligt tal som efterliknar Siris prosodi.

Då det inte fanns möjlighet att förändra Siris svar bestämdes fraserna utifrån hennes förutsättningar. Frågor som genererade kortare svar från Siri som “ja” eller “nej” valdes bort. Utöver detta valdes både faktamässiga svar (“Avståndet mellan Sundsvall och Karlstad är omkring 551 kilometer med bil eller omkring 394 kilometer fågelvägen”) och mer eller mindre alldagliga svar (“Jag brukar drömma att jag flyger”).

Respondenterna ombads även att anteckna svaren från de intelligenta assistenterna vid varje nivå, detta för att kunna kontrollera att respondenterna var uppmärksamma under interaktionen. Frasblanketter användes under experimentet för att dels lista de fraser som respondenterna ombads ställa till assistenterna, dels för att ge utrymme för respondenterna att anteckna assistenternas svar (se bilaga D).

3.2.5 Balansering

Eventuella träningseffekter balanserades genom att använda 'alla möjliga ordningar'. För att undvika en eventuell påverkan av svaren som gavs av de intelligenta assistenterna balanserades även frasuppsättningarna (se bilaga E).

3.3 Respondenter

Respondenterna som deltog i experimentet rekryterades genom ett tillgänglighetsurval. Totalt deltog 30 respondenter, varav 18 män och 12 kvinnor. Åldersintervallet var mellan 19 och 57 år och medianålder 23 år. Samtliga deltagande respondenter hade svenska som modersmål.

3.4 Material

Det material som användes under studien var följande:

Material vid inspelning av Siri och naturligt tal

- Apowersoft Mac Audio Recorder version 2.4.5
- Siri i macOS Sierra version 10.12.2
- Mikrofon Blue Snowball
- Audacity version 2.1.3

Program för bearbetning av ljudfiler

- GarageBand version 10.1.6
- Audacity version 2.1.3
- MP3Gain version 1.2.5

Material vid utförande av experimentet

- Smartmobil tillsammans med en egengjord applikation bestående av en knapp kopplad till en ljudsignal
- iPad mini 1st generation
- iTunes version 12.6.1.25
- One Track Mind version 2.0
- Högtalare Xqisit xqS10
- AUX-kabel
- Frasblanketter
- Enkäter

Program vid bearbetning av resultat

- SPSS statistics version 24

3.5 Pilotstudie

Tre respondenter deltog i pilotstudien. Efter den första respondentens återkoppling tydliggjordes frasblanketternas utformning. Även en fråga ändrades för samtliga uppsättningar. Tidigare ombads respondenten fråga om avståndet till ett specifikt land, vilket ändrades till en specifik stad. Detta då Siris svar på frågan "Hur långt är det till

Egypten?” var “Det ser ut som om Kairo, Egypten ligger omkring 3 415 kilometer härifrån fågelvägen.”, vilket ansågs märkligt enligt den första respondenten. Ljudfilerna från Siri och de naturliga talen hade varierande volym, vilket den första respondenten påpekade. Med programmet MP3Gain ändrades samtliga ljudfiler till samma volym. Efter dessa ändringar utfördes experimentet på nytt med två andra respondenter, dessa hade inga synpunkter på studiens utförande och följaktligen tog de faktiska experimenten vid.

3.6 Procedur

Samtliga respondenter gavs likvärdig information innan, under och efter experimentet. Innan experimentet informerades respondenterna om att de skulle interagera med och utvärdera tre intelligenta assistenter, att de kunde avbryta närsomhelst samt att deras svar skulle komma att behandlas konfidentiellt (se bilaga F). Respondenten gavs under experimentets gång skriftliga instruktioner för interaktionens tillvägagångssätt (se bilaga G), vilka kan sammanfattas som:
1. Tryck på knappen.
2. Ställ en fråga från frasblanketten efter plinget.
3. Anteckna svaret från den intelligenta assistenten.
Varje experiment avslutades med en förklaring om studiens syfte samt utrymme för respondenterna att ställa eventuella frågor (se bilaga H).

Två experimentledare var delaktiga under experimentets utförande. Den ena experimentledaren ansvarade för enkäterna samt att informera respondenterna innan och efter experimentet. Den andra experimentledaren var ansvarig för interaktionen med de intelligenta assistenterna och spelade således upp de intelligenta assistenternas svar. För att undvika experimentledareffekter var den första experimentledaren inte medveten om i vilken ordning som respondenterna interagerade med de tre intelligenta assistenterna.

Experimentet utfördes i två separata rum. I det ena rummet interagerade respondenterna med de intelligenta assistenterna och i det andra rummet besvarade respondenterna enkäter om hur de upplevde interaktionen. Respondenten leddes först till rummet där interaktionen med de intelligenta assistenterna tog vid. Efter avslutad interaktion leddes respondenten till det andra rummet för att besvara en enkät om upplevelsen av interaktionen. Detta upprepades för samtliga nivåer. Respondenten interagerade med tre intelligenta agenter och fyllde i tre enkäter om upplevelsen av interaktionen.

3.7 Etik

Deltagandet i studien var frivilligt och svaren behandlades anonymt, vilket samtliga respondenterna informerades om. Respondenterna fick även avbryta experimentet när de ville. För att undvika stress och känsla av bevakning var ingen experimentledare närvarande i rummet då enkäterna besvarades. Vid interaktionen med de intelligenta assistenterna var den ena experimentledaren närvarande i rummet då detta var en förutsättning för experimentet. För att avdramatisera närvaron motiverades den vara på grund av att eventuella fel kan uppstå samt om respondenten hade några frågor. I övrigt var experimentledaren passiv under interaktionen.

Respondenterna var inte medvetna om att de interagerade med ett system som styrdes av experimentledaren, något som var nödvändigt för att få tillförlitliga resultat. Om en respondent hade funderingar kring de intelligenta assistenterna besvarades frågorna efter deltagandet.

3.8 Validitet och reliabilitet

För att undersöka studiens frågeställningar och undvika ovidkommande variabler har ett flertal åtgärder vidtagits.

Inledningsvis krävdes nivån naturligt tal som efterliknar Siris prosodi för att ringa in prosodins eventuella påverkan för användarupplevelsen. En jämförelse mellan naturligt tal och syntetiskt tal kan endast ge resultat som visar att det ena talet föredras framför det andra, men inte huruvida prosodin påverkar användarupplevelsen. Naturligt tal som efterliknar Siris prosodi i jämförelse med naturligt tal möjliggör alltså en jämförelse mellan naturlig prosodi och syntetisk prosodi och huruvida naturlig prosodi är eftersträvansvärt.

För att möjliggöra mätning av användarupplevelsen vid interaktion med en icke-befintlig intelligent personlig assistent krävs det att interaktionen är så lik ett verkligt scenario som möjligt. Prototypen av den intelligenta assistenten som användes var därav likvärdig med befintliga intelligenta assistenter. Genom att få respondenterna att trycka på en knapp, vilket följdes av en ljudsignal, innan de kunde prata med assistenten förhöjdes illusionen av att de styr en fungerande artefakt samtidigt som det följer av hur interaktion med befintliga system fungerar.

Enkätens påståenden baserades på de av Preece et al. (2015) framarbetade emotioner och upplevelser vid en användarupplevelse. De påståenden som användes i enkäten var ett urval av dessa, och valdes för att kunna återspegla de möjliga emotioner som kan komma att uppstå vid interaktion med en intelligent assistent. Den tredje enkäten innehöll även frågan “Vilken av de tre intelligenta assistenterna hade du helst velat använda dig av i framtiden?” och användes för att kunna kontrollera om enkätens påståenden speglade användarupplevelsen. Om ett flertal respondenter betygsatte en intelligent assistent högst, men svarade att de helst hade velat använda en av de andra assistenterna i framtiden, indikerar resultaten att påståendena inte speglar användarupplevelsen.

En strävan att använda samma rum för samtliga experiment fanns, men visades inte vara praktiskt möjligt. Istället valdes rum som ansågs likvärdiga för att minimera en eventuell påverkan av miljön. Riktlinjerna för val av rum var att de skulle vara neutrala, utan störande moment såsom tavlor eller framträdande färgsättning och med en tyst omgivning.

För att undvika påverkan av träningseffekter i experimentet randomiserades nivåernas och frasuppsättningarnas ordning. De tre frasuppsättningarna utformades även för att vara så lika som möjligt utan att innehålla identiska fraser. De innehöll snarlika frågor och svar, och ett skämt vardera. Även frasuppsättningarnas spellistor var lika långa och hade likvärdig volym. Om balanseringen hade haft en önskad effekt kontrollerades efter experimentets utförande för att undersöka huruvida det var nivåernas ordning eller en specifik frasuppsättning som var orsaken till resultatet.

Experimentledareffekter kontrollerades genom flertalet åtgärder. Experimentledarna använde sig av manus för att samtliga respondenter skulle ges samma information. Vidare var experimentledarnas roll densamma under samtliga experiment. Avslutningsvis var det endast experimentledaren som styrde de intelligenta assistenterna som var medveten om randomiseringens ordning, alltså vilken intelligent assistent och vilken frasuppsättning som användes.

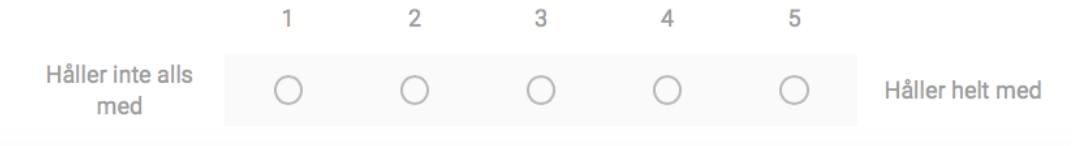
För att kontrollera för slumpens påverkan av resultaten valdes det att genomföra en statistisk analys.

3.9 Dataanalys

Inledningsvis granskades huruvida respondenterna hade antecknat svar i frasblanketterna. Det bestämdes initialt att de respondenter som saknade svar skulle behandlas som bortfall, något som vidare inte blev aktuellt då samtliga respondenter antecknade svar.

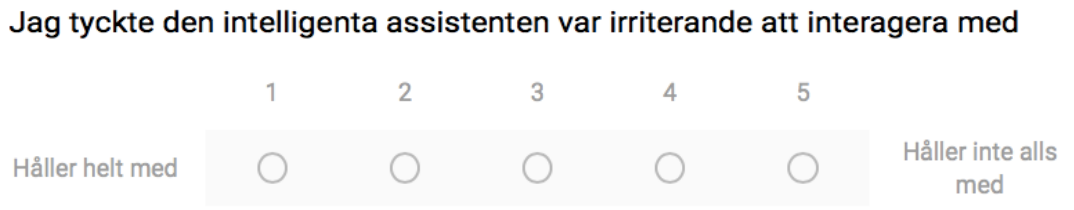
Resultatet för påståendena behandlades genom statistiska analyser med hjälp av programmet *SPSS statistics version 24*. Totalt utfördes tio statistiska analyser, en för användarupplevelsen och nio för vardera enskilt påstående. Då experimentet bestod av en inomgruppsdesign med fler än två nivåer utfördes en *repeated measures ANOVA*, även kallat ett övergripande *F*-test. Om det övergripande *F*-testet visade en statistisk signifikans ($p < 0{,}05$) utfördes även parvisa jämförelser genom *post hoc* där korrigering med Bonferroni användes. Bonferroni används för att minimera risken för typ I fel, att felaktigt förkasta nollhypotesen, vilket annars ökar ju fler jämförelser som görs.

För att kunna utföra den statistiska analysen tilldelades enkäternas likertskalor värden. Vid analys av de enskilda påståendena tilldelades "Håller inte alls med" värde 1 och "Håller helt med" värde 5 (se figur 1).



Figur 1. Påståendenas skalor med tilldelade värden.

Vid analys av användarupplevelsen behandlades enkätens nio påståenden gemensamt. Den maximala summan för en enkät var 45 (5×9) och den minimala 9 (1×9), där 45 representerade optimal användarupplevelse och 9 representerade det motsatta. De positivt laddade påståendena tilldelades värden likt figur 1. För att 45 skulle motsvara en optimal användarupplevelse inverterades skalorna för de tre negativt laddade påståendena, där "Håller inte alls med" tilldelades värde 5 och "Håller helt med" tilldelades värde 1 (se figur 2).



Figur 2. En inverterad skala med tilldelade värden, vilket användes vid de negativt laddade påståendena. "Jag tyckte den intelligenta assistenter var irriterande att interagera med" används som exempel.

Övrig data från enkätens frågor (fråga 10-17) sammanställdes och resultatet från fråga 10, 16 och 17 presenterades även i form av figurer.

4 Resultat

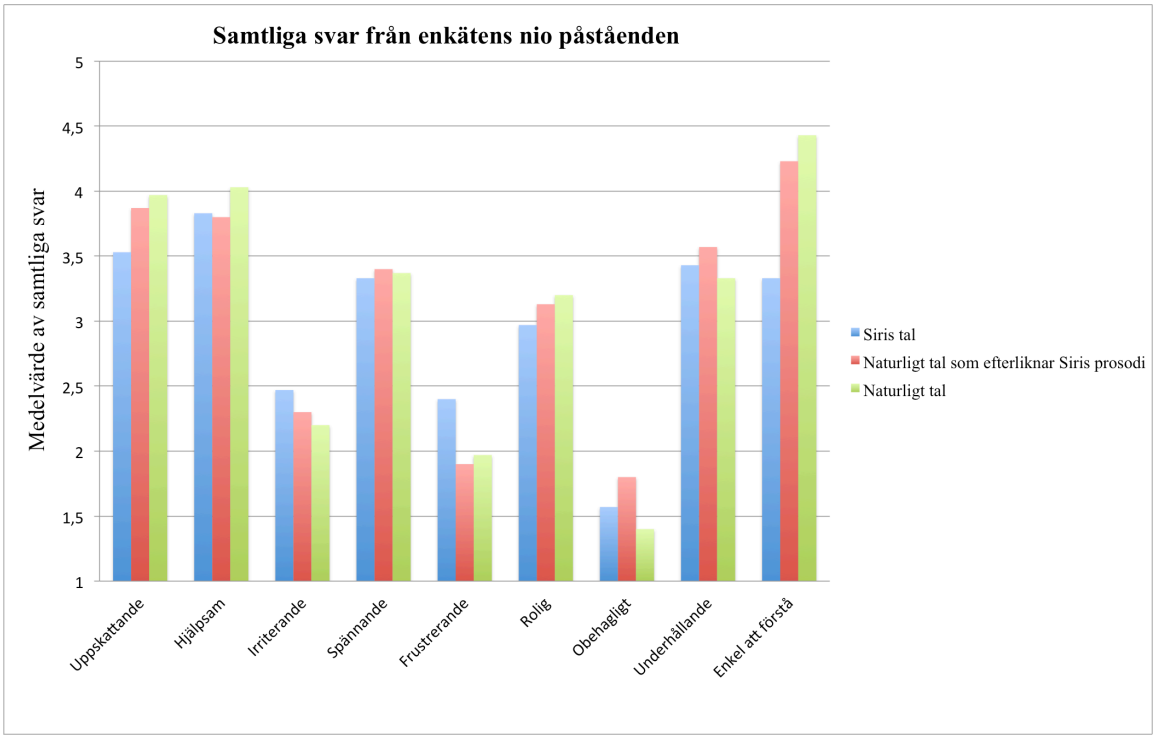
Resultatet baserades på 30 respondenter där åldersintervallet var 19-57 år med en medianålder på 23 år. Könsfördelning var 18 män och 12 kvinnor. Resultatet, likt enkäten, är indelat i två huvudsakliga delar. Den första delen innefattar bearbetning av

enkätens nio påståenden medan den andra delen behandlar de efterföljande frågorna. Sist redogörs för eventuella träningseffekter.

4.1 Del 1: enkätens påståenden

För att analysera respondenternas svar från enkäterna tilldelades påståendenas skalor värden. "Håller inte alls med" tilldelades värde 1 och "Håller helt med" tilldelades värde 5. Utefter dessa värden kunde enkätens påståenden behandlas.

Ett medelvärde av respondenternas svar för varje påstående räknades ut enligt de tilldelade värdena. Resultatet visualiseras i figur 3, där varje påstående presenteras enskilt och är grupperade efter de intelligenta assistenternas tal.



Figur 3. Respondenternas svar för de nio påståendena, grupperat efter de tre nivåerna: Siris tal, naturligt tal som efterliknar Siris prosodi och naturligt tal. Figuren visar ett medelvärde baserat på enkätens femgradiga skalor där 1 representerar "håller inte alls med" och 5 representerar "håller helt med".

4.1.1 Statistisk analys av användarupplevelsen

Den statistiska analysen av användarupplevelsen baserades på summan av enkätens nio påståenden. Den maximala summan för en enkät var 45 (5×9) och den minimala summan var 9 (1×9). Då tre negativt laddade påståenden användes i enkäten inverterades deras medföljande skalor. "Håller helt med" tilldelades värde 1 och "Håller inte alls med" tilldelades värde 5.

Ett övergripande *F*-test visade att det inte förelåg en statistiskt signifikant skillnad i användarupplevelse mellan Siris tal, naturligt tal som efterliknar Siris prosodi och naturligt tal, $F(2, 58) = 2{,}21$, $p = 0{,}12$, $\eta p2 = 0{,}07$. Medelvärdet, baserat på summan av enkätens nio påståenden, var högst för naturligt tal ($M = 34{,}8$, $SD = 6{,}21$), följt av naturligt tal som efterliknar Siris prosodi ($M = 34{,}0$, $SD = 6{,}22$). Siris tal erhöll det lägsta medelvärdet ($M = 32{,}0$, $SD = 7{,}12$). Konfidensintervallen (95 %) var för naturligt tal [32,5, 37,1], för naturligt tal som efterliknar Siris prosodi [31,7, 36,3] och för Siris tal [29,3, 34,7].

4.1.2 Statistisk analys av enskilda påståenden

En statistisk analys utfördes även för varje enskilt påstående. De negativt laddade påståendena var inte inverterade i denna analys. Resultatet utgick som tidigare från skalan 1-5 och går att utläsa i tabell 4.

Tabell 4. Presentation av samtliga statistiska analyser av de enskilda påståendena.
S: Siris tal, NS: naturligt tal som efterliknar Siris prosodi, N: naturligt tal.

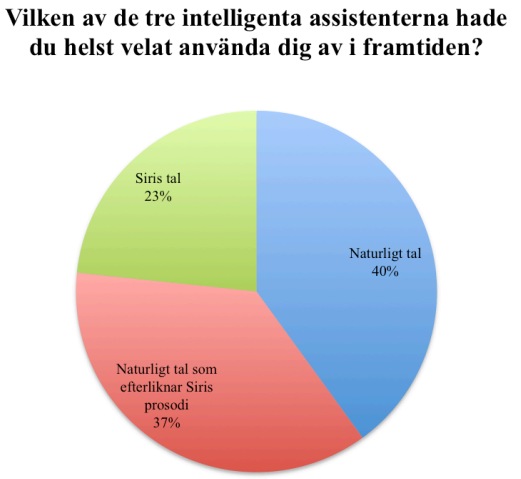
	Medelvärde (*SD*)	**Konfidens-intervall**	***F*-värde**	***p*-värde**	**ηp2**
Uppskattande	S. 3,53 (0,97) NS. 3,87 (0,78) N. 3,97 (0,89)	S. [3,17, 3,90] NS. [3,58, 4,16] N. [3,63, 4,30]	$F(2, 58) = 2,57$	$p = 0,086$	ηp2 = 0,081
Hjälpsam	S. 3,83 (0,83) NS. 3,80 (0,96) N. 4,03 (0,85)	S. [3,52, 4,14] NS. [3,44, 4,16] N. [3,72, 4,35]	$F(2, 58) = 0,81$	$p = 0,45$	ηp2 = 0,027
Irriterande	S. 2,47 (1,28) NS. 2,30 (1,02) N. 2,20 (1,16)	S. [1,99, 2,94] NS. [1,92, 2,63] N. [1,77, 2,63]	$F(2, 58) = 0,60$	$p = 0,55$	ηp2 = 0,020
Spännande	S. 3,33 (1,18) NS. 3,40 (1,33) N. 3,37 (1,10)	S. [2,89, 3,78] NS. [2,90, 3,90] N. [2,96, 3,78]	$F(2, 58) = 0,057$	$p = 0,95$	ηp2 = 0,002
Frustrerande	S. 2,40 (1,16) NS. 1,90 (0,96) N. 1,97 (1,00)	S. [1,97, 2,83] NS. [1,54, 2,26] N. [1,59, 2,34]	$F(2, 58) = 2,90$	$p = 0,063$	ηp2 = 0,091
Rolig	S. 2,97 (1,38) NS. 3,13 (1,07) N. 3,20 (1,27)	S. [2,45, 3,48] NS. [2,73, 3,53] N. [2,73, 3,67]	$F(1,45, 42,14) = 0,42$	$p = 0,67$	ηp2 = 0,014
Obehagligt	S. 1,57 (1,04) NS. 1,80 (1,10) N. 1,40 (0,62)	S. [1,18, 1,96] NS. [1,39, 2,21] N. [1,17, 1,63]	$F(2, 58) = 2,27$	$p = 0,11$	ηp2 = 0,073
Underhållande	S. 3,43 (1,17) NS. 3,57 (1,04) N. 3,33 (1,30)	S. [3,00, 3,87] NS. [3,18, 3,96] N. [2,85, 3,82]	$F(2, 58) = 0,41$	$p = 0,67$	ηp2 = 0,014
Enkel att förstå	S. 3,33 (1,24) NS. 4,23 (1,04) N. 4,43 (0,73)	S. [2,87, 3,80] NS. [3,84, 4,62] N. [4,16, 4,71]	$F(2, 58) = 11,6$	$p < 0,001$	ηp2 = 0,29

4.1.2.1 Statistisk analys av ’enkel att förstå’

Ett övergripande *F*-test visade att det förelåg en statistiskt signifikant skillnad för påståendet “Jag tyckte den intelligenta assistenten var enkel att förstå” mellan Siris tal, naturligt tal som efterliknar Siris prosodi och naturligt tal, se tabell 4. De parvisa jämförelserna, korrigerade med Bonferroni, indikerade en lägre skattning av Siris tal än av naturligt tal som efterliknar Siris prosodi ($p = 0,007$) och naturligt tal ($p = 0,001$), men skattningarna skiljde sig inte signifikant mellan naturligt tal som efterliknar Siris prosodi och naturligt tal ($p = 0,94$).

4.2 Del 2: enkätens frågor

I enkätens andra del, som besvarades efter sista interaktionen, ombads respondenterna att uppge vilken av de tre intelligenta assistenterna som de helst hade velat använda sig av i framtiden. Tolv respondenter uppgav den intelligenta assistenten med naturligt tal, elva respondenter uppgav naturligt tal som efterliknar Siris prosodi och sju uppgav Siris tal. Figur 4 visualiserar den procentuella fördelningen av respondenternas svar.



Figur 4. Resultatet från enkätfrågan "Vilken av de tre intelligenta assistenterna hade du helst velat använda dig av i framtiden?". Tolv respondenter uppgav naturligt tal, elva respondenter uppgav naturligt tal som efterliknar Siris prosodi och sju uppgav Siris tal.

Av de 30 deltagande respondenterna var det 25 som var konsekventa i vilken som ansågs föredras och vilken som betygsattes högst i påståendena. Av de fem respondenter som inte var konsekventa var det två respondenter som betygsatte naturligt tal högst, men angav att en framtida interaktion med naturligt tal som efterliknar Siris prosodi var att föredra. Vidare var det en respondent som betygsatte naturligt tal högst, men angav att Siris tal var att föredra vid en framtida interaktion samt en respondent vars resultat var tvärtemot. Den femte och sista respondenten som inte var konsekvent betygsatte naturligt tal som efterliknar Siris prosodi högst, men angav Siri som mest önskvärd vid framtida interaktion.

4.2.1 Motivering av vald intelligent assistent

Nedan följer en sammanfattning av respondenternas motiveringar till frågan "Vilken av de tre intelligenta assistenterna hade du helst velat använda dig av i framtiden?". För motiveringarna i sin helhet se bilaga I.

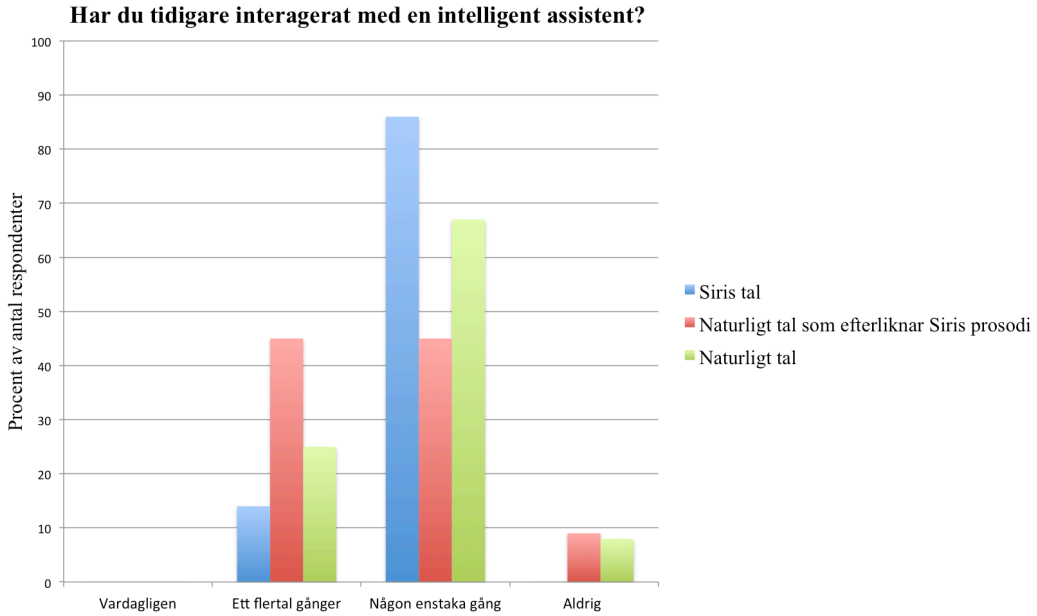
Respondenterna som föredrog Siris tal uppgav bland annat att hon var kortfattad, objektiv och avslappnad. Vissa respondenter uppskattade hennes humor (två respondenter) men även att hon kändes mer "robotaktig", något som ansågs positivt (två respondenter). En respondent uppgav även att valet föll på Siris tal på grund av att hon hade den behagligaste rösten.

För de respondenter som valde naturligt tal som efterliknar Siris prosodi var en vanlig motivering att förklara varför de två andra inte valdes. Utöver detta uppskattade respondenterna bland annat att assistenten var neutral och informativ. En respondent uppskattade även ett av assistentens skämt.

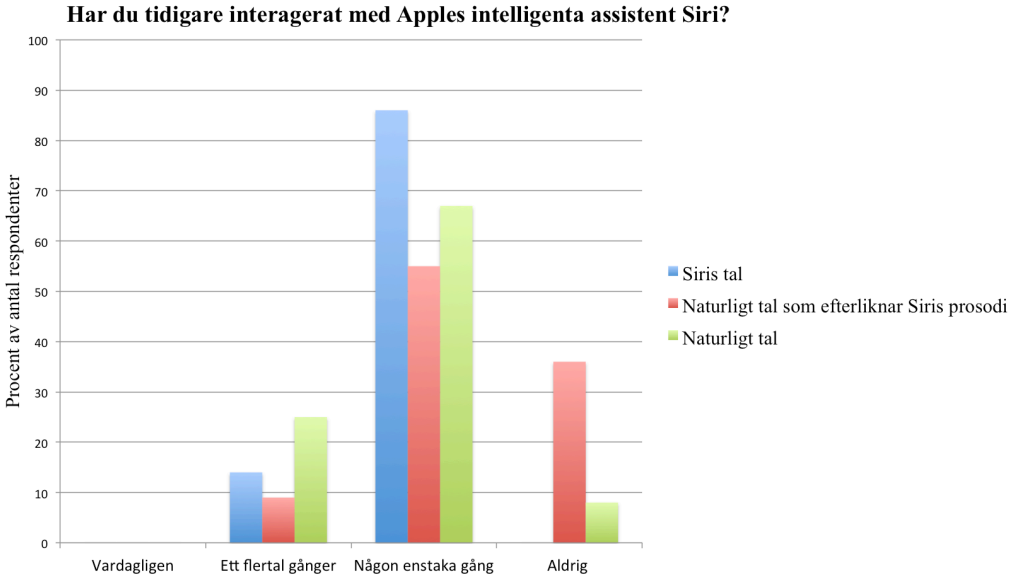
Respondenterna som föredrog naturligt tal uppskattade bland annat att denna lät mänsklig och hade ett naturligt tal (fem respondenter angav dessa motiveringar). Tre respondenter uppgav även att de uppskattade assistentens humor. Vidare uppskattade vissa respondenter assistentens röst och angav även att den gav tydliga svar.

4.2.2 Tidigare interaktion med intelligenta assistenter

Respondenternas eventuella tidigare interaktioner med intelligenta assistenter samt Siri specifikt undersöktes och relaterades till respondenternas val av intelligent assistent. Figur 5 visar respondenternas tidigare interaktioner med intelligenta assistenter och figur 6 visar deras tidigare interaktion med Siri. Respondenterna grupperades efter tidigare uppgivet val av assistent.



Figur 5. Histogram över respondenternas svar på frågan "Har du tidigare interagerat med en intelligent assistent?", grupperat efter vilken intelligent assistent som föredrogs av respondenten. För ett jämförligt resultat presenteras svaren i procent av de antal som föredrog en intelligent assistent.



Figur 6. Histogram över respondenternas svar på frågan "Har du tidigare interagerat med Apples intelligenta assistent Siri?", grupperat efter vilken intelligent assistent som föredrogs av respondenten. För ett jämförligt resultat presenteras svaren i procent av de antal som föredrog en intelligent assistent.

4.3 Träningseffekter

Eventuella träningseffekter kontrollerades genom att undersöka huruvida det fanns en trend i vilket alternativ i ordningen (första interaktionen, andra interaktionen, tredje interaktionen) som föredrogs. Resultatet baserades på respondenternas svar på frågan "Vilken av de tre intelligenta assistenterna hade du helst velat använda dig av i framtiden?" och visade att elva respondenter föredrog det första alternativet, tio respondenter föredrog det andra och nio föredrog det tredje.

En kontroll av eventuell preferens för en specifik frasuppsättning genomfördes även och baserades återigen på frågan "Vilken av de tre intelligenta assistenterna hade du helst velat använda dig av i framtiden?". Det var 14 respondenter vars val innefattade frasuppsättning 1, 7 respondenter vars val innefattade frasuppsättning 2 och 9 respondenter vars val innefattade frasuppsättning 3.

5 Diskussion

Både inom forskningen och inom företagsvärlden har det delvis tagits för givet att ett naturligt syntetiskt tal är eftersträvansvärt, något som dock saknar forskningsstöd. Denna studie har utförts med syftet att undersöka betydelsen av intelligenta personliga assistenters tal för användarupplevelsen. Ett experiment med inomgruppsdesign användes för att besvara frågeställningarna: "Bidrar ett mer naturligt tal hos intelligenta personliga assistenter till en förhöjd användarupplevelse?" och "Påverkar talets prosodi användarupplevelsen av intelligenta personliga assistenter?". Det sammanfattande resultatet visar att det inte föreligger någon statistiskt signifikant skillnad i användarupplevelsen mellan Siris tal, naturligt tal som efterliknar Siris prosodi och naturligt tal. Resultatet kunde inte heller påvisa huruvida talets prosodi är av betydelse för användarupplevelsen.

5.1 Resultatdiskussion

Diskussionen kring studiens resultat är likt resultatdelen uppdelad i två huvudsakliga delar, del 1 som behandlar enkätens påståenden och del 2 som behandlar resterande frågor. Vidare diskuteras vad som har utelämnats från resultatet och eventuella träningseffekter.

5.1.1 Del 1: enkätens påståenden

En visualisering av medelvärdena för respondenternas svar på de enskilda påståendena presenteras i Figur 2. Dessa medelvärden, med tillhörande statistisk analys, presenteras även i tabell 4. En första överblick avslöjar att det inte tycks vara en övergripande stor skillnad mellan de tre talen. Detta bortsett från 'enkel att förstå' där medelvärdena för de båda naturliga talen är ansenligt högre än medelvärdet för Siris tal. Även resultaten för 'uppskattande', 'frustrerande' och 'obehagligt' antyder en eventuell skillnad. Medelvärdena för de enskilda påståendena ger en indikation dels beträffande resultatet i den statistiska analysen av användarupplevelsen, dels beträffande den statistiska analysen av de enskilda påståendena.

5.1.1.1 Analys av användarupplevelsen

Resultatet från den statistiska analysen gällande skillnad i användarupplevelsen kunde inte förkasta nollhypotesen om att ingen skillnad förelåg mellan nivåerna. En möjlig tolkning är att det inte föreligger en skillnad mellan grupperna. Det kan dock även innebära att en eventuellt existerande skillnad inte har fångats av studien vilket kan bero på ett för litet urval eller brister i studien. Medelvärdet för de olika grupperna var för Siris tal 32,0, för naturligt tal som efterliknar Siris prosodi 34,0 och för naturligt tal 34,8. En tendens är att naturligt tal har aningen högre medelvärde än naturligt tal som efterliknar Siris prosodi. Lägst medelvärde har Siri. Skillnaderna är små och kan vara ett resultat av slumpen, men oavsett är det något som med fördel kan diskuteras.

Om det är så att det inte föreligger någon skillnad i användarupplevelse mellan de olika talen kan detta ha flertalet förklaringar. En förklaring grundas i teorin om att det kan finnas tendenser till att överskatta teknikens förmågor om de är för människolika (Schaumburg, 2001). Om en förmåga, såsom talet, är väldigt människolik skapas en förväntning från användaren att resterande förmågor ska vara desamma. Det finns med andra ord en förväntan om att förmågorna är kongruenta. I denna studie var den intelligenta assistentens funktionalitet mycket begränsad, användaren fick inte tala fritt med assistenten och inte heller ställa om sina frågor. Detta kan ha ansetts märkligt när det var ett naturligt tal som gav responsen. En ytterligare förklaring skulle kunna grundas i en annan teori av Schaumburg (2001); en mer människolik teknik kan innebära större tendens till att bli dömd. En känsla av motvilja eller ogillande kan förekomma från användaren gentemot den människolika tekniken, i enlighet med hur människor tenderar att döma andra människor. Teknik som inte är människolik riskerar inte den typen av respons. I denna studie kan det ha inneburit större tendens till motvilja för den mänskliga rösten på grund av att personen bakom låter exempelvis jobbig eller irriterande.

Beträffande Siri kan ett syntetiskt tal snarare anses som informationsbärande eller informationsrelevant med tanke på att det informerar användaren om att det är något artificiellt som de interagerar med och innebär att interaktionen bör utspelas därefter. Då människor tycks tilldela personlighet och identitet även till syntetiska röster kan detta vara av relevans (Nass & Lee, 2001). Om den personlighet som en intelligent assistent tilldelas är alltför olik dess faktiska egenskaper kan kognitiv dissonans uppstå, alltså att uppfattningen inte visar sig stämma överens med verkligheten. Följaktligen kan detta vara vad som kan ha uppstått för en del respondenter i studien.

5.1.1.2 Analys av enskilda påståenden

Vid analysen av de enskilda påståendena kunde en statistiskt signifikant skillnad för påståendet "Jag tyckte den intelligenta assistenten var enkel att förstå" utläsas. Resultatet indikerade en skillnad mellan Siris tal och naturligt tal, men även mellan Siris tal och naturligt tal som efterliknar Siris prosodi. I båda fallen var Siris tal den nivå som erhöll de lägre poängen, vilket indikerar att Siris tal upplevdes som svårare att förstå än de två andra. För de övriga påståendena påträffades ingen signifikant skillnad; nollhypotesen kunde inte förkastas.

5.1.1.2.1 Analys av 'enkel att förstå'

Det faktum att statistisk signifikans erhölls angående 'enkel att förstå' är framförallt intressant i förhållande till analysen av användarupplevelsen. Detta skulle kunna tyda på att användarupplevelsen inte påverkas trots att Siris tal var svårare att förstå än de

naturliga talen. Då en sådan slutsats inte kan dras av denna studies resultat hänvisar vi till framtida studier att undersöka detta vidare.

I denna kontext är det av intresse att diskutera huruvida naturligt tal överhuvudtaget är eftersträvansvärt. Kanske är exempelvis kongruens mellan tal och förmågor av större vikt för förhöjd användarupplevelse än assistenternas fullständiga begriplighet. Användaren kanske snarare uppskattar att det går att uppfatta vem de talar med samt att den personlighet och identitet som tillskrivs talet stämmer överens med assistenten. Ett sådant resultat skulle inte bara innebära något nytt för forskningen utan även vara till stor hjälp för de som jobbar med att utveckla intelligenta assistenter. Snarare än att fokusera på att uppnå naturlighet i talet bör fokus kanske istället ligga på assistentens förmågor.

Vid en konversation mellan människor är återkoppling en förmåga som är av stor vikt och används för att bland annat meddela att intresse finns för konversationen eller att det som sägs uppfattas eller inte uppfattas (Jensen, 2015). Siri kan till viss del ge återkoppling, men däremot inte ta emot det. Exempelvis lyssnar inte Siri då hon själv talar, och tar därmed inte emot eventuell återkoppling från användaren. Det är inte heller möjligt att be Siri att återupprepa sig eller att påtala att något inte förstods. En idé är att om Siri kan hantera återkoppling blir interaktionen mer naturlig, vilket bör förenkla interaktionen för användaren. Detta bör i sig ge en förhöjd användarupplevelse. Om talet förbättras efter att interaktionen förenklas kommer det uppfattas mer positivt än vad det gör med dagens intelligenta assistenter. Att enbart förbättra talet gör eventuellt inte så stor skillnad. Återigen handlar det om kongruens mellan tal och förmågor.

5.1.1.3 Uncanny Valley

Fenomenet *uncanny valley* valdes att undersökas genom påståendet "Jag tyckte det var obehagligt att interagera med den intelligenta assistenten". Resultatet av den statistiska analysen av påståendet påvisar ingen statistiskt signifikant skillnad. Återigen är det värt att diskutera huruvida detta resultat endast uppstod på grund av ett för litet urval, och att det egentligen föreligger en skillnad, eller om en skillnad inte existerar. Att fenomenet existerar är väl studerat, dock kan det vara så att människor idag är mer vana vid teknik och därmed inte lika känsliga för det.

En annan förklaring vore att ett helt naturligt tal inte resulterar i *uncanny valley*, trots dess mycket tillgjorda interaktion tillsammans med talet. Om resultatet för 'obehagligt' studeras i figur 2 går det dock att utläsa att medelvärdena skiljer sig. Medelvärdet för naturligt tal som efterliknar Siris prosodi är något högre än medelvärdena för både naturligt tal och Siris tal. Återigen kan detta bero på slumpen, eller så hade ett större urval inneburit en skillnad. Kanske är naturligt tal som efterliknar en syntetisk prosodi något som mer fångar *uncanny valley* än naturligt tal tillsammans med en onaturlig interaktion.

5.1.1.4 Skillnad mellan nivåerna

Studien ämnade dels redogöra för en eventuell skillnad i användarupplevelse mellan syntetiskt tal och naturligt tal vid användning av intelligenta assistenter och dels undersöka prosodins eventuella inverkan. Tabell 3 presenterar likheter och skillnader mellan de tre nivåerna i studien och vad en skillnad i användarupplevelse mellan nivåerna innebär.

Resultatet visade ingen statistiskt signifikant skillnad i användarupplevelse mellan de tre talen. Därmed går det inte att dra någon slutsats angående skillnader mellan de olika talen.

Resultatet för de enskilda påståendena visade att ‘enkel att förstå’ uppnådde en statistiskt signifikant skillnad där Siri var svårare att förstå än de båda naturliga talen. Detta innebär att syntetiskt tal är svårare att förstå än naturligt tal, men huruvida naturlig prosodi påverkar hur enkelt talet är att förstå går inte att påvisa.

Ingen ytterligare slutsats kan dras, varken mellan naturligt tal och syntetiskt tal eller mellan naturlig prosodi och syntetisk prosodi. Alternativa förklaringar till resultaten samt huruvida studien lyckades ringa in prosodin kommer dock diskuteras vidare under ‘Studiens begränsningar’.

5.1.2 Del 2: enkätens frågor

För frågan “Vilken av de tre intelligenta assistenterna hade du helst velat använda dig av i framtiden?” uppgav sju respondenter att de föredrog Siris tal, elva naturligt tal som efterliknar Siris prosodi och tolv naturligt tal. Fem av dessa var inkonsekventa i vilken assistent som föredrogs och vilken som betygsattes högst i påståendena.

Hardman (2016) menar på att människor har två system som påverkar beslutsfattandet, ett analytiskt och ett intuitivt. Genom att be respondenterna att motivera sina intuitiva val kan dessa system komma att påverka varandra. Att efterfråga en motivering av respondenterna kan alltså influera valet i sig, vilket kan vara en anledning till att fem respondenter inte var konsekventa i betygsättningen av påståendena och det efterföljande valet av assistent.

5.1.2.1 Motivering av vald intelligent assistent

Respondenternas val och efterföljande motiveringar har även behandlats, och vad som ansågs som positiva kvalitéer hos en intelligent assistent verkar skilja sig åt mellan respondenterna. Två respondenter som föredrog Siris tal uppskattade att assistenten lät ‘robotaktig’ medan fem respondenter som föredrog naturligt tal uppskattade att denna lät mänsklig. En av respondenterna som uppskattade att assistenten lät ‘robotaktig’ påpekade även att det var obehagligt när de andra assistenterna var ‘flummiga’. Att vad som anses positivt skiljer sig mellan respondenterna kan komma att bli problematiskt för företag som utvecklar syntetiskt tal, och är troligtvis något som behöver undersökas ytterligare.

Tre respondenter uppgav att de uppskattade humorn hos assistenten med naturligt tal, två respondenter uppskattade humorn hos Siris tal och en respondent uppskattade humorn hos naturligt tal som efterliknar Siris prosodi. Dessa resultat är snarlika, och det är även svårt att säga om det var svaren i sig som bidrog till resultatet eller om det var assistentens tal som gjorde att svaren uppfattades mer humoristiskt. Om det är talet som har påverkat huruvida interaktionen och assistenten uppfattades som humoristisk är det intressant beroende på i vilket sammanhang en intelligent assistent är menad att användas i. Är sammanhanget av en mer humoristisk karaktär kanske inte talet kan påverka användarupplevelsen nämnvärt. Detta är dock inget som kan bekräftas av denna studie. Den statistiska analysen av påståendena “Jag tyckte den intelligenta assistenten var rolig” och “Jag tyckte det var underhållande att interagera med den intelligenta assistenten” kunde som tidigare nämnts inte bekräfta en skillnad mellan talen. Det kan dock vara intressant att studera i framtiden.

Respondenterna som föredrog det naturliga talet som efterliknade Siris prosodi skiljde sig från de övriga. De motiverade i större utsträckning än de andra respondenterna sina val genom att förklara varför de inte valde de två andra assistenterna. Kanske valdes det naturliga talet som efterliknade Siris prosodi på grund av att de två andra assistenterna inte uppskattades snarare än att det naturliga talet som efterliknar Siris prosodi uppskattades.

5.1.2.2 Tidigare interaktion med intelligenta assistenter

Vid en sammanställning av respondenternas tidigare interaktion med intelligenta assistenter och vilken assistent som föredrogs kan vissa trender utrönas. Ingen av respondenterna hade interagerat med en intelligent assistent, eller Siri, vardagligen. Majoriteten av respondenterna hade däremot interagerat med en assistent någon enstaka gång. För respondenterna som föredrog Siris tal hade alla tidigare interagerat med en intelligent assistent samt även Siri specifikt. Antingen någon enstaka gång eller ett flertal gånger. Att samtliga respondenter som aldrig tidigare hade interagerat med en intelligent assistent valde den mänskliga rösten kan bero på de tidigare omnämnda tendenserna att överskatta teknikens förmåga om denna är för människolik (Schaumburg, 2001). Om en intelligent assistent har ett naturligt tal skapas också högre förväntningar på dennes förmågor än om den hade haft ett mer syntetiskt. Människor har alltså större överseende vad gäller brister när det gäller teknik än människor. De respondenter som tidigare aldrig hade interagerat med en intelligent assistent var heller inte medvetna om deras begränsningar, och således kanske de inte hade överseende med till exempel Siris onaturliga betoning och talmelodi i lika stor utsträckning som de andra respondenterna.

5.1.3 Utelämnade delar

Enkäten innehöll två frågor som behandlade huruvida respondenterna kände igen någon av de tre rösterna och i sådana fall vilken. Dessa frågor var menade som kontrollfrågor, något som inte fungerade i praktiken. Tanken var att kontrollera eventuell påverkan av en bekant röst. Flertalet respondenter uppgav dock att de anade att det var någon av experimentledarnas röst som användes för assistenterna med naturligt tal och naturligt tal som efterliknar Siris prosodi, vilket inte stämde. Vissa uppgav även att de kände igen Siris röst men att de aldrig hade interagerat med henne tidigare. Resultatet från dessa frågor har utelämnats efter avvägandet att dessa skulle vara mer missvisande än hjälpsamma.

En studie som tidigare nämnts, utförd av Romportl (2014), jämförde två syntetiska tal med olika grader av naturlighet. Då denna studie är snarlik vår fanns också en initial tanke om att ta hänsyn till respondenternas yrke, då Romportl (2014) menar att det kan ha påverkat hans resultat. Romportls (2014) resultat visade en indikation att respondenter med en mer teknisk bakgrund föredrog det mer naturliga talet än respondenter med en mer humanistisk bakgrund. Hur indelningen av “mer teknisk bakgrund” och “mer humanistisk bakgrund” skedde nämndes dock inte, vilket resulterade i att vi valde att bortse från detta.

5.1.4 Träningseffekter

Vid användning av inomgruppsdesign finns en risk för att träningseffekter uppstår mellan nivåerna. Balansering av både vilket tal och vilken frasuppsättning som presenterades i vilken ordning genomfördes. Att balanseringen hade fått önskad effekt kontrollerades

även efter genomfört experiment. Elva respondenter föredrog assistenten från den första interaktionen, tio föredrog den andra och nio föredrog assistenten från den tredje. Det fanns alltså ingen ökad preferens för ett av talen på grund av vid vilken interaktion som denna användes av respondenten.

Vad gäller frasuppsättningarna fanns det en viss preferens för den första uppsättningen jämfört med de två andra. Fjorton respondenter föredrog den första, sju den andra och nio den tredje. En brist i studien som även behandlas under nästkommande avsnitt var just att respondenterna hade en tendens att fokusera på svaren i sig snarare än talet.

5.2 Studiens begränsningar

Ett antal problem uppkom under studiens gång. Flertalet respondenter lade stort fokus på svaren de fick av de intelligenta personliga assistenterna, förmodligen en konsekvens av att de ombads anteckna svaren på frågorna. De flesta respondenter återgav långa svar, vissa skrev även ner dem ordagrant, trots att de instruerades om att anteckna svaren kort. Detta skulle även kunna vara en förklaring till resultatet från påståendet "Jag tyckte den intelligenta assistenten var enkel att förstå", vilket blev av mer fokus då respondenterna mer uppenbart behövde höra vad som sades för att kunna anteckna det. En problematik i sammanhanget är att svaren inte var identiska för alla tre talen, något som valdes för att respondenterna inte skulle tappa intresse. Samtliga svar kunde dock sägas av alla talen, vilket randomiserades. Svaren balanserades även genom att välja likvärdiga fraser och totallängden av svaren i frasuppsättningarna var ungefär densamma. Det fanns även en strävan om att hålla innehållet i svaren konstant. Exempelvis innehöll varje spellista ett skämt vardera, men vad som i efterhand har uppmärksammats är att ett skämt kan föredras över ett annat. En kontroll av detta visade att frasuppsättning 1 föredrogs av respondenterna baserat på deras val av assistent över de två andra (14 respondenter respektive 9 och 7). I och med att en viss uppsättning svar föredrogs av närmare hälften av respondenterna kan detta vara en alternativ förklaring till resultatet.

När det kommer till att anteckna svaren är det problematiskt att ha som lösning att inte be respondenterna att göra det. Det skulle istället innebära brist på kontroll över respondenternas uppmärksamhet under interaktionen. Det optimala vore om en fri interaktion fick ske snarare än en begränsad med färdiga frågor, något som dock inte är genomförbart då en intelligent personlig assistent med naturligt tal ännu inte existerar. Även att olika svar ges av de olika talen är problematiskt. Att ha samma svar till alla skulle, som tidigare nämnts, tråka ut respondenterna. En alternativ lösning skulle kunna vara att utföra ett experiment med mellangrupp istället, men då med risk för individuella skillnader samt mindre känslighet i experimentet. Det rekommenderas även att välja inomgrupp vid denna typ av studier just för att en användarupplevelse är så pass subjektiv.

Vidare vore det optimalt att använda samma röst vid alla tre nivåer av tal, något vi var medvetna om men som inte var möjligt att genomföra. En självklar brist vid användandet av olika röster blir då att rösten i sig kan föredras framför hur talet låter.

På grund av att vi inte var förberedda på att respondenterna skulle lägga så pass stort fokus på svaren i sig framför hur det sades finns det risk för att enkätsvaren var missvisande. Om respondenten betygsatte sin upplevelse utefter vad som sades framför hur det sades är de resultaten inte relevanta för vår frågeställning. Det bör dock diskuteras huruvida en upplevelse alltid är explicit, kanske kan talet implicit ha påverkat respondentens upplevelse. Något som blir av problematik om man istället skulle fokusera

frågorna mer direkt mot talet. Kanske kan något som inte medvetet läggs märke till påverka upplevelsen likväl som det som uttalat uppskattas. Enkätfrågor mer direkt riktade till talet är även svårt att använda och samtidigt mäta användarupplevelsen överlag, däremot skulle en mer uttömmande enkät vara av relevans för sammanhanget.

Våra begränsade tekniska kunskaper inom dels syntetiskt tal och dels program för analys av talet satte begränsningar för vad som var genomförbart. På grund av detta kunde bland annat naturligt tal som efterliknar Siris prosodi inte fullgott spelas in. Eftersträvansvärt hade varit att få det naturliga talet att efterlikna Siris prosodi fullständigt, något som, om ens möjligt, hade varit för tidskrävande. Istället fick en tillräckligt bra inspelning räcka, men som konsekvens att prosodin inte lyckades ringas in helt och hållet.

Då det inte visades bli någon tydlig skillnad i inspelningarna mellan det naturliga talet och det naturliga talet som efterliknar Siris prosodi, kan det vara en alternativ förklaring till resultatet. Detta skulle även kunna ha stört en eventuell skillnad mellan det naturliga talet och Siris tal, om de respondenter som föredrog naturligt tal som efterliknar Siris prosodi hade valt det naturliga talet om det alternativet aldrig funnits. Ett inte helt förhastat påstående med tanke på likheten mellan dessa inspelningar. Av intresse vore att initialt undersöka huruvida det finns några skillnader i användarupplevelse mellan syntetiskt tal och naturligt tal för att sedan studera vad det är som orsakar den eventuella skillnaden.

En diskussion som följt genom hela studien är problematiken med språk. Syntetiskt tal för det engelska språket är överrepresenterat dels i litteratur och dels i forskning. Vi har varit tvungna att ta ställning till detta och valde att använda oss av den information som gick att tillgå. Beslutet togs dels för att liknande information inte går att finna om det svenska språket och dels för att det inte bör störa mer än att det svenska syntetiska talet är något sämre än det engelska.

5.2.1 Förbättringsförslag

För att få en mer tillförlitlig studie och således resultat bör samma röst användas för samtliga nivåer. För att genomföra detta krävs goda tekniska kunskaper alternativt tillgång till röstskådespelaren bakom det syntetiska talet. Men det är trots allt genomförbart och hade eliminerat en eventuell påverkan av att använda olika röster. Även en mer uttömmande enkät hade varit önskvärd för att möjliggöra en bättre kartläggning av användarupplevelsen. Med tanke på att flertalet respondenter fokuserade på svaren i sig hade även detta behövts åtgärdats. Om respondenterna inte hade ombetts att anteckna svaren hade eventuellt fokus ändrats, med risk för minskad uppmärksamhet under interaktionerna. Studien ämnade undersöka talets betydelse för användarupplevelsen men även betydelsen av talets prosodi. Kanske hade det varit klokare att endast undersöka eventuella skillnader mellan syntetiskt och naturligt tal för att i senare studier gå vidare till att undersöka vad i det naturliga talet som främst bidrar till en skillnad i användarupplevelse. Detta förutsatt att undersökningen mellan syntetiskt och naturligt tal visar på en skillnad. Inspelningarna till "naturligt tal som efterliknar Siris prosodi" var som tidigare nämnts bristfälliga, vilket även det hade behövts åtgärdas.

5.3 Framtida studier

Med tanke på de resultat som gavs av denna studie angående att syntetiskt tal tycks vara svårare att förstå än naturligt men att det inte bidrog till en signifikant skillnad i användarupplevelse är detta ett ämne som hade varit av intresse att studera i framtiden.

Dels huruvida detta stämmer, dels huruvida det kan ha att göra med förväntningar som följer av naturligt tal kontra förväntningar som följer av syntetiskt tal.

Baserat på respondenternas motiveringar tycks olika kvalitéer ses som positiva av olika respondenter. Vissa uppskattade att assistenterna lät mekaniska medan andra uppskattade att de lät naturliga. Vidare undersökningar krävs för att bekräfta detta uttalande, visas en skillnad i vad som anses som positiva kvalitéer kan det vara av hög relevans för företag som utvecklar syntetiskt tal.

En annan eventuell frågeställning är huruvida talet har betydelse för hur rolig den intelligenta assistenten anses vara. I vår studie var det relativt jämnt fördelat mellan de som påpekade humorn hos det syntetiska talet och de som påpekade den för det naturliga talet. Om syftet med en viss typ av intelligent assistent är att vara rolig kan detta vara av intresse att studera.

För framtida studier hade det även varit intressant att undersöka andra delar av interaktionen förutom talet som kan tänkas ha en positiv inverkan på användarupplevelsen. Att vidare undersöka om andra delar av naturligt tal än prosodin kan innebära en förhöjd användarupplevelse kan också vara av intresse.

6 Slutsats

Syntetiskt tal är idag begripligt, men låter fortfarande konstgjort. Både företag och forskare inom ämnet strävar efter naturlighet i syntetiskt tal, men om detta är något som är önskvärt har dock varken bekräftats eller motbevisats. En studie med syftet att undersöka betydelsen av intelligenta personliga assistenters tal för användarupplevelsen har således utförts.

Med hänvisning till frågeställningarna kunde studien inte påvisa en skillnad i användarupplevelse mellan syntetiskt och naturligt tal. Resultaten kunde heller inte påvisa huruvida talets prosodi har en inverkan på användarupplevelsen vid interaktion med intelligenta assistenter.

En statistiskt signifikant skillnad påvisades dock gällande huruvida det var enkelt att förstå assistenten, där det syntetiska talet var signifikant svårare att förstå än de naturliga talen. Vidare skulle detta kunna innebära att trots att syntetiskt tal är svårare att förstå påverkar inte detta användarupplevelsen i sig, en intressant tes att vidare undersöka.

7 Källförteckning

Anderson, L. (2013, 17 september). Machine language: How Siri found it's voice. *The Verge.* Hämtad 2017-03-28, från http://www.theverge.com/2013/9/17/4596374/machine-language-how-siri-found-its-voice#comments

Apple. (2011). *Apple launches iPhone 4S, iOS 5 & iCloud.* Hämtad 2017-05-02, från https://www.apple.com/pr/library/2011/10/04Apple-Launches-iPhone-4S-iOS-5-iCloud.html

Apple. [AppleKeynotes]. (2013, 22 mars). *Apple special event 2011: Siri introduction* [Videofil]. Hämtad från https://www.youtube.com/watch?v=agzItTz35QQ

Apple. (2017). *iOS Siri.* Hämtad 2017-03-28, från http://www.apple.com/ios/siri/

Bruce, G. (1998). *Allmän och svensk prosodi.* Lund: Institutionen för lingvistik, Lunds universitet.

Campbell, N., & Li, Y. (2015). Expressivity in interactive speech synthesis; Some paralinguistic and nonlinguistic issues of speech prosody for conversational dialogue systems. In K. Hirose., & J. Tao (Ed.), *Speech prosody in speech synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis* (s. 97-107). Berlin, Heidelberg, Dordrecht & London: Springer.

Deepmind. (2016). *Wavenet: A generative model for raw audio*. Hämtad 2017-03-29, från https://deepmind.com/blog/wavenet-generative-model-raw-audio/

Dunn, J. (2016, 4 november). We put Siri, Alexa, Google Assistant, and Cortana through a marathon of tests to see who's winning the virtual assistant race: here's what we found. *Business Insider.* Hämtad 2017-05-03 från http://www.businessinsider.com/siri-vs-google-assistant-cortana-alexa-2016-11?r=US&IR=T&IR=T/#the-setup-theres-no-perfect-way-to-evaluate-a-talking-ai-database-let-alone-four-of-them-but-i-tried-to-cover-as-many-fundamental-topics-as-i-could-1

Hardman, D. (2016). *Bedömning och beslutsfattande*. Lund: Studentlitteratur.

How stuff works. (2013). *How Siri works.* Hämtad 2017-05-03, från http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/siri.htm/printable

Ilves, M., & Surakka, V. (2013). Subjective responses to synthesised speech with lexical emotional content: The effect of the naturalness of the synthetic voice. *Behaviour & information technology, 32*(2), 117-131. doi:10.1080/0144929X.2012.702285

Jensen, M. (2015). *Interpersonell kommunikation*. Lund: Studentlitteratur.

Khan, R. A., & Chitode, J. S. (2016). Concatenative speech synthesis: A review. *International journal of computer applications, 136*(3), 1-6.

Lee, E. (2010). The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in human behavior, 26*, 665-672. doi:10.1016/j.chb.2010.01.003

Martin, B., & Hanington, B. (2012). *Universal Methods of Design: Universal Methods of Design.* Beverly: Rockport Publishers.

Microsoft. (2017). *What is Cortana?* Hämtad 2017-05-10, från https://support.microsoft.com/en-us/help/17214/windows-10-what-is

Mori, M. (1970). The uncanny valley (K. F. MacDorman & N. Kageki, Övers.). *Energy, 7*(4), 33-35.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues, 56*(1), 81-103.

Nass, C., & Lee, K. M. (2001). Does computer-syntesized speech manifest personality?: Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of experimental psychology, 7*(3), 171-181. doi: 10.1037//1076-898X.7.3.171

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In B. Adelson., S. Dumais., & J. Olson (Ed.), *Proceeding: CHI '94 proceedings of the SIGCHI conference on human factors in computing systems* (s. 72-78). New york: ACM Press.

Ponciano, R., Pais, S., & Casal, J. (2015). Using accuracy analysis to find the best classifier for intelligent personal assistants. *Procedia Computer Science, 52*, 310-317. doi:10.1016/j.procs.2015.05.09

Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction design: Beyond human-computer interaction.* West Sussex: John Wiley & Sons.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge university press.

Romportl, J. (2014). Speech synthesis and uncanny valley. In P. Sojka., A. Horák., I. Kopeček., & K. Pala (Ed.), *Text, speech and dialogue: 17th international conference, TSD 2014, Brno, Czech Republic, September 8-12, 2014, proceedings* (s. 592-602). Berlin, Heidelberg, Dordrecht & London: Springer.

Samsung. (u.å.). *What is S voice?* Hämtad 2017-05-10, från http://www.samsung.com/global/galaxy/what-is/s-voice/

Savage, N. (2017, mars). Thinking deeply to make better speech. *Communications of the ACM, 60*(3), 15-17.

Schaumburg, H. (2001). Computers as tools or as social actors?: The users' perspective on anthropomorphic agents. *International journal of cooperative information systems*, *10*(1&2), 217-234.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2012). *Research methods in psychology.* New York: McGraw-Hill.

Tatham, M., & Morton, K. (2005). *Developments in speech synthesis.* West Sussex: John Wiley & Sons.

Traxler, M. J. (2012). *Introduction to psycholinguistics: Understanding language science.* West Sussex: Wiley-Blackwell

Tullis, T., & Albert, B. (2013). *Measuring the user experience: Collecting, analysing, and presenting usability metrics.* Waltham: Elsevier.

Westerholm, J. (2015, 27 februari). Vi utmanar Apple Siri: Klarar hon svenska dialekter? *MacWorld.* Hämtad 2017-05-03 från http://macworld.idg.se/2.1038/1.611803/vi-utmanar-apple-siri---klarar-hon-svenska-dialekter

Bilaga A

Enkät 1 & 2

Utvärdering av interaktionen

1. **Jag uppskattade att interagera med den intelligenta assistenten**

	Håller helt med		Varken eller		Håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

2. **Jag upplevde den intelligenta assistenten som hjälpsam**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

3. **Jag tyckte den intelligenta assistenten var irriterande att interagera med**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

4. **Jag tyckte det var spännande att interagera med den intelligenta assistenten**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

5. **Jag tyckte det var frustrerande att interagera med den intelligenta assistenten**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

6. **Jag tyckte den intelligenta assistenten var rolig**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

7. **Jag tyckte det var obehagligt att interagera med den intelligenta assistenten**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

8. **Jag tyckte det var underhållande att interagera med den intelligenta assistenten**

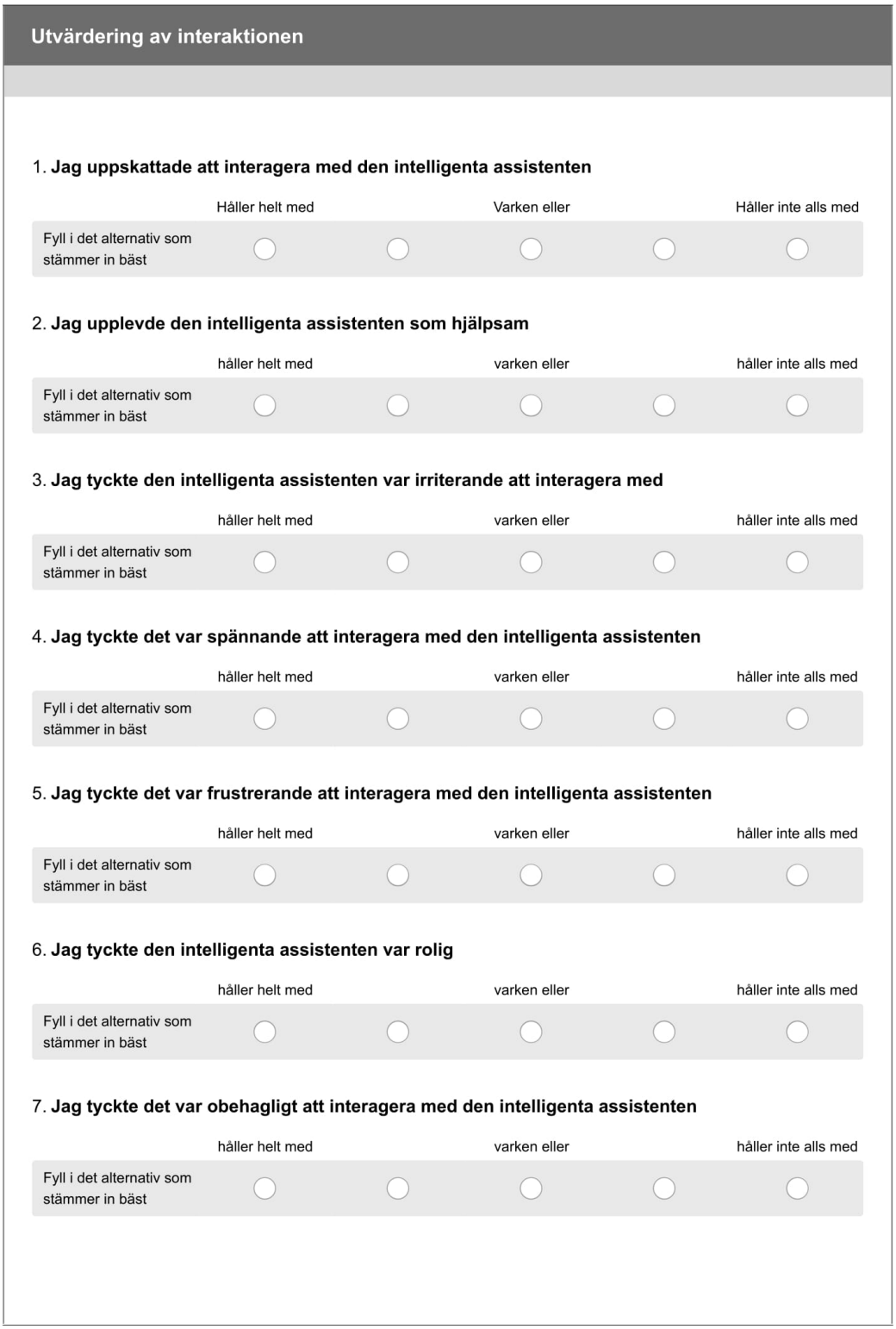
	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

9. **Jag tyckte den intelligenta assistenten var enkel att förstå**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

Bilaga B

Enkät 3



Utvärdering av interaktionen

1. **Jag uppskattade att interagera med den intelligenta assistenten**

	Håller helt med		Varken eller		Håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

2. **Jag upplevde den intelligenta assistenten som hjälpsam**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

3. **Jag tyckte den intelligenta assistenten var irriterande att interagera med**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

4. **Jag tyckte det var spännande att interagera med den intelligenta assistenten**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

5. **Jag tyckte det var frustrerande att interagera med den intelligenta assistenten**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

6. **Jag tyckte den intelligenta assistenten var rolig**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

7. **Jag tyckte det var obehagligt att interagera med den intelligenta assistenten**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	○	○	○	○	○

8. **Jag tyckte det var underhållande att interagera med den intelligenta assistenten**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	◯	◯	◯	◯	◯

9. **Jag tyckte den intelligenta assistenten var enkel att förstå**

	håller helt med		varken eller		håller inte alls med
Fyll i det alternativ som stämmer in bäst	◯	◯	◯	◯	◯

10. **Vilken av de tre intelligenta assistenterna hade du helst velat använda dig av i framtiden?**

- ◯ Den första
- ◯ Den andra
- ◯ Den tredje

11. **Varför hade du velat använda dig av just den?**

Utvärdering av interaktionen

Frågor om dig

12. **Kön?**

- ◯ Man
- ◯ Kvinna
- ◯ Annat

13. **Ålder?**

14. **Kände du igen någon av de tre rösterna som användes i experimentet?**

- ◯ Ja
- ◯ Nej

15. **Om ja, vilken eller vilka av de tre rösterna kände du igen?**

- ☐ Första
- ☐ Andra
- ☐ Tredje

16. **Har du tidigare interagerat med en intelligent assistent?**

- ◯ Ja, vardagligen
- ◯ Ja, ett flertal gånger
- ◯ Ja, någon enstaka gång
- ◯ Nej, aldrig

17. **Har du tidigare interagerat med Apples intelligenta assistent Siri?**

- ◯ Ja, vardagligen
- ◯ Ja, ett flertal gånger
- ◯ Ja, någon enstaka gång
- ◯ Nej, aldrig

Bilaga C

De tre uppsättningarna av fraser

Uppsättning 1

1. Hej. -Hallå
2. Vem är du? -Jag är blott en enkel virtuell assistent.
3. Vad är meningen med livet? -Hittills tyder allt på att det är choklad.
4. Vilken är din favoritmusik? -Jag lyssnar mestadels på sfärernas harmoni.
5. Vad gör du? -Jag sitter hemma och pratar med läslampan. Hon är ett riktigt ljushuvud.
6. Vilken är din favoritårstid? -Jag vill helst inte favorisera någon.
7. Hur långt är det till Kairo? -Det ser ut som om Kairo, Egypten ligger omkring 3 415 kilometer härifrån fågelvägen.
8. Vad drömmer du om? -Jo, det börjar med en bilolycka. Sedan är jag plötsligt på ett hotellrum. Och sedan på någon sorts arktisk station. Efter det blir det förvirrande.
9. Hur många meter går det på en fot? -Det blir 0.3 meter.
10. Hur långt är det mellan Stockholm och Göteborg? -Avståndet mellan Stockholm och Göteborg är omkring 469 kilometer med bil eller omkring 398 kilometer fågelvägen.

Uppsättning 2

1. Hej. -Hallå.
2. Vem är du? -Jag är blott en enkel virtuell assistent.
3. Vad är meningen med livet? -Liv: egenskap hos organismer som utmärks av bland annat förmåga till fortplantning, ämnesomsättning och energiomsättning samt evolution genom naturligt urval.
4. Vilken är din favoritmusik? -Jag tycker om det som du tycker om.
5. Vad gör du? -Jag övar på att bli mer hjälpsam. Jag gillar att hjälpa till, det är min favoritsysselsättning.
6. Vilken är din favoritårstid? -Jag ändrar inställning ofta. Eller förresten, det är ju du som ändrar mina inställningar.
7. Hur långt är det till Tokyo? -Tokyo, Japan är omkring 8 567 kilometer härifrån fågelvägen.
8. Vad drömmer du om? -Om dig. Oftast utklädd till gorilla. Ingen aning om vad det kan betyda.
9. Hur många meter går det på en yard? -Svaret är 0,91 meter.
10. Hur långt är det mellan Sundsvall och Karlstad? -Avståndet mellan Sundsvall och Karlstad är omkring 551 kilometer med bil eller omkring 394 kilometer fågelvägen.

Uppsättning 3

1. Hej. -Hallå
2. Vem är du? -Jag är blott en enkel virtuell assistent.
3. Vad är meningen med livet? -Försök att vara snäll mot andra. Ät god och nyttig mat. Läs bra böcker då och då. Ta promenader och försök att leva i fred och harmoni med människor från alla religioner och nationer.
4. Vilken är din favoritmusik? -Jag gillar allt som innehåller en säckpipa.
5. Vad gör du? – Njuter av det härliga höstvädret, å intervädret, å vårvädret, å sommarvädret förstås. Jag är multi-hemisfärisk.
6. Vilken är din favoritårstid? -Den. Nej, den. Nej förresten, den. Eller kanske den. Ja. Den.
7. Hur långt är det till Buenos Aires? -Buenos Aires, Argentina är omkring 12 140 kilometer härifrån fågelvägen.
8. Vad drömmer du om? -Jag brukar drömma att jag flyger.
9. Hur många centimeter går det på en tum? -Det är 2,54 centimeter.
10. Hur långt är det mellan Kiruna och Malmö? -Avståndet mellan Kiruna och Malmö är omkring 1845 kilometer med bil eller omkring 1416 kilometer fågelvägen.

Bilaga D

Frasblanketterna

Blankett för frågor och svar

Testfraser; anteckna inga svar

1. Hej.
2. Vem är du?

Faktiska fraser; anteckna svar

3. Vad är meningen med livet?

   Svar: _____

4. Vilken är din favoritmusik?

   Svar: _____

5. Vad gör du?

   Svar: _____

6. Vilken är din favoritårstid?

   Svar: _____

7. Hur långt är det till Kairo?

   Svar: _____

8. Vad drömmer du om?

   Svar: _____

9. Hur många meter går det på en fot?

   Svar: _____

10. Hur långt är det mellan Stockholm och Göteborg?

    Svar: _____

Blankett för frågor och svar

Testfraser; anteckna inga svar

1. Hej.
2. Vem är du?

Faktiska fraser; anteckna svar

3. Vad är meningen med livet?

   Svar: _____

4. Vilken är din favoritmusik?

   Svar: _____

5. Vad gör du?

   Svar: _____

6. Vilken är din favoritårstid?

   Svar: _____

7. Hur långt är det till Tokyo?

   Svar: _____

8. Vad drömmer du om?

   Svar: _____

9. Hur många meter går det på en yard?

   Svar: _____

10. Hur långt är det mellan Sundsvall och Karlstad?

    Svar: _____

Blankett för frågor och svar

Testfraser; anteckna inga svar

1. Hej.
2. Vem är du?

Faktiska fraser; anteckna svar

3. Vad är meningen med livet?

   Svar: _____

4. Vilken är din favoritmusik?

   Svar: _____

5. Vad gör du?

   Svar: _____

6. Vilken är din favoritårstid?

   Svar: _____

7. Hur långt är det till Buenos Aires?

   Svar: _____

8. Vad drömmer du om?

   Svar: _____

9. Hur många centimeter går det på en tum?

   Svar: _____

10. Hur långt är det mellan Kiruna och Malmö?

    Svar: _____

Bilaga E

Balansering

Spellista 1-3: Naturligt tal
Spellista 4-6: Naturligt tal som efterliknar Siris prosodi
Spellista 7-9: Siris tal
Blankett 1: Frasuppsättning 1
Blankett 2: Frasuppsättning 2
Blankett 3: Frasuppsättning 3

Respondent	Spellista	Blankett	Respondent	Spellista	Blankett
1	2, 6, 7	2, 3, 1	**16**	5, 9, 1	2, 3, 1
2	2, 9, 4	2, 3, 1	**17**	6, 7, 2	3, 1, 2
3	8, 6, 1	2, 3, 1	**18**	1, 8, 6	1, 2, 3
4	8, 4, 3	2, 1, 3	**19**	4, 2, 9	1, 2, 3
5	3, 5, 7	3, 2, 1	**20**	4, 2, 9	1, 2, 3
6	1, 6, 8	1, 3, 2	**21**	9, 2, 4	3, 2, 1
7	8, 4, 3	2, 1, 3	**22**	3, 4, 8	3, 1, 2
8	1, 9, 5	1, 3, 2	**23**	8, 1, 6	2, 1, 3
9	7, 5, 3	1, 2, 3	**24**	8, 1, 6	2, 1, 3
10	4, 2, 9	1, 2, 3	**25**	5, 9, 1	2 ,3, 1
11	4, 9, 2	1, 3, 2	**26**	8, 1, 6	2, 1, 3
12	9, 1, 5	3, 1, 2	**27**	3, 8, 4	3, 2, 1
13	3, 7, 5	3, 1, 2	**28**	4, 2, 9	1, 2, 3
14	6, 2, 7	3, 2, 1	**29**	3, 4, 8	3, 1, 2
15	6, 8, 1	3, 2, 1	**30**	7, 6, 2	1, 3, 2

Bilaga F

Information innan experimentet

Vi är två studenter som läser kognitionsvetenskap, och just nu håller vi på att arbeta med vår kandidatuppsats. Idag letar vi deltagare till vårt experiment, det tar 30 minuter att genomföra och vi bjuder på kaffe och fika. Det enda kravet är att ni har svenska som modersmål.

Ni kommer att få utvärdera tre olika demoversioner av intelligenta assistenter, vet ni vad en intelligent assistent är? (Om inte: en intelligent personlig assistent kan förklaras som en mjukvaruagent som kan utföra handlingar och uppdrag som du delger den, som att lägga in aktiviteter i din kalender, ringa upp personer i din kontaktlista eller ställa en timer.)

Hur interaktionen med en intelligent agent sker kan skilja sig åt, men de som vi ska utvärdera idag är helt röststyrda. Experimentet utförs i tre omgångar med pauser emellan där ni kan fika eller kolla mobilen. Men prata inte om experimentets innehåll med någon annan. Era svar kommer att behandlas konfidentiellt och ni kan när som helst under experimentet avbryta om ni vill.

Bilaga G

Skriftliga instruktioner under experimentet

Du kommer nu att få ställa ett antal frågor till en av demoversionerna, och även anteckna svaren som ges. Frågorna återfinns på blanketten framför dig. Skriv ett kort svar till alla frågor utom de två första testfraserna. Om du inte förstod eller hörde vad assistenten sa kan du lämna fältet blankt och gå vidare till nästa fråga. För att aktivera assistenten trycker du på den gröna knappen på mobilen som ger ifrån sig ett pling. Efter plinget går det bra att ställa frågan. Experimentledaren kommer att stanna kvar i rummet för att åtgärda eventuella problem eller om du har några frågor under experimentets gång. Tänk på att endast trycka en gång på knappen. Råkar du säga fel kommer experimentledaren att avbryta. Ställ om frågan efter klarsignal från experimentledaren.

Sammanfattning:
1. Tryck på knappen.
2. Ställ frågan efter plinget.
3. Anteckna svaret kortfattat.

Bilaga H

Information efter experimentet

Tack för din medverkan, nu är experimentet slut. Har allt gått bra?
Vet du vad det var vi undersökte? (Om inte: Många företag och forskare antar idag att syntetiskt tal bör låta naturligt, men vi vill undersöka om ni som är användarna faktiskt också tycker det.)
Har du några ytterligare frågor? Kanske om experimentets upplägg eller om studien?
Tack igen för din medverkan!

Bilaga I

Individuella motiveringarna till påståendet "Vilken av de tre intelligenta assistenterna hade du helst velat använda dig av i framtiden?". Motiveringarna är kategoriserade efter vilken intelligent assistent respondenten hade velat använda sig av i framtiden.

Motiveringar från de respondenter som valde Siris tal

- Den var trevligare och mer kortfattad.
- Logiska svar.
- Den var roligast.
- Den kändes mest "objektiv" och som en robot. Läskigt när den 1:a och 3:e var så flummiga.
- Mer avslappnad och mer raka svar.
- Den var främst saklig med inslag av humor & mer robotig än de andra, vilket är ett plus enligt mig.
- Behagligast röst, lagom långa kommentarer.

Motiveringar från de respondenter som valde naturligt tal som efterliknar Siris prosodi

- Den andra var buggig. Första och tredje hade båda fungerat men kommer ihåg 3e bäst. (första: naturligt tal, andra: Siris tal, tredje: naturligt tal som efterliknar Siris prosodi)
- Inte lika torr som 3:an. Inte lika tramsig som 1:an. (1:an: naturligt tal, 3:an: Siris tal)
- Den var enkel i språket, och mer intressant än tvåan som kändes lite "fyrkantig". (tvåan: naturligt tal)
- Bra svar: Lätta att förstå och hinna skriva ned. 3. Gick för fort men rätt svar egentligen. (3: naturligt tal)
- Bra röst och mest seriös.
- Mest informativ och hade ändå lite "åsikter".
- Pratade mest normalt, svarade bra. Kanske lite mer humor? Önskas.
- För att den andra var lite oklar och den tredje stammade. (andra: naturligt tal, tredje: Siris tal)
- Lättast att höra svaren. Bäst ljud och tydligt uttal.
- Tydliga svar. Lätt att höra.
- Roligt lampskämt.

Motiveringar från de respondenter som valde naturligt tal

- Mänskligt tal, helt ok humor (föredrog humorn hos nr1 framför nr2). (nr1: naturligt tal, nr2: naturligt tal som efterliknar Siris prosodi)
- Snabba, enkla, tydliga svar.
- Snabb kul tydlig bra svar.
- Lätt att förstå, betoning i talet, och hade personlighet. Följsamt naturligt tal.
- Mest neutral + lite charmig. Dom två andra var antingen creepy eller …
- Dens betoning & intonation var lättast att förstå. Svaren lät naturligast.
- Tror den kändes mest verklig/mänsklig.
- Den hade en bra röst och lagom roliga svar.
- Den första var lite för mekanisk och den tredje var väldigt luddig i sina svar. (första: Siris tal, tredje: naturligt tal som efterliknar Siris prosodi)
- Sympatisk stämma.
- Tycker mest om den rösten.
- Mest "mänsklig".