Onboarding Users to a Voice User Interface

Comparing Different Teaching Methods for Onboarding New Users to Intelligent Personal Assistants

Filip Eriksson

June 24, 2018
Master's Thesis in Interaction Technology and Design, 30 credits
Supervisor at Umeå University: Kalle Prorok
Supervisor at Bontouch: Isabella Gross

Examiner: Thomas Mejtoft

Umeå University
Department of Applied Physics and Electronics
SE-901 87 UMEÅ
SWEDEN

Abstract

From being a fictional element in sci-fi movies, voice user interaction has become reality with intelligent personal assistants like Apple's *Siri* and Google's *Assistant*. The development opens up for new exciting user experiences and challenges when designing for these experiences. This thesis has aimed to investigate the user experience of different ways of onboarding new users to intelligent personal assistants. The process has included interviews with experienced users, a test of a Google Home for three months and a wizard of oz (WOZ) test. The interviews and the long term test was done in correlation with a literature study to determine how users interact with an intelligent personal assistant (IPA) their flaws, benefits, what added value they have etc. The goal of the WOZ test was to compare two different teaching methods during the onboarding of a new user. The methods were a voice tutorial by the IPA and a visual interaction on a mobile device. The outcome was to see if the users memory retention was different between the two methods for features learned during the test as well as the users opinions of the two different methods. The results from the interviews show that the benefits of using an IPA is in situations where it reduces friction, e.g when both hands are occupied. They also showed that there are still issues with IPAs and there is a long way to go before they can a accomplish a more human-to-human like conversation. In the WOZ test the results showed that there were no significant difference in user remembrance of learnt features between the two teaching methods. However the user insights showed that the majority of users would like to have a multimodal interaction, a combination of voice and visual interaction when being taught to use an IPA.

Användaronboarding för ett röststyrt användargränssnitt

Sammanfattning

Från att enbart ha varit ett element i sci-fi filmer har röststyrd interaktion blivit verklighet med intelligenta personliga assistenter som Apples *Siri* och Googles *Assistant*. Utvecklingen öppnar upp för nya spännande användarupplevelser och utmaningar vid design för dessa upplevelser. Denna avhandling har som mål att undersöka hur användarupplevelsen skiljer sig åt mellan olika användaronboarding metoder med intelligenta personliga assistenter. Processen har inkluderat intervjuer med expertanvändare, ett längre test av en Google Home och ett Wizard of OZ-test, där de två första gjordes i samband med litteraturstudien för att komma fram till hur användare interagerar med en intelligent personlig assistent dess fördelar, brister, vad för värde de medför m.m. Målet med Wizard of OZ testet var att jämföra två olika metoder att lära nya användare funktioner under första onboardingfasen. Metoderna som testades var en rösttutorial som gjordes av assistenten och en visuell interaktion via en mobil enhet. Målet var att se hur många funktioner som användarna kom ihåg efter att ha lärt sig på de två olika sätten och sen jämföra dessa mot varandra. Testet bidrog också till användarnas åsikter om de olika metoderna. Resultatet visade att fördelarena med att använda en intelligent personlig assistent är i situationer den tar bort friktion och gör saker lättare, e.g när båda händerna är upptagna och en timer behöver sättas. Slutsatsen från intervjuerna var också att det finns mycket brister med assistenterna fortfarande och det är en bit kvar innan de kommer klara av en felfri konversation likt en mellan två människor. I Wizard of Oz testet visade resultaten att det fanns ingen signifikant skillnad i hågkomst av funktionerna mellan de olika metoderna. Däremot skapades en bild av vad användarna ville ha som metod för att lära sig. De hade fördragit en multimodal interaktion, dvs en kombination av röst och visuell interaktion för att initialt lära sig att använda en intelligent personlig assistent.

Acknowledgements

The author would first of all thank Isabella Gross for being a great supervisor at Bontouch. A big thanks to the people at Bontouch for the support and encouragemen and a huge thanks to the peer-review group, *Jane Bertheim*, *Arvid Långström* and *Sandra Waldenström* for giving great feedback, good discussions and pushing each other forward. Thank you to my supervisor at Umeå University, Kalle Prorok, for constructive feedback as well.

Last but not least a thank you to *Oscar*, *Emma*, *Sandra*, *Patrik* and *Matilda* for the support and company at the office during these months.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In 2013 the movie Her was released [4]. The movie takes place in a near future, and is about a lonely writer that develops a relationship with an operating system that is designed to meet his every need. More than 30 years earlier [5] the first Star Wars movie was released, in which a golden protocol robot named C-3PO appears, designed to communicate in a conversational way with organics [5]. Due to the new advances in machine learning and artificial intelligence (AI) voice control is no longer a fictional element, some of these elements have become reality [6]. Voice control have become reality through intelligent personal assistants (IPA) like Apple's Siri[1], Microsoft's Cortana[2], Amazon's Alexa[3], and Google's Assistant[4] making their way into peoples everyday lives. And although they are not nearly as good as the operating system played by Scarlet Johansson in Her, 10 million users (25%) of Microsoft's IPA Xiaoice in China have told the IPA that they love it [7].

On the annual CES (Consumer Electronics Show) in 2017, the year was deemed to be the year of voice recognition [8]. The prognosis was that IPAs would double from around 5 million to 10 million active assistants. Adobe [9] reports that in the last quarter of 2017, Google Home and Amazon Alexa sales doubled compared to the same period in 2016. The total sales increased with 103% from the year before. The increasing development opens up for new exciting user experiences and challenges when designing for these experiences. Just as there are differences when designing a user interface for a desktop app and a mobile app, a voice controlled interface needs its own approach and set of design rules, in the design process depending on the designed medium.

An IPA is an AI agent that is designed to support humans in their everyday

[1] Read more - https://www.apple.com/ios/siri/
[2] https://www.microsoft.com/en-us/windows/cortana
[3] https://developer.amazon.com/alexa
[4] https://assistant.google.com/#?modal_active=none

lives, doing banal, boring or time-consuming tasks to increase human productivity [10]. IPAs are mostly designed to learn from the user over time and become more personalised e.g Google Assistant knows which user is talking to it and will know what playlist to put on when asked to "play my dance playlist"[5]. This thesis will use the term IPA, however there are other terms used as well, e.g virtual personal assistants, digital personal assistants, mobile assistants, or voice assistants when using voice when interacting [11].

As voice control grows and improves, it will be a more natural part of day to day life for people. With it comes the need for great user experience design and a need to understand the challenges of voice controlled interfaces. One of these challenges is user onboarding and how to create a great experience from set-up to teaching users how to use the assistant and become confident users. A few years back Microsoft got to experience the importance of user onboarding. They asked their users what features they were missing in Microsoft Office[6]. The result showed that 90% of the requested features was already available, hence they had done a terrible job at teaching their users how to use their product [12, p. 106].

1.1 Problem Statement

One of the main differences with voice user interface (VUI) design from visual design is the lack of visual cues of how to use the product. In a graphical interface, possible features and actions are shown on the screen. With a VUI the user have to remember what can be done or use trial and error and ask the assistant [3]. Picture a smartphone with a lot of apps installed but with no screen to show them. By leaving the user to explore on their own and use trail and error, which can be risky since the user needs to be motivated to learn [13, 14]. This is what onboarding can be used for, to teach the user the value of the product by showing them what is possible to do [15].

Onboarding of new users are currently being done by Apple, Google and Amazon in two different ways [16–18]. Google and Amazon are using a similar approach by introducing their IPAs to the user in their companion app, rather than letting the assistant introduce itself [17, 18]. Apple on the other hand are doing the initial setup with an iOS[7] device and then do a combination of visual and voice interaction. By letting Siri introduce itself and present some features by voice and simultaneously showing it visually on an iOS device. Siri also encourages the user to perform one task and then guides the user to complete it. When completed the users are free to explore on their own [16].

[5] Command your audio — Google Home; https://www.youtube.com/watch?v=HN–45sGtw4

[6] https://www.office.com/

[7] https://www.apple.com/se/ios/ios-11/

By looking at how these three companies have designed their user onboarding there does not seem to be one best practise way to do it. The focus on this thesis will be put on evaluating the process of onboarding a new user to an IPA, and how different onboarding approaches affects the user experience.

1.2 Objective & Goals

This study aims to investigate the user experience of different ways of onboarding a new user to an intelligent personal assistant (IPA) on a device that only supports voice as a mean for interaction. The part of the onboarding that will be investigated is teaching first time users to use an IPA. To achieve this, research is going to be conducted to determine the limitations and possibilities of voice interaction. This includes a literature study, interviews and two different tests. The literature study and interviews will be carried out to get a deeper understanding of VUIs, IPAs and user onboarding. A long-term test of a Google Home will be carried out as a complement to the interviews. And finally a test will be performed on ten different people to compare users memory retention of taught features and get users insights of two different teaching methods, through voice interaction and by an external graphical interface.

1.3 Outline

The thesis has 7 chapters. **Chapter 1** gives an introduction to the subjects of this thesis as well as the objectives and goals. **Chapter 2** presents background information to voice user interface (VUI) design, intelligent personal assistant (IPA) and user onboarding. **Chapter 3** describes different design principles and guidelines for both VUI design and user onboarding as well as the concept of memory retention. **Chapter 4** describes which methods are used to achieve the objective and goals of this thesis, interviews, a test of an IPA (a Google Home) and a WOZ test. **Chapter 5** shows the results of the thesis and in **Chapter 6**, the results are discussed. **Chapter 7** is the conclusion of the thesis.

Chapter 2

Background

The following chapter consists of three sections: 2.1 Bontouch, 2.2 Voice User Interface (VUI) and 2.3 User Onboarding. The background introduces theory that is valuable to understand the rest of this thesis.

2.1 Bontouch

This thesis was written in collaboration with Bontouch[1] at their office in Stockholm in the Spring of 2018. They are interested in learning more about how users can interact with IPAs and this thesis aims to give more insight for future projects involving IPAs, conversational user interfaces and voice interaction.

Bontouch is a digital design agency, founded in 2008, that partners with brands such as SJ, PostNord, Coca-Cola and SEB to create world-class smartphone, tablet and wearable applications for their consumers. With 90+ designers and engineers on staff, Bontouch serves clients in Europe, North America and China from three studios in Stockholm, London and New York. Today, the products that Bontouch helps creating are used by 50 million people in 196 countries.

Bontouch won the Swedish Design Awards in 2017 as "Designer of the year", as well as first (PostNord app) and second place (SJ app) in the category "Digital - Smartphone". Bontouch has developed a Siri integration within the Taxi Stockholm mobile app so their customer can just say "book a cab" and a taxi will arrive.

[1] https://www.bontouch.com/

2.2 Voice User Interface (VUI)

A voice user interface (VUI) is what a user interacts with when communicating with a spoken language application [19]. The ability to understand what the users say and then take action based upon it is a combination of two important technologies: automated speech recognition (ASR) and natural-language understanding (NLU) [3]. ASR can be viewed as writing down the phonetics to a language you do not understand, and have no idea what it actually means. And NLU, in neuropsychology, linguistics and the philosophy of language, is any language that has naturally evolved in humans by repetition and use without actively tending to it. Natural languages can be expressed in different ways, e.g through speech or signing (sign language). They differ from designed and formal languages such as those used for programming or to study logic [20].

Having a conversation is a lot more than the exchange of information. In conversation, participants usually share natural assumptions about a topic of how a conversation should develop. The participants have expectations about the quality and quantity of the contributions that each person should make. On top of that, there are a set of natural rules of conversation that include politeness, consistency, etc. Plus, people instinctively know to disregard superficial meanings if they're unclear or abstract and search for deeper, non-literal interpretations. While we all do it naturally, conversation is a rather complicated process [21].

Linguistics philosopher Paul Grice [22] said that to be understood, people need to speak cooperatively. He called this the Cooperative Principle. With this principle he introduces Grice's conversational maxims which consists of four basic rules to have a cooperative conversation [3, 21]:

- **Quality** - Say what you believe to be true.
- **Quantity** - Say as much information as is needed, but not more.
- **Relevance** - Talk about what is relevant to the conversation at hand.
- **Manner** - Try to be clear and explain in a way that makes sense to others.

In conversations where these rules are not followed it is common to feel confused or frustrated. VUIs that do not follow these maxims will cause similar issues. Pearl [3] gives some examples of ways VUIs might effect the user's experience in a negative way (see Table 2.1)

2.2.1 Conversational User Interface

Before discussing conversational user interface (CUI) giving a definition to what a conversation is feels appropriate. The Oxford English Dictionary[2] defines a

[2]https://en.oxforddictionaries.com/definition/conversation

Table (2.1) Example of breaking Grince's maxims [3]

Quality	Advertising things you can not live up to, such as saying, "How can I help you?" when really all the VUI can do is take hotel reservations.
Quantity	Extra verbiage, such as "Please listen carefully, as our options may have changed." (Who ever thought, "Oh, good! Thanks for letting me know")
Relevance	Giving instructions for things that are not currently useful, such as explaining a return policy before someone has even placed an order.
Manner	Using technical jargon that confuses the user.

conversation *as a talk, especially an informal one, between two or more people, in which news and ideas are exchanged.* This definition suggests that initiative belongs to both sides of the conversation, Radlinski and Craswell [23] calls this mixed initiative.

According to Radlinski and Craswell [23] a conversational system is a system for retrieving information which permits a mixed-initiative back and forth between user and agent, where the agent's actions are performed in response to current user needs based on the conversation, using both short- and long-term knowledge of the user. Furthermore they discuss that the system should at least include the following properties:

- **User Revealment** - The system helps the user express, and potentially discover, their true needs.
- **System Revealment** - The system reveals its capabilities, building the user's expectations of what it can and cannot do.
- **Mixed Initiative** - Both the user and the system can take initiative conversation.
- **Memory** - The user can reference to past conversations and statements, which remain true unless told otherwise.
- **Set Retrieval** - The system can reason about the utility of sets of complementary items.

2.2.2 Intelligent Personal Assistants (IPAs)

IPAs are one step closer to realization of the science fiction dream of interacting with systems and devices by talking to them. Apple's Siri, Microsoft's Cortana, Amazon's Alexa, and Google's Assistant are all systems that run on different devices e.g purpose built speakers and smartphones. When activated

(by keyword or physical interaction) it records the user's voice and sends it to a receiving server, that processes it as a command. Depending on the input, the assistant will receive the appropriate information to read back to the user, play requested media, or complete tasks with different connected devices and services [24].

An IPA is defined by its ability to handle natural language as input and output, as well as the ability to process the input to answer questions and perform actions given to the system [25]. The systems are built with several interconnected subsystems. Two of these subsystems are speech recognition and natural-language processing (NLP) that translate what the user says into requests that the computer can understand. The requests are then processed and executed [11].

IPAs have, in recent years, started to surface on the market thanks to progress in artificial intelligence and deep learning. Neural networks can now process vast amounts of data which make computers a lot better at NLP, speech recognition and answering questions [11]. Combined with faster, cheaper, more accessible computer power and high amounts of available data have made it possible to use deep learning to train AIs that are good enough to be useful in consumer products. The possibility to build IPA products have resulted in that many of the leading technology companies have followed the trend and released IPAs [11].

In the book *Designing Voice User Interfaces*, Pearl [3] argues that current IPAs, on the market today (begining of 2018), that rely on voice when interacting, do not fully use a conversational interaction but rather a command-and-control voice user interface. It is more an exchange of "one-offs" than an actual conversation. According to Pearl, IPAs are partly lacking the ability to remember the past. There are two ways in which the past is key to a conversation. There's the past from previous conversations, e.g what you ordered yesterday, most requested song and most called Sara in your phone book. The other past is what have been said within the same conversation - if not in the last turn. Engaging in conversation with a human being could involve some key components: contextual awareness (pays attention to you and the environment), remembering previous interactions and conversations, and an exchange of appropriate questions. All of these contribute to a feeling of common ground. If VUIs do not include context and memory, the usability will be limited and might not live up to their possible potential.

2.2.3 Discoverability

Norman [26] describes the term discoverability of a product as: It is possible to determine what actions are possible, where and how to perform them as well as the current state of the device. If a product has high discoverability it is easy for the user to know which actions are possible and the current state

of the product. One of Norman's more famous examples of an object that lack discoverability has become known as the *Norman Door*. The concept was introduced by Norman in 1988 and is a door where people struggle to figure out if they should push or pull to open it. Norman's explanations that it is because the door lacks affordances, i.e. affordances exist to make the desired actions possible. To get even higher discoverability, signifiers can also be used. A signifier gives cues to possible actions [26].

One of the bigger challenges for voice only interaction is that users are unaware of available features [27], and Corbett and Weber [28] concludes that it is very challenging to design a VUI with good and high discoverability. Corbett and Weber [28] concluded in 2016 that there have always been issues with discoverability and learnability when interacting with a voice interface. Discoverability of voice commands is tricky because it is harder to give hints and cues compared to a graphical interface. They also conclude that users will often have inconsistent and inaccurate mental models of what the system can and can not do.

Interacting with a purely voice activated system comes with several issues. One reason is that language is our main way of communicating as humans. With the consequence that expectations of a voice interfaces become very high, and these expectations have not been reached yet [29]. Speech is also asymmetric, meaning that speaking is faster than typing, but listening is slower than reading and VUIs are also less scannable than a graphical user interface (GUI) [29]. An other big issue is that the user can not see the interface and which options/action that are available, after the system outputs through speech it is gone. This puts pressure on the user to remember what to say next as well as where in the conversation they are, which can be tricky since humans short term memory is limited. Therefore it is not a good idea to to give users long instructions to what is available and expect the users to remember everything. All of these issues are limitations that should be thought of when designing VUIs.

2.2.4 Multimodal Interaction

Multimodal interaction is part of humans everyday life, we gesture, speak, move, shift gaze, touch (senses or modalities) etc. as part of an effective flow of communication [30]. A multimodal interaction system aim to support these modalitites and use the recognition of naturally happening forms of human language and behaviour through the use of recognition-based technology. There are several benefits with multimodal systems and interaction with them. One of the key benefits of multimodal interaction is the support of a more natural way for humans to interact with each other and their surroundings. Using multiple modalities makes it more flexible for the user to pick the best way to interact with a system, depending e.g on the situation, and could meet the needs for a more diverse range of users [30].

The IPA Google Assistant that runs on a phone supports multimodal interaction [3, 31]. Interaction with the assistant can be done by speaking to the assistant and have it respond using voice as well. The phone will present the entire conversation on the screen of the phone. There is also the possibility to type questions to the assistant instead of speaking, which can be useful in certain situations. The support of different combinations of interaction methods therefore enables the Google Assistant on phones to become a multimodal interaction.

2.3 User Onboarding

The term onboarding, also called organisational socialisation, is originally used in the organisational world and is defined as *"the process that helps new employees learn the knowledge, skills, and behaviours they need to succeed in their new organisations"* by Bauer and Erdogan [32]. It has been adapted as a term in software development for bringing new users onboard, from the one-time setup, conveying the value of the product as well as teaching users to use it [15]. The UX designer, and user onboarding Guru [12] Samuel Hulick, says that *"User Onboarding is the process of increasing the likelihood that new users become successful when adopting your product"* [1]. An important note when it comes to user onboarding is that there are no real best practises that can be applied in every situation. There are guidelines and design principles but every situation and product need to be evaluated ceparatly [12].

Chapter 3

Theory

The theory consists of three main sections: 3.1 Key Design concepts of Voice User Interfaces, 3.2 User Onboarding and 3.3 Memory Retention & Ebbinghaus Forgetting Curve.

3.1 Key Design Concepts of Voice User Interfaces

To avoid some of the issues presented in the Background (chapter 2) with e.g discoverability. There are some design guidelines and concepts to follow. This section will present a few of these concepts that are relevant to this study.

3.1.1 Command-and-Control or Conversational

Currently the most common way of interacting with a VUI is by *command and control* [3] , which means that the user must explicitly indicate when they want to speak. An other type of design, which is becoming more common is conversational, that uses a more natural turn-taking approach. Most IPAs today can be classified as *command-and-control*, the indication to speak can be a wake word (e.g "Hey Google", or "Alexa") it can also be a button or a tap on the device. When this event occurs the typical response from the VUI can be a nonverbal audio, and/or visual feedback (a wavy line, animated dots, a glowing light on the device). This indicates to the user that the system is listening and it can be spoken to. When the system has decided that the user is done speaking, it indicates this in some way, e.g nonverbal audio or flashing lights in a different way, and then responds. At this point the conversation can be over. The system

is not expecting the user to speak again and the user must thereforee indicate that they want to speak again [3].

A system where the user is more involved in a closed conversation where there is a beginning and an end, these systems are more conversational, e.g a chatbot. Applying the mandatory indication to speak in a longer back-and-forth conversation between the user and VUI is not necessary and can be cumbersome and unnatural. Pearl [3] gives the example of when being in a conversation with a real person there is no need to give indicators at every turn.

At times it makes sense to switch between *command-and-control* and *conversational* modes. However it is important to only do this in situations where the users will understand that the mode has changed. An example of this is the game *Jeopardy!* game on the Amazon Echo. To enable the game the user have to say "Alexa, start *Jeopardy!*" [3]. In the game the user do not need to say "Alexa" when saying something to the assistant. This is known as a conversational structure, and users have no trouble figuring it out [3].

3.1.2 Keep It Simple

One of the goals when designing visual experiences are to try to limit the number of clicks a user must take to complete an action [3]. With more clicks the experience feels less effective and often dull and boring. The same principle can be applied to voice interaction. E.g asking for the users complete address and having them go through four call and responses before completing a single task. The same task could be completed in one single interaction.

In a visual experience users can quickly move their attention to the information that is most relevant to them, skipping parts of the interface that is uninteresting for them [3]. A VUI, on the other hand, are linear. It is not possible to scan for the most valuable information, nor is skipping around the interface. The user is forced to listen to everything the application has to say. Since this is the case, keep it short. Present only the most important information and with the most important parts first. Given the example of a shopping app, and the user listening to search results for a given product. Things to present can be title of the product, the price and rating. Everything else can be presented later and letting the user ask more specifically what the want to know. Compared to a graphical interface where a lot more information can be presented, even on a small screen [3].

3.1.3 Context

To explain how Google handles context Pearl [3] uses the following example:

USER
OK Google. Who was the 16th president of the United States?

GOOGLE
Abraham Lincoln was the 16th president of the United States.

USER
How old was he when he died?

GOOGLE
Abraham Lincoln died at the age of 56.

USER
Where was he born?

GOOGLE
Hogenville, KY.

USER
What is the best restaurant there?

GOOGLE
Here is Paula's Hot Biscuit.

In the example Google manages to continue the conversation, and remember the context making it feel more like a conversation. The alternative being that the user have to say the name of the president when asking questions about him. The use of pronouns does the trick in this example. Google knew that "he" refered to Abraham Lincoln and that "there" meant Hogenville, Kentucky. A system that can not handle this type of information will not be able to do anything besides one-turn actions. The usage of two different terms to refer to the same thing is called *coreference*, and is a vital part of communication. And without it, conversations will quickly break down. [3]

3.1.4 Providing Context

One of the biggest challenges when it comes to VUI design is teaching users what they can do. Compared to graphical interfaces, where this is less of a problem since everything is there on the screen. Interactable objects are visually available and the user can see available buttons and menus. With voice interfaces, the visual discovery of features is not available. thereforee the design of the VUI should be done is such a way that it informs the users how to interact with the system, what actions is available and how to respond. [3]

Pearl [3] suggest prompting the user, and help to inform the user how they should respond, or what specific actions they can take. One example is "Here are four recipes for ratatouille. You can ask me for more information on any of them" [3]. However a user might still forget what they were doing. And a good VUI design can handle the user getting lost by allowing the user to ask for help

at any time. This by reorient them by providing their current context (given there is one) and give hints in the right direction. [3]

3.1.5 Disambiguation

There are situations when the user provides some but not all of the necessary information to perform a certain action [3]. An example being that a user might ask for the weather for a location that exists in more than one place: "What is the weather in London?". By using contextual information it is possible for the system to make a qualified guess from the context, in this case the location. For instance if the user has registered its home address in London, United Kingdom (UK) (or anywhere in the United Kingdom) it is fair to assume that this is the London the user is referring to and not the one in Canada. It is also possible to use other contextual cues. If the user has looked up a restaurant in London, UK and then asks for the weather, it is a fair assumption to make that it is the same one [3].

If no contextual information is available the system needs to ask for the missing information [3]. By simply asking "Do you mean the one in the UK or the one in Canada?" In some situations it might be needed to ask the user to narrow down their answers, since current IPAs have constraints that make them unable to handle to much information.

3.1.6 Confirmations

An important part of VUI design is to make sure that the user feel understood [3]. The purpose of this is also to let the user know when they were not understood. However over-confirmation can be bad for the user experience and create a very unnatural interaction. It might ensure greater accuracy, but it will drive the user crazy. When designing confirmations the situation is to be taken into account. What would the consequence be if something went wrong? When transferring money confirmation becomes really important. But with a media service like playing music, the stake is much lower and one mistake does not come with high consequences.

When a user asks a Google Home a question, lights on the top of it lights up to confirm that it is listening [3]. In reality it is always listening but the device does not light up until it knows that the user wants to interact with it. When a question have been asked the device indicates with the lights in a different pattern that it is processing. This to confirm to the user that something is happening.

According to Pearl [3] there many different ways to confirm information to the user. The methods are explained below:

- **Three-Tired Confidence** - The system will explicit confirm information between a certain threshold (e.g 45-80% accuracy), reject anything with a lower confidence and implicitly confirm anything above 80%.
- **Implicit Confirmation** - This method confirm things implicitly, without rejecting the user to take action. An example is when asked "What is the world's tallest mountain" the systems response could be "The world's tallest mountain is Mount Everest".
- **Nonspeech Confirmation** - In this case a spoken response is not performed when an action is completed. For example, an application that turns the lights on an off, the confirmation becomes the lights turning on and off. A spoken confirmation is not needed in this situation.
- **Generic Confirmation** - In a more conversational system other types of confirmations might be needed. This can be useful in situations were the user is doing more open ended chatting. For instance if a system is asking how a user is feeling. And the user is asked how they slept the night before, the user might say poorly. The response from the system can confirm this my saying "I'm sorry to hear that".
- **Visual Confirmation** - In multimodal interfaces that uses both a screen and voice to interact, e.g a mobile device. It is common to include visual confirmation as well as voice. When asking Google Assistant a question the reply will be provide both an audio and visual confirmation.

3.1.7 Novice & Expert Users

It is important to keep in mind if users are going the use the VUI system often, since it should include different strategies in the design. Pearl [3] gives an example of an healthcare app where the user logs in and takes their blood pressure. From the start the prompts include more instructional details so the user understands what and how to take their blood pressure. After a while and when the user has become familiar with the app there is no need to have these instructions. A better approach is to give enough information so the user still understands what to do but with less instruction [3].

Pearl continues to say that if the user returns and uses the app only one or two times a month, these instructions might be needed always or at least for a longer period of time than for someone that uses it a few times a week [3]. thereforee only counting the times the user have used the app is not the best approach. It is more better to see to the need of the user and how each individual user use the application, rather than just a specific time they have used it.

The concept of *priming* can also be used in these situations. Priming means that when someone is exposed to a particular stimulus (e.g a word or an image) it will influence their response to a later stimulus [3]. For instance if watching a nature

documentary about bears, and then being asked to name an animal that starts with the letter "B", the answer will more likely be bear than beaver.

3.1.8 Error Handling

In an interactive voice response (IVR) system, if a user was not heard or understood, the system prompts the user to repeat themselves. An IVR system are for instance a host system used to book an appointment at a doctors office [3]. Prompting the user to repeat themselves is important because by doing so the user knows that the call was not cut off or the system is not functioning. Since the system uses a fixed script it is also required to make sure that the user provides input so that the conversation can move forward. And therefore when the system times out it asks the user to speak again. If these timeouts are not planned carefully, the user and the system can end up cutting each other off creating a rather awkward interaction [3].

When designing for any VUIs device it is not always mandatory to reprompt the user when there is an error. For instance The Amazon Echo does nothing if the user stays quiet after the wake word [3]. According to Pearl [3] when a users speaks to a device (especially with a name), users are prone to respond to silence in the same way that they would do if they were talking to a human, by repeating themselves. An IPA does not wait for the next input from the user like a IVR, simply because it is often a one-off command. And there is no need since it is only that one transaction that fails and not an entire conversation [3].

It is important to remember that although that speech recognition has improved immensely over the past 10 years (more than 90% accuracy given the right conditions). However this does not guarantee a smooth interaction for users when interacting with a VUI. A great design should always be able to handle when things go wrong, which it will [3].

There are a few different ways a VUIs can make mistakes, which need to be designed for and handled, these are:

- No speech detected
- Speech detected, but nothing recognised
- Something was recognised correctly, but the system uses it in the wrong way
- Something was recognised incorrectly

Escalating error is a strategy for cases when speech is expected [3]. These can be used if for instance a user does not provide all information needed or is unclear in the response. A VUI system can ask again and even add information to make it clearer what the user should include in their response. E.g If wanting to know the weather, the input needed is city and state, and the user provides

just one. The system can apologise for not getting it and ask for the city and state again [3].

3.2 User Onboarding

This section covers important design guidelines and concepts of user onboarding.

The One-time Setup

The one-time setup in onboarding is often when the user sign up for a new product or service [15]. In this phase there are a few things to think about. The first is to remove friction, one way can be to use social login for sign up, e.g Google or Facebook, to make it faster and easier for users to sign up [12]. Other examples of friction is receiving an e-mail requesting that the user verifies their e-mail address. If it is not a 100% necessary it is better to let the user finish the sign up and then they can leave the app to verify their email.

It is also important to only ask information about the user that is actually needed. Asking for more only adds more friction to the sign up which decreases the user experience. One way to do it is to use Facebook or Google log-ins where most users already have a log-in. However there should always be an option to do a normal sign up as well. One other big reason is that storing all that unnecessary user data costs storage space and does not really fill a function [15].

An other great way to create a better user onboarding experience is to personalise it. By adding a simple "Hello" or if the users name is available "Hello Anna" creates a more personal touch to the product and adds to a better onboarding experience [15].

Convey the Value

The next part of creating a great onboarding experience is to really convey the value of the product. With this approach it is very important to remember that people do not use software because they have a lot of spare time [15]. They use software to significantly improve their lives in some way.

As Hulick [1] puts it, it is helpful to think of onboarding in terms of how the product makes its users successful, instead of in terms of activating features. By getting the users engaged and contributing to making them better people. Hulick [1] continues with an example, a note-taking app like Evernote[1] can say "we make people better at writing, storing, and searching for notes," however

[1] https://evernote.com/

this is really just describing the use of the product. A better way might be to say "we make people better at remembering things" which is closer to their users' main aspiration.

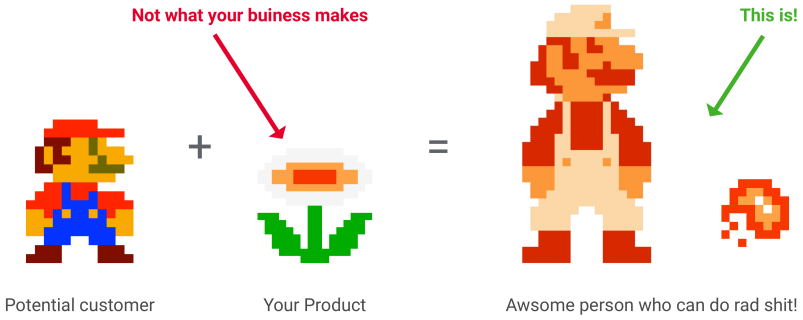

Figure (3.1) Figure based on one by Samuel Hulick[1], explaining the difference between just promoting a product and something that changes the users life.

To explain this way of thinking even further Hulick [1] uses Mario, and especially the game Super Mario Bros[2]. In the game Mario is small at first, but there are these fire flowers that both make him big and is given the ability to shoot fireballs. What is it that makes the user excited? It is not the green stem or because you could walk over it to get it, the excitement is the possibility to hurl fireballs. The product (the fire flower) and its characteristics is not what causes excitement. That comes with knowing you can throw fireballs by picking it. This is illustrated in figure 3.1.

Teach Users to Use

Another part of onboarding is to teach the users how to use the product in an effective way. Depending on the product or system this can be done in various ways that are more or less effective. An overall go to is like Rolander [15] puts it "Use the app to explaining the app", meaning that the software shows the user how to use it, a learn by doing approach. An example of an application that does this well is Slack[3]. They have developed a chatbot named "Slackbot" that the user can ask questions of how to use Slack in a conversational tone which gives a personal touch to the experience. It also gives examples of what the user can ask, if users feel unsure what to ask. An other way is to use content as a guide for new users. In Trello[4] users are given a welcome board with pre-made

[2]http://www.ign.com/articles/2010/09/14/ign-presents-the-history-of-super-mario-bros?page=2
[3]https://slack.com/
[4]https://trello.com/

to-do cards (see figure 3.2), where each of the cards explains a different function or interaction. They use the interface and content to teach the user how to use the product [12].

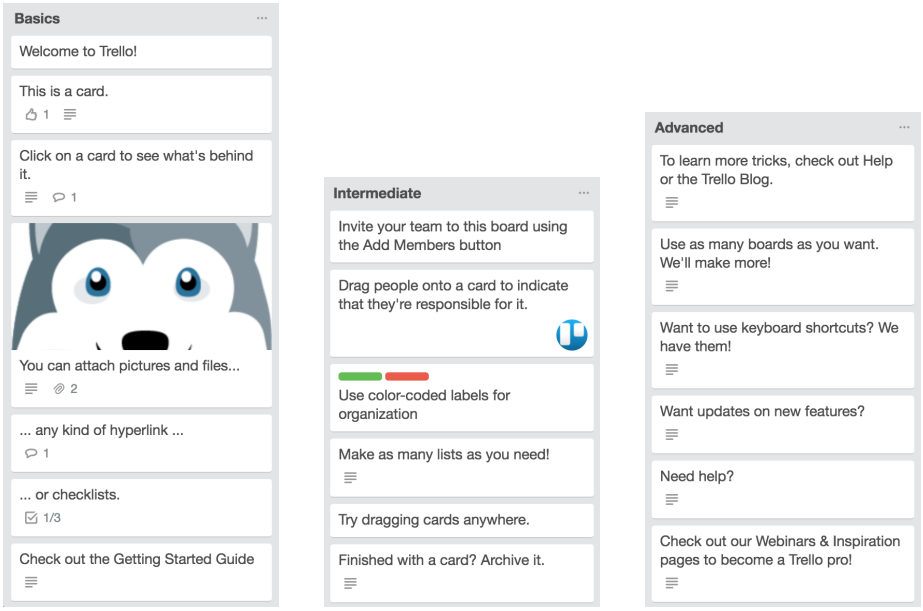

Figure (3.2) Trello uses a welcome board to teach users how to use their product

They are both providing a warm and human experience of their products, it encourages action. It is according to Hulick [1] a far better way than saying "Your campaign is empty and that is all I am going to say". Saying that to a user in person is something no one would do, so why do it in a product?

Progressive growing [15] is also an effective way of introducing a new user to a system. By not showing everything that is available at once, which can be overwhelming for the user, focus the users attention to the most important features of the application [15]. This can be done by fading out certain parts of the application or colouring e.g an icon grey, which gives the impression that it is not interactable. What is important is that the user still feels in control, there is a difference between hijacking and forcing the user to do something, and making them believe that it was their own decision [15]. When it is actually the design that points them in the right direction.

Duolingo[5] uses progressive growing in their application. After the user have picked a language to learn, they are presented with two options, see far left in figure 3.3. They ask if the user is a beginner or if they already know some of the language and want to take a test to be able to jump ahead. These are really

[5]https://www.duolingo.com/

the only two options available, except the bottom menu. Pressing the "beginner button" takes the user to the first lesson, and after completing it they are asked to create a profile to be able to save their progress. When completed the user is returned to the home page, see middle image figure 3.3. Duolingo could have skipped the onboarding image (the left one) and just used the standard page. However by using onboarding Duolingo guides their users to what they which the users first action should be. Instead of them clicking around the app until they actually start using the app for what it is intended for, hence show them the value of the app. By moving sign up to after showing the value of the app they have also successfully removed friction for a first time user. When the user is done with their first lesson they will gladly sign up if they like the experience and want to keep using it [15].

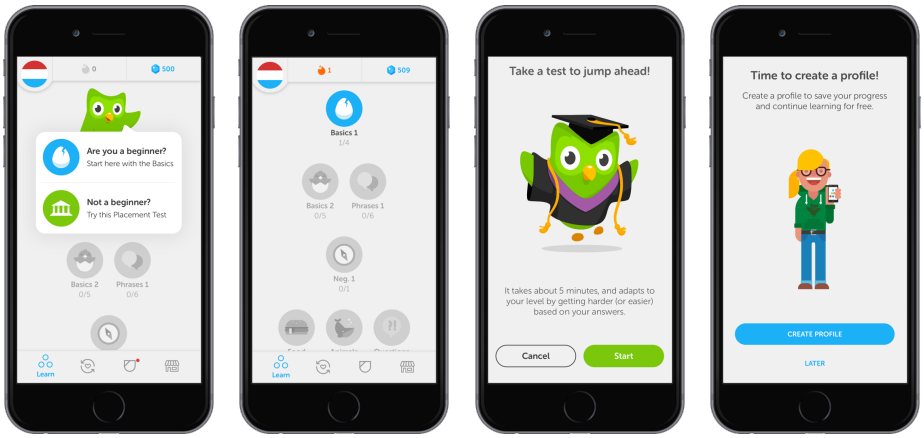

Figure (3.3) Examples from the Duolingo app

An other important part of user onboarding is timing[15]. When to show something to the user, send out an email or a notification. Rolander [15] gave an example involving Starbucks[6]. In the Starbucks app users are able to pay for their coffee. Starbucks had tried to use pop ups with information and promos in their application, but according to their data most users just closed them immediately. The reason could be that they were sent out around the time that most users was standing in line to get coffee, and was about to pay with the app. In that situation everything that is in the way for doing just that is discarded and seen as less important. A better approach would be to find a time where users have time to browse the application and are open to view information in the app [15].

[6]http://www.starbucks.se/

Context

Context and giving the user certain information at the right moment is a good way to do onboarding [15]. By showing the user tips when they are using the product, "Are you trying to do this" and showing the user where to do it. It is also a great way to teach users new features in a product, either a brand new one or one the user have not found yet. This can be done by using subtle tool tips in the interface when the user actually needs to execute the action [15]. Google does this in a very good way in Google Photos[7]. E.g They tell the user to press Shift to mark more than one image at once and another telling the user to press "Shift + D" as a shortcut for downloading.

3.2.1 Current User Onboarding for IPAs

This section will present and compare in which ways Apple, Google and Amazon onboards their new users to each of their individual IPAs.

(a) Amazon Echo [33] (b) Apple HomePod [34] (c) Google Home [35]

Figure (3.4) IPAs from Amazon, Apple and Google.

Amazon Echo (2nd Generation)

In the initial setup of a new Amazon Echo [33] (see Figure 3.4a), she greets the user and refers them to the Alexa app to manage the set up, settings and preferences for the device and the users account [18]. After the set up is complete a video is presented inside the app to give the user a sense of what is possible to do with Alexa, controlling a smart devices or presenting the users day as two examples. When the video is done the user is left to explore on their own [18]. To learn more about how to use Alexa, what can be said and which features that are available, examples are presented in the Alexa app [36].

[7] https://photos.google.com/

Apple HomePod

To set up a new HomePod [34] (see Figure 3.4b) an iOS device is required [37], the device is placed close to the HomePod and the setup is displayed on the device. The user can pick where the speaker is placed, give enable personal request and transfer settings to automatically set up the HomePod (e.g iCloud, home Wi-Fi network, Apple Music and more) [37]. When the HomePod is set up Siri presents herself and gives the user some examples of what can be said to her. This is done both by using voice and on the iOS device simultaneously. She also wants the user to interact with her and try one feature, one example is Music and she asks the user to play music [16]. The comes with an app as well, where the user can control the speaker without speaking to it [37].

Google Home

When setting up a new Google Home device [17, 35], (see Figure 3.4c), the Google Home Assistant greats the user and tell them to download the Google Home app to their smartphone or tablet. The user is then guided through the set up in the app on the phone. When completed the assistant lets the user know that it is ready to use and tells the user to continue in the Google Home app to learn more about how to use the assistant. The app shows four different headings as part of the onboarding, with things that can be said to the assistant. These include music, setting timers and alarms, asking what noise a panda makes and much more. There is a more complete list of features available that can be viewed in the Google Home app at all times [17].

Comparison

Comparing the three (Amazon Echo, Apple HomePod and Google Home) Amazon and Google are doing things fairly similar with Apple sticking out from the two others in their approach. Apple is the only one of them that lets the assistant, in this case Siri, introduce itself and inform the user of some available features as well as introducing and letting the user perform one of them. This is done by simultaneously showing the same information on the iOS device as a compliment [37]. Whereas the other two rely heavily on the companion app for the initial user onboarding, when learning the users how to use the assistant [17, 18].

3.3 Memory Retention & Ebbinghaus Forgetting Curve

Ebbinghaus Forgetting Curve describes the brains ability to retain memory over time [2]. The hypothesis was stated by Hermann Ebbinghaus in 1885. The theory is that humans start to loose memory of learned knowledge over time. To retain knowledge, and not forget in a matter of days or weeks, conscious repetition is needed. An other related concept is strength of memory, which states that the time period to recall a certain memory is based on the strength of the particular memory [2].

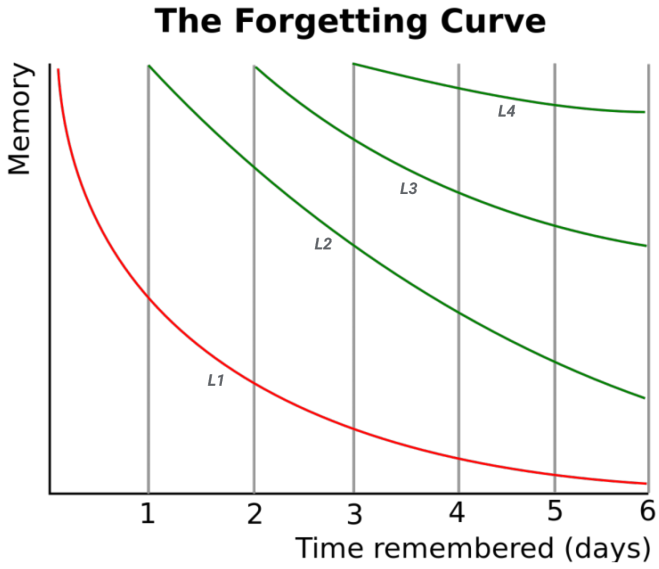

Figure (3.5) Ebbinghaus Forgetting Curve - Memory retention without (L1) and with repetition (L2 - repetition twice, L3 - repetition three times & L4 - repetition four times) [2]

Ebbinghaus [2] did a series of test on himself, which consisted of memorising and forgetting of meaningless three letter words. These were complete nonsense words and more a random combination of letters such as "WID", "ZOF" and "KAF". He then tested if he could retain the information after different periods of time. The results that was obtained is plotted in a graph, refereed to as the Forgetting Curve, see Figure 3.5 [2].

Ebbinghaus drew the conclusion that the forgetting curve is exponential because there is a huge memory loss within the first few days of learning. However the rate of memory loss decreases and the rate of forgetting is much slower. Memory

retention is said to be 100% at the time of learning any piece new information. Within the first few days of learning it drops rapidly down to below 40%, before the declination of memory retention slows down. However the memory loss can be slowed down by repeating the learned information the following days. The more days the information is repeated the slower the decrease in memory retention [2].

There are other factors than repetition that can affect the rate of forgetting. Some of these are:

- Meaningfulness of the information
- The way it is represented
- Physiological factors (stress, sleep, etc)

The rate of forgetting is also individual, humans capacity for remembering varies [2].

Chapter 4

Methods

The methodology is presented in this section and consists of five main sections: 4.1 Literature Study, 4.2 Interviews, 4.3 Testing of a Google Home and 4.4 Wizard of Oz (WOZ) Test. The method was conducted into two phases: *Understanding* and *Exploration*.

Understanding

In the first phase, which includes section 4.1, 4.2 and 4.3, the focus was to understand the research area, current (spring 2018) market situation, user experiences and current usability issues. This was done by conducting a literature study resulting in extensive knowledge of areas touched on in this thesis. Followed by a series of interviews with experienced users of voice controlled IPAs as well as multimodal assistants to understand their usage, needs and challenges using assistants available on the market. To get even more insights a long-term test of a Google Home Assistant [35] was done as well.

Exploration

The goal of the second phase was to use the knowledge and research found during the first phase to explore new areas. The areas that was explored further were comparing two different onboarding methods to new users of an IPA running on a home speaker device. The onboarding methods that were tested are a voice tutorial and a visual interaction. This was done by doing a qualitative test [38, 39] to get a glimpse of the performance of the onboarding methods as well as user insights, some of the data in the test however is considered to be quantitative rather than qualitative [38, 39]. The test was designed using the wizard of oz (WOZ) methodology and interviewing the participants in conjunction with the test. The goal of the test were to see the difference in memory retention between

users exposed to the different onboarding methods as well as getting insights in how users experienced the different interactions and what they preferred. The memory retention part of the test generated quantitative data whereas the interviews is considered to be qualitative [39]. The results are presented in section 5.3 and discussed further in section 6.3.

4.1 Literature Study

To get a better understanding of the fields touched on in this thesis, relevant literature and research was gathered. These varied from books, articles, posts on credible design blogs, podcasts and videos. The reviewed research was found using Umeå University's online library[1] with help of Google Scholar to make the search process easier. Keywords using during the search includes "User Onboarding", "Conversational User Interface", "Voice User Interface", "Multimodal Interface", etc. After gathering, the literature was analysed and sorted according to relevance to this thesis. The included research is used in both the Background (chapter 2) and Theory (chapter 3) of this thesis.

4.2 Interviews

To get a better understanding of IPAs and how they are being used, five qualitative interviews were conducted. Focus was put on IPAs, but also included the use of assistants in other devices such as smartphones and wearables. The participants were between the ages of 25 and 32, three men and two women. They are all experienced users and have used IPAs more than six months and on more than one device. The goal of the interviews was to find out how this group of users use IPAs, which problems they may have experienced and the added value for them. Since early adopters are ahead of the curve when adopting new technology, their experiences and problems give an insight to possible issues the rest of the population might experience when starting to use IPAs. The interviews took place at Bontouch's office in Stockholm and were all conducted in Swedish, the participants native tongue. The interviewees had all been using IPAs in English, with the exception of Siri that is available on iOS in Swedish.

Before the actual interviews some preparations were done. First establish who to interview as well as a set purpose of the interviews, and then an interview guide (see Appendix A) was created [40, p. 85]. The guide was used during the interviews to make sure that everything got answered and that the interview stayed on topic. The content of the guide includes a brief explanation and goal of the study, which is shared with the participant, and the questions that are

[1] For more information, see http://www.ub.umu.se/

the focus of the interview. As part of the preparation a Google Home Assistant was tested at home for a few weeks as well as reading about the subject. This was done according to Hall and Zeldman [40, p. 85] and are an important part of the interview preparation.

The interviews structure was based on a concept called *three boxes, loosely joined* [40, p. 85-89]. It is divided into three parts, the introduction and warm-up, the body of the interview, and the conclusion.

Introduction - The introduction to an interview is used to inform the participant about the interview, brief the subject with all formal information and to make sure that the interviewee feels comfortable taking part in the interview. It is also important to express gratitude towards the participant.

Body - The body is the main part of the interview. The goal of this part is to get all the prepared questions answered and as much as possible of what the participant has to say about the subject. Questions asked should be open, to get as much out if it, avoid closed questions that can easily be answered with Yes/No. As interviewer the key is to only speak when actually needed. Allow pauses, since silence is uncomfortable participants might fill them with valuable information. An other aspect is to look out for vague answers and try to make the participant expand their answer by asking questions like *"Tell me more about that"* and *"Why is that?"* to get clearer and more defined answers.

Conclusion - Ending the interview in a friendly and polite manner is very important. And thanking the participant for taking their time no matter how the interview went. If the interview becomes unproductive it is O.K to end it early even though the interviewee has not answered all the prepared questions.

To minimise the error risk by taking notes and missing valuable information all the interviews (with the participants permission) were recorded and later on transcribed. Which is also proposed by Hall and Zeldman [40].

4.3 Long-term Test of a Google Home

As a complement to the interviews a Google Home Assistant [35] was tested in a home environment for more than two months during the process of writing this thesis. The purpose of the test was to get a deeper understanding of how a current IPA performs, its flaws and benefits, which features that are available and how well they work, as well as in which ways Google uses user onboarding for their device. The assistant was used around a month before the interviews took place, which made it easier to ask valid questions as well as understanding the answers in an easier way. A Chromecast[2] was available during the testing, which makes it possible to control features like Netflix[3] and YouTube[4] using

[2]https://store.google.com/product/chromecast_2015
[3]https://www.netflix.com/
[4]https://www.youtube.com/

voice.

By testing the assistant over a long time the hope was to get insights in how actual users use the device on a daily basis and compare the findings with the results of the interviews to validate the results. Test sessions with the IPA was also performed by asking it various questions to get a deeper understanding of available features, what is possible to say and how well it performs. One simply being, asking the assitant what is possible to say to it. Focus was also put on in which way new users are onboarded, by doing the set up more than once and using the Google Home App, as instructed by Google as part of their user onboarding.

4.4 Wizard of Oz (WOZ) Test

The Wizard of Oz is a musical fantasy adventure film that was produced in 1939 [41]. The film is based on the novel "The Wonderful Wizard of Oz" by L. Frank Baum from 1900. The story can be summarised by the adventures of a little girl by the name of Dorothy in the colourful and magical Land of Oz. An important character in the plot is the wizard, who is an ordinary man who hides behind a curtain and acts as a powerful wizard. This deception is what forms the basis for the evaluation technique that is called wizard of oz (WOZ) [41][26, p.227].

In a WOZ test, subjects interact with an application that seems to be fully functional but in reality it is an unimplemented system that are being simulated by a human operator, also known as the wizard. Participants are observed or measured to serve the evaluation purpose as if they are working with a real system. Most common is that the participants are not informed of the involvement of an other human, to encourage a more natural reaction [41][26, p.227].

For the WOZ technique to be effective, it must be carefully planned because the wizard cannot improvise and have to behave exactly like the system would [11]. In order to do that, the current design for the system behaviour must be clear and the wizard must be trained to execute it. This to make sure users believe that they are interacting with a system, the wizard needs to be able to select the system's response in real time to make this possible.

The WOZ methodology was picked because both Pearl [3] and MacTear et al. [11] mentions it as a good option for testing a conceptual system that has not yet been developed and is also ideal to test VUI design. The ideal WOZ test requires two researches a wizard and an attendant. The wizard keeps track of what the user is saying and enables the next action, the attendants job is to take notes and do the interviewing [3]. According to Norman [26] the wizard should not be in the same room as the user to add to the illusion of a functioning system. However do to limitations the tests were done only with a wizard that was in the same room as the user. An attendant was not available so the wizard

had to do the interviewing as well, but the tests were recorded to make sure that nothing the user said was lost.

A total of 10 tests was performed with 10 participants and the goal to test the users memory retention as well as getting user insights. The participants were 7 females and 3 males between the ages of 18 - 64. The goal was to find participants with none or very little experience with a Google Home Assistant. A few of the participants have tried one before but no one owns one. They were divided into two groups with equal distribution of 5 in each group. Group 1 consisted of participant A - E and Group 2 of F - K, see Table 4.1. The groups was presented with 5 different features and was asked to perform different task involving these. In the first part of the test, which was presented in a different way for the groups. The first group was presented with a voice tutorial that was designed using the WOZ methodology through which they were presented with the features. And the second group was presented with the features through an app (see Figure 4.1). The features the users go to use was picked from the results of the interviews (5.1) and these were deemed the most relevant. They were *Music, Timer, Alarm, Weather and Reminders*. In part 2 of the test all of the participants were asked which features they had been asked to perform in the first part. They were also asked in which way they would prefer to be onboarded in, when learning to use an IPA.

Table (4.1) Test participants

Participant	**Gender**	**Age span**
A	Female	55-64
B	Male	55-64
C	Male	18-24
D	Female	18-24
E	Female	25-34
F	Female	18-24
G	Female	25-34
H	Female	25-34
J	Female	25-34
K	Male	25-34

4.4.1 Design of the Test

The test was divided into to parts and into two test groups where each group got to do the same tasks but the presentation was different. In part one both groups was first guided through the setup on an for a new Google Home[17]. It was

done using Marvel's app *POP*[5] and importing print-screens to *POP* and making them interactable to simulate the actual Google Home app. The goal was to create more context and try to make the experience as realistic as possible. A Google Home and a Google Home Mini was used during the tests as well as an iPhone 6s. The wizard was using a MacBook Pro to control the voice for the tutorial.

In part one the first group got to do a voice tutorial where the assistant greeted the user and presented various features step by step and instructed them to try to perform the tasks along the way. This was done using a WOZ environment where the wizard was controlling the voice tutorial by using the speaker of the Google Home to play different sentences for the user. Since a Google Home was used it was fully capable of answering an executing the actions that the user asked it to do. The voice used for the wizards commands were picked to be the same voice as the assistants. Since the wizard needed to imitate a real system as much as possible and can not improvise during the test a script for the conversation was designed. The goal was to design the tutorial to have a conversational tone rather than telling the user exactly what to say (command-and-control). This was done by letting the assistant present features and letting the user formulate a response on their own, instead of presenting the feature and a way to execute that action for the user to then repeat. The actual design of the conversation was done by creating things for the assistant to say by thinking in reverse. By starting from the desired response from the user and coming up with one or more sentences to lead up to that response. The assistant was also included in the process to include its responses to make sure that the overall conversation felt as fluent and coherent as possible. Error handling was thought of as well, by including encouraging sentences for instance "Why don't you give it a go?" to prompt the user to try to perform a certain action. Other errors that was caused by the user for example saying something the assistant did not understand was handled by the assistant itself.

Instead of doing a voice tutorial the second group got to do a similar onboarding to the one Google uses [17]. After the setup they were given a redesigned version of the companion app for Google Home as the last picture in *POP* (see Figure 4.1). The app displayed the same features that the first group was presented with as well as an example of how to express themselves to perform these actions. They were asked to go through all of them, one by one. The order of which the app displays the features and how to perform them are in the same order as in the voice tutorial. This to keep the tests as similar as possible and to minimise the risk for errors in the results. The features that were presented to both groups were *Music*, *Weather*, *Timer*, *Alarm* and *Reminders*. These were picked based on the result from the interviews (see 5.1), they were the most commonly used features that did not require any external device.

The second part of the test was done in the same way for both of the groups. The goal was to find out if the presentation method made any difference to

[5] https://marvelapp.com/pop/

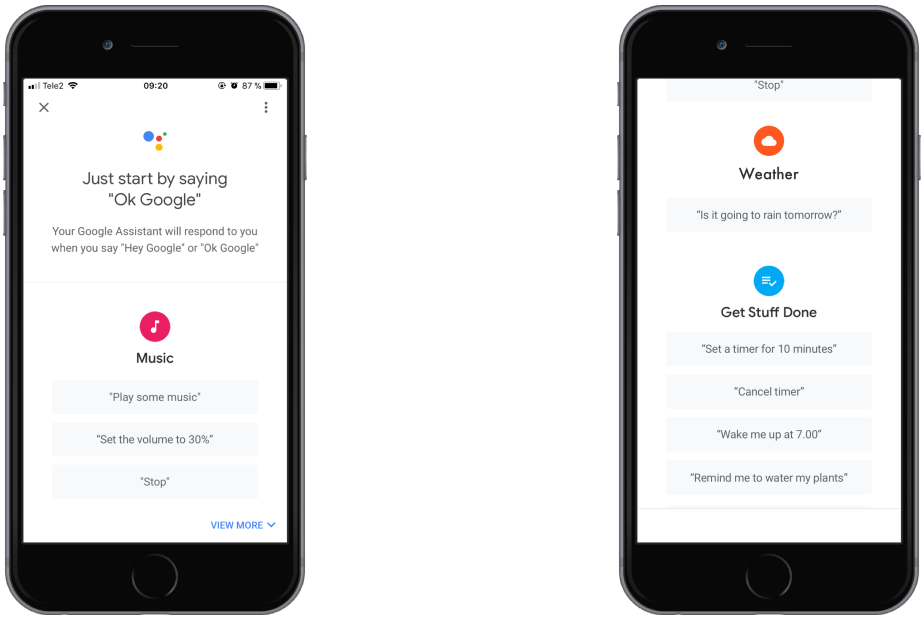

Figure (4.1) Modification of the Google Home companion app, teaching new users a few features, the version used in the test.

how many of the features the participants could remember from part one. And then compare the two groups to see if one excelled from the other. Part two was done two days after the first part. This was decided based on Ebbinghaus Forgetting Curve [2] and that in general humans forget around 60% of new knowledge learned over the first couple of days. The idea was to try to simulate a real situation where the user first learns a few features, and then it might go a period of time before they find a need to use it again. In this time gap they will either remember or forget specific features. For the test the process was to ask the participant "Which features do you remember performing in the first test?" and them telling the once they remembered. They were also asked about a preferred way to be initially taught to use the assistant. If they preferred to explore on their own with the help of an app or if they wanted the assistant to present itself and a few features. The questions was written down beforehand to make sure that the participants answered the exact same questions.

Development of the Wizards Test Environment

When the design was set the next step was to develop the parts needed for the voice tutorial to be able to work. The goal was to create an experience that felt as real as possible and that felt robust enough to handle real time interaction,

therefore Meteor[6] was picked. Meteor is a full-stack framework for building real-time apps with JavaScript, here real-time is referred to as a web-based application where the transactions between client and server take place at such a speed that users experience direct feedback on their interactions.

The system that was developed was used by the wizard to create the illusion that the IPA was always the one speaking but in reality it was a mix. The wizard sent prompts from the system to the Google Home that was played on the device's speaker and when the user interacted with it the assistant gave the responses. The prompts were designed beforehand and were a part of the script that was created during the design phase. For the illusion to work the voice had to be the same that Google uses for their assistant, otherwise the entire illusion would fall flat. This was accomplished using the browser based JavaScript SpeechSyntesis API[7], an experimental technology that is a part of the drafted Web Speech API specification[8]. It includes the voice that Google uses for their assistant and can also handle text-to-speech.

The created system consists of one view that displays inserted prompts on buttons, when a button is pressed by the wizard the SpeechSynthesis API reads the prompts out loud (see Figure 4.2). To add new prompts a text-field is available where the wizard can simply type in a new prompt and it appears as a button. Removal of prompts are available as well if needed. The system is developed to run on a desktop screen, since this was what the wizard was using during the tests.

4.4.2 Testing

The first step after completing the design and development of the test was to run pilot tests for both versions of the test, one for the voice tutorial and one for the app. This was done to make sure a situation that was not thought of beforehand could be taken care of before the actual tests. In both cases the pilot tests went well and nothing had to be changed for the final versions of the test. Part 1 of each test was performed at Bontouch's office in Stockholm and the second part was mainly performed at the office with a few exceptions that were done over phone. In total there were 10 participants in the test that was divided into two groups of 5 each. The reason for the number five in each groups is based on an article by Nielsen [42] that explains that only five participants is needed to find around 80% of usability issues.

[6] https://www.meteor.com/
[7] https://developer.mozilla.org/en-US/docs/Web/API/SpeechSynthesis
[8] https://dvcs.w3.org/hg/speech-api/raw-file/tip/webspeechapi.html

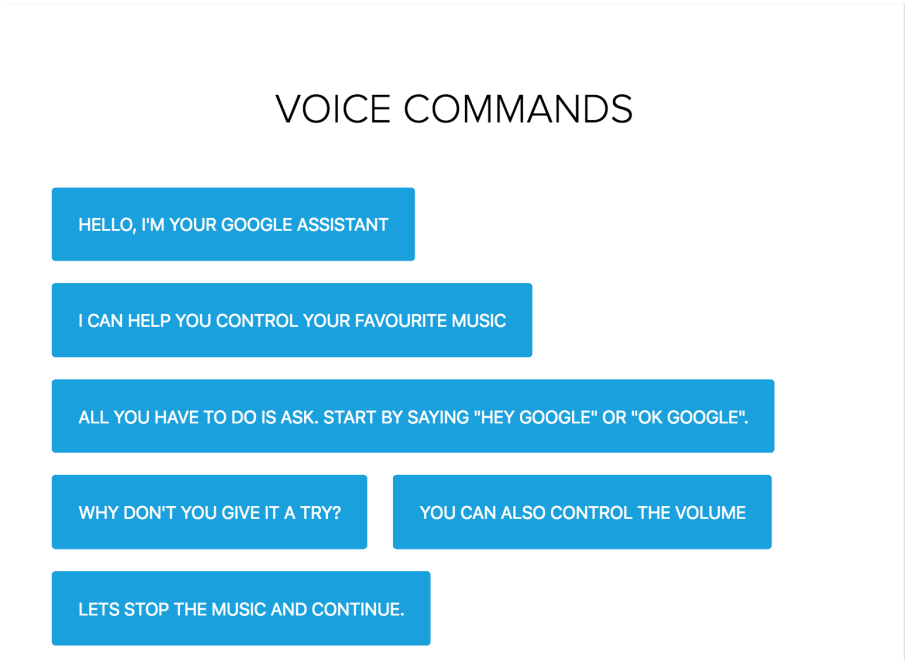

Figure (4.2) Screen shot from the system view used by the wizard in the test.

Part 1

In the first part of the test the participants of both groups were presented with print-screens from the Google Home App for setting up a new Google Home device. It was presented on an iPhone 6s and made interactive with the app POP by Marvel. The participants were asked to click through and read on each image. The last (second to last with group 2) image told the user that the assistant was ready to use. From this point the test became different for the two groups.

Group 1 - After the last image Group 1 was told that they would do a voice tutorial and the wizard started it by sending the first voice prompt to the assistant which greets the user. The prewritten script was gone through and if the users struggled there were prompts to guide the user to try. The entire script used for the voice tutorial can be viewed in appendix B.

Group 2 - The last view for the second group (see figure 4.1) differed from Group 1 which did not get to see it. It shows the same features that the voice tutorial presented to the first group as well as some ways to ask the Google Home to perform these features. The users were told to ask the assistant and perform all of them from top to bottom.

When both groups had finished they were told that there would be a second

part of the test in two days time. They were also told that it would involve answering a few questions and that it would not take up much of their time. Finally they were thanked for their participation and help.

Part 2

The second part of the test was performed two days after the first test for all participants. In part 2 both groups were asked the same questions. The first being how many features they could remember from the first part. The second question was about how they enjoyed the way they were taught to use the IPA or if they would have wanted to learn in the other way. E.g a participant in Group 1 that did the voice tutorial got to compare that to having an app instead, and vice versa for Group 2. They were also asked if they would have wanted a multimodal interaction with a combination of voice and app instead of only one of them. To avoid vague answers follow up questions was made to make the user expand their answers and get a clearer picture. After being satisfied with the participants answers they were again thanked for their participation. They were also informed about the test and what the goal was with it. If they had any more questions they were free to ask them as well.

4.4.3 Evaluation

After the tests were completed they were all gone through and transcribed. The outcome of the test were to find out how much each participant remembered between the test sessions, the usability of the assistant as well as their opinions on the first time interaction with the Google Home. The data from the test was compiled for each participant.

T-Test for Two Independent Means

To be able to compare the groups memory retention against each other, a t-test was performed to compare the differences of the two groups means values and to see if the result is considered of statistical importance.

Assuming that the means are independent from each other the t-test was calculated in the following way. First a hypothesis was defined as [43]:

$H_0 : \mu_1 = \mu_2$ (*the two group means are equal*)
$H_1 : \mu_1 \neq \mu_2$ (*the two group means are not equal*)

Followed by the definition of these variables:

$\overline{x}_1 = Mean\ of\ first\ group$
$\overline{x}_2 = Mean\ of\ second\ group$
$n_1 = Number\ of\ observations\ first\ group$
$n_2 = Number\ of\ observations\ second\ group$
$s_1 = Standard\ deviation\ of\ first\ group$
$s_2 = Standard\ deviation\ of\ second\ group$
$s_p = Pooled\ standard\ deviation$

The arithmetic mean for each group was then calculated. Which is defined as the sum of the numerical values of each and every observation divided by the total number of observations. With a data set of: $a_1, a_2, ..., a_n,$, the arithmetic mean A is defined by:

$$A = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 +, ..., +a_n}{n}$$

There are different ways to perform a t-test it depends on the standard deviation for both sample groups and if the standard deviations is assumed to be equal or not. To be able to tell which one to use the standard deviation for both groups was calculated [43].

The standard deviation is defined by[9]:

Let X be a stochastic variable with mean value μ, hence:
$E[X] = \mu$

Where E is the expected value of X. Which gives the standard deviation σ:

$$\sigma = \sqrt{E[(X - \mu)^2]}$$

And the variance[10] of X is defined by:

$Var(X) = E[(X - \mu)^2]$
$Var(X) = \sigma^2 = E((X - E(X))^2) = E(X^2) - E(X)^2$

After calculations (see the result 5.3.3) the standard deviation was considered equal and therefore the following formulas was used:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

and

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

[9] http://mathworld.wolfram.com/StandardDeviation.html
[10] http://mathworld.wolfram.com/Variance.html

The calculated t-value is then compared to a critical t-value from the t distribution table[11] with degrees of freedom value of $df = (n_1 + n_2 - 2)$. If the calculated t-value is greater than the critical t-value, then the null hypothesis is rejected [43].

User Insights

Lastly the participants opinions of their first time interaction was evaluated and summarised as well as their behaviour, questions, struggles and things they said during the test. The results are presented in chapter 5 and the results are discussed in chapter 6.

[11] http://www.maths.lth.se/matstat/kurser/tabeller/tabeller.pdf

Chapter 5

Results

The result section consists of three main sections: 5.1 Interviews, 5.2 Testing of a Google Home and 5.3 Wizard of Oz (WOZ) Test.

5.1 Interviews

The outcome of the five interviews gave an insight into how users use their voice controlled IPAs as well as other assistants they use, see Table 5.1 for all IPAs used. All of the interviewees had a Google Home at home that they use everyday. Three of the users use Google Assistant on their phone and one of them has used it on their android wear as well. Four of them have tried Siri on iPhone, with one that uses Siri everyday with their iPhone, with or without AirPods and has just purchased an Apple Home Pod which is used at home. One has previously used Siri on iPhone and Apple Watch, but has recently switched to an Android phone and therefor stopped using it. Four of the users have tried Alexa briefly and one of them has an Amazon Echo Dot at home. A summary of IPAs and hardware used can be viewed in table 5.1.

5.1.1 Most used Features

The most commonly used features are timers and controlling the lights at home with either Philips Hue[1] or Telldus[2], Tellstick. These features is used by everyone except P2 who bought a Google Home Assistant to be able to control the lights, but have not yet bought the Philips Hue system. P3 also said that controlling the lights was the main reason for purchasing an IPA. The timer feature is mainly used in the kitchen, because it does not require hands to do and is

[1] Read more: https://www.philips.se/c-m-li/hue
[2] http://telldus.com/

Table (5.1) IPAs and hardware used by Interviewees

Interviewee	IPA	Hardware
P1	Google Assistant, Alexa	Android Wear, Google Home Assistant, Google Pixel 2, Amazon Echo
P2	Google Assistant, Siri	Google Home Mini, iPhone, Apple Watch
P3	Google Assistant, Alexa, Siri	Google Home Mini, Amazon Echo Dot, iPhone, Apple Home Pod, Siri for Car, Google Pixel 1 & 2
P4	Google Assistant, Alexa, Siri	Google Home, Amazon Echo, Pixel 2, iPhone
P5	Google Assistant, Alexa, Siri	Google Home Mini, Google Home, Pixel 2, Amazon Echo, Apple Home Pod

much faster than through a phone. Controlling media is used by everyone as well, interviewee P1, P2, P3 and P4 uses their Google Home Assistant to control Netflix[3], which is done with the help of their Chromecast[4]. Three of them use their IPAs to control music at home, P2, P3, P4. Where P2 and P5 uses an Chromecast Audio[5] connected to their sounds system and interviewee P3 uses an Apple Home Pod for music. An other feature, that is used by everyone, are reminders. Interviewee P4 and P5 set reminders on their Pixel phones to do something when getting home. Another feature that is used is checking the weather, mostly done by P1, P3 and P5. P3 check the weather before leaving every morning to figure out which jacket to wear. P4 and P5 have sent messages through their assistants, P4 misses the T9 keyboards and does not like touch-screen keyboards, and therefore like this feature. P3 has used it as well, but does not enjoy the experience since there is no way to edit a message or add to it. It works for short messages but for longer messages it gets difficult since one need to say and remember the entire message without taking any pauses.

5.1.2 Learning to Use

All of the users used trial and error to learn how to use their IPA. Interviewee P3 also watched YouTube[6] videos and read new release notes form Apple about Siri to learn new features. Most of the users tend to search on Google when they

[3]https://www.netflix.com
[4]https://store.google.com/product/chromecast_2015
[5]https://store.google.com/product/chromecast_audio?hl=sv
[6]https://www.youtube.com/

want to learn a new feature or to find out if that feature is actually available. This is usually done by first trying to figure it out on their own and if they do not get it to work, search on Google. Two of the users checks the Google Home app which has a list of available features and one simply just ask if something can be done (with mixed results).

After some time of usage, one learns what you can say and how you should say it to get the result you want is something that is mentioned by P1, P2 and P3. P2 and P3 also talks about the difference between Siri and Google Assistant. The main one being that Google Assistant can handle more different formulations to execute one command, whereas Siri is more limited. Giving Siri the feeling of being less intelligent than Google Assistant.

5.1.3 Uncomfortable Usage

Using a voice assistant in public amongst other people is overall something the interviewees are reluctant to do. More specifically in situations where it can appear as a disturbance towards other people, e.g on the tube, in the office etc. In these situations they would also be reluctant to talk on the phone since it might disturb the people around them. Since Google's Assistant is multimodle when used on a phone, P4 and P5 types instead of talks to it in these situations. P3 uses Siri sometimes on the go, usually to play music, the interaction is done with AirPods and almost never with other people close by. In this situation is mostly because it is more convenient than removing gloves and using physical interaction with the phone. Interviewees P1 and P2 however, do not mind using Siri in public at all, P2 used to make calls with an Apple Watch using Siri. P1 do not feel a need to use the assistant in public situations, as on the tube, but states that if there was it would not be an issue.

5.1.4 Added Value

When asked about the added value and why they use their assistants the overall response was that it is easy to use, and in situations where both hands are occupied or it is easier than finding and doing it with their phone. One interviewee said that it reduces friction for commonly done tasks and therefore adds value. Being able to control their lights is also something that adds value for them. It makes a boring task less dull and also much easier. There is a possibility to create different preset alternatives, e.g morning lighting, saying goodnight to their IPA will shut off all the connected lights automatically.

5.1.5 Experienced Errors and Issues

Even though the users experience that the assistants improve and become better and better each day. They still experience problems when using their IPAs.

One issue that more than one interviewee experienced was with their Google Assistant and that after giving a command it says that it will execute the action, but nothing happens. It can be "Playing music on Spotify" but no music starts playing, the issue is experienced with Netflix and Philips Hue as well. One of the Philips Hue users have started to use their phone again since the issue happened too often. For this user it was months ago (fall 2017) and for an other user it recently started to happen (winter 2018) both are using a Google Home Assistant. A similar problem is when the user has spoken to the IPA and there is no response, it stays quiet and stop indicating that is listening. Interviewee P4 would want better error handling in these cases, for instance "I do not understand" is usually enough.

Another issue is something that one user referred to as "The Cocktail Party Problem" which is when the assistant can not hear the user since there is too much ambient noise, or to high music (experienced with Google Home). An issue that Apple seems to have solved with their Home Pod according to P2. Overall the assistants are performing quite well in understanding the user, all interviewees say that it is seldom that they have to repeat themselves. A more resent issue is when someone has more than one assistant at home, and therefore more units that listens. Resulting in two or more assistants responding to the same command. This issue (experienced with Google Home) has improved tremendously over the past few months according to interviewee P5, which has experienced this issue. Instead it simply says that it will respond on another unit.

Since all users are Swedish and live in Sweden all have experienced shortcomings and limitations because of this. Currently Siri is the only assistant that speak Swedish, but there are still major issues with her understanding the user correctly when for instance giving an address to navigate to. Simpler addresses works fine, but if they get more complicated in their pronunciation she usually gets it wrong according to interviewee P3. General features is also restricted in Sweden e.g there are no Swedish news channels and trying to play Swedish songs or artists are real challenge on Google Assistant. Interviewee P5 had issues with her address not being available with the weather feature. Telling E that there were no weather data at the location.

Issues regarding discoverability was also discussed. The main discussion revolved around how to "browse" available features by using the voice assistant. Since it does not give the user any visual ques the user have to ask and ask correctly. The issue becomes that the user needs to be creative and think of ways it might be possible to use the assistant. With a possible scenario of some users never discovering valuable features for them.

5.1.6 Positive Experiences

The interviewees was asked about situations where the assistant had surprised them by performing very well. Media control was one of the mentioned situations, and more specifically how well it handles interaction with Netflix and being able to play complicated playlist from Spotify. One mentioned the feature of being able to stop playing music after a set time, which made a great difference. The same interviewee was impressed that it can handle context to some extent and that it constantly improves and becomes better.

Most of the interviewees did not have very high expectations when they started to use their assistants. Now they all of them has a positive attitude towards their assistants and find it valuable.

5.1.7 Siri Versus Google Assistant

A major difference, that was pointed out by interviewee P3, is that Google Assistant can communicate with other devices in a way that Siri can not. A Google Home Assistant can communicate and control actions with a Chromecast, a phone running Google Assistant and more, since it is a cloud based service. It communicates actions through the cloud and can therefore communicate with other devices running Google Assistant. Siri on the other hand can not talk to other devices running Siri. Therefore a user can not talk to Siri on their iPhone and start music on their Home Pod, only Siri running on the Home Pod can do that. There is still the possibility to stream music from the phone to the HomePod and use it as a remote speaker.

5.2 Testing of a Google Home

User Onboarding

The first part of starting to use a Google Home is the set up. It is done by plugging it in to a power outlet Google Assistant greats the user after a few seconds and tells the user to download the Google Home app and use it to set up. After completing the steps in the app, which includes connecting to Wi-fi as well as adding account information to Spotify[7] and Netflix. The assistant says that it is ready to use and to continue in the Google Home app to learn more about available features and what can be said to it. The app at this point can be seen in Figure 5.1.

The features presented by the Google Home app as part of the onboarding was all tested, starting with playing music, increasing the volume and turning it off again. Next part is about the user asking questions and getting answers for the

[7] https://www.spotify.com/se/

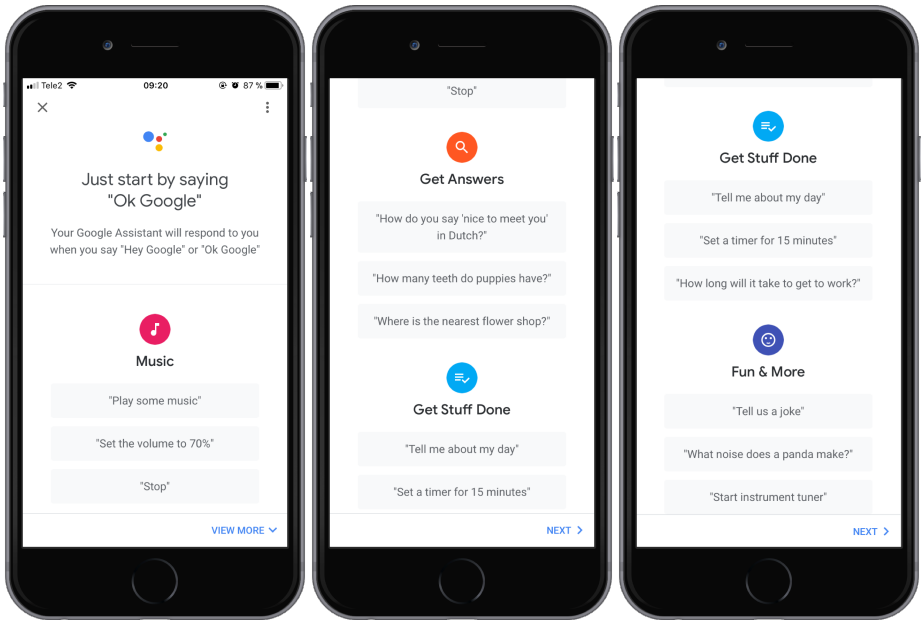

Figure (5.1) Part of Google's User Onboarding towards new users, presenting features and possible commands in the Google Home app.

IPA, some of the examples is learning to say "Nice to meet you" in Dutch, how many teeth a puppy have and where the nearest flower shop is. After learning that puppies have no teeth at first, but quickly gaining 28 teeth, it is time to move on to the third part which is getting stuff done. Here Google presents features that can be useful to make everyday tasks easier and more accessible. First off Google presents tell feature "Tell me about my day", which presents the weather, the latest news (no Swedish news right now), commute information, important things to do, meetings etc and more. Users are also being taught to set a timer and presented to the possibility to ask the assistant how long it will take for them to get to work. When asked the later, the assistant said that no work address had been added and gave the alternatives of adding it using the app or telling it straight away using voice. A few attempts were done using voice, however since it does not understand Swedish this made entering a Swedish address nearly impossible using voice, and was therefor added using the Google Home app. The last section of the onboarding is perhaps the least important but it might add some fun elements to the assistant. With questions like "What noise does a panda make" and asking the assistant to tell a joke.

The next part of the onboarding is when the user have used the product for a while and wants to learn new features. Google presents some of these in the Goggle Home app under a menu called "Explore". Here features are shown with some things to say, correlated to that specific feature. For instance under "Play

Music" examples are "Raise the Volume", "Pause" and "Play Queen". The IPA can also handle different formulations for the same action e.g when asking about the weather, it is possible to say "What is the weather?" but it is also possible to say "Do I need an umbrella?". The reply will be slightly different but the content of the answer fairly similar if not exactly the same. Some other features that was done was to play radio and podcasts, which did not work very well at all. When asking the assistant to play radio (like it says in the app) music will start playing on Spotify. After some trial and error however it was accomplished to play radio, this after figuring out how to actually say it. In this case it was to ask the assistant to play a specific radio station and not just to play radio. Currently there is a similar issue when wanting to play podcasts. Right now the assistant tells the user how to play a podcast, which is an improvement from a few months back when it said that it did not understand the command "Play Podcast". However so far it has not worked playing a podcast using the assistant.

It is also possible to ask the Google Home Assistant "What can I say to you?". This results in the assistant telling two random features that can be asked for and if the user wants to hear more. In total six features are portrayed before it refers the user to the Google Home app to learn more features. If asked again, different possible things to say are presented than the first time, same goes for a third time. During one of these tries the assistant said that it is possible to pair the speaker to Bluetooth by telling the assistant to "Pair Bluetooth". Something that was later looked for in the app, but without success.

Continuous Usage

After a few months of testing an insight of what a current IPA can be used for, what it is good at and what needs improving. The most used feature was using the assistant and Chromecast to control Netflix on the TV. The interaction is much faster and simpler than with the app or through Chrome on a computer, if the user knows what to watch, if the goal is to browse the assistant falls flat and can not compete with a graphical interface. Other controls like changing between episodes, fast forward etc. works really well. Playing music on Spotify using the built in speaker have been done a lot as well. It works quite well at playing and controlling music, even at finding playlists with more complex names. However it sometimes struggles to hear the wake word ("Hey Google") if there is too much ambient noise or if the music is too high. Since the assistant does not understand Swedish it have a hard time with Swedish artists and songs, giving quite interesting suggestions. The last feature that have been used is timers, usually in association with cooking. Since there are no hands needed for the interactions it makes it so much simpler and easier than using a smartphone.

After Interviews

After completing the interviews (see Figure 5.1) a few new features and ways of thinking when interacting with the assistant was attained. First one being that there are other smart assistants available on the Google Home besides Google Assistant. All the user have to do is ask to speak to them, e.g "Talk to Swedish Radio (SR)" will let the user speak to SR's assistant. At this point the SR assistant can not play radio on the Google Home, it has yet to be developed. There are other assistants available as well however they are limited at this point since the technology is fairly new with a limited amount of users. In some cases the assistant will suggest an assistant to talk to. For instance if the assistant is asked to play "ambient" sounds it will give suggestions to talk to an assistant that can do that. The mindset of just asking for things had been thought of before but gained more traction after this discussion. The other new feature was that it is possible to stop playing e.g music after a certain period of time and that the opposite feature might be available soon as well, starting music at a certain time.

5.3 Wizard of Oz (WOZ) Test

The results of the test are presented below. The memory retention of each group are shown in sections 5.3.1 and 5.3.2. In section 5.3.3 the calculations and the result from the t-test are presented. Section 5.3.4 gives user insights as well as observations made during the part 1 of the test.

5.3.1 Group 1 - IPA Voice Tutorial

The memory retention for Group 1 can be viewed in Table 5.2 and shows that everyone remembered 3 or more features. With A and C that remembered 3, D and E that remembered 4 and B that remembered them all. Looking at each individual feature (see Table 5.3) *Weather* seems to be the feature that most struggled with remembering where as *Alarm* and *Timers* were remembered by everyone.

5.3.2 Group 2 - Instructed by App

Moving on to Group 2 and how the group performed which can be viewed in Table 5.4. The results shows that that everyone remembered 3 or more features. Looking to the memory retention of each individual participant F, G and H remembered 3, K remembered 4 and J remembered all of the features. When looking at each individual feature *Alarm* and *Music* was remembered by everyone, both *Weather* and *Reminders* more than half of the participants

Table (5.2) Participant memory retention for Group 1 that was intructed by a voice tutorial.

Participant	**Remembered Features**	**Memory Retention**
A	Alarm, Reminder, Music, Timer	4/5
B	Alarm, Reminder, Music, Timer, Weather	5/5
C	Alarm, Music, Timer	3/5
D	Alarm, Reminder, Music, Timer	4/5
E	Alarm, Music, Timer, Weather	4/5

Table (5.3) Memory retention for different features in Group 1, shown in the order they were presented to the participant.

Feature	**Memory retention**
Music	5/5
Weather	2/5
Timer	5/5
Alarm	5/5
Reminder	3/5

remembered and *Timers* was the feature that most struggled with remembering.

5.3.3 T-test for Two Independent Means

The calculations and the results of the T-Test are presented below.

The hypothesis was defined as [43]:
$H_0 : \mu_1 = \mu_2$ (*the two group means are equal*)
$H_1 : \mu_1 \neq \mu_2$ (*the two group means are not equal*)

the following variables are defined:

Table (5.4) Participant memory retention of features presented by the app in the test for Group 2.

Participant	**Remembered Features**	**Memory Retention**
F	Alarm, Music, Timer	3/5
G	Alarm, Reminder, Music	3/5
H	Alarm, Music, Weather	3/5
J	Alarm, Reminder, Music, Timer, Weather	5/5
K	Alarm, Reminder, Music, Weather	4/5

Table (5.5) Memory retention for different features in Group 2, shown in the order they were presented to the participant.

Feature	**Memory retention**
Music	5/5
Weather	3/5
Timer	2/5
Alarm	5/5
Reminder	3/5

$\overline{x}_1 = Mean\ of\ first\ group$
$\overline{x}_2 = Mean\ of\ second\ group$
$n_1 = Number\ of\ observations\ first\ group$
$n_2 = Number\ of\ observations\ second\ group$
$s_1 = Standard\ deviation\ of\ first\ group$
$s_2 = Standard\ deviation\ of\ second\ group$
$s_p = Pooled\ standard\ deviation$

With the data from both groups in the tests the means were calculated:

$$\overline{x}_1 = \frac{(\frac{4}{5} + \frac{5}{5} + \frac{3}{5} + \frac{4}{5} + \frac{4}{5})}{5} = \frac{4}{5} = 0.8$$

$$\overline{x}_2 = \frac{(\frac{3}{5} + \frac{3}{5} + \frac{3}{5} + \frac{5}{5} + \frac{4}{5})}{5} = \frac{18}{25} = 0.72$$

The number of observations for each test are 5, which gives:

$n_1 = 5$
$n_2 = 5$

The standard deviation of Group 1 is calculated by first calculating the variance which is done by calculating the mean of difference between the data points values and the mean value which are then squared:

(1) $(0,6 - 0,8)^2 = 0,04$
(2) $(0,8 - 0,8)^2 = 0$
(3) $(0,8 - 0,8)^2 = 0$
(4) $(0,8 - 0,8)^2 = 0$
(5) $(1 - 0,8)^2 = 0,04$

$$Var = \frac{0,04 + 0 + 0 + 0 + 0,04}{5} = 0,016$$

The standard deviation s_1 is gained by taking the square root of the variance:
$s_1 = \sqrt{0,016} \approx 0,13$

The same is done for Group 2, which gives:

(1) $(0,6 - 0,72)^2 = 0,0144$
(2) $(0,6 - 0,72)^2 = 0,0144$
(3) $(0,6 - 0,72)^2 = 0,0144$
(4) $(0,8 - 0,72)^2 = 0,0064$
(5) $(1 - 0,72)^2 = 0,0324$

$$Var = \frac{0,0144 + 0,0144 + 0,0144 + 0,0064 + 0,0324}{5} = 0,0164$$

From the variance the standard deviation s_2 can be computed:

$$s_2 = \sqrt{0,016} \approx 0,13$$

Since $s_1 = s_2 = 0,13$ the standard deviations are assumed to be the same and the t-test [43] is therefore calculated as follows:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}} = \sqrt{\frac{(5-1)0,13^2 + (5-1)0,13^2}{(5 + 5 - 2)}} =$$

$$= \sqrt{(\frac{0,1352}{8})} = \sqrt{0,0169} = 0,13$$

And:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0,8 - 0,72}{0,13\sqrt{\frac{1}{5} + \frac{1}{5}}} = 0,9730... \approx 0,97$$

From the distribution table the value $2,31$ is retrieved, which gives $0,97 \not> 2,31$. Therefore the conclusion is that the null hypothesis is not rejected and the result is of no statistical importance. Hence the means are considered equal and the two groups memory retention as well.

5.3.4 User Insights

All participants were asked about in which way they preferred the first interaction to be. In the same way that they had done during the test, in the way that the other group had interacted with the assistant or a combination of them both.

Group 1

Participant C would like both an app and a voice tutorial. For the initial onboarding a learning process C wants the tutorial and think it is a fun and good way to get started. C would like to have the app mainly to learn possible new features over time. D gave a similar answer and liked the tutorial because without the tutorial, D feared that they would feel lost and not sure what to do. Participant E would have wanted an app as a complement to be able to get a better overview. E liked that the tutorial gave an introduction to the assistant. A and B gave similar answers as E and would have wanted an app to show some information parallel to the tutorial and that the tutorial felt a bit to long.

Group 2

In group 2 participant F said that they liked the setup in the app and would not want that to be any different. F also liked to have the app instead of a voice tutorial because F wanted to be in control and not have to wait for the assistant to finish talking. However F also said that it would have been nice if the assistant introduced itself a little bit, but not more than that. G also like the setup but struggled a bit with knowing that they had to say the wake word ("Hey Google") for it to listen, this could have been clearer. G found it good to be presented with a few features but wanted it to be voluntary and not mandatory. Preferred the app over the idea of a voice tutorial although G would have wanted to see that the assistant introduced itself more. And that being instructed to try one command might be a good idea to be shown that it works. H had similar thoughts as G and liked to be able to have an overview of the initial features with the app and would have like that the assistant introduced itself more. H felt that just starting to talk to it without it introduction itself first felt strange. An example that H gave was for it to say "Hello, what can

I help you with" and then looking at the app to see what is possible to ask for.

Test person J said that it would have been nice if the assistant greeted the user. J also said that it would be nice if it helped to get going with one feature and then left to explore on your own. The idea of a voice tutorial participant J did not think was a good idea since they would want to explore on their own and a full tutorial would be more annoying than helpful. J also said that just using the app would feel less personal and said that it felt more logical that the assistant had a bigger part in the learning since it is the device that is going to be used and not the phone. Test person J reflected about different target groups and for some, less technical, it might be good to have a longer tutorial. The last participant K liked to be presented with a few examples of what can be said. And would like more to be available for later. However when asked about the voice tutorial K said that they preferred only an app since K wanted to be able to read instead of hearing. K also said that a better confirmation that the assistant have heard what they said.

Observations

During part 1 of the tests for both groups a few observations were made. Either things the test participants said during the test or observations made during the tests. One of the most prominent were that most of the participants struggled with the concept of the wake word, "Hey Google or OK Google" to enable the assistant to listen. After saying it the first time most of them waited for a response from the assistant, but none came. What happens is that it lights up when spoken to as an indication that it is listening and waits for the user to speak. However in these situations the participants were waiting for the assistant to say something. After the first trial they usual go to was to try again and asking it to do something. A fairly common issue as well was that the participant waited to long between saying the wake word and the command so the assistant had stop listening. A reflection that a few had and asked as well was that the wondered if they had to say the wake word everytime they wanted to say something to the assistant. In this situation they said many commands at once as well which might not be the normal situation a user is in. Most of the participants forgot to say the wake word atleast once during the test and many of them more than once. Everytime looking puzzled over not getting any response from the assistant.

Another observation that was correlated to the fact that all participants were Swedish, is that some of the participants struggled with formulating questions and commands to the assistant. In some cases the situation was that the participant said the wake word and then "Uhm, how do I say that?". Leaving them to have to think one more time and trying again. In a few situations they received help with how to ask a specific question. One of the participants in Group 1 exclaimed "Yes please!" after the assistant said "Wondering if it is going to rain

tomorrow?" during the voice tutorial. Having the assistant responding with "Indeed" and nothing more. After explaining that the assistant does not understand unless a question is asked. The participant said "But she asked me a question!".

5.3.5 Errors Caused by the Google Home

During the tests the assistant behaved incorrectly in some situations when a command was given by the test participants. Which caused some confusion and in most cases they tried again or was encourage to do so. The error prompt that Google uses that caused the most confusion was in situations where the participant was asked to perform a task and the response from the assistant was *"I do not know how to do that yet"*. Other examples were when the assistant completed a different task then was asked for. This happened more during testing of the voice tutorial where the participants were given more freedom in their formulation than with the app. However these errors could be found in the app as well, where the assistant misinterpret what was said to it. In one of the tests the assistant got stuck in a loop, asking the user over and over at what time they would like to be woken up, when a participant asked the assistant to wake them up the following morning.

Chapter 6

Discussion

This chapter has three main parts: 6.1 Interviews, 6.2 Testing of a Google Home and 6.3 Wizard of Oz (WOZ) Test.

6.1 Interviews

From the interviews knowledge was gained which provided a deeper understanding of current IPAs and how they are used. As well as how users learned to use their assistant, which added value it has, what they use their assistant for and usability issues they have experienced with the assistant or in some cases assistants. Other valuable insights were if they were situations where they would avoid using a VUI, which assistant is currently most commonly used and also the current differences between how Apple's and Google's assistants are able to communicate with other devices or not.

Learning to Use

Looking at how current users have learned to use their IPAs seems exclusively to be trail and error. Where they either just try something in hope that it might work or search online to see if that specific feature is available and possible to perform. One option that is available, that only one of the interviewees takes advantage of, is to ask the assistant what is possible to do. Which can be because the mindset of asking an interface how it should be used does not come natural. Currently it seems to be very much up to the user to learn and find new features. A discussion with the interviewees came to the same conclusion. The lack of visual cues, discoverability in a VUI and the fact that there is no good way to browse features. The issue becomes that the user have to be creative and think of ways it is possible to use the assistant. Therefore there is always

the possibility of users never discovering features that could have been valuable to them.

It would be interesting to look into how to continuously onboard users and to teach them new features after they have had their device for a while. A possibility could be for the assistant to tell the user if it have gained a new feature after an update. However the interviewees stated that they would not like the assistant to just start speaking. There are other ways though, a light indicator on the device or something similar that indicates that the assistant has something it wants to tell the user. This is currently the way Google Home does it when the user has set a reminder and the assistant wants to remind the user of it.

Usage and Features

All of the interviewees had a Google Home and only one had an other devices (besides on their phones) which is a Apple HomePod and an Alexa. The reason could be that the HomePod was just released as well as the limited availability in Sweden. Currently it seems that Google Home is mainly used together with other devices. All of the interviewees had or had the intention of buying Philips Hue or similar to be able to control the lights at home. A few of them used Google Chomecasts and connected the assistant to be able to control their TV and music with it. It seems that home automation is one of the main fields that assistants will be used for in the future as well. One of the interviewees stated that they were looking into more sensors to be able to control their window blinds at home etc.

Other features that was commonly used on Google Home were weather, setting alarms, timers and reminders. For all of the interviewees that used the assistant for these features stated that it was easier and faster than doing it on their phone. Especially in certain situations when both hands were occupied with other things, e.g while cooking. With all of the interviewees which was one of the reasons that the assistant gained value to them. The main theme seams to be as one of the interviewees put it, that it reduces friction for commonly performed tasks and therefore it adds value.

Uncomfortable Usage

An important aspect to discuss is that talking in public to an assistant still feels awkward and something most of the interviewees is reluctant to do. One plausible reason could be that the technology is fairly new and society have yet to catch up to the technology and make it socially acceptable. An other reason that the interviewees gave were that they avoided talking to the assistant (on their phone) in situations where it could disturb other people. These situations are similar to those where they would avoid talking on the phone.

Errors and Issues Experienced

The interviewees had experienced some problems with their IPAs. Some of these errors seems to be bugs in the software rather than a fault in the actual design. For instance an error that was experienced was telling the assistant a command, the assistant confirming it but nothing happened. This happened mostly with third party applications, e.g Spotify, Netflix and Philips Hue, it might have been a connection issue that the Google Home (assistant) system did not realise. An other reported issue was after a question had been asked, the Google Home indicates that it is processing the command but then nothing happened. The frustrating part of this error is that there is no indication of what went wrong. Did it not understand? Did the request time out for some reason? or what did actually happen.

The cocktail party effect is something that was mentioned as well during the interviews and the issue of to much ambient noise, e.g conversation or music, which leads to the assistant to being able to hear the user. One of the interviewees discussed the issue and made the proposal that when playing loud music, the assistant should know what is playing and being able to block out the music. Which sounds possible and would be a major improvement. However the issue of uncontrolled ambient noise coming form conversation etc is harder to do something about. In loud environments a voice assistant will struggle and in these situations a graphical interaction might be preferred. An example situation is talking a phone call during a loud concert, it is usually impossible to hear anything. In this situation text messaging is much better since the surrounding noise has no effect on the conversation.

An other issue that one of the interviewees reported is having more than one Google Home. If both of them listens and hears the wake word initially both wanted to respond. Currently it one speaker will say "I will respond on an other device" and the other device handles the request from the user. However if someone decides to use Google Home as a sound system and places speakers all over the house. Having multiple speakers saying "I will respond on an other device" does not seem to be an optimal solution.

All of the interviewees are in Sweden and are using their Google assistants in English since it is not available in Swedish until later this year (2018)[1]. Siri is available in Swedish but it seems like it sufferers some issues still. Mostly understanding the user in some situations. One of the interviewees have tried to use Siri on an iPhone for navigation with various results. Some addresses works fine but some it struggles a lot with others. The result will probably be that users do not use the feature after struggled the first times. If it adds friction rather than remove it there little reason to use it. With Google Assistant trying to use navigation to Swedish addresses, trying to play Swedish artist or podcasts is nearly impossible because it does not understand what the user is trying to

[1] https://m3.idg.se/2.1022/1.697995/google-assistant-kommer-till-sverige-under-2018

say. Hopefully these issues will be removed when the assistant is available in Swedish.

6.2 Testing of a Google Home

The test of a Google Home gave valid and good complimentary information to the one gained from the interviews. The first being a more in depth understanding of how the onboarding of new users are done, as well as how the onboarding is done over a longer period of time. It also gave insights into how well it performs and what is currently possible to do with it.

A new way of thinking towards how to interact with the Google Home was also gained. Which was that Google Assistant is the only assistant available on the device. However to learn that this is available is either done by learning it from someone else or by stumbling on to it by asking the right questions. There might be other external ways as well, but in the Google Home app there is no indication that this is available. On Google Home everything is available to everyone, there is no installation needed for different external applications the user just have to ask for them. The issue becomes to know that they are available. The overall issue that VUI have had historically and it seems like they still suffer from it, is low discoverability. There are no real clues to what can be done. The option to ask is of course available, but it puts pressure on the user to know what to ask for. There were features that were found after asking the assistant what it could do. These were later looked for in the Google Home app without result. Which raises the question of how many of these features there are? Features that could give value to the user but since the presentation of them are badly executed with low discoverability users might never find them.

6.3 Wizard of Oz (WOZ) Test

The test gave insights into users memory retention when taught to use features by a voice tutorial or an app. Other valuable insights were given by interviewing the test participants and receive knowledge of how they would have wanted the initial onboarding to work. A few observations was made during the test that was also found to be valuable.

Memory Retention

Group 1 - all in group one remembered 60% or more with *Music*, *Alarm* and *Timer* that was remembered by everyone. *Weather* was the feature that only 2 out of 5 remembered and reminders 3 out of 5. The average memory retention for Group 1 was $0,8$.

Group 2 - from the second group the memory retention was 60% or higher for all the participants. With *Music* and *Alarm* being features that was remembered by all participants. For Group 2 *Timer* was the feature that only 2 participants remembered with *Weather* and *Reminders* was remembered by 3. The overall memory retention for Group 2 was $0,72$.

Comparing the two groups all participants remembered 60% or more with everyone remembering *Music* and *Alarm*. *Reminders* was the same for both groups as well with 3 participants remembering the feature. The remaining two performed a bit differently for the both groups. With *Timer* being remembered by everyone in Group 1 but only by 2 in Group 2. The reason might be relevance to the different participants, that timers adds more value to participants in Group 1 than in Group 2 and therefore they remember it. The last feature *Weather*, had low remembrance rate in Group 1 and only slightly better in Group 2. The reason for the low memory retention might be the way the feature was presented. In the voice tutorial the user was asked "Wondering if its going to rain tomorrow?" added by "I can check that for you" and in the app it said "Is it going to rain tomorrow?". The formulation of these might have had an impact on the users memory retention. An alternative that might have performed better would be to use "Weather forecast" in the prompts to use the name of the feature and therefore the remembrance rate might have increased.

The overall performance of the groups with the mean values of $\frac{4}{5} = 0.8$ for Group 1 and $\frac{18}{25} = 0.72$ for Group 2. The results gives an indication that Group 1 and the voice tutorial performed slightly better than the app and Group 2 when looking to memory retention. With Group 1 remembering 2 features more than Group 2. The indication is therefore that a voice tutorial is a slightly better way to teach new users when looking to remembrance rate. However the t-test showed that there is no statistical significant difference between the two groups and a conclusion that one is better than the other can not be drawn. To examine if a statistically significant result can be achieved a quantitative should be performed with a lot more participants involved in the study to see if the two groups will differ more or stay similar to the results of this test. A possible scenario could be to do a two version test (an A/B-test) on newly bought devices to get a higher rate of participants as well as a more real life scenario. A conclusion that can be drawn from the results is that both a voice tutorial and an app are valuable options that performs similarly when onboarding new users. From the perspective of memory retention and how well users remembers features.

User Insights

From the tests valuable user insights were gained. Primarily in which ways that the participants liked the interaction that they got to try as well as a reflection on how they would have wanted it to be. Both of the groups liked the setup that was performed in the app, and no one wanted it to be done differently. In

Group 1 the overall take was positive towards a voice tutorial when starting to learn to use the assistant. In the group three of the participants would have wanted an app as well during the tutorial as a complement. And an additional participant wanted the app for further learning.

With Group 2 there were some what similar answers to Group 1. Two of the participants in Group 1 wanted app only since they they preferred to be able to read on the screen, be able to skip features that were of no interest. One of them also said that they did not want to wait around for the assistant to finish speaking but rather try it themselves. Although this participant also said that it would have been nice if the assistant presented itself a bit but did not explain any features. The other of the two liked to be presented with a few features and then be able to look up more in the app later.

One participant liked the app as well, but would have wanted the assistant to present itself more than the two previous participants. That not only present itself but prompt the user to try one feature to make sure it works and then be able to try for themselves. Which was similar to another participant who wanted to be presented with one feature by the assistant as well as an introduction. The idea of a full voice tutorial was expressed to be more annoying than helpful. This participants also liked to have an overview of the features in the app. The next participant had a similar approach, where they wanted a more personal introduction and being asked by the assistant what it could help them with.

With both groups it seems as a multimodal interaction would be preferred by most. By combining a voice introduction by the assistant and having it present one feature as well as showing a few features in a mobile application. For the assistant to be completely quiet seems to be considered impersonal and a bit awkward. A too long voice introduction does not seem to be the best way either. The ones that preferred a voice tutorial were all from Group 1 and actually got to try it. The responses might have been a bit slightly different in Group 2 if they were the once that got to try it.

Observations

One of the main struggles that nearly all the participants had difficulties with was the concept of the wake word "Hey Google". Nearly everyone forgot to say it before addressing the assistant atleast once during the test. An other observation was that some of the participants said "Hey Google" to wake the assistant, but no command. They waited for the assistant to indicated by voice that it had heard them. After realising that it only lights up as an indication they tried the wake word and a command instead. An issue that involved the wake word as well was that some of the participants waited to long between saying "Hey Google" and a command, resulting in nothing happening. Some of the participants asked during the test if they had to say the wake word every

time before they said a command.

The struggle with the wake word probably has many different explanations. The easiest being that the users are not used to addressing an IPA and are a bit unsure how to do it, with an insecurity and being afraid to do something wrong. An other explanation can be that the confirmation that it have heard, the blinking lights, is not enough for the new users to understand that it is listening. When addressing a real human there are all sorts of signs that they are listening, eye contact, verbal confirmation etc. Using both the lights and a sound could therefore be an idea to increase the indication that the assistant is listening. An other would be that the assistant gave a verbal confirmation, atleast with new users. That the users forget the wake word between commands is probably only a beginner mistake. Prompting the user to say something when they take to long to say something to the assistant might help new users. It might take a turn in the opposite direction an not become helpful and only irritating instead.

The issue that one participant had in correlation to the design of the voice tutorial and one of the prompts. The situation was that the assistant said "Wondering if it is going to rain tomorrow?" which is directly asking a question. The response the user gave was quite natural, by simply answering the question with a "Yes!". With the assistant responding with "Indeed". A conclusion here is that asking direct questions in a situation were it might cause errors can be tricky. However if the entire design of the voice tutorial was made by Google, it would be easier to design for errors in situations like these. By actually being able to handle them.

In this situation the design of the voice tutorial might have been better off with another voice prompt and a different way of presenting the feature than with a question that can be misinterpreted and might lead to confusing the user.

Chapter 7

Conclusion

Development in voice interaction is growing and improving everyday. Voice interaction is perfect in situations were it can remove friction and make life easier for users. Although intelligent personal assistants (IPAs) are not nearly as good as the operating system in Her or the protocol droid C-3PO from Star Wars, the possibilities are many. This thesis aimed to investigate the user experience and memory retention of different ways of onboarding a new user to an IPA. Findings also show an insight into the usability of IPAs on the market right now (spring 2018).

The process of this thesis was divided into two phases, understanding and exploration. The understanding phase included a literature study, five interviews with experienced users of IPAs and a long-term test of a Google Home. The purpose was to get an insight into IPAs, getting a better understanding of how to design for VUI as well as how users experience IPAs and how they are used. The second phase, exploration consisted of a WOZ test with the goal of finding out the difference in memory retention between two different onboarding methods. A voice tutorial performed by the IPA and getting taught by an app. An other goal of the test was to find out the users opinions of the onboarding methods and which one they preferred.

The findings of this study shows that there is no difference in remembrance of features when using a voice tutorial performed by the IPA or through an app on a mobile device. The results also show that the majority of users would prefer to have a multimodal interaction when being taught to first use an IPA. A multimodal interaction is achieved by combining voice interaction with an visual screen interaction on a mobile device, an app. Furthermore the IPA should introduce itself and prompt the user to try one feature and show an overview of a few more available features on the mobile device. Other findings from expert user interviews and a long term test of a Google Home show usability issues with IPA as well as added value to the users, experienced issues. The overall

conclusion from them was that in situations where friction is removed a voice interaction is preferred. This can be to put on a timer while cooking or be able to turn on and off lights at home using their voice.

7.1 Future Work

To see if the difference in memory retention between the two test groups will remain the same or show a significant difference more tests need to be performed. The result from the interviews also show that a multimodal interaction is something that the users prefer. Therefore an inclusion of a multimodal version of the user onboarding to the test and adding a third group to the test is very relevant. It could also be interesting to include more than memory retention as a parameter to the test. Examples could be error rate, tasks completed without help, average time-on-task etc.

References

[1] Samuel Hulick. *The elements of user onboarding*. 2014.

[2] Psychestudy. Ebbinghaus forgetting curve, 2018. URL `https://www.psychestudy.com/cognitive/memory/ebbinghaus-forgetting-curve`. [Online; accessed April 1, 2018].

[3] Cathy Pearl. *Designing Voice User Interfaces: Principles of Conversational Experiences*. O'Reilly Media, Inc., 1st edition, 2016. ISBN 1491955414, 9781491955413.

[4] IMDB. Her, 2013. URL `http://www.imdb.com/title/tt1798709/?ref_=rvi_tt`. [Online; accessed Februari 26, 2018].

[5] IMDB. Star wars, a new hope, 1977. URL `http://www.imdb.com/title/tt0076759/characters/nm0000355?ref_=tt_cl_t6`. [Online; accessed Februari 26, 2018].

[6] Google Development. The conversational ui and why it matters, 2017. URL `https://developers.google.com/actions/design/`. [Online; accessed Februari 26, 2018].

[7] Max Slater-Robins. People are falling in love with microsoft's 'little bing' virtual companion. 2015. URL `http://www.businessinsider.com/people-in-china-love-microsoft-xiaoice-little-bing-2015-11?r=UK&IR=T&IR=T`. [Online; accessed Februari 26, 2018].

[8] Amy Nordrum. Ces 2017: The year of voice recognition - there's game-changing technology in our midst, and it's probably not connected cat bowls, 2017. URL `https://spectrum.ieee.org/tech-talk/consumer-electronics/gadgets/ces-2017-the-year-of-voice-recognition`. [Online; accessed Februari 27, 2018].

[9] Adobe. Adi consumer electronics report, 2018. URL `https://www.slideshare.net/adobe/adi-consumer-electronics-report-85929760?ref=http://www.cmo.com/adobe-digital-insights/articles/2018/1/8/`

the-future-of-consumer-electronics-is-voice-activated.html. [Online; accessed Februari 27, 2018].

[10] Michelle A Hoyle and Christopher Lueg. Open sesame!: A look at personal assistants. In *Proceedings of the International Conference on the Practical Application of Intelligent Agents (PAAM97), London*, volume 51, page 1997. Citeseer, 1997.

[11] Michael MacTear, Zoraida Callejas, and David Griol. *The Conversational Interface: Talking to Smart Devices*. Springer, 2016.

[12] Intercom. *"Intercom on Onboarding"*. Intercom Inc, 2016. ISBN 978-0-9861392-4-6.

[13] Psychestudy. Trial and error learning, 2018. URL https://www.psychestudy.com/behavioral/learning-memory/trial-error-learning. [Online; accessed April 19, 2018].

[14] Eric Nishio. The effective learning method of trial and error, 2018. URL http://www.self-learner.com/effective-learning-method-of-trial-and-error/. [Online; accessed April 19, 2018].

[15] Fiona Rolander. User onboarding, seminar presented at Bontouch, Stockholm, February 2018.

[16] SlashGear. Setting up apple's homepod, 2018. URL https://www.youtube.com/watch?v=9IzMhwUh7Rk. [Online; accessed April 5, 2018].

[17] DHTV Dan. Google home mini - unboxing, setup & review 2018 pros & cons, 2018. URL https://www.youtube.com/watch?v=UZlrXeBJLjI. [Online; accessed April 5, 2018].

[18] Jordan Keyes. All-new amazon echo (2017, 2nd generation) unboxing and first impressions, 2017. URL https://www.youtube.com/watch?v=Qg4H9YDFdh8&t=405s. [Online; accessed April 5, 2018].

[19] Michael Harris Cohen, James P Giangola, and Jennifer Balogh. *Voice user interface design*. Addison-Wesley Professional, 2004.

[20] John Lyons. *Natural Language and Universal Grammar: Essays in Linguistic Theory*, volume 1. Cambridge University Press, 1991. doi: 10.1017/CBO9781139165877.

[21] Google. Be cooperative...like your users, 2017. URL https://developers.google.com/actions/design/be-cooperative. [Online; accessed Februari 11, 2018].

[22] H. Bowe and K. Martin. *Communication Across Cultures: Mutual Understanding in a Global World*. Cambridge University Press, 2007. ISBN 9780521695572. URL https://books.google.se/books?id=Ds-MdFG2hlMC. [Online; accessed Februari 11, 2018].

[23] Filip Radlinski and Nick Craswell. A theoretical framework for conversational search. pages 117–126, 2017. doi: 10.1145/3020165.3020183. URL http://doi.acm.org/10.1145/3020165.3020183. [Online; accessed Februari 11, 2018].

[24] Matthew B Hoy. Alexa, siri, cortana, and more: An introduction to voice assistants. *Medical reference services quarterly*, 37(1):81–88, 2018.

[25] Johann Hauswald, Michael A Laurenzano, Yunqi Zhang, Cheng Li, Austin Rovinski, Arjun Khurana, Ronald G Dreslinski, Trevor Mudge, Vinicius Petrucci, Lingjia Tang, et al. Sirius: An open end-to-end voice and vision personal assistant and its implications for future warehouse scale computers. In *ACM SIGPLAN Notices*, volume 50, pages 223–238. ACM, 2015.

[26] Donald Norman. *The design of everyday things: Revised and expanded edition*. Basic Books, 2013.

[27] Nicole Yankelovich. How do users know what to say? *interactions*, 3(6): 32–43, 1996.

[28] Eric Corbett and Astrid Weber. What can i say?: addressing user experience challenges of a mobile voice user interface for accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 72–82. ACM, 2016.

[29] Dirk Schnelle, Fernando Lyardet, and Tao Wei. Audio navigation patterns. In *EuroPLoP*, pages 237–260, 2005.

[30] Zeljko Obrenovic and Dusan Starcevic. Modeling multimodal human-computer interaction. *Computer*, 37(9):65–72, 2004.

[31] Google. Google assistant, 2018. URL https://assistant.google.com/platforms/phones/. [Online; accessed May 15, 2018].

[32] Talya Bauer and Berrin Erdogan. Organizational socialization: The effective onboarding of new employees. 3:51–64, 01 2011.

[33] Amazon. Echo (2nd generation) - charcoal fabric, 2018. URL https://www.amazon.ca/Echo-2nd-Generation-Charcoal-Fabric/dp/B0749YXKZ1?th=1. [Online; accessed April 5, 2018].

[34] Apple. Welcome homepod., 2018. URL https://www.apple.com/homepod/. [Online; accessed April 5, 2018].

[35] Google. Google home, 2018. URL https://store.google.com/gb/product/google_home?hl=en-GB. [Online; accessed April 5, 2018].

[36] Amazon. Amazon alexa app basics, 2018. URL https://www.amazon.com/gp/help/customer/display.html?nodeId=201549940. [Online; accessed April 5, 2018].

[37] Apple Inc. Set up homepod, 2018. URL https://support.apple.com/en-us/HT208241. [Online; accessed April 5, 2018].

[38] Christian Rohrer. When to use which user-experience research methods, 2014. URL https://www.nngroup.com/articles/which-ux-research-methods/. [Online; accessed May 21, 2018].

[39] Ditte Mortensen. Best practices for qualitative user research, 2018. URL https://www.interaction-design.org/literature/article/best-practices-for-qualitative-user-research. [Online; accessed May 21, 2018].

[40] Erika Hall and Jeffrey Zeldman. *Just enough research*. A Book Apart, 2013. ISBN 978-1937557102.

[41] Jiannan Li and Sutapa Dey. Wizard of oz in human computer interaction, 2013. URL http://pages.cpsc.ucalgary.ca/~jiannali/course/wizard_of_oz.html. [Online; accessed April 16, 2018].

[42] Jakob Nielsen. Why you only need to test with 5 users, 2000. URL https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/. [Online; accessed May 21, 2018].

[43] Kent State University. Independent samples t test, 2018. URL https://libguides.library.kent.edu/SPSS/IndependentTTest. [Online; accessed May 21, 2018].

Appendices

Appendix A

Interview Guide

Tack för att du ställer upp och tar dig tid. Jag Filip Eriksson och som du kanske vet skriver mitt examensarbete om intelligenta personliga assistenter och onboarding. Den här intervjun görs med dig för att jag vill veta mer om hur en användare använder sig av röststyrning, vilka problem som finns, vad den har för mervärde för dig. Jag kommer också ställa lite frågor om vilka typer av kommandon du använder och om du minns hur du lärde dig att använda assistenten första gången. Det sista är om det är okej att jag spelar in intervjun? Inspelningen kommer endast att användas för studien, och att det blir lättare för mig att få med det viktigaste från intervjun, vad som sägs och för att jag inte ska behöva skriva ner allt som sägs vilket även gör att intervjun går fortare. Är det något du inte vill svara på är det självklart okej, det är även okej att när som helts avbryta intervjun.

1. Berätta om vilka röstassistenter (Alexa, Google Assistant, Siri, Cortana) du har använt och på vilka plattformar du använt dessa.
2. Vilken platform lärde du dig använda röstassistenten på först? (Mobil före IPA etc)
3. Vad använder du dina olika assistenter till?
4. Hur skiljer sig din användning av dem?
5. Var använder du den/dom?
6. Vad brukar du säga till assistenten?
7. Brukar du ofta behöva upprepa vad du säger till den?
8. Hur vet du vad du kan säga till assistenten?
9. Hur tror du att du har lärt dig att använda assistenten?
10. Hur fungerade det när du lärde använda dig av assistenten i början?
11. Vad brukar du göra för att lära dig fler saker assistenten gör?

12. Hur brukar du gå till väga för att formulera en fråga eller ett kommando till röstassistenten?
13. Har du någon gång behövt tänkta till lite extra om hur du formulerar dig?
14. Hur känns det att prata med en röstassistent, hur mycket konversation tycker du att det är? kontra ge ett kommando och få ett svar.
15. Vad är din inställning till att prata med en röstassistent?
16. Är det någon situation du väljer bort det?
17. Hur var din bild av röstassistenter innan du började använda dem?
18. Vad är din bild just nu?
19. Varför köpte du din Google Assistent? Vad gör den värd att använda?
20. Har du något exempel på när du har blivit imponerad över hur bra röstassistenten är?
21. Har du något exempel på när du har blivit besviken på hur röstassistenten fungerar?

Appendix B

Voice Script

VOICE ONBOARDING - SCRIPT