UMEÅ UNIVERSITY

Chatbot or voice assistant in a help desk application?

A study of users' experiences and preferences

Christina Metcalfe

M.Sc Interaction Technology & Design, 300 credits

Master thesis 30 credits
Department of Applied Physics & Electronics
Spring 2021

Supervisor: Mattias Andersson

External supervisor: Victor Edsund

Examiner: Thomas Mejtoft

Umeå University, Sweden

Abstract

Companies across a wide range of business areas are working hard to fulfill users wishes to speak to digital voice assistants. The trend of replacing chatbots in favour for voice assistants carries a risk of companies not considering which applications will actually benefit from getting a voice user interface (VUI) resulting in poor user experience.

This thesis aims to investigate which help desk support task will benefit from being implemented in a VUI. By following the Service Design methodology, research on the topic has been conducted and a prototype has been build and tested on a target audience. The results from a user study were evaluated and conclusions have been drawn about which tasks are best suited for being handled by a digital voice assistant.

Two kinds of help desk tasks were evaluated in a user study to compared users experience of the current text based digital assistant with a prototype of a voice based assistant. The aim of the user study was to find which task would benefit from becoming voice based by looking at users acceptance level and over all experience.

The results from the user study showed that employees who use the current text based assistant for help desk tasks, will not choose to speak to a digital voice assistant because they are happy with the service available today. However, employees who don't use the current text based assistant, will find the digital voice assistant useful. It was also found that short executing tasks such as unlocking accounts, are a better fit for the VUI compared to longer interactions providing information. Two conclusions were drawn, peoples' preferences are different, meaning that it should be possible to interact with both a text based and voice based assistant when performing help desk tasks. Secondly, the voice based assistant should be implemented as a function in the help desk phone queue instead of being implemented in a browser. Because the users argued that they would be more comfortable speaking to a phone then to a screen.

Chatbot eller röstassistent i en kundtjänst applikation. En studie om användarnas upplevelser och preferenser

Sammanfattning

Företag i alla branscher jobbar hårt för att uppfylla sina användares önskan om att interagera med röstassistenter. Trenden att byta en chattbot mot en röstassistent medför en risk att företag inte tar hänsyn till huruvida en tjänst faktiskt drar nytta av att göras om till ett röstbaserat användargränssnitt, vilket kan resultera i en försämrad användarupplevelse.

Denna uppsats undersöker vilka funktioner i en kundtjänst som skulle gagnas av att implementeras i ett röstbaserat användargränssnitt. Genom att använda Service Design modellens forsknings- och idé-genererings fas har en röstbaserad prototyp tagits fram och testats på målgruppen. Resultaten från användarstudien har utvärderats och slutsatser har formulerats.

Två typer av kundtjänstfunktioner har undersökts i en användarstudie som jämfört användarnas upplevelse av den befintliga chatbotten och en röstassistentsprototyp. Målet med användarstudien var att definiera vilka kundtjänstfunktioner som skulle gynnas av att bli röstbaserade genom att titta på användarnas acceptansnivå och övergripande upplevelse.

Resultaten visar att, användare som idag använder, och är nöjda med, chatbotten förmodligen inte kommer att ersätta denna med röstassistenten. Samtidigt som användaren som idag inte använder chatbotten kan tänka sig att använda röstassistenten istället för att ringa till kundtjänsten.

En annan upptäckt från användarstudien var att funktioner som utför en uppgift, så som att låsa upp ett konto, passar bättre i ett röstbaserat sammanhang i jämförelse med när längre information ska förmedlas.

Slutligen formulerades två slutsatser. För det första, olika personer har olika preferenser, det borde alltså vara möjligt att interagera med både chatbotten och röstassistenten för kundtjänstärenden. För det andra, röstassistenten borde implementeras som en plugin som användaren kan utnyttja när denne sitter i telefonkön till kundtjänsten snarare än en egen funktion i på den befintliga hemsida. Detta på grund av att användarna uttryckte att det är mer bekväma med att prata i telefon snarare än till en skärm.

Acknowledgements

First of all, the author would like to thank Victor Edsund at SEB for being a great supervisor in these strange times of working from home. Always being available in the Teams chat and providing support and guidance throughout the spring. A special thanks goes to Greta Krivickaite and Erik Svensson at SEB for providing relevant data and sharing their knowledge on the topic of Aida and writing dialogue.

Secondly, thanks to Mattias Andersson the supervisor from Umeå University, for the weekly meetings and for reading the thesis almost as many times as the author has done herself.

Thirdly, special thanks to Thomas Lindqvist at SEB for providing the contact details needed to get in touch with SEB in the first place, this thesis would not have been written without you. And to Dr. Margaret Metcalfe, thank you for helping the author improve her written language and for the inspiring pep-talks though out the process of writing such a long piece.

Last, but not in any way least, thank you Lisa Fjellstörm, Evelina Malmqvist and Amanda Engström for the weekly meetings, thanks for the constant support and encouraging words. This spring would have been a lot tougher and longer without you.

Glossary

- **AI - Artificial Intelligence**: the study of machines that have similar qualities of the human mind, such as the ability to solve problems, understand language, and learn.
- **Aida**, also referred to as **textAida** in this thesis: SEBs digital assistant, a web based chatbot that can handle IT and HR related questions.
- **ASR - Automatic Speech Recognition**: an independent, machine-based process of decoding and transcribing speech used together with NLP in IPAs
- **Computational intelligence**: the theory, of biologically and linguistically motivated computational paradigms.
- **CUI - Conversational User Interface**: a user interface for computers that imitates a conversation with a human being. Providing opportunity for users to communicate with computers in natural language rather than in syntax specific commands.
- **Conversational structure**: methods used by speakers to structure conversation efficiently and manage turn taking. The purposes of arranging conversational structure are keeping the flow of conversation and avoiding 'overlap'.
- **Digital assistant**: an advanced computer program that simulates a conversation with the people. Is sometimes referred to as ”chatbot”.
- **Discoverability**: the degree of ease with which a user can find all elements and features when they first encounter a system. It is an important aspect to consider in user interface and user experience design for hardware devices, software applications and websites.
- **Entity**: the type of information that is extracted from user input. One of the parameters in Dialogflow.
- **GUI - Graphical User Interface**: a user interface for hardware devices, software applications and webpages that includes graphical elements, such as images, icons and buttons.
- **Hi-fi prototype - High-fidelity prototype**: a computer-based interactive representation of the product or a close resemblance to the final design in terms of details and functionality.
- **Intent**: categorizes a user's intention for one conversation turn. One of the parameters in Dialogflow.
- **IPA - Intelligent Personal Assistants**: a software agent that can perform tasks or services for a person, based on questions or commands.

- **IVR - Interactive Voice Response**: an automated phone system technology that allows incoming callers to access information via a voice response system of pre recorded messages. Is allows users to get information without having to speak to an agent. The menu options are utilised via touch tone keypad or speech recognition to have the call routed to specific departments or specialist agent.

- **Learning effect** (in an online controlled experiment): a positive or negative effect from an intervention that only becomes pronounced after a certain time has passed. The effect might be gradually increasing over time or there might a more drastic level shift after a certain amount of time.

- **Learnability**: a quality of interfaces that allows users to quickly become familiar with them and able to make good use of all features. Learnability is a component of usability and is often heard in the context of user interface or user experience design.

- **Lo-fi prototype - Low-fidelity prototype**: a quick and easy way to translate high-level design concepts into testable prototypes. The most important role of lo-fi prototypes is to check and test functionality rather than the visual appearance of a product. An example of a lo-fi prototype is a paper prototype.

- **Natural Language**: refers to the way humans communicate with each other. Namely, speech and text.

- **NLP - Natural Language Processing**: a collective term referring to automatic computational processing of human languages. This includes both algorithms that take human-produced text as input, and algorithms that produce natural sounding text as outputs.

- **NLU - Natural Language Understanding**: a branch of AI that uses computer software to understand input made in the form of sentences in text or speech. NLU enables computers to understand commands without the formalized syntax of computer languages and for computers to communicate back to humans in natural language.

- **UI - User Interface**: the means in which a person controls a hardware device or software application. A good user interface provides a user-friendly experience, allowing the user to interact with the hardware or software in an intuitive way.

- **UX - User Experience**: is the overall experience a user has with a product or service. In the Usability field, this experience is usually defined in terms of ease-of-use. However, the experience encompasses more than only function and flow, but the understanding compiled through all of the user's senses.

- **VoiceAida**: the voice based prototype of SEBs digital assistant Aida, created as a Dialogflow prototype for the user study in this project.

- **Voice assistant**: a voice based piece of software that can supply information and perform certain types of task.

- **VUI - Voice User Interface**: allows the user to interact with a system through speech commands. The primary advantage of a VUI is that it allows for a hands-free or eyes-free way that users can interact with a product while focusing their attention on somethings else.

Contents

List of Figures

List of Tables

1 Introduction

The idea of interacting with a computer through spoken natural language was first introduced to the public by Stanley Kubrick in 1968. His blockbuster film 2001: A space odyssey [12] features H.A.L 9000 who “assists” astronauts in their quest to find the origin of a mysterious artifact found on the moon. This cinematic experience of what human-computer communication could look like inspired many filmmakers to continue exploring the seemingly endless possibilities of human-computer interaction. In the film Iron Man (2008) [13] Tony Stark uses his supercomputer assistant, JARVIS, both to help in his everyday life and on his many adventures. The use of a personalized digital assistant was explored further in Spike Jonze’s film Her (2013) [14] in which a lonely writer falls in love with an operating system designed to meet his every need. Filmmakers are not the only ones inspired by the idea of intelligent personal assistants; researchers and developers have in the last decades worked on teaching computers to understand and communicate through spoken language.

When Bell Labs built the first single-speaker digit recognition system in the 1950’s the system was of little use outside the lab, but the ideas of what computers could offer, had sparked [15, p. 1]. The vision was that users could communicate with a computer through natural language and therefore not have to learn any specific language or prompts [16]. However, it turned out to be quite complex to understand spoken language. It was believed by many that only entities (human beings) living in the real world could effectively understand language, arguing that without context, the meaning of a word is impossible to understand [15, p. 1]. Research and development continued and by the 1990’s the first speaker-independent system hit the market, enabling anyone (who spoke English) to talk to it [15, p. 1]. In the early 2000’s, Interactive Voice Response (IVR) systems were capable of understanding and acting on commands given through natural language. Anyone (speaking English in the USA) could use a telephone (connected to the land line) to call the IVR to get stock quotes, book flights or get traffic information [15, p. 1]. IVR systems have in the last decade transformed and are today integrated in mobile apps, smart speakers and wearable technology such as smart watches and google glasses. The most popular tasks users assign to IVR systems today are, asking general questions, asking for directions and streaming music [17]. With this transformation, the term Voice User Interface (VUI) was coined by a few of the big tech companies to describe the interface of the new voice controlled technology.

Films such as 2001: A space odyssey, Iron Man and Her have set a high standard for what users expect of Intelligent Personal Assistants (IPAs) and today’s technology is not capable of handling language in such a way yet. Several studies have shown that first time users build an incorrect mental model of what the IPA can and cannot do and most importantly, how they should interact with it [16]. This incorrect mental model tends to have a negative effect on the use of the system and often results in use ceasing [16].

De Boer [18] points out that humans (and their predecessors Homo heidelbergensis) seem to have had 400,000 years worth of developing language and speech. Pearl [15] asks users to be patient with computers, since they have only been around for five decades and so they might need some time to get the hang of spoken language. Computational intelligence has grown exponentially over the last decades [19] and different researchers have predicted that artificial intelligence (AI) will reach the singularity[1] by 2025 [19].

However, if the interface through which AI communicates continues to be poorly designed, it won't matter if it reaches the singularity. Speech has a temporal nature and is demanding on the user's memory [20] and it is not possible to "design away" the high impact on a user's attention. There are several parameters that need thought when designing VUIs and in this thesis, two aspects will be covered: firstly, identifying appropriate tasks to ensure that they are beneficial to users. Secondly, adjusting conversational structures to fit the spoken conversation format.

1.1 Problem statement

Companies across a wide range of business areas have in the last few years replaced their text-based chatbots in customer support applications with voice based smart assistants. A mistake made by many companies implementing the use of voice assistants is to assume that a voice assistant is a speaking chatbot, resulting in them hiring a voice actor to read the existing chatbot responses and adding them to the existing system. However, written language and spoken language are structured differently and this becomes painfully clear when interacting with a poorly designed voice assistant. An example taken form the authors own experience was when calling a carrier service about a missing parcel delivery, the digital voice assistant asked for the 15 digit long tracking-ID. The letter and number recognition was poor, resulting in a lot of frustration and finally ending the call and visiting the carriers web page to type the tracking ID instead. When upgrading from a text-based chatbot to a voice assistant it is important to evaluate which functionalities are worth keeping as they are and which need to be redesigned; information cannot be presented through audio in the same way as it is presented on a screen. The conversational structures need to be customized to fit the spoken language. Conversation design is a fairly new concept in the area of User Experience (UX) but is becoming more and more important as the use of voice assistants continues to grow in customer support applications [21].

[1] When a computer's intelligence outgrows the intelligence of the human brain

1.2 Objective

This thesis aims to investigate which kind of help desk functionalities, would benefit from being implemented in a VUI. In turn the objective has been rephrased in the following four research questions:

1. How do SEB employees use the current digital assistant, Aida?
2. Which are Aida's most frequently used processes?
3. Which of these processes will, based on the current research, fit a VUI?
4. How do SEB employees experience interacting through voice commands?

1.3 SEB

This work was carried out in collaboration with SEB (Skandinaviska Enskilda Banken AB) during the spring of 2021. SEB have an interest in investigating how their existing intelligent personal assistant (IPA), called Aida, can be equipped with a VUI to improve the user experience for their employees. SEB was founded in 1856 as Stockholm's first private bank and one of the first commercial banks in Sweden [22]. Today (spring 2021) SEB have approximately 15 000 employees around the world with the shared vision of "delivering world-class services to their customers" [23]. In order to deliver according to their vision SEB work around four core statements; Customer first, Commitment, Simplicity and Collaboration. They want to play an active part in developing a society built on strong customer relationships by offering financial services to both corporate and private customers [23].

SEB have previously participated in master's thesis work on the topic of voice interaction where design guidelines for increasing the trustworthiness when paying invoices through a VUI were produced. Lundqvist's [8] thesis can be found on Diva.

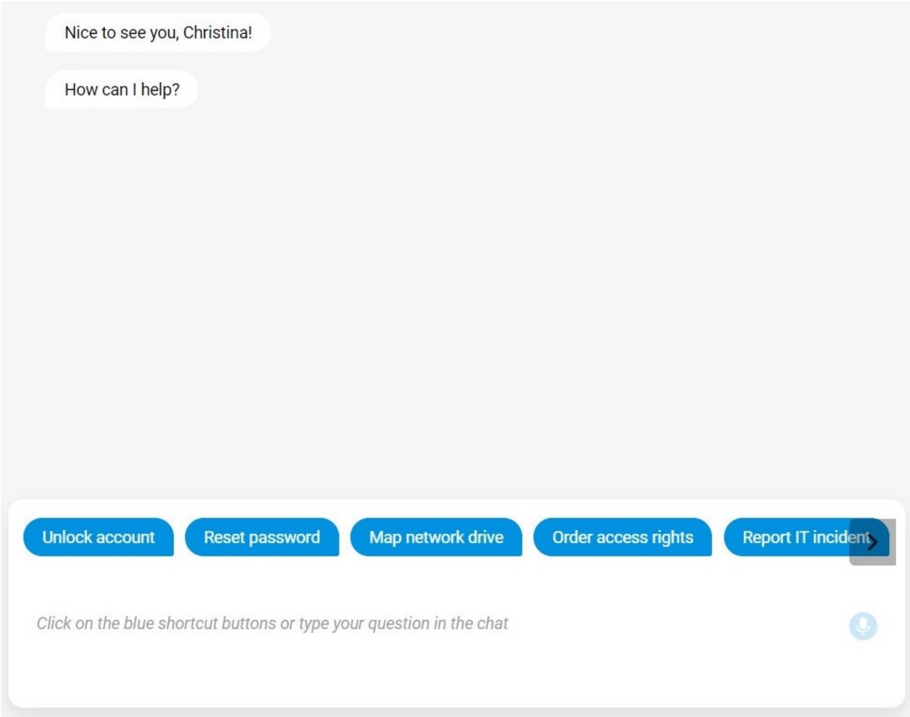
In order to keep delivering world-class services to their customers SEB needs to develop voice integrated services. This project aims to be one of the first stepping stones in that transition.

1.3.1 Aida

Aida is SEB's text based digital assistant, shown in figure 1. Aida is an automated service agent aiming to handle simpler tasks and relieve the help desk personnel from tedious monotonous tasks. Today there are two versions of Aida, an external and an internal. The external version provides guidance to SEBs customers. The internal version provides services for SEBs employees on questions related to HR and IT support. This study will focus on the internal version of Aida. Aida is available to all SEB employees regardless of position.

Aida can execute actions or provide information for three things,

- Providing guidance on HR and IT related questions [2], either by offering long answers or by redirecting to other pages and suggesting where additional information can be found.
- Changing user information: unlocking accounts, updating passwords and phone numbers.
- Viewing user information: employee codes, vacation days, services ownership.



Nice to see you, Christina!

How can I help?

Click on the blue shortcut buttons or type your question in the chat

Figure 1: SEBs digital assistant Aida is a chat application.

[2] Examples of these will be given later in the thesis

1.4 Delimitation

This research is focused on converting functionalities in an existing chatbot to voice based commands. The intelligent assistant created for this study is based on SEBs current digital assistant, and communicates strictly through speech. Security, privacy, log-in and speaker verification functionalities are not part of this research. It has been assumed that the technology used when implementing these design suggestions (prototype, sample dialogues, conversational structures) has a sufficient quality and accuracy, so problems regarding understanding and hearing the users have not been addressed.

2 Theory

This section aims to conceptualize the three cornerstones of the theory behind conversation design and the tools that have been used while working with this thesis. This theory section starts by describing the first cornerstone, user experience and how it relates to usability. It then goes on to describing what an intelligent personal agent is and why personas are important when designing voice applications. The second cornerstone, interface design, is briefly introduced together with the challenge that the invisible nature of speech inflicts on VUI design. The third cornerstone consists of tools and concepts that have been created to resolve the previously mentioned challenges. The last part of the theory section introduces the well known design model, Service Design and specifically looks at within-group tests and the evaluation tool System Usability Scale.

2.1 User experience (UX)

User experience, often referred to as UX, is a term used when describing a user's experience of a service, product, application or system. UX has many different application areas, one example is a website where the users' experience is dependent on how useful the interface is to complete a given task. While aesthetics are important, there is more to UX than ”something looking nice”. Another term frequently used in this field is Usability, ISO defines usability as; ”extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.” [24]. Usability and UX go hand in hand, aiming to make tasks easier for the user by making the interaction with a product as simple and effective as possible. Designing for usability is as much about user-friendliness, e.g. simple and logical systems, as it is about benefiting the user, e.g. providing the user with sufficient information in order to fulfil a particular task.

There are many methods for evaluating and measuring UX. Research has shown that the first step in securing a high level of user experience is to get the user's acceptance of the system [25]. In 1967 Ajzen and Fishbein created the Theory of Reasoned Action (TRA) [26] to determine the relationship between attitudes and behavior within human action. In 1989, Davis extended the TRA model and created the Technology Acceptance Model (TAM) [25] aiming to measure user's acceptance of digital systems. Davis said that when designing a product or service, it is vital to consider *perceived usefulness* and *perceived ease of use.*

2.1.1 Perceived usefulness and perceived ease of use

People are more likely to use an application which they believe will improve their performance on a given task [25]. It is more about what the user believes the product can be used for, rather than what the product actually can be used for. Davis defines *perceived usefulness* as “the degree to which a person believes that using a particular system would enhance his or her job performance” [25]. This is based on the definition of the adjective useful being “able to be used advantageously” [27]. In a work environment good performance is often rewarded with a pay raise or a promotion motivating the user to perform well [25].

Regardless of how useful an application might seem, a user is less likely to use it, if she believes it to be too hard to use and therefore assumes that the performance benefits are outweighed by the efforts required to get the task done. Davis defines *perceived ease of use* as “the degree to which a person believes that using a particular system would be free of effort” [25]. This is based on the definition of the noun ease being “freedom of difficulty or effort” [8]. An application needs to have a high perceived ease of use as well as perceived usability to ensure user’s acceptance [25].

Lundqvist [8] found that a user might overlook some level of poor usability if an application provides a critical function. She goes on and points out that there are many reasons for a user not to use an application. One example discussed is complexity - if an application is too complicated to use, the user might not choose to use it, however useful it could be. A second example discussed is that of an application missing a useful function, which reduces motivation to use the application. These examples show that the challenge of designing for user experience goes well past designing “something that looks nice” [8].

2.1.2 User experience in the context of voice applications

A questions researchers have discussed for many decades could be stated as: ”How do human beings want to speak to machines and computers? Do they expect it to be similar to the ways that they speak to each other?” Studies show that people tend to talk slowly and articulate clearly when talking to voice assistants, in some cases they even change their choice of words in order to better be understood [20]. Therefore some researchers argue that conversation design shouldn’t be based on the conversational structures of spoken language [20]. On the other hand, big tech companies [3], are convinced that humans don’t want to speak “computer” when interacting with a voice assistant and are pushing technology forward to allow computers to understand natural language [28, 29].

Kamm [20] states that a successful interaction, regardless of the mediums in-

[3] Such as Apple, Amazon and Google.

volved, is one where the task at hand is accomplished in an easy and efficient way from the users point of view. Designing VUIs and building conversations inspired by natural language is what the big tech companies are doing and consider to be a good way to ensure good user experience and successful interaction between a human being and a voice assistant [21].

2.2 User Interface

Users interact with digital devices through an interface, usually referred to as User Interface (UI) or Graphical User Interface (GUI). A GUI consists of a combination of pictures, graphical elements, colours, brightness and fonts, all related to the website or digital system. The aim of a GUI is to capture the user's interest and act as a guide through the application by providing the right amount of information at the right time. In other words GUI promotes an experience and improves the flow of information [30].

2.3 Voice User Interface (VUI)

Communicating through speech is something that comes naturally to people and as technology becomes smarter and more integrated in our daily life it seems right to use speech as an input. The aim of designing a VUI is to make the interaction between human beings and computers smoother and more efficient compared to the classic GUI implementations [28]. Kamm, and many others, argue that using speech as a tool for interaction is more natural than pressing, dragging and tapping on a screen or keyboard [20]. It is assumed that a computer that communicates in a natural manner, though natural language, is beneficial because it can take advantage of human communication skills and thus create an efficient way for information to be transferred and interpreted [20]. However, human speech is unpredictable and what the user says, is not always what the user means. There are many reasons for this and a few examples are the many different ways to express an opinion or formulate a question and the language use by a person is influenced by their surroundings and situation. VUI designers need to keep these aspects in mind while designing to cover as many dialogues as possible.

In order for speech based applications to be successful they need to be beneficial to users [31]. An application needs to be easier to use with speech for users to choose it over traditional interaction frameworks [28]. Kamm suggests three key factors that need to be considered when designing a VUI [20].

- The information requirements of the task.
- The limitations and capabilities of the voice technology.

- The expectations, expertise, and preferences of the user.

In order to gain a good result, the VUI needs to be an integral and early component in the overall iterative design process of the product or service [20].

Pearl [15, p. 3] lists a number of of advantages of using a VUI; hands-free interaction, intuitiveness, a sense of empathy, and speed. With voice interaction users can multitask when they have their hands full, like asking for a recipe while cooking or send a text while driving. A study performed by a research group at Stanford University showed that speech-to-text technology is faster than typing written messages by hand [32]. Other reasons for developing voice based applications are users wish to multitasking and interact with a device hands-free and that there are many devices with small screens (e.g. wearable technology like smart watches) or even no screens (e.g. smart speakers, such as Google Home). Studies have shown that speech is the preferred way to interact with these small screens [15, p. 4-5].

There are certain situations when voice based interaction can be a disadvantage, an example is that users are uncomfortable talking to a device in a public place. Studies have shown that it is the audible response from a voice assistant that is the most frequent reason why users decide not to use it in public. Users are afraid that private information might be overheard [15, p. 5].

Mapping where, when, why and how users use voice assistants are important aspects to investigate and evaluate when designing VUIs. Since there are no visual components to a VUI, tools such as the law of proximity or visual perception used by designers of GUIs are not applicable to VUI development [33]. Companies like Google have created new guidelines [34] and universities around the world are changing their curricular for degrees in Design to include the challenges VUI designers face [35]. One challenge is the fact that speech is invisible and transient, making editing and reviewing past commands difficult [36, 37].

2.4 Intelligent personal assistants (IPA)

An intelligent personal assistant (IPA) is an application with the ability to provide assistance through user input such as audio (voice), visual (image) and contextual information [38]. The tasks vary from answering questions and making recommendations through natural language, to performing actions like turning lights on and off [38]. An IPA consists of multiple subsystems, where different entities handle different requests. Take registering user input as an example, the subsystems Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) are run to translate the acoustic input into information a computer can process [39]. In turn the processed information that consists of the coded acoustic input, allows the system to execute tasks [39]. The outputs used in today's IPAs are synthetic voices based on neural networks and artificial intelligence and not prerecorded prompts like the ones used in phone-based assistants [20].

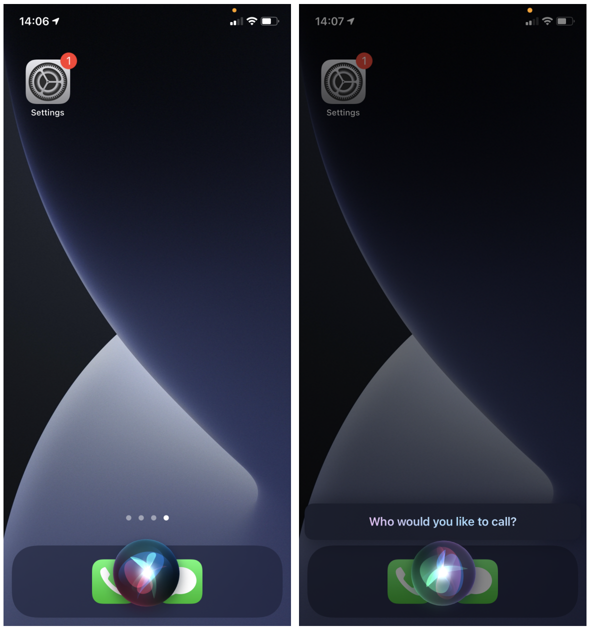
Advances in neural networks, deep learning and artificial intelligence enable the use of IPAs in consumer products such as the Google Home, figure and Apples Siri shown in figures 2 and 3. Many of these smart assistants interact with users in a "questions-answer" format with a strong task orientation, rather than being social or conversational [40]. When communicating in speech humans infer personality traits and social information from the voice they hear and start to build a mental image of the person they are talking to [2, p 75]. Interacting with an IPA is no different and from a design point of view this is why it is important to use a persona when building conversations [2, p 77].

Figure 2: A Google Home is an IPA without a visual interface.

2.4.1 VUI Persona

In conversation, people make assumptions about the personality of their conversation partner based on linguistic cues and vocabulary. It has been found that this occurs even with an automated conversation partner [2, p 75]. Creating a VUI persona is a tool used early in the design process, to ensure that the IPA perceived personality fits the organisations brand. Many VUI designers agree that there is no such thing as VUI without a personality [2,40, p 75]. Chohen et al. [2, p 77] define a persona as "a rough equivalent of a character, as in character in a book or film". In the world of VUIs the conversation style, intonation and pitch are decided based on the VUI persona. Cohen et al. [2] continues by saying "Take every chance you get to steer the users opinion towards what you want them to think, don't leave it to chance."

A well designed VUI persona should strengthen the users mental model of the system. The users interaction behaviours are heavily influenced by their mental model of the IPA [40, 41].



Who would you like to call?

Figure 3: Siri is an IPA with a visual interface.

2.5 Spoken language

Spoken language differs from written language in many ways, and keeping them apart when writing sample dialogue takes practice [42]. When grammatical errors occur they have a negative impact on the user's cognition [43]. When reading written sentences out loud they can sound overly polite and stilted. It is therefore useful to read sample dialogue out loud and not just write it down. The following section will provide guidance on how to phrase an appropriate VUI dialogue.

To assist developers, Google have put together language guidelines for VUI design [44]:

Focus on the user: The dialogue should be user-centred. This makes it easier to stay on track and makes the conversation crisp and to-the-point.

Skip the long monologues: Design the responses to be informative, but concise. Let the users have their turn in the conversation and stay away from providing too much information, if the users hasn't specifically asked for it. Consider the examples in figures 4 and 5.

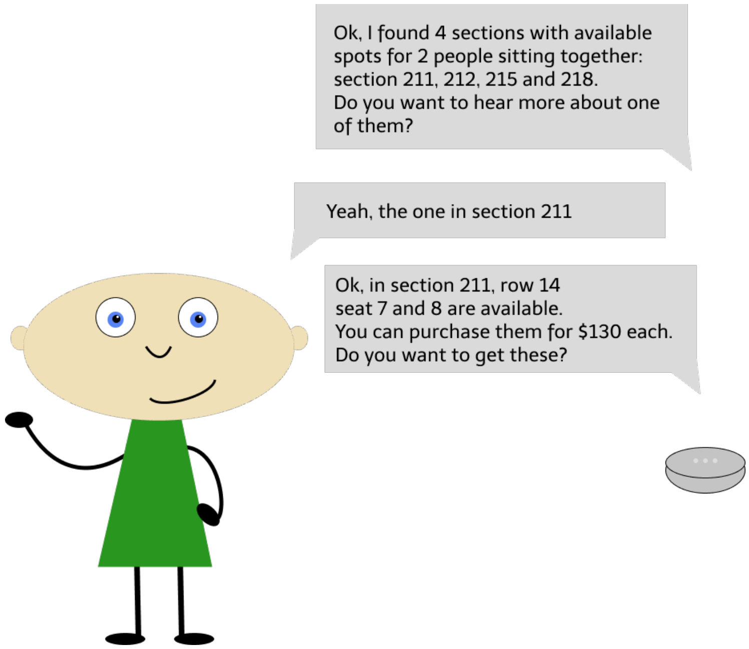

Ok, I found 4 sections with available spots for 2 people sitting together: section 211, 212, 215 and 218. Do you want to hear more about one of them?
Yeah, the one in section 211
Ok, in section 211, row 14 seat 7 and 8 are available. You can purchase them for $130 each. Do you want to get these?

Figure 4: Informative and concise responses the user can choose to listen to.

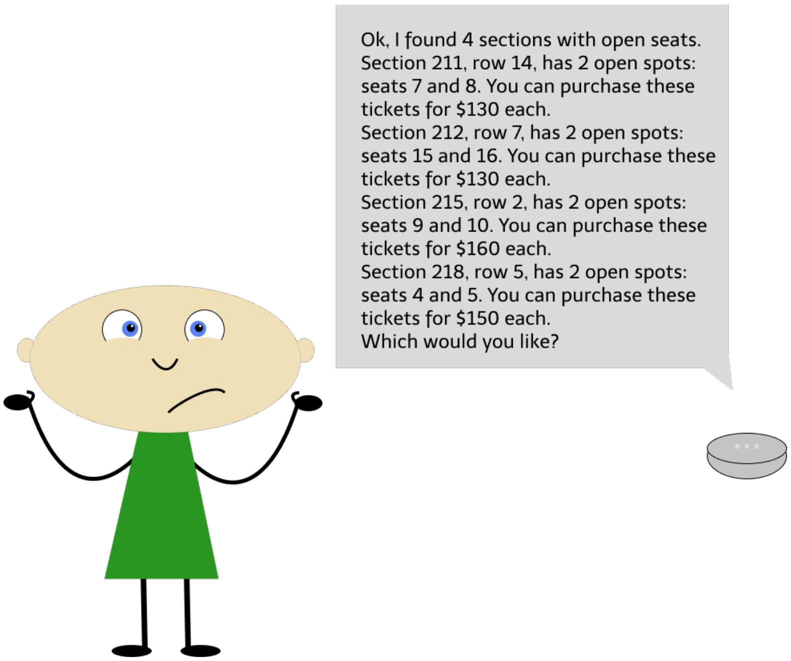

Ok, I found 4 sections with open seats. Section 211, row 14, has 2 open spots: seats 7 and 8. You can purchase these tickets for $130 each. Section 212, row 7, has 2 open spots: seats 15 and 16. You can purchase these tickets for $130 each. Section 215, row 2, has 2 open spots: seats 9 and 10. You can purchase these tickets for $160 each. Section 218, row 5, has 2 open spots: seats 4 and 5. You can purchase these tickets for $150 each. Which would you like?

Figure 5: Skip long informative monologues.

Not only would the second prompt take a long time to listen to, but the user also has no chance of remembering the different options and would be likely, either ask to hear the whole thing again or leave the conversation.

Common language without jargon and legalese. Using common terminology will most likely appeal to the broader audience making the VUI accessible to people of different backgrounds. There is a risk of eliciting mistrust and misunderstanding if the VUI uses jargon and specialist expressions.

Randomize similar responses. By randomly choosing from a number of similar responses the conversation feels more natural and human-like. A person very seldomly expresses themselves identically when asked to repeat themselves.

Skip the niceties. Overly polite and nice responses make the VUI feel formal and distant. Consider the following examples:

> **VUI:** Sure, I can help you with that. But first there's one thing you need to do: accept our terms of service.
>
> or
>
> **VUI:** Please accept the terms of service in order to proceed.

Use a friendly and familiar, but not too formal tone.

Use contractions. One doesn't say cannot or do not, one says can't and don't.

2.5.1 Conversational components

Responses are constructed of a number of conversational components, Google has put together a list of examples [45].

Acknowledgements: Okay.

Apologies: Sorry, I can't send eCards yet.

Chips: Add to cart.

Commands: Create a bouquet of yellow daisies and white tulips

Confirmations: Got it. The men's running shoes in royal blue and neon green. In what size?

Discourse markers: By the way,...

Earcons: ”welcome chime when Google Home powers on”

Endings: Anything else I can help you with right now?

Errors: Sorry, for how many?

Greetings: Welcome.

Informational statements: 42 is an abundant number because the sum of its proper divisors, 54, is greater than itself.

Questions: What kind of flowers would you like in your bouquet?

Suggestions: I can tell you more about I/O. For example, you might like to know about the keynotes, codelabs, or app reviews. I can also help you find sessions, or office hours. So, what do you want to know?

2.5.2 Cohesion

Consider the following sentence inspired by an example from Cohen et al. [2, p 137]:

Christina couldn't wait to get to England. However, she didn't stay there long.

“She” refers to “Christina”, “there” refers to “England” and “however” establishes a relationship between the proposition that Christina was eager to get to England and that the visit was short-lived.

“Cohesion is the glue of discourse” Cohen et al. [2, p 137].

Cohesion devices in dialogue design are discourse markers, special pointer words such as “this” and “that” and pronouns. Cohesion devices enable and strengthen the over all comprehension, but are not by themselves responsible for creating meaning in a sentence. They are, rather, cues that the communicating parties use to enable shared understanding within the immediate linguistic context [2, p 138].

Shiffirin defined discourse markers as “sequentially dependent elements which bracket units of talk” [46] Simply put, discourse markers connect utterances by relating what has just been said to what will be said. Quirk and Greenbaum [47] put together a list of discourse markers, a handful are listed below.

- Enumerative: first, second, third; for one thing, and for another thing, to begin with, for starters; in the first place, in the second place; one, two, three . . . ; a,b,c ...; next; then; finally; last; lastly; to conclude.
- Apposition: namely, in other words, for example, for instance, that is, that is to say.
- Inferential: else, otherwise, then, in other words, in that case
- Concessive: anyhow, anyway, besides, else, however, nevertheless, still, though, yet, in any case, at any rate, in spite of that, after all, on the other hand, all the same, admittedly.

Pronouns such as “it”, “one” and time adverbs such as “then” are used in spoken conversation to let the listener know that the current referent is the same as the previously used referent. In written conversation however pronouns sometimes refer to referents that come up later in the text. This is another significant difference between spoken and written discourse. Consider the following examples from M. Cohen et al. [2, p 138]:

1. I saw the cat the other day. It was still wandering around without a home
2. I saw a robin the other day. It was the first one I saw that spring.
3. We moved there in 1982. We didn’t even have jobs then.
4. System: You have five bookmarks. Here’s the first bookmark. . . Next bookmark. . . That was the last book mark.
   Caller: Delete a bookmark
   System: Which bookmark would you like to delete?
   . . .
   System: Do you want to delete another bookmark?
5. System: you have five bookmarks. Here’s the first one. . . Next one.. That was the last one.
   Caller: Delete bookmark.
   System: Which one would you like to delete?
   . . .
   System: Do you want to delete another one?

In (1), (2) and (3) it is easy for the listener to fill in the missing information from an earlier point in the discourse to make sense of the meaning of the pronouns. Compared to the natural flow in (1), (2) and (3) the prompt sounds stilted and formal in (4). But when the noun “bookmark” is replaced with a pronoun “one” the dialogue in (5) flows as naturally as (1), (2) and (3).

2.5.3 Universals

Universals are commands that should always be available. For universals to be effective, users need to be able to access them using a variety of words, for example help, repeat, go back, live agent, quit, good bye, cancel [2, p 72, 122]. Universals need to be callable with different commands, for example;
repeat: ’pardon’, ’repeat’, ’huh’, ’what’,
help: ’I’m not sure’, ’help’, ’what can I say’,
quit: ’I have to go’, ’bye’, ’quit’, ’get me out of here’ [48].

Universals are there to assist at any moment of the conversation, figure 6. The designers need to categorise the user’s input (their intent) in order to know what response is appropriate in each situation.

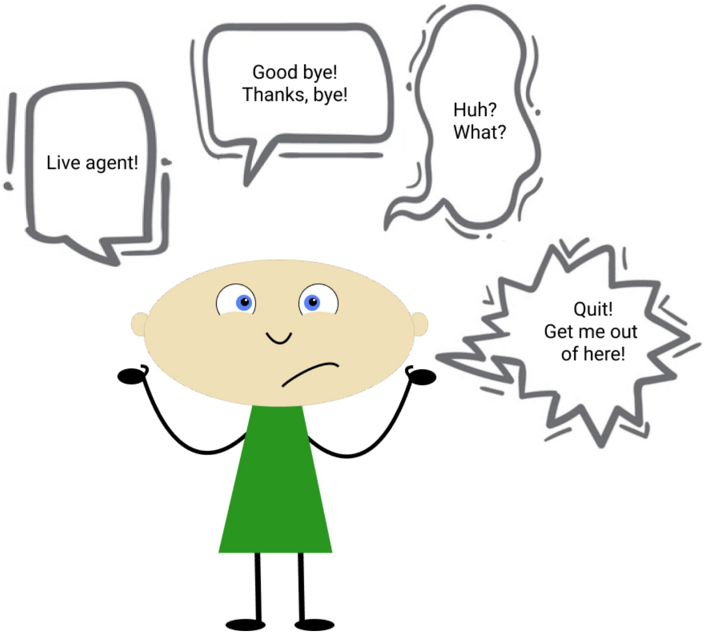


Figure 6: Universals are commands that should always be available.

The following section lists a few important aspects that need to be taken into consideration when designing a VUI.

2.6 Conversation design

Conversation- and voice user interface design is a relatively new field of research and, as this thesis is written, there is little peer-reviewed work published. Most of the information available to date is in the form of whitepapers, blog posts or conference talks by the big tech companies active in this area. To avoid biased opinions and misleading information, sources used in this thesis range from recently published material from Google and Amazon to older impartial studies performed on telephone-based conversation design.

When designing spoken conversation for synthesized voices, two things are important: designing the prompts (what the system says) and the prosody (how the system says it) [2, p 133, 171], [20].

Huang summarizes the ultimate goal of conversation design as: solving problems for people, providing value in people's lives and sparking joy and delight [1]. According to Google's design guidelines "conversation design is a design language based on human language" [34]. The more the interface resembles human-to-human interaction the less the users have to learn how to use it.

Urban, a conversation designer at Google, has put together a diagram to explain the different layers there are to understanding and designing conversations [1], as shown in figure 7.

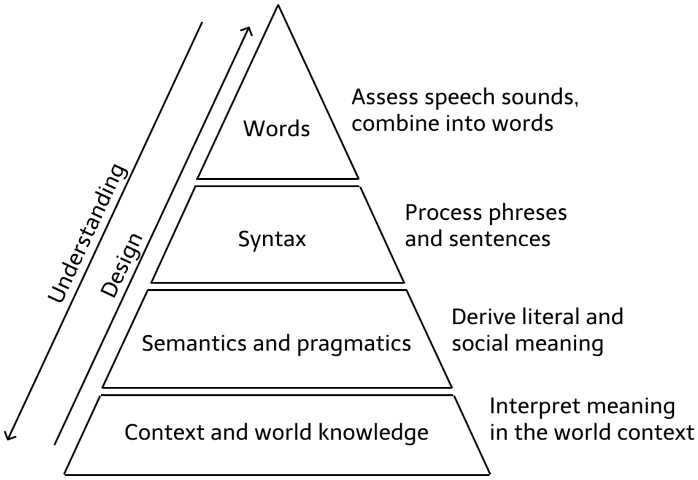

Understanding
Design
Words
Assess speech sounds, combine into words
Syntax
Process phreses and sentences
Semantics and pragmatics
Derive literal and social meaning
Context and world knowledge
Interpret meaning in the world context

Figure 7: A representation of the various layers of a conversation. Based on a model created by Urban [1].

2.6.1 Cooperative principle

In order to have a successful conversation Grice [49] has defined the Cooperative Principle. This states that the listener and speaker act cooperatively and mutually accept one another to be understood [50]. The principle can be broken down into four maxims, making it easier to define what "good communication" is. The four maxims are [50]:

- Maxim of quality - the truth of what we say.
- Maxim of quantity - the quantity of information that we provide.
- Maxim of relevance - the relevance of what we contribute.
- Maxim of manner - the way we strive to communicate clearly, without obscurity or ambiguity

When the cooperative principle is not followed "bad communication", confusion and frustration can arise between the conversing parties.

Pearl [15, p. xiii-xiv] provides a few examples of how the VUI can abuse Grice's maxims resulting in a negative impact on the user's experience.

- The maxim of quality is broken if the VUI suggests actions it cannot live up to, e.g. saying "how can I help you?", when all it can do is make a hotel reservation.
- The maxim of quantity is broken if the VUI adds unnecessary verbiage, such as "Please listen carefully, as our options may have changed." (Whoever thought, "Oh, good! Thanks for letting me know??").
- The maxim of relevance is broken if the VUI gives instructions for things that are not currently useful, such as explaining a returns policy before someone has even placed an order.
- The maxim of manner is broken if the VUI starts to use technical jargon that confuses the user.

Another example of why the cooperative principle is a good complement to the regular grammatical rules is shown in the examples presented by Giangola at Google IO 2017 between two fictional people Carla and John [50].

Carla: Do you know how to get to room 105?

John: Yes.

In this first interaction John's response is technically grammatically correct and fulfils the maxim of quality but it is possible to argue that none of the other maxims are fulfilled and that his response is not helpful or efficient.

Carla: Do you know how to get to room 105?

John: Sure, it's down the hall to your left.

In this second interaction John's response is informative and helpful to Carla and it is possible to argue that all Grice's maxims are fulfilled.

2.6.2 Conversation implicature

Giangola also discussed conversational implicature, the "shared library" of world knowledge between conversing parties. In the example below Sammy makes a statement [50].

Sammy: I really need a drink.

Alex: Have you been to the Eagle?

Alex responds with a question, which could be interpreted as rude but if there is shared knowledge of the world, provides information about the Eagle [1], [50].

Alex listens "between the lines" and picks up that Sammy is asking for suggestions of places nearby to buy an alcoholic drink rather than just stating the fact that he is thirsty [1]. Conversing is not just about what the user says, but also what the user means [1]. Humans have practiced the skill of recognizing intent for generations, computers on the other hand have not. While advances in Automatic Speech Recognition (ASR) mean that a voice assistant almost always knows what a user says, determining what a user means can still be a challenge. An utterance can be hard to understand in isolation, which is why context is so important. Understanding the meaning of a sentence is essential when providing a natural and relevant response to keep the conversation moving forward [1].

With the recent development of Natural Language Understanding (NLU) computers are starting to understand the intent in a request rather than just identifying the words used. When the user says "Alexa, what's it like outside", NLU enables the voice assistant to recognize that the user wants the weather report and not the state of the political climate or to be told what to expect when going out into a city or country environment [51].

2.6.3 Turn taking in conversation

According to the Cambridge dictionary, the definition of conversation is “a talk between two (or more) people in which thoughts, feelings, and ideas are expressed, questions are asked and answered, or news and information is exchanged” [52]. Regardless of medium, communication between humans consists of turn-taking between participants [53], as shown in figure 8. Each turn is designed to “do” something, it is the interplay of what one speaker is doing as a reaction to what the other speaker said and did in their prior turn [53]. Turn taking is also useful to reduce interruptions, minimize silences and keep track of what has been said previously [1].

There is more to a conversation than just the words being said. Body language, eye-gaze and facial expressions are only a few examples of the non-verbal signals that contribute to indicating who is in control of the turn and therefore gets to speak [1, 20]. Pauses at phrase breaks and intonation are examples of cues available in VUI design to indicate when the turn shifts.

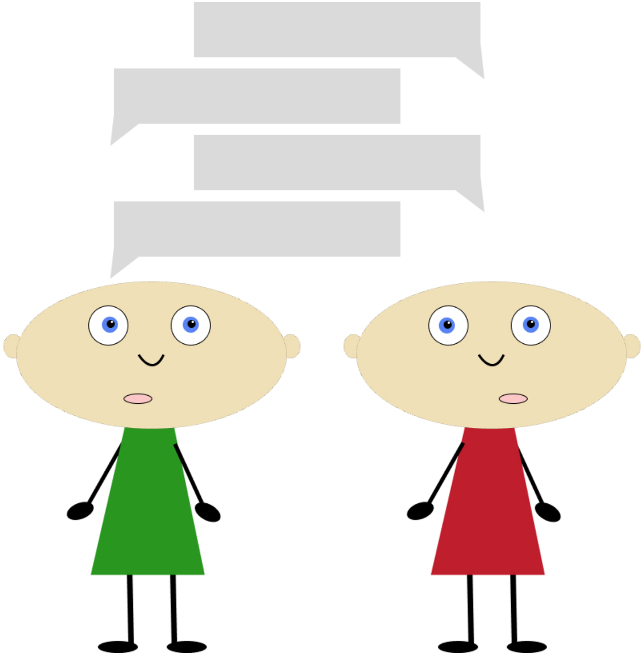


Figure 8: Turns in a conversation.

2.6.4 Context

Context helps people associate new information with something familiar and therefore lessen the cognitive load [2, p 125].

Human beings are born with the ability to understand and use context, but this ability is not something that can be replicated in computers [54]. Human beings are able to construct a reply not only based on the asked questions, but also on the surrounding situation, past interactions, who asked the question and the shared world knowledge [54].

Two approaches to understanding the role of context in conversation will be described - the conversation context and the user's physical situation.

The conversational context
Conversational context relates dialogue states to each other preventing the application from sounding stilted and awkward [2, p 108]. What this means is that answers previously given by the users should be stored and used later in the conversation if needed. When humans ask each other a yes/no question, the receiver will understand that the response is connected to the previous question. However, computers need to be programmed in order to recognise the response as part of the conversation rather than the start of a new topic [54]. Programming for conversational context makes the application feel like it has a conversational awareness and makes the flow of information feel natural.

Physiological studies have shown that people are more likely to understand and remember information if it is presented in a context [2, p 125], [55]. Consider the following passage [55];

> "The procedure is actually quite simple. First you arrange things into different groups. Of course, one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities, this is the next step; otherwise you are pretty well set."

When asked to remember as many ideas from this instruction as possible, people remember about three. However, when told beforehand that these are instructions for doing laundry, recall rates rise to about six. This shows that context helps people associate information with familiar concepts, easing the cognitive load and allowing them to recall more information [2, p 125], [55].

In their study, Dutton, Forest and Jack [56] investigated whether or not metaphors improve users' experience of a VUI. Three different systems were constructed, one with no metaphor, simply describing the merchandise through a menu. The second had a department store metaphor in which the users used a virtual lift (with sound effects) to move between floors containing different merchandise. The third system used a catalogue magazine metaphor describing the merchandise presented in pictures. The users rated the system with the department store metaphor more favourably than the magazine metaphor, which in turn was rated more favourably than the system without a metaphor, as shown in figure 9. The findings show that users are able to better navigate a system where a metaphor is present, hence, context setting through metaphors can improve user satisfaction and effectiveness [56].

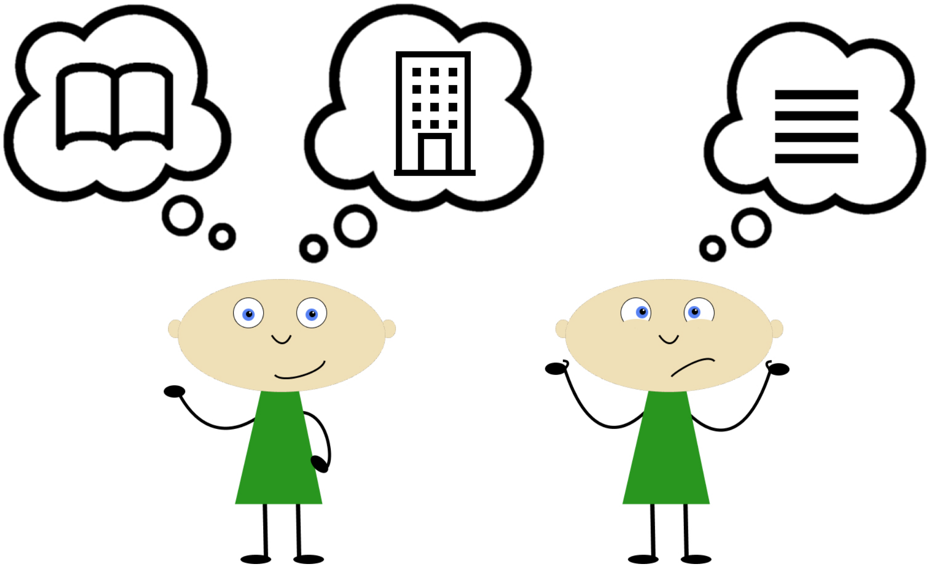


Figure 9: Metaphors are when familiar objects are used to help facilitate understanding in another domain [2, p 125].

The user's physical situation

Studies [57, 58] have shown that it is important to consider in which situations the users will be interacting with the VUI. Participants were asked to rate different interaction scenarios where different kinds of information, private or non-private, were being transmitted. Using the IPA in a private place, such as at home and disclosing non-private information was ranked as more likely than disclosing any kind of information with it in a public place. The studies also showed that disclosing private information in general was ranked very low [40].

Predicting and designing dialogue that fits every kind of situation is a task future developers will have to tackle. For now, creating dialogue that can follow a set of given paths and handle simpler contexts is as good as current IPAs can muster.

Just like creating a VUI persona makes designing a dialogue easier, setting a context is equally important when creating a well designed voice experience.

2.6.5 Error handling

“Error handling is the make it or break it difference of your users having a successful experience or not“ - Pearl [59]. Errors are inevitable and need to be handled so that the user doesn’t feel at fault [20]. Preventing and handling errors have always been of high importance when designing user interfaces. The reason for error handling being such an important aspect in VUI design is argued to be because a person’s voice is a primary attribute of their being. Using a VUI therefore makes errors feel personal [48].

There are many reasons that an error message might be triggered. It can be a result of the user providing unexpected or inappropriate input (no matches e.g. something outside the systems vocabulary), the system misrecognizing or misinterpreting the user’s speech (no input e.g. noise in the background preventing the system from interpreting the user correctly), or a combination of the two [20, 48].

Pearl [59] underlines the importance of having a strategy how to handle unexpected user input, because users will be inconsistent, hesitate and say unexpected things when speaking. Unlike writing, it is not possible to take back or change something that has been said.

Conversation repair
Stocker and Nicholls, the former a conversation designer and the latter a developer at Google, talk about changing the view of errors and how to handle them at Google I/O 2017. One way of changing the negative view of errors is to call them ‘conversational repairs’ to shift the focus from the fact that something has gone wrong and instead see them as a way to get the conversation back on track [48]. “There are no errors in human conversation, or at least we do not see them as errors” [48]. Hesitations and pauses are a natural part of human interaction, resulting in humans being experts at repairing and getting a conversation back by taking cues from each other [48].

The goal is to design the conversational repairs in such a way that the users does not even realise an error has occurred and been handled [60]. The possibility of triggering the conversational repair mechanism needs to be available from both sides of the conversation [20]. The VUI should allow the users to rephrase their statement or question, so that they feel like they are the ones getting the conversation back on track.

Direct and indirect costs
The cost[4] of understanding or recognizing errors can be defined as direct or indirect, the amount of cost is also an aspect that needs to be considered [20]. An example of a direct cost is an incorrect deposit in an automated banking transaction, while indirect costs occur for example when a user is dissatisfied with the system, resulting in reduced usage.

Depending on the type and amount of cost the error handling can be designed in different ways [20]. If there is a risk of an error having a huge impact, an explicit confirmation of the user's input prior to the action being executed is desirable [20]. An example of this kind of high risk task could be transferring money between bank accounts.

2.6.6 Confirmation

Confirmations provide users with feedback on how the VUI has interpreted their input. Confirmations allow users to correct mistakes immediately and establish common ground by reassurance [3].

In a dialogue there are two things that need to be confirmed; Parameters and Actions. Parameters are key principles. Actions are something the VUI is about to complete or has completed. An example of a Parameter being confirmed would be:

User: I want the blue and green ones.
VUI: Got it. The men's running shoes in royal blue and neon green. In what size? [3]

An example of an Action being confirmed would be:

User: Add it to my schedule.
VUI: Alright, I've added it to your schedule. Now, would you like to hear about the other sessions? [3]:

There are three ways a VUI can handle confirmations; Explicit confirmation, Implicit confirmation or no confirmation. An explicit confirmation requires a response from the user, usually a yes/no question.

VUI: A table for three people. Did I get that right?
User: Yes

[4] Cost of interaction failures

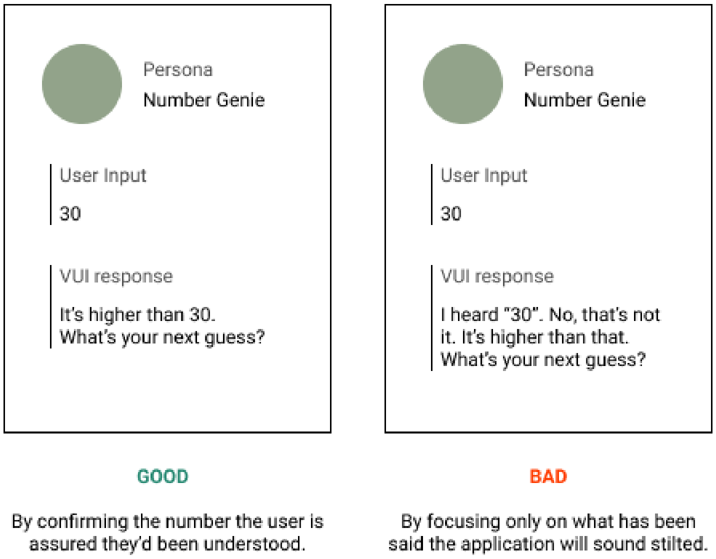
The implicit confirmation doesn't require the users to respond, but the user is given the option to correct the VUI if the confirmation is incorrect.

VUI: Alright, there is a table for four people outside. Would you like to book it?
User: No, three people.

No confirmation, does exactly that, moves the conversation forward without confirming the user's input.

User: Can I book a table for three.
VUI: At what time are you arriving?

Different confirmation models are suitable in different situations, a few examples of common scenarios from [3] can be found in figures 10, 11 and 12.



It's higher than 30. What's your next guess?

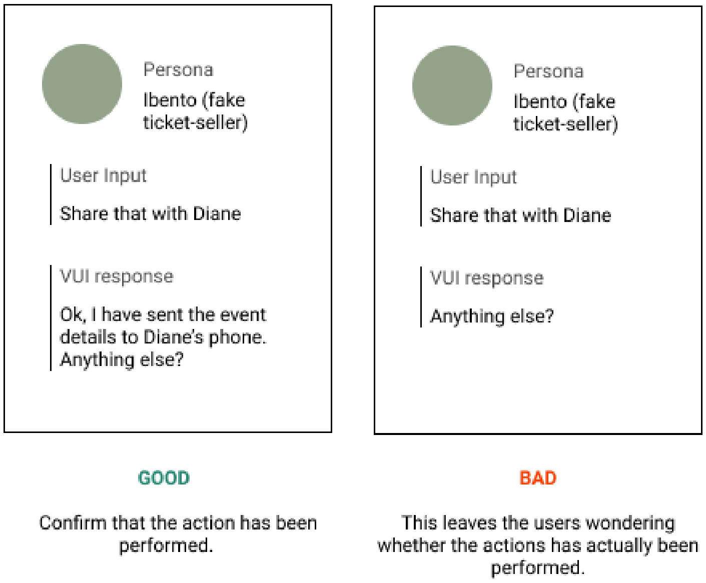
I heard "30". No, that's not it. It's higher than that. What's your next guess?

By confirming the number the user is assured they'd been understood.

By focusing only on what has been said the application will sound stilted.

Figure 10: Implicit confirmation of parameters. Image inspired by [3].

Similar to the discussion of error handling, keeping the cost of failed interaction low is important when discussing confirmation models [20]. If the cost of a failed interaction is high, for example when transferring money between accounts, booking tickets or cancelling a subscription, an explicit confirmation is needed before the action is executed. For low cost situations implicit confirmations are adequate.



Persona

Ibento (fake ticket-seller)

User Input

Share that with Diane

VUI response

Ok, I have sent the event details to Diane's phone. Anything else?

Persona

Ibento (fake ticket-seller)

User Input

Share that with Diane

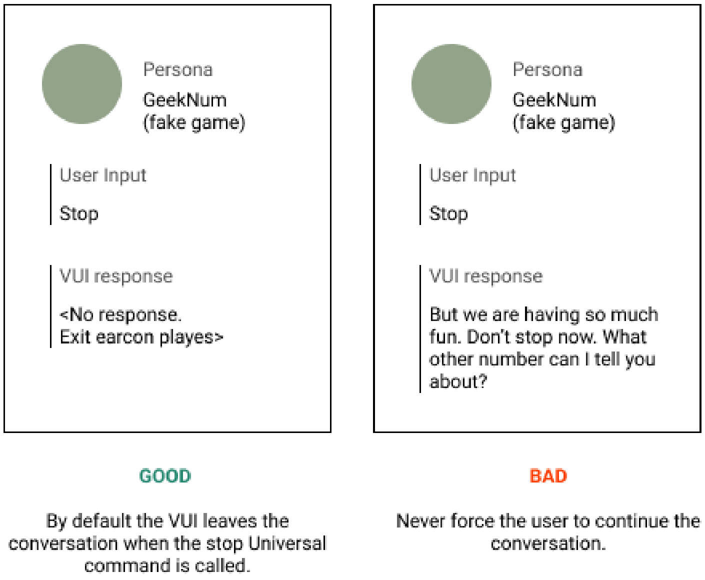
VUI response

Anything else?

Confirm that the action has been performed.

This leaves the users wondering whether the actions has actually been performed.

Figure 11: Implicit confirmation of actions. Image inspired by [3].



Persona

GeekNum (fake game)

User Input

Stop

VUI response

<No response. Exit earcon playes>

Persona

GeekNum (fake game)

User Input

Stop

VUI response

But we are having so much fun. Don't stop now. What other number can I tell you about?

By default the VUI leaves the conversation when the stop Universal command is called.

Never force the user to continue the conversation.

Figure 12: No confirmation of actions. Image inspired by [3].

As with all things, different users expect different levels of confirmations, some may experience confirmations as repetitive and unnecessary while others want every step to be confirmed before moving forward.

2.6.7 Cognitive load

Cognition is the study of mental processes such as problem solving, attention, language processing, memory, decision making and perception, simply put, cognition is how people think [61], [2, p 119]. The psychological term cognitive load describes the amount of mental effort it takes for a person to process and learn new information [61]. In UX design, cognitive load is seen as a user's "mental processing power" [61] or as Cohen et al. describes it, the amount of mental resources needed to perform a given task [2, p 119]. With good design the mental processing power required from a user can be reduced [61], the aim being not to overwhelm the users with too much information or too many options to choose from.

There are three cognitive challenges VUI designers should consider in their design process [2, p 120];

- Conceptual complexity - how complex is it to learn a new task?
- Memory load - will the user's short term memory need to hold lots of information in order to complete the task?
- Attention - will the users attention be divided while using the application? What happens if the user is interrupted halfway through the interaction, is it possible to pick up where they left off?

Memory load
In the influential article "The magical number seven, plus minus two" Miller found that humans' short term memory is severely limited when it comes to receiving, processing and remembering [62]. The average person has the ability to remember seven, plus minus two things. Later studies have however shown that remembering seven things, when they have been communicated over the telephone, is a challenge for most people [63]. In an experiment set up to test users' short term memory capacity when using a completely auditory approach, the results showed that, on average, the users remembered three things [63]. Hence, Millers [62] golden rule of "the magical number seven" should be reconsidered when designing for VUIs.

Studies have shown that it taxes user's memory less if the thing they need to remember is the last thing they hear, this is called the recency effect [64].

To spare users' memory load, just-in-time instructions are an efficient way to give users just the right amount of information to ensure they can complete the next task [2, p 127].

Attention
When building GUIs, designers reduce cognitive load by making paths intuitive and logical, using well known design patterns that users recognise [4]. When browsing a well designed website the ratio between feedback and input is kept fairly constant, this has a low but constant effect on the user's cognitive load [4]. In UX, attention pattern is a term used for measuring users attention when interacting with an interface [4]. Even when the users don't actively click a button or scroll up and down the page, their attention keeps moving between reading bits of text, looking at images and navigating the interface mentally, keeping their attention level fairly constant [4].

The attention pattern looks different when interacting with a VUI because when the user uses the activation phrase [5], the VUI responds almost immediately [4]. Speech has a temporal nature and is demanding on the user's memory [20]. The transient nature of voice means that the user will have to be fully alert to receive and understand the response [4]. Because listening is linear and the user cannot skip, go back or re-read prompts she is required to direct their full attention to the VUI and what it says, resulting in peaks in the attention pattern [4].
The graph in figure 13 shows the difference in a user's attention while using GUI and VUI.

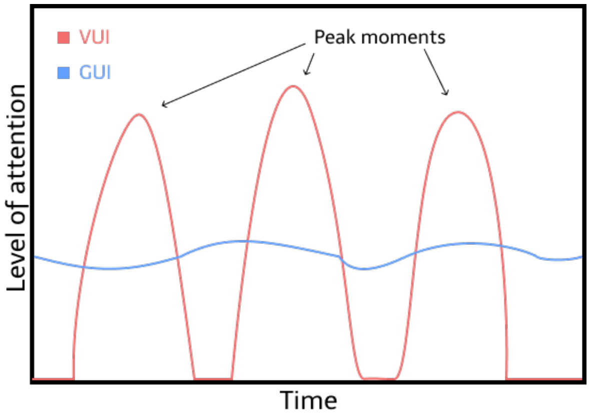


Figure 13: Users' attention peaks comparing interaction with GUI and VUI. Image inspired by Westerlund [4].

When interacting through a VUI the time frame is limited and this puts a high pressure on the users' cognitive load and the user has little control over the speed of the information flow [4]. Unfortunately attention peaks cannot be "designed away", listening will always have a high impact on the user's attention and therefore it is important that the peaks are kept short to stop the user from getting overwhelmed [4], [20]. Long responses are not only tedious to listen to

[5] E.g. Hey Google... or Siri... or Alexa...

but there is also a significant risk of the user not remembering what options are available [65].

A solution to this is to keep the VUI responses as short and to the point as possible [4]. Cohen et al. point out the importance of the use of universals to achieve constancy [2, p 121]. Users also need to be able to ask the system to repeat a previous command or piece of information [20]. The menu should consist of only a few options [20] and should have some kind of context in order for the user to draw a correct mental image and to ensure they have plausible expectations [2, p 125]. Synthetic speech has a high demand on cognition [20].

When a user's cognitive load is too high it can lead to a poor UX an increased amount of user errors, resulting in decreased usage [20].

2.7 Sample dialogue

According to Pearl [15] and Cohen et al. [2] writing sample dialogue is the first step when designing a VUI. The dialogue will be the base of how the user interacts with the application. Sample dialogue is a way to traverse the possible paths through a conversation. It specifies both the system side and the user side in a conversational sequence [2, p 108]. Sample dialogue should consist of a large set of prompts mapping as many potential paths as possible to ensure a natural flow of the conversation, capture different angles in the conversation, prevent errors and ensure that the message is getting across to the user.

Sample dialogue consists of intents and entities. Intents are ”what the user means”, entities are ”the keywords of the conversation”. For example, imagine a user wants to book a time to repair their broken bicycle (intent), in order to book a time the VUI needs to know what is wrong with the bike (entity = eg. flat tire) and when the user wants to hand it in (time = eg. tomorrow after lunch).

By adding colloquial words the dialogue sounds more natural and value is brought into the conversation, making it more cohesive and comprehensible and the application is perceived to have an awareness of what is happening [2, p 109-110]. Dene Earley-Cha [59], an action developer at Google, talks about the difference between traditional web development where the designing the UI is the last step in the development process, compared to building VUIs where the cornerstone of the application is the dialogues. The whole application is dependent on those sample dialogues.

A good first step when writing sample dialogue is to observe two people having a conversation. Though observations is becomes clear that people don't converse in a consistent way, everyone has their own way of formulating statements and questions. A study found that there are no “obvious” or “natural” words for a majority of objects, and therefore it is impossible to name functions and assume that users will know the right word or action command [66]. By listening and writing down the observed conversations it is possible to capture different ways of expressing things. In a later stage designers must come up with a wide range of versions of commands for the same action in order to build a system with high user recognition [66].

Researchers found that an efficient use of IPA would only be possible if the vocabulary used was restricted to a limited amount of utterances per action [16]. This view has however changed in recent years as the development of natural language understanding progresses. However, there are still issues that need to be addressed regarding free speech [59], for example, questions where the system expects a yes/no or a numerical answer. What happens when the VUI asks "do you like sweets" and the user answers "do you count chocolate". A human would understand the user's response as a "yes, if chocolate counts as a sweet", but the computer won't and will probably respond with an error prompt.

Another challenge is how to handle related questions, for example when a user calls to book a table at a restaurant and the VUI asks for the size of the party. The user might respond with the questions "do you have outdoor seating?" instead of the number of guests [59]. If not properly designed this response will cause the VUI to reply with an error prompt because it expected the answer to contain a number rather than a related question. The user does not ask these questions to make things difficult for the VUI, it might just be that the number of guests is dependent on if the dinner can be held outside or not.

Pearl [59] urges designers to avoid open ended questions as they tend to leave the user in a position of not knowing what to do or say to keep the conversation moving forward. Users like to be acknowledged, instead of having the VUI answer "I don't understand" when a user asks a question outside the scope of the application, prepare instead a response along the line of "I cannot help you with that yet, but I can do these things. . . " [59].

In conclusion, when writing sample dialogue it is important to design for the best practices, when the user acts as expected, but more importantly the sample dialogue should contain universal prompts, adjacent questions and related answers that the user can respond with.

2.8 Service Design

Moritz [67] defines service design as “service design helps to innovate (create new) or improve (existing) services to make them more useful, usable and desirable for clients and efficient as well as effective for organisations. It is a new holistic, multidisciplinary, integrative field”. Stickdorn et al. [5] suggest that service design can be explained in many different ways depending on the situation in which it is implemented. Service design can be explained as a mind set, a process, a toolkit or a management approach [5, p. 20-21]. The main purpose is to get people and groups to co-create solutions [5, p. 21]. This thesis uses service design as a process and toolkit to identify SEBs current situation, the needs of SEBs employees and from there a suitable solution and a set of guidelines. The different phases used in this thesis are based on Stickdorn et al.’s book This Is Service Design Doing: Applying Service Design Thinking In The Real World [5].

The phases have been illustrated in figure 14 and a description follows below:

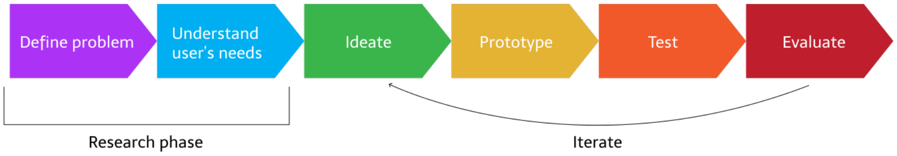


Figure 14: A representation of what a Service Design model can look like, based on the definition by Stickdorn et al. [5].

2.8.1 Research phase

The research phase is usually one the first phase in a service design project, but it is not uncommon that teams go back to the research phase after the ideation-, prototype- or test phase to evaluate newly found questions. The research phase ensures that designers base ideas on data rather than on assumptions. Arguably, there are three broad research methodologies: quantitative (gives insight into the “what” and “how” of an experience eg. surveys), qualitative (gives insight into the “why” - people’s motivations and needs eg. interviews and focus groups) and mixed-method which is a combination or the two. For this project an understanding of user’s needs was best carried out by conducting interviews, fitting into a qualitative methodology.

2.8.2 Ideation phase

In the ideation phase ideas are generated based on the knowledge collected in the research phase. Ideation contains many creative idea-generating activities such as; Brainstorming, How might we…?, 10 plus 10, Decision matrix and Mapping Journeys [68].

In service design, ideas are just starting points within the bigger evolutionary process and their real value often lies in the outcome that stems from them rather than in the ideas themselves.

Design thinking is a powerful tool to use in the ideation phase, it is a way to solve complex problems. In the classic linear approach, designers use tests to work their way forward in the process [69]. In design thinking the end user has a central position from the start and most design ideas are tested and evaluated by users through focus groups, interviews and user studies. This approach helps designers find problems early in the process, keeping costs and reworked ideas to a minimum [5].

2.8.3 Prototyping phase

The prototyping phase implements the ideas from the ideation phase to find which ones are most useful in an everyday business reality. Service design is an iterative model, this means that the prototyping phase, test phase and evaluation phase are run several times until a desired prototype is ready to either be tested on a lager scale or ready to be launched. The results of the prototyping phase can therefor be anything from a sketched lo-fi prototype to a fully functioning hi-fi prototype or finished product.

2.8.4 Test phase

Kamm [20] points out that in order to reach optimal VUI results the design and implementation need to be iterative processes with continuous users testing leading to revised prototypes and where deficiencies are detected and corrected.

Within-subject testing
Within-subject testing (also known as within-group testing) requires all subjects to be exposed to multiple versions of what is being tested (multiple conditions) [6, p. 50]. It requires only one test group, which means less participants are needed because every participant contributes to all the scenarios.

Advantage of using within-subject testing: It requires a smaller sample size meaning that less test participants are needed, usually resulting in shorter processing times and less costs. When analyzing the test result from within-subject tests, data comes from the same participants acting under different conditions. This means that individual differences can be isolated and can be seen to have a relatively low impact on the test results [6, p. 51]. When investigating complicated tasks such as reading, comprehension and information retrieval individual differences can have a larger effect on the result [6, p. 54, 67].

If for example, the test consists of four different conditions and every condition should have 16 users, the within-subject test would need 16 participants compared to the other well known model, between-subjects testing, where 64 participants would be needed. In the field of Humans-Computer interaction (HCI) sample sizes are important because it might not be very easy to recruit people with the “right” background/ level of knowledge / skill set to test the prototype.

Limitations when using within-subject testing: The learning effect is one of the biggest limitations to within-subject testing, since the participants will complete the same task several times, even if it is done in different ways with different conditions. There is a risk of the participant learning what to expect and also what to do and therefore performing better under the later conditions.

Another limitation with within-subject testing is that the participants need to perform several tests with different conditions, resulting in long tests and the risk of participant becoming tired and bored. In contrast to learnability [6] that can increase the score of the later tests, long test can have a negative effect on the conditions appearing towards the end of the experiment.

Nielsen suggests that a test should not be longer than 60-90 minutes, and should preferably be shorter [70]. A short test is less likely to be affected negatively by fatigue. If a longer test is necessary it is helpful to offer the users a break during it.

To minimize the impact of learnability and tiredness it is helpful to change the order of the conditions between participants. By randomising the order in which the conditions appear, the testers cannot design the test to be successful for a particular or hoped for result.

Studies suggest that the learning curve of a new system is steeper when initial interaction occurs but as the users get more familiar with the tool, the inclination decreases and settles at a flatter level, visualised in figure 15. Letting the participants use the system before the test starts will therefore greatly decrease the effect of the learning curve on the test results [6, p. 55].

[6] Is a quality of products and interfaces that allows users to quickly become familiar with them and able to make good use of all their features and capabilities.

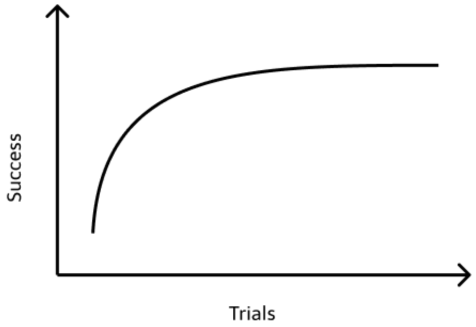


Figure 15: The learning curve for users when using a new system. Based on the design by Lazar et al. [6, p. 55].

2.8.5 Evaluation phase

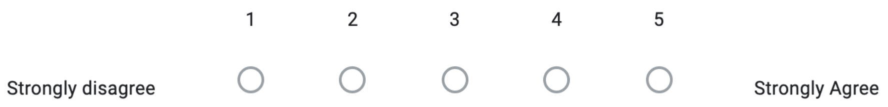
The choice of evaluation methods is based on what kind of prototype has been developed, the evaluation methods used in this project are Brooks System Usability Scale (SUS) [11] and Davis' Technology Acceptance Model (TAM) [25].

System Usability Scale (SUS)
A System Usability Scale (SUS) provides quantitative data on how usable the prototype is [10]. The questionnaire consists of ten statements that the user scores on a Likert five point scale ranging from (1) "strongly disagree" to (5) "strongly agree" [10,11], as shown in figure 16 . The statements are formulated so that alternative statements are positive and negative. SUS was originally created in 1986 by John Brooks [11,71]. SUS is a "quick and dirty" but reliable tool for measuring the usability of a service or system [71]. For usability it is important to evaluate whether the service is fit (or appropriate) to the required task [11].

The statements found in a classic SUS questionnaire are:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.

7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.



	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly Agree

Figure 16: Likert five point scale ranging from (1) “strongly disagree” to (5) “strongly agree”.

The scoring system in SUS

As stated above, the even numbered statements express positive attitudes while the odd ones express negative attitudes [10]. When calculating the SUS score, one is subtracted from the odd-numbered responses and the even-numbered responses are subtracted from five. At this point the scoring scale will range from zero to four, with four being the most positive. By adding all responses from the participants and multiplying the total with 2.5 the range will be converted from 0-40 to 0-100.

$$SUSscore = 2.5 * ((odd.nr - 1) + (5 - even.nr))$$

Over the years SUS has been used freely in many situations and has thereby been tweaked to fit different situations better [72, 73]. The score ratings can be found in table 1 and image 17.

The score sheet created by Brook [11] is shown in table 1:

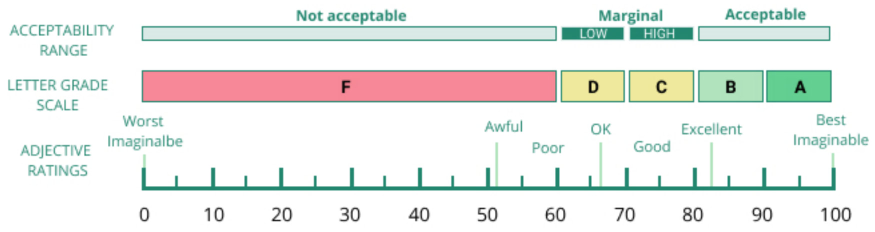


Figure 17: SUS score graphical scale based on the scoring table (image based on [7] and [8]).

SUS score	Letter grade	Adjective rating
Above 80.3	A	Excellent
Between 68 and 80.3	B	Good
68	C	OK
Between 51 and 67	D	Poor
Bellow 51	F	Awful

Table 1: The SUS scoring table. SUS scores can be translated into letters and adjectives, making explaining the score to stakeholders and others outside the organisation easier [10].

Technology Acceptance Model (TAM)
The Technological Acceptance Model (TAM) models how users come to accept and use technology [25]. The model suggests that when users are faces with new technology there are a number of factors that influence their experience and decision of how and when to use it [74]. These factor are *perceived usefulness* and *perceived ease of use* and have been discussed in section 2.1.1.

Statistical evaluation
When evaluating statistics it is common to use some kind of statistical hypothesis tests to determine if there is a significant difference between two or more groups. There are several questions that need to be answered in order to choose which hypothesis test should be used to evaluate the data. Examples of these questions are; Is the data the whole population or a sample of the population? Can the data be assumed to be normally distributed? To determine whether data is normally distributed a Shapiro-Wilk Test can be run.

Terminology:
Null hypothesis $\mathbf{H_0}$: The hypothesis that is tested.
Alternative hypothesis $\mathbf{H_1}$: The opposite of H_0.
p-value: is a measure of the probability that an observed difference could have occurred just by random chance.
Significance level: is the probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.
Critical value: can be found in the tables assigned to the specific test eg. [75, p. 437-496]
Test statistics: The smaller of the two (positive and negative) rank sums.

Shapiro-Wilk Test (Normal distribution) [75, p. 332]
The Shapiro-Wilk Test examines if a sample is normally distributed in a popu-

lation. The null hypothesis used in this case is "The values are samples from a population that follows a normal distribution". To determine whether to keep or reject the null hypothesis the probability (p-value) of finding the sample in the data is calculated. If the p-value is small, but the sample was found in the population anyway, it can be assumed that the null hypothesis was wrong and therefore should be rejected. The rule of thumb is that the null hypothesis (H_0) should be rejected if the p-value < 0.05. In easier words this means that if the p-value > 0.05 the population can be assumed to be normally distributed.

t-test [75, p. 325]

Data sets that are assumed to have a normal distribution, are collected in a within-subject test structure and have less than 30 data points, can be analysed with a paired t-test.

Wilcoxon Signed Rank Test [75, p. 393]

Data sets with paired and continuous values that can not be assumed to have a normal distribution can be analysed with Wilcoxon Signed Rank Test. By calculating and ranking the difference between the paired values as positive or negative, a test statistic is found by taking the absolute value of the smallest (usually the negative) rank sum. By checking the Wilcoxon Signed Rank Test table the critical value can be found at the specified significance level (often $95\% = 0.05$). If the test statistics is less than the critical value, there is enough evidence to reject the null hypothesis (H_0).

3 Method

The focus of this thesis has been to determine which of Aida's current processes will benefit from becoming voice based, the following section describes the tools that have been used during the investigation.

Throughout the 20 weeks of the project, meetings have been held with the supervisors from Umeå University and SEB on a weekly basis, these meetings have contributed to the decisions that have been made in the various stages. In week six a meeting was set up with a few applications specialists at SEB to provide insight to what could be done and thereby define the scope of the thesis further. In week 15 the results were presented to the responsible team at SEB to determine how the research results should be packaged and delivered.

3.1 Research phase

The research phase spanned the first six weeks of the project. A literature research was conducted to gain an understanding of the scope of previous research and to define the questions for this work. Interviews focusing on the current use of Aida were carried out with the target audience of SEB employees as well as analysis of the usage statistics of the Aida application.

3.1.1 Literature research

The literature research was carried out by reading published literature and research papers, found in the Google Scholar database and the resources available at Umeå University library, to understand the terminology and what has been done previously in the area of conversation design, voice interaction and VUI design. Technological development has only recently supported natural language processing to make spoken conversation an effective way to communicate with technology. The publications in this developing area tend to be in the form of blog posts written by individuals and white papers published by companies. A lot of content was also available through presentations at Google's annual I/O events[7] and through SEBs digital courses. These courses provide guidelines on brand identity and what one should think about when writing dialogue for their chatbot, Aida. These guidelines have been taken into consideration and modified slightly to fit spoken language [42]. The aim of the literature research was also to create a checklist for VUI design to follow throughout the project. Different design processes and test structures were researched to enable informed decisions to be made throughout the project. During the research phase it became clear that the initial thought of setting up a Wizard of Oz test environment

[7] https://events.google.com/io/

[8] would not give satisfactory results, as a result several other prototyping tools were researched and evaluated. A design model and test structure was chosen after discussions with both supervisors about the findings from the literature research.

3.1.2 Interviews with SEB employees

Because Aida, SEBs digital assistant, is used by all employees at SEB it was important to obtain a range of perspectives on the use of the tool.

Thirty interview questions were created in a brainstorming session. The session was run together with another student from Umeå University and started with discussions about why the interviews would be conducted and what they aimed to find. The next step was to identify different user personas in the target audience. A persona template [41] created by Stickdorn et al. [5] was filled in to create the various user personas. With a clear aim of what the interview should result in and a defined target audience of SEB employees, a version of Knapp's crazy 8s [76, p. 112] was conducted. To evaluate and decrease the number of questions, pilot interviews were conducted by using the personas created to represent the target audience [41]. The pilot interviews were iterated twice, both as a role playing session where a student acted as the different personas to see which questions provided the most useful data and therefore should be kept. The second iteration of the pilot interview was performed during a meeting with the SEB supervisor were the questions were discussed based on his understanding of the target audience. As a result of the pilot interviews the questions were then merged and re-written. Finally, ten open-ended questions were used in a semi-structured interview [77, 78], with room for follow-up questions, see Appendix A. The questions focused on the participants previous experience with IPAs in general, how and what they used the current Aida application for and what they would like a voice assistant to sound like.

3.1.3 Usage statistics

To gain deeper insight of the current use of Aida, statistics of the application usage from the previous six months (sep 2020 - feb 2021) were analysed. The data was divided into five groups to pinpoint which kind of processes were most frequently accessed.

[8] A test environment in which the test subject interacts with a computer system they believe to be autonomous, but which is actually being operated, or partly operated by an unseen human.

The five groups were as follows:

> **Group one** – Processes with short input from the user, resulting in short plain text responses from Aida in the chat. (e.g. Account unlocks, Password changes, Asking for information on weather).
>
> **Group two** - Processes with rich input from the user, resulting in rich and or long responses from Aida. (e.g. Filling in a form, Presenting a table, Input of long string of characters e.g. request ID, email).
>
> **Group three** - Processes in which the user's request result in Aida redirecting the user to an external source. (e.g. by providing a hyperlink to instructions/intranet).
>
> **Group four** - Processes which clarify the user's intention and redirects to other processes (asking clarifying questions and providing choice of buttons e.g. "did you mean...?").
>
> **Group five** - Processes where the user requests results and Aida provides long plain text answers/instructions in the chat.

Together with the interviews, this data provided a valuable understanding of users attitude towards the application and which kinds of processes had the potential of being implemented in a VUI.

3.2 Ideation phase

The ideation phase spanned over the following four weeks, the aim was to produce ideas to be further developed in the prototyping phase. Several service design tools such as Journey mapping [79], How might we...-notes [68] and crazy 8s [68] were used. The user personas from the research phase were reused to ensure that the design ideas were consistent and in line with the needs of the target audience. Use cases were created with correlating dialog flowcharts and drafts of sample dialogue.

3.2.1 Use cases

Based on the trends seen in the analysis of the usage statistics provided by SEB, two use cases were defined and user scenarios were created around these. The use cases and user scenarios were created by using a Mapping Journey template [79]. The use cases were used in the design process and as the core structure of the user study.

3.2.2 Dialogue flowchart

For each user scenario a dialogue flowchart[9] was created to structure and plan the flow of the conversation. Inspiration for the dialog flowcharts was taken from the current Aida application, Lundqvist's [8] and Hellman's [80] master's theses as well as the Dialogflow[10] documentation.

By identifying several different paths a user can take, it is easier to design to prevent errors, and in the cases where this is not possible provide fallbacks / error handling, minimizing the impact the errors have on the users' experience. The flowchart consists of examples of user input, the VUIs responses, universal commands and fallback intents. An example of a flowchart is shown in figure 18.

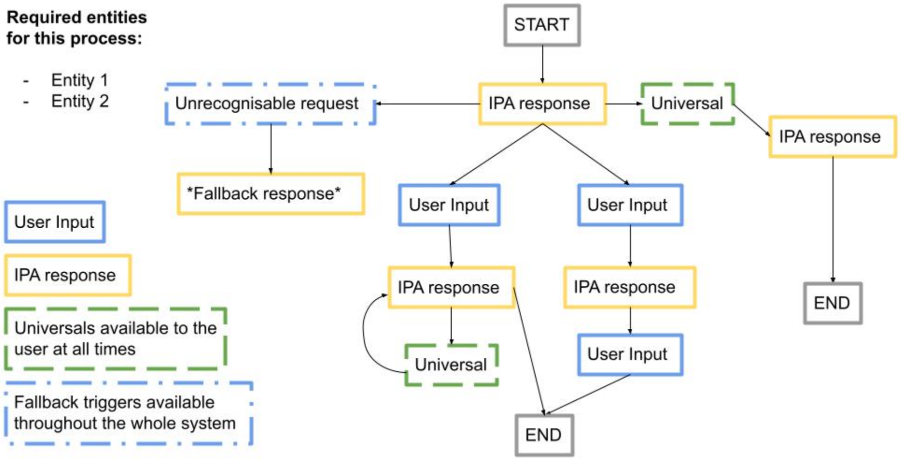


Figure 18: A dummy flowchart.

3.3 Prototyping phase

The prototyping phase lasted for two weeks and resulted in a hi-fi prototype ready for the user study. Before the implementation started the sample dialogue was created and tested in several iteration through the method ”Around the table reading”. An ”Around the table reading” can be compared with lo-fi prototype testing [81]. The next step in the prototyping phase was to implement the sample dialogue. Googles prototyping tool Dialogflow was used and

[9] https://cloud.google.com/dialogflow/es/docs/tutorials/build-an-agent/create-customize-agent

[10] https://cloud.google.com/dialogflow

two prototypes were created and tested. Before launching the prototype and initiating the test phase several pilot tests were run.

3.3.1 Sample dialogue

Using the dialogue flowcharts, sample dialogue[11] was written. Writing sample dialogue has similarities to writing the dialogue in a screenplay [59]. Sample dialogue aimed to predict as many scenarios as possible. When the sample dialogue was produced several conversational aspects were considered,

- The user's cognitive load and available working memory were considered when choosing the number of available options in a menu and the length of the response.
- The implicit and explicit confirmation models were considered and used in different versions.
- The length of the responses varied to find an appropriate length.
- The vocabulary was varied between formal and chatty.
- The response content varied between being either factual and to the point, or explanatory and wordy.
- The level of follow-up, varying helping the users forward in the conversation or letting the user guess what to do next.
- Placing the important information in the beginning, middle or at the end of a response.

SEBs vision is to design Aida with two ways of presenting information to users, verbally or suggesting to send the instructions in written form via email or through the Microsoft Teams chat. The user is given the two options and the dialogue continues down different conversational tracks depending on how the user wants the information delivered. SEBs language guidelines [42] were taken into consideration while writing and designing the sample dialogue.

3.3.2 Around the table reading

To get a feeling of what the dialogue would sound like it was important to act it out, or at least read it out loud together with other people. Because no technical effort has been added to the prototype by this early stage, redesign is quick and

[11] `https://developers.google.com/assistant/conversation-design/write-sample-dialogs`

easy. The activity ”Around the table read” is usually preformed in pairs. One person reads the users questions and the other reads the VUIs responses. In this case several table reads were performed with different people reading the users part. Several dialogues were read each time and the wording were changed to sound more natural and be easier to understand based on the participants comments.

3.3.3 Dialogflow

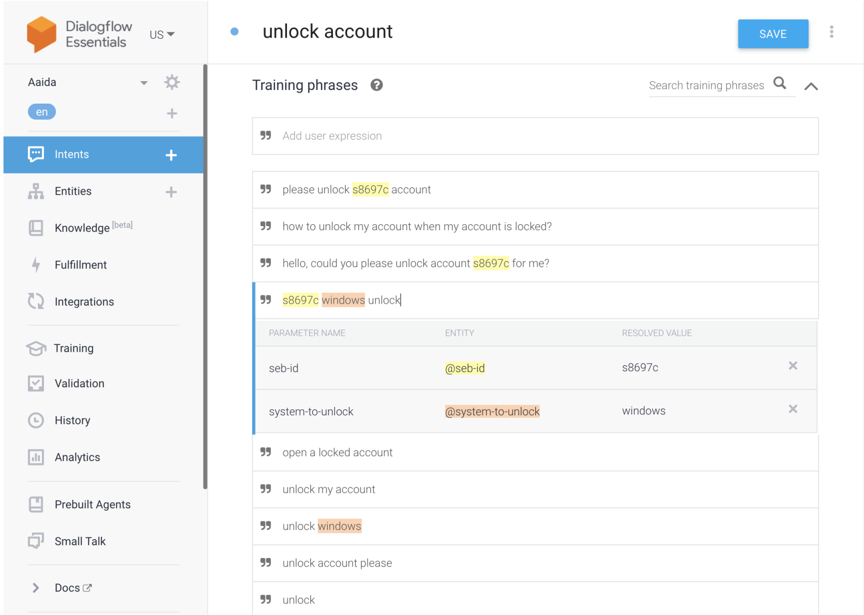
Dialogflow (also known as Dialogflow Essentials) is a natural language understanding platform developed by Google [9]. This platform allows developers to design and integrate VUIs for mobile apps, web applications, smart devices and so on. A IPA can comprehend and analyse both written and audio input and also respond to the user through a number of different outputs, including audio.

Limitations in Dialogflow:

- The IPA cannot be interrupted, and the users cannot give additional information as the input is being processed or a response is being read out. This can become problematic when the IPA takes longer then 10 seconds to process what the user has said or when the response from the IPA is long.
- It is not possible to tweak the IPAs voice or even add a customised voice, making it difficult to test different VUI personas.
- Connecting different intents with context was hard because there was no overview where the connections were visible.

Dialogflow uses machine learning to train the IPA. The logic behind the GUI is, a decision tree allowing the system to navigate down through the tree depending on what input is given by the user and therefore providing the agent with relevant responses [8]. There are three main parameters that need to be added for the IPA to learn and provide value to the conversation, these are; intents, entities and context. Dialogflows GUI is shown in figures 19 and 20.

Intents are what the user wants to do. Entities capture specific things the user says. Context is very important in conversation design and Dialogflow enables the designer to specify the context for each response. The user input provides training material for the system to ensure that the right intent is triggered. The machine learning part of Dialogflow allows users to express their intents with different words. The more variation in user input provided to Dialogflow as training material the better it becomes at predicting what intent the users wants to trigger even when the request is worded slightly differently to the expectation of the system [9].



unlock account

Training phrases

Add user expression

please unlock s8697c account

how to unlock my account when my account is locked?

hello, could you please unlock account s8697c for me?

s8697c windows unlock

PARAMETER NAME	ENTITY	RESOLVED VALUE
seb-id	@seb-id	s8697c
system-to-unlock	@system-to-unlock	windows

open a locked account

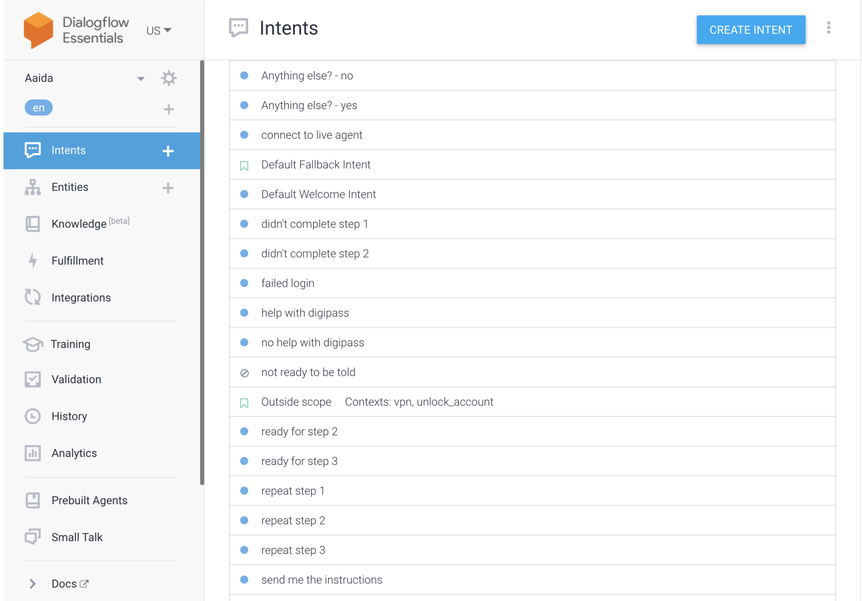
unlock my account

unlock windows

unlock account please

unlock

Figure 19: Dialogflow GUI where entities are set [9].



Intents

- Anything else? - no
- Anything else? - yes
- connect to live agent
- Default Fallback Intent
- Default Welcome Intent
- didn't complete step 1
- didn't complete step 2
- failed login
- help with digipass
- no help with digipass
- not ready to be told
- Outside scope Contexts: vpn, unlock_account
- ready for step 2
- ready for step 3
- repeat step 1
- repeat step 2
- repeat step 3
- send me the instructions

Figure 20: Dialogflow GUI where intents are set [9].

3.3.4 Pilot test

Pilot tests are important to ensure that the test answers the given research question and works as intended. Conducting a pilot test also shows whether or not the elements of the test will work together and checks that the technical aspects work effectively.

In this project several pilot tests were conducted. The use of Microsoft Teams for the interviews made it particularly relevant to check the technical aspects of the test. Instead of writing a test instruction to provide the participant with, a PowerPoint presentation was created to give an introduction to the test and the various tasks.

The pilot test group consisted of three participants, they went through the test process to troubleshoot the layout and structure of the test and the follow-up questions. The improvements made after the pilot tests resulted in a test design in the format of two tasks performed in two scenarios which was deemed suitable for this research.

3.4 Test phase

The test phase was initiated when an acceptable test layout and dialogue design had been implemented in Dialogflow. Continuous discussions with the supervisors gave an indication to when an acceptable test layout had been reached and the author made the decision to initiate the test phase by contacting participants from the target audience. The test phase spanned over three weeks, about three quarters into the project.

The first step of the test pilot test design was to determine which conditions were available and how they could be used to answer the research question. Depending on the type of conditions, such as the number of participants, task related parameters, learnability rate [12] and so on, different test models could be used. Finding qualified participants is a problem frequently occurring in the HCI field. This study aims to test employees at SEB, most of them are very busy. This means booking a meeting with them needs to be done several weeks in advance and the tests and interviews need to be kept short and efficient. In this study it was decided, by the author, to use a within-subject testing approach.

[12] How quickly can a user become familiar with and use the interface and its features and capabilities

3.4.1 Test setup

To gain insight into the users' experience of the prototype, a user study was conducted comparing users' experience of the existing chat version of Aida (from hereon referred to as textAida) with the voice prototype (from hereon referred to as voiceAida). Because English is the official language at SEB and because it was easier to find a prototyping tool that used English compared to Swedish, the decision was made to implement both Aida versions in English.

Because of covid-19 the user study had to be conducted online. A Microsoft Teams meeting was set up between the participants and test leader (on both phone (Iphone 11) and computer (HP Elitebook)).

The script that was followed to ensure that all participants received the same information can be found in Appendix B. A PowerPoint presentation was shared before the test started while the test leader provided necessary information about the study and how the answers and collected data would be handled. The participant was then introduced to the two scenarios and two tasks in the test. The test was divided into two versions, version A started the two tasks by interacting with textAida first and then with voiceAida. Version B did the opposite.

The tasks in version A were ordered as follows:

- Unlock an account using textAida
- Unlock an account using voiceAida
- Learn to connect to VPN using textAida
- Learn to connect to VPN using voiceAida (guidance)
- Learn to connect to VPN using voiceAida (send instructions)

In the two scenarios the participants were asked to imagie that: The two scenarios the participant was introduced to were as follows:

1. They were working from home and had an issue with first unlocking an account and then connecting to VPN and had to log on and chat to textAida in the current chat.
2. They were working from home and had an issue with first unlocking an account and then connecting to VPN and had to call help desk and speak to voiceAida.

Between every task the participants were asked to fill in a System Usability Scale (SUS) [11] questionnaire for each task. Since the tests were performed over

Microsoft Teams, a paper copy of the SUS questionnaire was not appropriate, a digital version was instead created using Google Form [13], see figure 21.

VoiceAida was run on the test leader's computer, a MacBook Pro through Dialogflow. TextAida was run on the participants computer and were asked to share their screen in order for the test leader to follow the interaction.

To collect the participants final thoughts and experiences, the test leader asked a number of open-ended questions, found in Appendix C. Every user study was recorded and later transcribed.

3.5 Evaluation phase

The evaluation phase consisted of two evaluations, firstly how the users evaluated the prototype during the user study and secondly an evaluation of the task completion times.

3.5.1 System Usability Scale (SUS)

One of the SUS questionnaire can be found in figure 21.

The scores for each task were calculated according to the equation

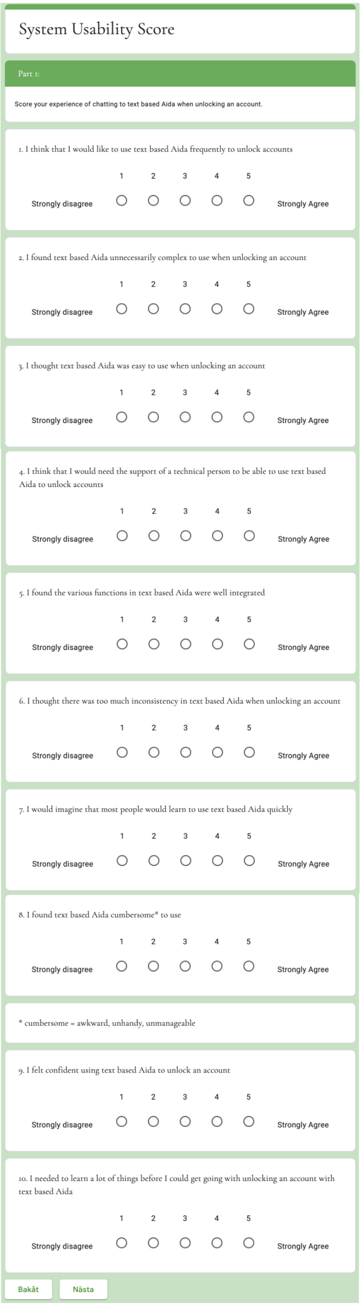
$$SUSscore = 2.5 * ((odd.nr - 1) + (5 - even.nr))$$

and compared to the scoring table 1 created by Booke [11].

3.5.2 Follow-up questions

Follow-up questions were asked to collect reactions and thoughts from the users after performing the tasks in the user study. The questions were a complement to the SUS questionnaire and to gather qualitative data from the users. The questions can be found in Appendix C. The participants' answers were analysed through TAM to determine whether voiceAida performs well for the selected tasks.

[13] `https://www.google.se/intl/sv/forms/about/`



System Usability Score

Part 1:

Score your experience of chatting to text based Aida when unlocking an account.

1. I think that I would like to use text based Aida frequently to unlock accounts

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

2. I found text based Aida unnecessarily complex to use when unlocking an account

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

3. I thought text based Aida was easy to use when unlocking an account

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

4. I think that I would need the support of a technical person to be able to use text based Aida to unlock accounts

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

5. I found the various functions in text based Aida were well integrated

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

6. I thought there was too much inconsistency in text based Aida when unlocking an account

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

7. I would imagine that most people would learn to use text based Aida quickly

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

8. I found text based Aida cumbersome* to use

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

* cumbersome = awkward, unhandy, unmanageable

9. I felt confident using text based Aida to unlock an account

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

10. I needed to learn a lot of things before I could get going with unlocking an account with text based Aida

1 2 3 4 5

Strongly disagree ○ ○ ○ ○ ○ Strongly Agree

Figure 21: The SUS questionnaire about unlocking an account with textAida in Google Forms.

3.6 Evaluation of results from user study

In order to know which statistical model to use when evaluating the results, the data sets need to be identified as normally distributed or not. A Shapiro-Wilk Test was run to ensure normal distribution. Thereafter a t-test and a Wilcoxon Signed Tank Test were used to evaluated whether the difference between the two Aida versions task completion times were statistically significant. These test results further aimed to strengthen the recommendation of which functions were suited to be implemented as VUI by showing which Aida version allows the user to complete the task faster.

3.6.1 Shapiro-Wilk Test (Normal distribution)

To ensure normal distribution the Shapiro-Wilk Test was run with the following hypotheses:

> **Hypothesis $\mathbf{H}_0$** (Null hypothesis): The values are samples from a population that follows a normal distribution.
>
> **Hypothesis $\mathbf{H}_1$**: The values are not samples from a population that follows a normal distribution.

3.6.2 t-test

Two hypothesis were defined for this evaluation;

> **Hypothesis $\mathbf{H}_0$** (Null hypothesis): There is no statistically significant difference between the time it takes to complete the task of unlocking an account with the different applications.
>
> **Hypothesis $\mathbf{H}_1$**: There is a statistically significant difference between the time it takes to complete the task of unlocking an account with the different applications.

3.6.3 Wilcoxon Signed Rank Test

If the test statistics is less than the critical value, there is enough evidence to reject H_0. Two hypotheses were defined for this evaluation;

> **Hypothesis $\mathbf{H}_0$** (Null hypothesis): There is no statistically significant difference between the time it takes to complete the task of learning how to connect to VPN with the different applications.

Hypothesis H$_1$: There is a statistically significant difference between the time it takes to complete a takes to complete the task of learning how to connect to VPN with the different applications.

4 Results

The following section presents all the results of the project, from the research phase with initial interviews to the final prototype and user study.

4.1 Research phase

This section will present the results from the activities in the design process' first phase. The most important findings from the literature research have been summarised, insight from the interviews and the results from the statistical analysis of Aida's usage statistics.

4.1.1 Literature research

The literature research gave insight to what has previously been done in the field of conversation design and examples of tools used by the leading tech companies. The biggest insights to take away from the literature research have been summarised in the following list.

- Creating a VUI persona early in the design process is essential [2].
- Let the user decide on how much information they want to hear [44].
- Errors occur frequently in all kinds of conversation, therefore the VUI needs a well thought though error handling set up [48].
- Confirmations in the form of sounds can replace the visual confirmations in a VUI [45].
- Studies show that users can remember three (plus minus two) instructions they have been told [63].

Service design was chosen to be the thesis' overarching methodology because it is a familiar process for SEB and as a means of problem solving, fits a project in which the objective and scope are finally defined only in the process of carrying out the work. When communicating a project's progress, service design has a user friendly framework that is easy to understand for those outside the project group as well as those within it [5, p 19]. Many of the tools suggested by the authors of the book "This is service design doing: applying service design thinking in the real world" [5] have been used for the planning and execution of this user study. Additionally, service design keeps the needs of the user at the core of the process, hence suiting the approach of this research which does not seek a finished product or service but to provide a roadmap for the further development of SEBs digital assistant Aida.

Google's conversation builder tool, Dialogflow [14] was chosen as the prototyping tool, because it had an easy to use interface and with only a few introductory video tutorials it was possible to design and build a functional VUI. Another advantage Dialogflow had compared to other tools, was the fact that it uses the same NLU-models as the Google assistant, providing a high quality speech recognition. It is also possible to sign in and create an account for free as long as the IPA does not leave the prototyping stage. Information about Dialogflow can be found in section 3.3.3.

Based on the insights from literature research the following decisions were made together with the supervisors, to move the project forward.

- Service design would be the overarching methodology throughout the project.
- The prototype would be built with Dialogflow.
- The user study would compare users' experience of interacting through text and voice, focusing on different lengths of the IPAs responses.
- To limit the scope to only look at how to handle different amounts of information.

4.1.2 Interviews with SEB employees

Eight SEB employees participated in an initial interviews over Microsoft Teams. The participants were spread across the following positions at SEB: IT support specialist, Applications developer, Team manager, UX-writer, Test manager and Business developer.
Age and gender: Three women and five men between the ages of 25 - 57.
Nationalities: Swedish and Lithuanian.

Many valuable insights were gained through the interviews, as summarised below.

Attitudes towards voice assistance in general
Two out of eight did not use a voice assistant at all, arguing that the systems was not good enough yet and that they just got frustrated when they could not get the system to understand their requests.

Three out of the remaining five participants used multiple Google assistants [15]

[14] `https://cloud.google.com/Dialogflow/docs`
[15] `https://assistant.google.com/`

in their homes and used them on a daily basis to control music, turn lights on and off and set alarms. These three participants also felt comfortable asking the Google assistant to search for answers on sites like Wikipedia. Despite this frequent usage none of the participants had particularly high expectations of the tasks a digital assistant could perform.

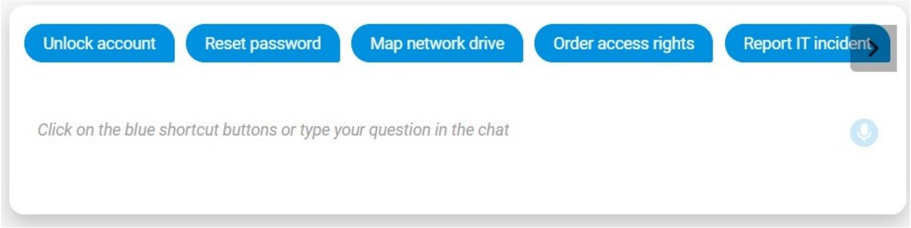
Attitudes towards SEBs digital assistant Aida
Two out of eight participants defined themselves as frequent users of Aida.

All participants thought that the most frequently used function provided by Aida was unlocking accounts.

Two out of eight participants used more than one function on a regular basis.

When asked why the participants did not use more functionalities, the most frequent answers were “I don’t know what Aida can do and therefore I don’t think she’ll be able to help me” or “I don’t know how complex questions I can ask, and instead of asking to find out I just look up the information by myself. It is usually quicker anyway”.

When asked about how the participants interact with Aida, five out of eight participants answered that they only used keywords or the pre-labeled buttons to give commands. See figure 22.



Click on the blue shortcut buttons or type your question in the chat

Figure 22: The pre-labeled buttons available in Aida today.

Three participants commented that Aida struggled with longer sentences and often missed the users’ intent, it was therefore preferable to keep the interaction short and quick and only provide one piece of information at a time.

When asked how the participants would like information to be presented they agreed that short and specific answers are preferable. The participants also agreed that giving web based Aida a voice today probably would not be very beneficial and that the tech department should instead focus on extending Aida’s scope and language understanding capabilities. However, the idea of implementing Aida as a “first line of support” in the help desk application seemed like a good idea to six out of eight participants.

The participants agreed that the best thing with Aida was that unlocking accounts was very fast and that the system was always available. One participant commented that ”there is no need to wait in a queue and it doesn’t feel like I am taking up anyone else’s time for simple tasks that I cannot perform by myself”.

However, Aida’s scope is limited, the system struggled to understand longer questions, had no memory and did not keep track of what had been said in previous steps in the conversation. Most importantly, the participants did not seem to know what Aida could and could not do.

Aida was implemented as a bookmark on every employee’s desktop, and still four out of eight participants found it cumbersome to get to the chat window. A comment from one participant was “the whole idea of a voice assistant is that it should be easy to access, you should be able to ask a question out loud and get an answer straight away. Having a web based voice assistant just doesn’t make sense to me. When I am sitting by my computer I want to type in a chat and not speak to my screen.”

Attitudes Aida’s VUI persona
The following perspectives describe which voice traits the participants want Aida to have.

Traits: Friendly, a combination between strict and chill, trustworthy, kind, recognisable from other voice assistants (following the same guidelines as other VUIs). Helpful, smooth and calming, making the users feel safe and confident.
Pitch: High pitch (it should be recognisable as a woman’s voice).
Language: The participants thought that it would be useful if Aida’s language could vary depending on where the user was calling from.
Dialect: If the language is set to English the dialect should be set to British English, or perhaps a English with a Swedish dialect because SEB is a Swedish company.
Gender: Female or gender neutral.

4.1.3 Usage statistics

Today Aida handles about 200 different processes that can be triggered by different user commands. The usage statistics from September 2020 to February 2021 (figure 23) consists of the five process groups and how frequently they have been triggered over a six month period.

The groups are as follows:

> **Group one** – Processes with short input from the user, resulting in short plain text responses from Aida in the chat.
> There are 32 processes/functions in this group.

Group two - Processes with rich input from the user, resulting in rich and or long responses from Aida.
There are 13 processes/functions in this group.

Group three - Processes in which the user's request result in Aida redirecting the user to an external source.
There are 28 processes/functions in this group.

Group four - Processes which clarify the user's intention and redirect to other processes.
There are 29 processes/functions in this group.

Group five - Processes where the user request results and Aida providing long plain text answers/instructions in the chat.
There are 96 processes/functions in this group.

The user statistics have been compiled in a graph in figure 23.

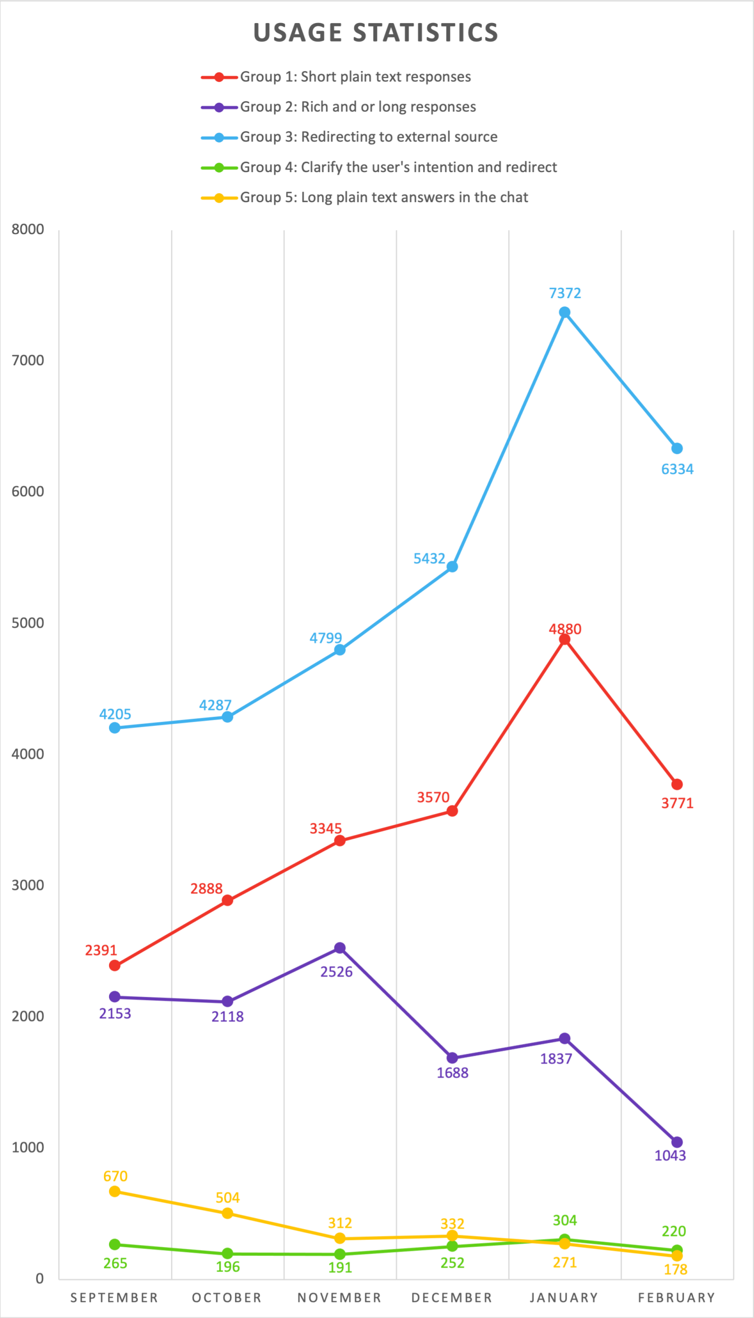


Figure 23: Usage statistics of the Aida application during the period of 2020.09.01-2021.02.28.

The two groups most frequently used were those processes which provided short answers (group one) or redirected the users to external sources (group three). Based on the findings a suggestion was made by the author, to investigate how two specific use cases from the two groups could be implemented as VUI. The tasks of unlocking an account and finding information on how to connect to VPN were chosen as representations of the two groups. The choice to investigate these two groups was purely based on the frequency which they were called by the users. The supervisor at SEB agreed that focusing on two frequently used tasks was a good angle for this project.

4.2 Ideation phase

In the following section the results from the ideation phase have been summarised.

4.2.1 User scenarios

Based on the conclusions drawn from the statistical analysis and interview answers, two user scenarios have been created.

1. **Scenario one:** User asking to unlock a windows account (group one)
2. **Scenario two:** User asking how to access Virtual Private Network (VPN) (group 3)

These scenarios were used when developing dialog flowcharts, testing sample dialogue and laid the basis for the user study. Initially the aim with the user study was to add Aida's existing dialogue to a voice application prototype and have as the baseline test. However, because it was not possible to conduct more than one larger user study with the target group at SEB this idea was discarded and replaced by a well iterated version of spoken dialogue. The dialog was still based on Aida's initial dialogue but changed to fit the spoken medium better.

4.2.2 Conversation flowcharts

Conversation flowcharts were created based on the user scenarios and inspired by the Dialogflow documentation. The different inputs in the conversation have been represented with different colours and outline styles to make the chart easier to understand. The flowcharts are shown in figures 24 and 25.

There are multiple ways to navigate through these conversations and therefore a 'right way' has not been marked. To prevent the chart from becoming difficult to read with too many arrows, the universals are only connected to one response, but the universals should, of course, be reachable from any part of the dialogue.

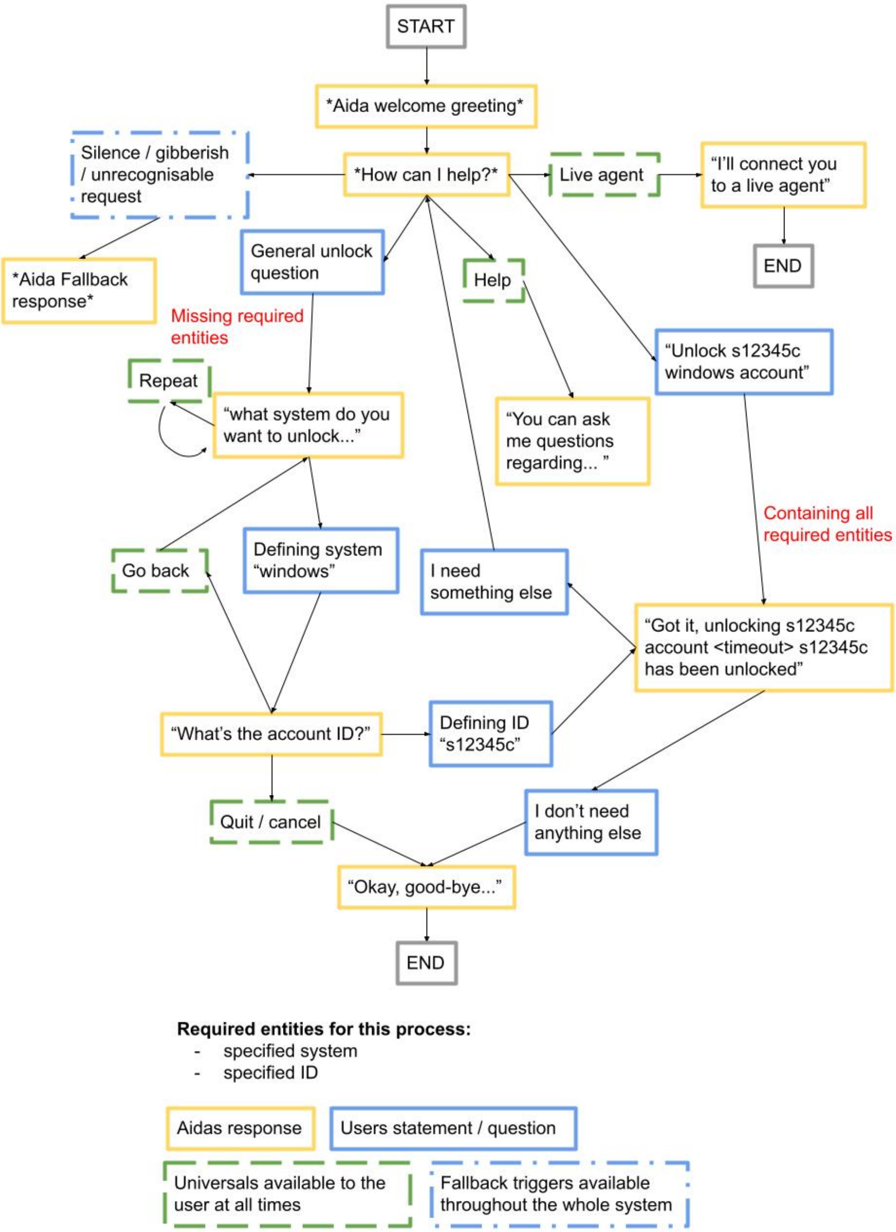

START
Aida welcome greeting
How can I help?
Silence / gibberish / unrecognisable request
Aida Fallback response
Live agent
"I'll connect you to a live agent"
END
General unlock question
Help
Missing required entities
Repeat
"what system do you want to unlock..."
"You can ask me questions regarding... "
"Unlock s12345c windows account"
Containing all required entities
Go back
Defining system "windows"
I need something else
"Got it, unlocking s12345c account <timeout> s12345c has been unlocked"
"What's the account ID?"
Defining ID "s12345c"
Quit / cancel
I don't need anything else
"Okay, good-bye..."
END
Required entities for this process:
- specified system
- specified ID
Aidas response
Users statement / question
Universals available to the user at all times
Fallback triggers available throughout the whole system

Figure 24: Flowchart representing the dialogue for unlocking a user's Windows account.

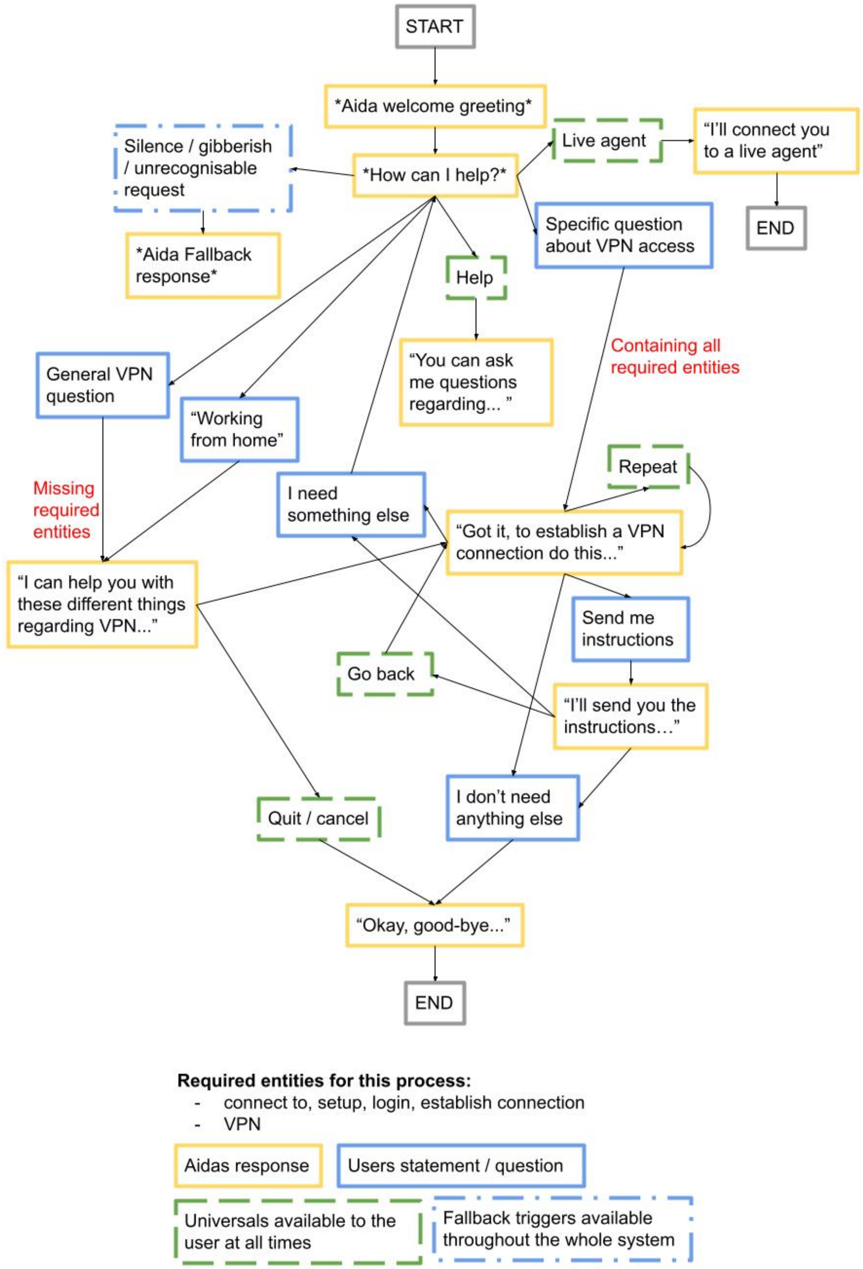

START
Aida welcome greeting
How can I help?
Silence / gibberish / unrecognisable request
Aida Fallback response
Live agent
"I'll connect you to a live agent"
END
Specific question about VPN access
Help
"You can ask me questions regarding... "
Containing all required entities
General VPN question
"Working from home"
Missing required entities
I need something else
Repeat
"Got it, to establish a VPN connection do this..."
"I can help you with these different things regarding VPN..."
Send me instructions
Go back
"I'll send you the instructions…"
Quit / cancel
I don't need anything else
"Okay, good-bye..."
END
Required entities for this process:
- connect to, setup, login, establish connection
- VPN
Aidas response
Users statement / question
Universals available to the user at all times
Fallback triggers available throughout the whole system

Figure 25: Flowchart representing the dialogue for gaining information on how to access the VPN.

4.3 Prototyping phase

Based on the flow charts and discussions with a language specialist at SEB several dialogues were written and evaluated multiple times. The sample dialogue was divided between the two tasks, unlocking an account and connecting to VPN. All communication with SEB employees in this section was done over Microsoft Teams.

4.3.1 First iteration:

The sample dialogue for unlocking an account was created by copying the dialogue used in the current text based Aida application, it can be found in Appendix D.1.1. Because the dialogue in the current textAida contains external links the dialogue could not simply be copied, this resulted in a first dialogue draft based on the dialog flowchart, Appendix D.1.2.

The two sets of sample dialogue were evaluated though "Around the table readings" together with an employees at SEB. The employee acted as the user and asked questions regarding accounts and VPN, and the author read Aida's responses to the different queries. Between the conversations the employees assessed the dialogues and commented on the level of information these can be found in Appendix E.1.

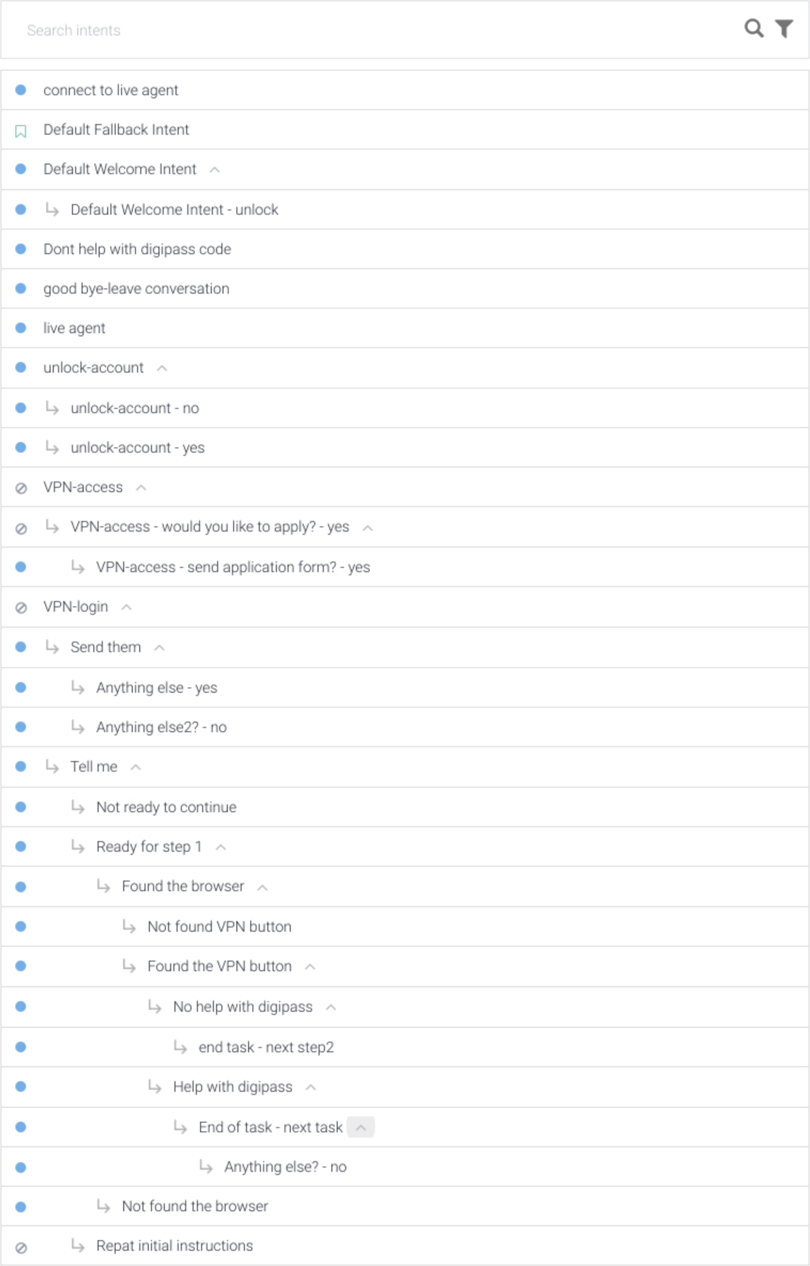
After the "Around the table reading" the dialogues were changed based on the comments, hence starting the second iteration.

4.3.2 Second iteration:

The sample dialog was changed based on the comments from the "Around the table reading" and written feedback from a language specialist at SEB. The sample dialogue for unlocking account can be found in Appendix D.2.2 and for connecting to VPN in Appendix D.2.3. In this iteration greeting phrases and endings to the dialogue were also written, see Appendix D.2.1. The sample dialog was once again evaluated though an "Around the table reading" with a second SEB employee, the comments can be found i Appendix E.2.

4.3.3 Third iteration:

After the second "Around the table reading" the decision was made that the dialogue was good enough to be implemented in Dialogflow and run though a pilot test. The first version of the Dialogflow prototype was created, see figure 26.



Search intents

- connect to live agent
- Default Fallback Intent
- Default Welcome Intent
  - ↳ Default Welcome Intent - unlock
- Dont help with digipass code
- good bye-leave conversation
- live agent
- unlock-account
  - ↳ unlock-account - no
  - ↳ unlock-account - yes
- VPN-access
  - ↳ VPN-access - would you like to apply? - yes
    - ↳ VPN-access - send application form? - yes
- VPN-login
  - ↳ Send them
    - ↳ Anything else - yes
    - ↳ Anything else2? - no
  - ↳ Tell me
    - ↳ Not ready to continue
    - ↳ Ready for step 1
      - ↳ Found the browser
        - ↳ Not found VPN button
        - ↳ Found the VPN button
          - ↳ No help with digipass
            - ↳ end task - next step2
          - ↳ Help with digipass
            - ↳ End of task - next task
              - ↳ Anything else? - no
      - ↳ Not found the browser
    - ↳ Repat initial instructions

Figure 26: The structure of the intents in the first Dialogflow prototype.

An initial pilot test was run with a student from Umeå University, to test the prototype. The student provided feedback on the dialogue and the fourth iteration was initiated, the comments can be found in Appendix F. Even though the student navigated though the prototype without any major difficulties the decision to reconstruct parts of the prototype was made as a result of the pilot test.

4.3.4 Fourth iteration:

Based on the comments from the first pilot test the Dialogflow prototype was rebuilt to avoid some of the pitfalls, such as not being able to go backwards in the conversation without starting from the begining, in the previous prototype design.

Another three pilot tests were conducted before the user study started. The pilot test group consisted of the supervisor at SEB and two SEB employees. As a result of the pilot tests the prototype was changed slightly again by changing and modifying words to fit the SEB guidelines and also to make it easier for the user to understand the VUI.

From the pilot test it became clear that running voiceAida and sharing computer audio on a computer while simultaneously holding the Microsoft Teams meeting would not work. Microsoft Teams is efficient in filtering the sound from the computer's speakers so it was not possible for voiceAida to hear what the participant said.

Different methods were tested such as, changing the test setup to be a Wizard of Oz test instead or simply have some sort of "Around the table reading". Finally, a solution which included connecting a phone to the Microsoft Teams-meeting, direct the sound to the phone speaker and thereby allowing the computer microphone to pick up the participants voice was used.

4.4 Test phase

Results and comments from the user study can be found below. All participants spend approximately the same amount of time using the prototype. Because it was only a prototype in Dialogflows test environment the machine learning functionality, where the system uses the users input as training data was not used. This means that the prototypes performance did not vary though out the test.

4.4.1 User study

The user studies were conducted over three weeks and the test duration varied between 35 minutes to 75 minutes. In total ten SEB employees with varying previous knowledge of voice assistance participated. None of the participants had English as their first language. Four females and six males participated, all in different professional roles varying from team lead to UX-writer and application developer.

All test were recorded enabling the test leader focus on the test rather then the comments from the participants. The work with transcribing the user test videos was ongoing though out the whole test phase. From the user study several observations regarding user's experience of both textAdia and voiceAida could be made. The following section presents the major findings. No changes were made in the prototype between the different tests and the machine learning function was not active during the test to prevent the prototype from improving towards the end of the user tests.

In table 2 the average completion times for each task can be found.

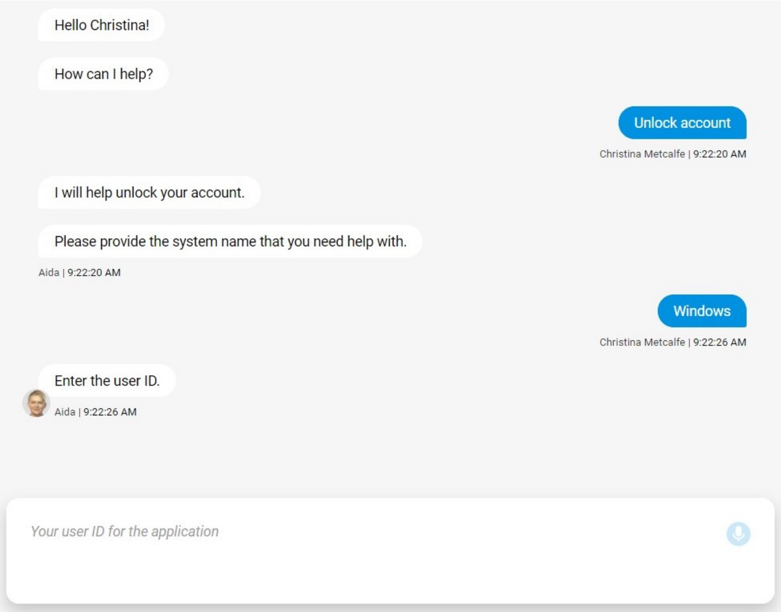
Unlock TA	Unlock VA	VPN TA	VPN (guide) VA	VPN (send) VA
28 s	61 s	166 s	106 s	59 s

Table 2: Average task completion time. TA = textAida, VA = voiceAida.

Task one - Unlocking a colleague's windows account
From the verbal instructions provided in the beginning of the test, the participants knew what to do in task one. Due to technical difficulties with conducting the user study over Microsoft Teams, the sound quality was poor resulting in voiceAida struggling to hear some of the users requests. Four out of ten participants had to repeat certain commands or in some cases, redo the whole interaction because voiceAida misheard the request and entered into an error-loop. This error-loop was not designed properly, thus making getting back on track close to impossible and therefore forcing the participant to restart the interaction. However, all participants successfully completed task one with both textAdia and voiceAida.

All participants agreed that unlocking an account felt just as fast with both Aida versions, even though the statistics showed that textAida was 33 seconds faster on average. A participant commented that they would rather use textAida because they found thinking of and saying their User-ID simultaneously cumbersome. Another participant was of a similar opinion and pointed out that textAdia is absolutely sufficient for unlocking accounts and has the advantage of visual feedback guiding the user through the process, which is not possible in voiceAida. Two participants thought voiceAida lacked appropriate feedback before executing the action. All participants had experience of using textAida to unlock accounts before and were therefore familiar with the process, this made completing the task easier.

Figure 27 shows the chat window used when unlocking an account with textAida.



Hello Christina!

How can I help?

Unlock account

Christina Metcalfe | 9:22:20 AM

I will help unlock your account.

Please provide the system name that you need help with.

Aida | 9:22:20 AM

Windows

Christina Metcalfe | 9:22:26 AM

Enter the user ID.

Aida | 9:22:26 AM

Figure 27: TextAida GUI, task: unlock windows account.

Task two - Finding out how to connect to the VPN

Establishing a VPN connection is a task that SEB employees have to carry out several times everyday since everyone has been working from home the past year. The task was divided into two different approaches, a guiding sequence and a suggestion of sending written instructions to the user. The interaction contained more information, causing higher levels of frustration and the potential for more errors. Here eight out of the ten participants caused voiceAida to go into an error-loop when the system misheard the request. The reason for most of the misunderstandings were due to poor sound quality but also the fact that there were several ways of asking for help compared to the previous task. Another element that caused some confusion and potential frustrations were the long pauses while voiceAida either listened or processed information, these pauses could last for up to ten seconds. The test leader kept quiet and instead used gestures to ask the participant to repeat themselves or wait as voiceAida was processing the request. All participants commented that some kind of feedback was required to inform the user of what was happening, earcons [16] in the form of beeps or a short sound were suggested to replicate the visual feedback provided by the three dots indicating “typing” in textAida.

Seven out of ten participants thought that the level of information in the guiding sequence provided by voiceAida was appropriate and divided into steps that were easy to follow. All participants appreciated being given the choice of either being guided or having the instructions sent in an email or in the Microsoft Teams chat. Three participants arrived at the same conclusion that three steps should be the maximum number in a guiding sequence. Any instructions consisting of more steps should only be provided as written instructions and be sent to the user.

One participant pointed out that they imagined many things that could go wrong when speaking to voiceAida. An example was the web address provided in the first step in the guiding sequence. In the scenario, the web address was short and all participants had heard and seen it before, since users visit it every day, however, as soon as a new or even just a slightly longer web address is used the user will probably ask for it to be repeated slowly. Another question asked by one participant was ”what happens if the user wants to back-track in the instruction because a previous step has gone wrong?”. Another point made by several participants was that humans do not speak in a consistent way, there are many ways of asking for the same things and that it is probably impossible to cover all the different ways of requesting information. A difference in interacting with textAida and voiceAida was that the users tended to use shorter commands and keywords more frequently when writing, and longer sentences while speaking. Two participants commented that they would ask a question in a sentence when speaking, because that feels more natural.

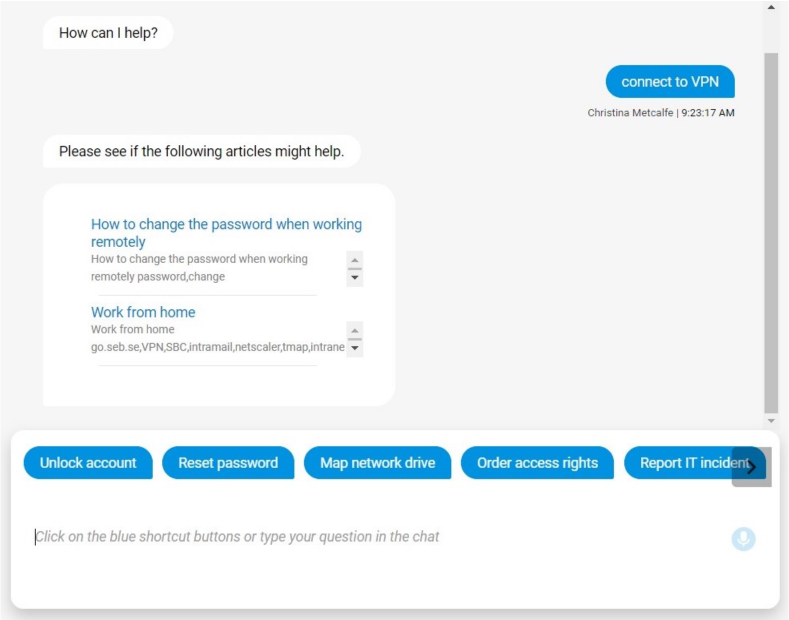
Figure 28 shows chat window when typing questions regarding VPN.

Comments from the follow-up questions

According to the interview questions at the end of the user study, six participants said they prefer to communicate in English using text while one participant said they prefer speaking English. Three participants claimed that their preference depended on the situation and the nature of the task, but when asked which way of communicating in English they would choose most often, all three answered ’text-based’.

Six out of ten participants were positive about having voiceAida as an asset while calling to help desk, supporting users with simpler tasks. Four participants thought voiceAida would work better as a filter, channeling the various tasks to a human service agent. One participant came up with the idea of letting users in the help desk queue interact with voiceAida while they wait, to see if the

[16] A sound / tone indicating something, providing feedback



How can I help?

connect to VPN

Christina Metcalfe | 9:23:17 AM

Please see if the following articles might help.

How to change the password when working remotely
How to change the password when working remotely password,change

Work from home
Work from home go.seb.se,VPN,SBC,intramail,netscaler,tmap,intrane

Figure 28: TextAida GUI, task: get information about how to connect to VPN. TextAida provides external links for the user to follow to get more information on the topic.

query can be resolved. If so, the user hangs up and leaves the queue and if not the user stays in the queue and is helped by a human agent.

Below follows an example of a conversation designed based on the participants suggestion of having Aida help users waiting in the help desk queue:

> **AIDA:** “Welcome to help desk, what do you need help with?”
>
> **USER:** “My windows account has been locked”
>
> **AIDA:** “There is a queue at the moment. Aida, our digital assistant can help you with unlocking accounts, would you like to speak to her? You won’t lose your place in the queue. So if you are not happy with the service Aida provides you can still speak to one of her human colleagues when it is your turn. Your place in the queue is 25, estimated time remaining ten minutes. Would you like to speak to Aida?”
>
> **USER:** “Alright, I’d like to speak to Aida”

AIDA: ”Which windows account would you like to unlock?”

(for a longer example dialogue see Appendix D.3.2)

None of the participants saw the benefit of implementing voiceAida in the textAida application, enabling a multimodal solution where the users speak to the web app.

All participants felt that voiceAida processed information too slowly and they became unsure what to do because of the long pauses. They all agreed that something needs to fill the silence that occurs when Aida is either listening or processing. Three participants pointed out that another area in voiceAida that needs improvement is the voice; it lacks a natural rhythm and some of the responses would improve if they contained a pause to let the users have time to perform the action, find the button to press or read the information in the guiding sequence.

In regard to the cognitive load of the two Aida versions, all participants said they felt that they needed to think carefully about what to say and speak more clearly than usual when speaking to voiceAida. In comparison to when they wrote to textAida they just wrote the first thing that popped into their minds, the comment being they felt more in control while writing and could go back and make changes if needed.

In conclusion all participants were positively surprised that voiceAida worked as well as it did, considering the technical setup of the user study and how far voice recognition software has come. Two participants said that part of the surprise is because of their lack of knowledge of what textAida currently can do. They would like to see textAida being marketed in a different way, believing that would have a positive impact on the use of textAida, regardless of the way of interaction. Part of this has to do with the user not knowing what to say in order to trigger the desired process.

4.5 Evaluation phase

The results from the user study were evaluated through different evaluation models and statistical test. The results from the evaluation phase have been summarised in the following section.

4.5.1 System Usability Scale from the user study

The four tasks in the user study were scored and evaluated according to SUS. The results were calculated as described in section 2.8.4 and the average scores are shown in table 3.

Task	Score	Grade
Unlock account (text based, textAida)	89	A (Excellent)
Unlock account (voice based, voiceAida)	74,75	B (Good)
Connect to VPN (text based, textAida)	62,75	D (Poor)
Connect to VPN (voice based, voiceAida)	72,75	B (Good)

Table 3: Summary and grading of the SUS scores from the user study. Grading according to Brook [11].

4.5.2 Shapiro-Wilk Test (Normal distribution)

The Shapiro-Wilk Test returned p-values for the four data sets in table 4 and concludes that all tasks but one have normally distributed samples.

	Unlock TA	Unlock VA	VPN TA	VPN (guide) VA	VPN (send) VA
Shapiro-Wilk test	True	False	True	True	True
p-value	0,720893	0,624941	0,003092	0,735441	0,216137

Table 4: Which data sets are normally distributed? TA = textAida, VA = voiceAida.

4.5.3 T-test

Results from the paired t-test suggests that there is a significant difference in the task completion time while using the text based application compared to using the voice based application. The p-value = 0,001508930023 $\leq$ 0,05 = alpha, provides enough evidence to reject H_0 and therefore conclude that textAida is significantly faster for unlocking accounts.

4.5.4 Wilcoxon Signed Rank Test

The results from the task VPN(Text based) cannot be assumed to be normally distributed, therefore the Wilcoxon Signed Rank Test was used when evaluating statistically significant difference in task completion times. The results from the Wilcoxon Signed Rank Test are shown in table 5, the significance level of 95% provided the critical values which sanctioned the following conclusions to be drawn.

There is evidence to suggest that there is a difference between the completion times of the text based and voice based applications when comparing VPN text based with VPN guide (the guiding sequence).

However, there is no sufficient evidence to suggest that there is a difference between the completion times of the text based and voice based applications when comparing VPN text based with VPN send (sending written instructions).

These results enable conclusion to be drawn that voiceAida is faster to use for the VPN guiding sequence task compared to textAida. However, it is not possible to draw the same conclusion about voiceAida in the VPN send instruction task.

	VPN TA vs. VPN (guide) VA, N=10	VPN TA vs. VPN (send) VA, N=9
Critical value	8	5
Test statistics	8	0

Table 5: Results from Wilcoxon Signed Rank Test for the task VPN. TA = textAida, VA = voiceAida.

5 Discussion

Voice assistants have in the last decade transitioned from being futuristic movie phenomenons to becoming a integrated part of our day to day lives. To ensure users acceptance and willingness to interact with machines in this new way it has become increasingly important to determine which application areas will actually benefit from being implemented as VUIs. This thesis aimed to answer the objective and the four associated research questions, by conducting a user study comparing users experience of a voice-base and a text-based version of SEB's digital assistant and discussing the results in the following section.

RQ1: "How do SEB employees use the current digital assistant, Aida"

The initial interviews provided answers to the first research question, "How do SEB employees use the current digital assistant, Aida". All interviewees use Aida for unlocking accounts, but only two of them use any of the other processes. A frequent reason given for not using Aida was the limited scope and that the user expected Aida to work in the same way as the search function on the intranet does. The usefulness and ease of use experienced by the users can because of this be understood to be quite low. A reason for the users low expectations can also be argued to be that the system isn't used to the extent of its capabilities because the users do not know how to use it.

RQ2: "Which are Aida's most frequently used processes"

The answer the second research question, "Which are Aida's most frequently used processes", can be found in the analysis of the usage statistics (found in section 4.1.3). The most frequently used processes are those where Aida either executes an action such as unlocking an account or provides more information by redirecting the user to an external page.

RQ3: "Which of these processes will, based on the current research, fit a VUI"

To answer the third research question, "Which of these processes will, based on the current research, fit a VUI", a few aspects need to be considered, in the following section these have been discussed.

Who will choose to use voiceAida?

The opinions from the user study regarding executing tasks were divided, some participants thought textAida was sufficient enough to perform simple tasks and therefore found little reason to use voiceAida. While a majority of the participants found the interaction with voiceAida to be more pleasant, the technical aspect of interacting with a VUI is still a limitation and has a negative effect in the users experience. The findings from the user study suggest that users, who today are content with using textAida to unlock accounts, would not be likely to change their behaviour and call to voiceAida to unlock an account. The users who do not use textAida on the other hand, would choose to speak to voiceAida as they wait in line to be connected to a human help desk agent. This finding strengthens the assumption made by Padgett [28] that in order for a VUI to benefit it needs to be more pleasant to use then the current chatbot.

How should longer instructions be handled?

A second finding from the user study suggests that longer guiding sequences would not benefit of being implemented as they were suggested in the user study. The participants liked the way the information was structured and presented, but they all agreed that it was too cumbersome to navigate through the VUI to get to the desired information. It was found that any conversation longer then three steps felt too long and required too much of the users memory. The findings further suggest that Millers [62] findings, that users' can remember seven things, should be revised to three things in the context of VUI.

A solution to decrease the amount of information which the user had to keep in mind, was implemented in the user study by, voiceAida suggesting to send written instructions instead of guiding the user through the steps. The interaction lasted for three conversational turns and resulted in a suggestion to send the instructions in the Microsoft Teams chat or as an email. This implementation got positive comments from the participants who, all except for one, stated that they would prefer writing and reading rather then speaking and listening to English.

The use of external pages

Interviewees' responses and the usage statistics point to the same conclusion, users prefer short and quick answers when interacting with textAida. One interviewee pointed out the importance of feeling confident that the information provided by Aida is correct, further commenting that being redirected to an external page increased the reliability of the information. The developers working with textAida are on the same track when reasoning about redirecting the user to external pages or the intranet instead of having information stored locally in textAida. This lessen the maintenance work of keeping textAidas information up to date. In a big company it is essential for information to be presented in a consistent way. Having the same information in several different places can be hard to manage. The maintenance of a system is an aspect that is often overlooked when a new solution is to be planned and implemented. Managing information in a VUI can be a cumbersome task and this might be a reason not to implement guiding sequences or informative conversations. Sticking to executing tasks might be the way forward with a VUI for SEBs help desk environment.

What does the SUS score say about the Aida versions' usability?
Although the SUS questionnaire is not a certified way of measuring ease of use and usefulness it is still a widely used usability measure and the results give an indication towards users' experience and therefore acceptance of a system. From the SUS scores it is clear that textAida is the most user friendly tool to use to unlock an account even though several participants also commented that they find it somewhat cumbersome to find and use textAida which is located on SEBs intranet. Even though the SUS score for voiceAida is more consistent in the sense of scoring ”good” on both tasks, compared to textAida scoring both ”excellent” (the task unlocking accounts) and ”poor” (the task connecting to VPN). More studies are needed before further assumptions can be drawn regarding the quality of service provided by voiceAida.

Which aspects would make the interaction even better?
The question of memory is not only in regard to the users memory capacity, it is also the question of the possibility of building a VUI with memory where Aida can remember what has been said in the earlier conversational turns or even previous interactions. The aspect of back-tracking in the conversation is an important part of providing the VUI with a memory to ensure an interaction as similar to a human conversation as possible. However, this aspect has not been covered by the scope of this thesis, but should be considered as the development of voiceAida continues.

Cognitive load and users preferences
Looking at the conversational aspect of cognitive load, all participants commented that they had to concentrate when speaking to voiceAida. They argued that they did not want voiceAida to do the wrong thing based on them saying something wrong or not articulating enough, this worry might have been eliminated if the participants had had the chance to get to know the system before hand. During the test there were several times where voiceAida did not hear the participant's request, resulting in the system returning an error-message to get the user back on track. This had an interesting impact on the participants, when the error message was triggered, they thought they were doing something wrong, resulting in them feeling uncomfortable or doubting their ability to speak to the VUI. This observation correspond with Stocker and Nicholls [48] theory about how being misunderstood when speaking has a greater negative impact on peoples self-confidence compared to being misunderstood while writing.

The reason for interacting with Aida, regardless if it is done through textAida or voiceAida, is to get something done, whether it is unlocking an account or retrieving information. This leads to the cognitive aspect of a persons ability to interpret information. People interpret information in different ways, some prefer reading written instructions and communicating in a written manner while others prefer being told what to do through spoken communication. In this study, 90 % of the participants claimed to prefer communicating in English in text, all saying the same thing “I prefer writing in a chat but I am sure that is only me, most people probably would prefer to speak”. Even though English is the official and well established language at SEB the observation that multiple participants found it uncomfortable to interact with voiceAida was interesting. This finding indicates that more user studies need to be conducted to determine whether building a VUI is the right thing to do in this case.

The lack of visual feedback
Looking at the conversational aspect of turn taking and feedback there are areas that need to be improved, several participants found it hard to understand when it was their turn to speak and when voiceAida had enough information to move forward in the conversation. TextAida has the advantage of visual feedback and a GUI that the participants are used to interacting with. There are buttons and help texts to instruct the user what to do and how to move forward in the conversation. Another advantage of using a GUI is the easy way to backtrack in the conversation, the user can take their time to read through the information and if necessary reread bits that are important or hard to understand. The information does not disappear in the same way as it does in a voice application.

To compensate the lack of visual feedback while using voiceAida, many participants requested some kind of earcon to indicate that voiceAida was listening or processing information. Most participants found the lack of visual feedback to be unfamiliar and thought it was harder to get an overview of where the conversation was heading and also what had happened previously.

Despite these insight about visual feedback, all participants, when asked if voiceAida should be implemented in a browser with a GUI, strongly disagreed. The argument most frequently used was that the users would be sitting in front of the computer already and therefore it would probably be easier to write in textAida for both privacy reasons and to prevent colleagues in the same office space to be disturbed. A majority thought voiceAida would work better as a filter or an additional agent in the phone based help desk environment. Pointing out that voiceAida could provide short and executing services or route the caller to a human agent with the right qualification to handle the callers query.

Human speech is inconsistent
Finally the aspect of inconsistency of user speech has an impact on the development of VUIs. Even though both Pearl [15] and Giangola [50] argue the point of the cooperative principle, users will still formulate questions in different ways. A textbased chatbot will need fallbacks and error-messages to handle unexpected user input and guide the user through the conversation. But because people are not consistent in the way they speak, have different dialects and cannot be assumed to be in a quiet environment while communicating with an IPA the VUIs needs many more fallbacks and error handling messages to guide the users through the conversation.

Comments of the test setup

The technical issues that accrued during the test had a negative impact on the users experience. Performing the user test over Microsoft Teams had the disadvantage of voiceAida not always hearing what the user said and therefor not triggering the correct user intent. This, in combination with slow processing, resulted in long pauses making the users unsure if voiceAida had heard their request.

Another technical aspect is the systems ability to correctly predict the users intents. By adding more training data the systems ability to predict users intent is enhanced, but because of the angle of the user study voiceAida only needed to handle two kinds of intents, unlocking accounts and connecting to the VPN. Additional training data would not have made any significant difference on the users experience in this case.

The users were not provided any time to interact with voiceAida before the test started to decrease the aspect of learnability [17]. However, looking back it might have benefited the user to "play around" with the prototype and the technical setup to allow then to feel more comfortable with the system and gain an understanding of the systems limitation. Because the participants had a limited understanding of how much voiceAida could understand none of them dared to provoke the system by saying something unexpected in fear of ruining the user study. On the other hand, the user study was designed to test the objective of this thesis and not the functionality of the prototype, for future VUI development the user study and tests need to have a different design to evaluate voiceAidas full functionality.

RQ4: "How do SEB employees experience interacting through voice commands"

To sum up and answer the fourth research question, "How do SEB employees experience interacting through voice commands", the participants in the user study enjoyed interacting with voice, commenting that it was fun to try something new and that it worked better then they had expected. However, all participants found the lack of feedback confusing and cognitively straining and only one of the participants preferred speaking to writing. Many questions were asked regarding how to navigate the VUI, how to know what to say to get the desired outcome and how to back-track the conversation.

The answer to the objective is therefore that both executing tasks, such as unlocking accounts, and providing longer instructions, such as getting information of how to connect to the VPN, can be seen to benefit from being implemented as VUIs but it is also dependent on users' preferences and the situation in which the system is used.

[17] Is a quality of products and interfaces that allows users to quickly become familiar with them and able to make good use of all their features and capabilities.

6 Conclusions

The findings from the initial interviews and the user study show that using voiceAida, a help desk VUI application for SEB employees, as an alternative to the help desk application to perform simpler task, such as unlocking accounts, would be beneficial to the users. The results from the user study shows that it would not be beneficial to have a VUI provide information in the format of longer guiding sequences. Previous studies [1] as well as results from the current study shows that voice recognition and a systems ability to predict users intents correctly is still a limitation that prevents further development within the VUI field. Hence technological developments are necessary in order to further integrate VUI in day to day life.

By internally marketing the capabilities of SEBs digital assistant, Aida, the interviewees believe that reasonable expectations of what they system can do will be established among the employees. Combined with the fact that voice assistants are becoming more widely used, they think that users' acceptance of voiceAida will increase and in time be possible to launch a voiceAida version with a wider and more diverse range of processes. The findings regarding users preferences from the initial interviews and the users study concludes that one Aida version should not exclude the other, people have different preferences and the two Aida versions should coexist, to reach as many employees as possible. Implementing Aida as a "first line of support" available as both text based and voice based will benefit the service personnel by lessening their workload of simple tasks, as well as the SEBs employees who will get quicker service and not have to wait in long phone queues.

6.1 Future work

While working with this thesis many related research questions have occurred and made funneling the topic challenging because there are so many interesting and important aspects of conversation design. The findings from the literature research indicate that the following aspects should be further explored before launching an IPA with a VUI.

First and foremost, questions regarding discoverability [18], how does the users know what the application can do and what to say to activate the desired functionalities? At present, there are no existing solutions to implement discoverability without a GUI. Attempts have been made, for example in Dutton, Forest and Jacks study where metaphors were used to help users navigate the VUI [56] but their findings did not result in any new design guidelines. A topic tightly

[18] Is the degree of ease with which the user can find all the elements and features of a new system when they first encounter it.

intertwined with discoverability is learnability [19], where the core question is; how VUIs should be designed to make them easy to learn to use. The invisible nature of speech is one of the things that makes designing for discoverability and learnability extra important to make the capabilities of the IPA apparent to users.

Secondly, the topic of the difference between the two conversational structures "command and control" and "conversational dialogue" needs to be investigated further. An evaluation of which structure is better suited for this use case. It might be that both should be used depending on the kind of process.

Thirdly, investigating how to implement memory in an IPA so that a user can jump back into a previous conversation or have the IPA remember what has been said previously in order to get a natural flow backwards and forwards in the conversation.

Fourthly, the aspect of privacy is important to consider. What information is OK to have spoken out loud? How should speaker verification work? The privacy aspect will become even more important if launching an external Aida version handling bank errands for SEBs customers.

Lastly, these topics are specifically aimed at the future development of voiceAida at SEB.

Firstly, all findings in the literature research regarding VUI personas point to the necessity of giving voiceAida a persona. There needs to be a strategy for what voiceAida should sound like in terms of vocabulary, language, intonation, dialect and many more. From the interviews a few traits were suggested for what Aida should sound like, the participants also pointed out the necessity of Aida having a familiar and consistent voice to strengthen the brand identity and increase the reliability of the system.

Secondly, further research is needed to understand this topic beyond the scenarios considered here. More processes such as updating user data, requesting access rights and filling in forms need to be tested in order for a decision to be made whether implementing a VUI is the right fit in this help desk environment.

Thirdly, evaluating the possibility of using voiceAida as a filter in the help desk queue, and if such a solution could be implemented in the current system.

[19] is a quality of products and interfaces that allows users to quickly become familiar with them and able to make good use of all their features and capabilities.

References

[1] M. Huang, “Designing actions on google,” 01 2020. https://uxdesign.cc/intro-to-conversation-design-ce3bd30e4385, acccessed: 2021-02-02.

[2] M. H. Cohen, M. H. Cohen, J. P. Giangola, and J. Balogh, *Voice user interface design*. Addison-Wesley Professional, 2004.

[3] GoogleDevelopers, “Confirmations,” 02 2021. https://developers.google.com/assistant/conversation-design/confirmations, acccessed: 2021-04-09.

[4] D. Werterlund, “How to deal with cognitive load in ux and voice design,” 08 2017. https://careerfoundry.com/en/blog/ux-design/voice-ui-design-and-cognitive-load/, acccessed: 2021-02-05.

[5] M. Stickdorn, M. E. Hormess, A. Lawrence, and J. Schneider, *This is service design doing: applying service design thinking in the real world.* ” O’Reilly Media, Inc.”, 2018.

[6] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[7] Usability.gov, “User testing metrics: The system usability scale,” 11 2017. https://www.userlytics.com/blog/system-usability-scale, acccessed: 2021-03-17.

[8] J. Lundqvist, “Designing a trustworthy voice user interface for payments and transactions-a study in user experience design,” 2019.

[9] GoogleDevelopers, “Dialogflow,” 03 2021. https://cloud.google.com/dialogflow/docs, acccessed: 2021-03-17.

[10] Usabilitest, “System usability scale (sus) plus,” 2021. https://www.usabilitest.com/system-usability-scale, acccessed: 2021-03-17.

[11] J. Brooke, “Sus: a “quick and dirty’usability,” *Usability evaluation in industry*, vol. 189, 1996.

[12] IMDb, “2001: A space odyssey,” 2021. https://www.imdb.com/title/tt0062622/?ref_=nv_sr_srsg_0, acccessed: 2021-02-03.

[13] IMDb, “Iron man,” 2021. https://www.imdb.com/title/tt0371746/?ref_=nv_sr_srsg_0, acccessed: 2021-02-03.

[14] IMDb, “Her,” 2021. https://www.imdb.com/title/tt1798709/?ref_=tt_mv_close, acccessed: 2021-02-03.

[15] C. Pearl, *Designing voice user interfaces: principles of conversational experiences.* ” O’Reilly Media, Inc.”, 2016.

[16] L. Karsenty, "Shifting the design philosophy of spoken natural language dialogue: From invisible to transparent systems," *International Journal of Speech Technology*, vol. 5, no. 2, pp. 147–157, 2002.

[17] Voicebot.ai, "Smart phone voice assistant customer adoption report," 11 2020. Available for download here: `https://voicebot.ai/2020/11/05/voice-assistant-use-on-smartphones-rise-siri-maintains-top-spot-for-total-users-in-the-u-s/`, acccessed: 2021-02-24.

[18] B. De Boer, "Evolution of speech and evolution of language," *Psychonomic bulletin & review*, vol. 24, no. 1, pp. 158–162, 2017.

[19] C. Baldwin, "Mediabrands arena ted talk on digital assistants, chatbots and future of humanity," 2017. `https://www.youtube.com/watch?v=5TDdc34UZTo`, acccessed: 2021-02-03.

[20] C. Kamm, "User interfaces for voice applications," *Proceedings of the National Academy of Sciences*, vol. 92, no. 22, pp. 10031–10037, 1995.

[21] SAIConference, "Everything you ever wanted to know about conversation design - cathy pearl, google," 04 2019. `https://www.youtube.com/watch?v=vafh50qmWMM`, acccessed: 2021-02-19.

[22] SEBgroup, "Our history," 2012. `https://sebgroup.com/about-seb/who-we-are/our-history`, acccessed: 2021-02-02.

[23] SEBgroup, "Who we are," 2012. `https://sebgroup.com/about-seb/who-we-are`, acccessed: 2021-02-02.

[24] "Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts," standard, International Organization for Standardization, 2018.

[25] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly*, pp. 319–340, 1989.

[26] I. Ajzen and M. Fishbein, "Attitude-behavior relations: A theoretical analysis and review of empirical research.," *Psychological bulletin*, vol. 84, no. 5, p. 888, 1977.

[27] Collins, "Useful (in british english)," 2021. `https://www.collinsdictionary.com/dictionary/english/useful`, acccessed: 2021-02-05.

[28] D. Padgett, "Finding the right voice interactions for your app (google i/o '17)," 05 2017. `https://www.youtube.com/watch?v=0PmWruLLUoE&list=PLJ21zHI2TNh9VkAu1EsOhpw92Wkm-XcuD&index=5`, acccessed: 2021-02-12.

[29] C. Pearl, "In dialogue—with our devices," 07 2019. `https://design.google/library/conversation-design-intro/`, acccessed: 2021-04-14.

[30] Fastdev, "Ux vs ui. what's the difference?," 10 2016. `https://www.fastdev.com/en/blog/blog/ux-vs-ui-whats-difference/`, accessed: 2021-02-05.

[31] J. Spitz, "Collection and analysis of data from real users: Implications for speech recognition/understanding systems," in *Speech and Natural Language: Proceedings of a Workshop Held at Pacific Grove, California, February 19-22, 1991*, 1991.

[32] A. Shahani, "Voice recognition software finally beats humans at typing, study finds," 08 2021. `https://www.npr.org/sections/alltechconsidered/2016/08/24/491156218/voice-recognition-software-finally-beats-humans-at-typing-study-finds`, acccessed: 2021-02-05.

[33] J. Giangola, "Conversation design: Speaking the same language," 08 2017. `https://design.google/library/conversation-design-speaking-same-language/`, acccessed: 2021-02-05.

[34] GoogleDevelopers, "Designing actions on google," 02 2021. `https://developers.google.com/assistant/conversation-design/welcome`, acccessed: 2021-02-02.

[35] C. Murad and C. Munteanu, "Designing voice interfaces: Back to the (curriculum) basics," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020.

[36] D. Schnelle and F. Lyardet, "Voice user interface design patterns.," in *EuroPLoP*, pp. 287–316, 2006.

[37] B. Shneiderman, "The limits of speech recognition," *Communications of the ACM*, vol. 43, no. 9, pp. 63–65, 2000.

[38] J. Hauswald, M. A. Laurenzano, Y. Zhang, C. Li, A. Rovinski, A. Khurana, R. G. Dreslinski, T. Mudge, V. Petrucci, L. Tang, *et al.*, "Sirius: An open end-to-end voice and vision personal assistant and its implications for future warehouse scale computers," in *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 223–238, 2015.

[39] M. F. McTear, Z. Callejas, and D. Griol, *The conversational interface*, vol. 6. Springer, 2016.

[40] B. R. Cowan, N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, D. Earley, and N. Bandeira, "" what can i help you with?" infrequent users' experiences of intelligent personal assistants," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–12, 2017.

[41] M. Stickdorn, M. E. Hormess, A. Lawrence, and J. Schneider, “Creating personas,” 2021. `https://www.thisisservicedesigndoing.com/methods/creating-personas-2`, acccessed: 2021-04-07.

[42] SEBCampus, “Written identity - learn how to write to our customers,” 2021. Internal course at SEB, completed: 2021-03-10.

[43] GoogleDesign. and M. Urban, “Method podcast, episode 8,” 01 2018. `https://design.google/library/margaret-urban-vui-google-assistant/`, acccessed: 2021-02-15.

[44] GoogleDevelopers, “Language,” 02 2021. `https://developers.google.com/assistant/conversation-design/language`, acccessed: 2021-04-09.

[45] GoogleDevelopers, “Overview of conversational components,” 02 2021. `https://developers.google.com/assistant/conversation-design/conversational-components-overview`, acccessed: 2021-04-09.

[46] D. Schiffrin, *Discourse markers.* Cambridge University Press, 1987.

[47] R. Quirk, S. Greenbaum, *et al.*, *A concise grammar of contemporary English.* Harcourt School, 1973.

[48] N. Stocker. and L. Nicholls, “In conversation, there are no errors (google i/o ’17),” 05 2017. `https://www.youtube.com/watch?v=oOLo071Pj1U&list=PLJ21zHI2TNh9VkAu1EsOhpw92Wkm-XcuD&index=4`, acccessed: 2021-02-09.

[49] H. P. Grice, “Logic and conversation,” in *Speech acts*, pp. 41–58, Brill, 1975.

[50] GoogleDevelopers, “Learn about conversation,” 02 2021. `https://developers.google.com/assistant/conversation-design/learn-about-conversation`, acccessed: 2021-02-05.

[51] AmazonAlexa, “Cwhat is natural language understanding?,” 2021. `https://developer.amazon.com/en-GB/alexa/alexa-skills-kit/nlu`, acccessed: 2021-02-18.

[52] CambridgeUniversityPress, “Conversation (noun),” 2021. `https://dictionary.cambridge.org/dictionary/english/conversation`, acccessed: 2021-02-02.

[53] P. Drew, “7 turn design,” *The handbook of conversation analysis*, p. 131, 2013.

[54] B. Ream, “6 ways to build context into your conversation designs,” 07 2020. `https://www.voiceflow.com/blog/6-ways-to-build-context-into-your-conversation-designs`, acccessed: 2021-04-08.

[55] J. D. Bransford and M. K. Johnson, "Considerations of some problems of comprehension," in *Visual information processing*, pp. 383–438, Elsevier, 1973.

[56] R. Dutton, J. Foster, and M. Jack, "Please mind the doors—do interface metaphors improve the usability of voice response services?," *BT Technology Journal*, vol. 17, no. 1, pp. 172–177, 1999.

[57] A. E. Moorthy and K.-P. L. Vu, "Voice activated personal assistant: Acceptability of use in the public space," in *International Conference on Human Interface and the Management of Information*, pp. 324–334, Springer, 2014.

[58] A. Easwara Moorthy and K.-P. L. Vu, "Privacy concerns for use of voice activated personal assistant in the public space," *International Journal of Human-Computer Interaction*, vol. 31, no. 4, pp. 307–335, 2015.

[59] GoogleCloudTech, "Conversational ai best practices with cathy pearl and jessica dene earley-cha: Gcppodcast 195," 10 2019. `https://www.youtube.com/watch?v=mXvx_JXBZBA`, acccessed: 2021-04-08.

[60] GoogleDevelopers, "Errors," 02 2021. `https://developers.google.com/assistant/conversation-design/errors`, acccessed: 2021-02-19.

[61] A. Margot, "Cognitive psychology in ux design: Minimising the cognitive load," 04 2019. `https://medium.com/design-signals/cognitive-psychology-in-ux-minimising-the-cognitive-load-d97ad8e3115b`, acccessed: 2021-02-05.

[62] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information.," *Psychological review*, vol. 63, no. 2, p. 81, 1956.

[63] M. Daneman and P. A. Carpenter, "Individual differences in working memory and reading," *Journal of verbal learning and verbal behavior*, vol. 19, no. 4, pp. 450–466, 1980.

[64] B. Ballentine, "Re-engineering the speech menu: A device approach to interactive list-selection," *Human factors and voice interactive systems*, pp. 205–235, 1999.

[65] N. Yankelovich, "How do users know what to say?," *interactions*, vol. 3, no. 6, pp. 32–43, 1996.

[66] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais, "The vocabulary problem in human-system communication," *Communications of the ACM*, vol. 30, no. 11, pp. 964–971, 1987.

[67] M. Stefan, "Service design practical access to an evolving field," *Mater dissertation, Köln International School of Design, Germany*, 2005.

[68] M. Stickdorn, M. E. Hormess, A. Lawrence, and J. Schneider, "Tisdd method library," 2021. https://www.thisisservicedesigndoing.com/methods, acccessed: 2021-02-03.

[69] Limetta, "Design thinking - metod för att lösa problem," 2021. https://limetta.se/tips-metoder-for-digitala-projekt/Vad-ar-Design-Thinking/, acccessed: 2021-02-22.

[70] J. Nielsen, "Time budgets for usability sessions," *Useit. com: Jakob Nielsen's web site*, vol. 12, 2005.

[71] Usability.gov, "System usability scale (sus)," 04 2021. https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html, acccessed: 2021-03-17.

[72] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009.

[73] J. Brooke, "Sus: a retrospective," *Journal of usability studies*, vol. 8, no. 2, pp. 29–40, 2013.

[74] I. Nair and V. M. Das, "Using technology acceptance model to assess teachers' attitude towards use of technology as teaching tool: A sem approach," *International Journal of Computer Applications*, vol. 42, no. 2, pp. 1–6, 2012.

[75] S. E. Alm and T. Britton, *Stokastik: sannolikhetsteori och statistikteori med tillämpningar*. Liber, 2008.

[76] J. Knapp, J. Zeratsky, and B. Kowitz, *Sprint: How to solve big problems and test new ideas in just five days*. Simon and Schuster, 2016.

[77] M. Stickdorn, M. E. Hormess, A. Lawrence, and J. Schneider, "Contextual interview," 2021. https://www.thisisservicedesigndoing.com/methods/contextual-interview, acccessed: 2021-03-16.

[78] M. Stickdorn, M. E. Hormess, A. Lawrence, and J. Schneider, "Interview guidelines," 2021. https://www.thisisservicedesigndoing.com/methods/interview-guidelines, acccessed: 2021-03-16.

[79] M. Stickdorn, M. E. Hormess, A. Lawrence, and J. Schneider, "Mapping journeys," 2021. https://www.thisisservicedesigndoing.com/methods/mapping-journeys, acccessed: 2021-05-21.

[80] D. Hellman, "Managing the expectations of voice-controlled access solutions," 2019.

[81] WomenTechmakers, "How to write a sample dialog in 60 seconds!," 02 2019. https://www.youtube.com/watch?v=sb75sitmPCc&list=PLJ21zHI2TNh-MpZ-7h1EPK9O7pjx109Bz&index=3, acccessed: 2021-04-06.

Appendix A Interview questions

The questions for the initial interviews in both Swedish and English.

1. Vad har du för tidigare erfarenhet av att använda en röstassistent?

   Do you have previous experience with voice assistants?

2. Använder du Adia ofta / skulle du säga att du är en van användare?

   Do you use Aida often? Do you consider yourself to being an experienced user?

3. Varför använder du Aida ofta / sällan?

   Why do you use Aida frequently/ seldom?

4. Hur många funktioner använder du?

   How many functions do you use?

5. Vilka funktioner använder du?

   What functions do you use?

6. Vanliga anrop / kommandon / formuleringar?

   How do you formulate the questions you ask? press the buttons or write sentences?

7. Vilken typ / mängd information förväntar du dig få av Aida?

   What kind of information do you expect to receive from Aida?

8. Vilka fördelar ser du med att ge Adia en röst? Vilka ärenden skulle bli lättare?

   What benefits do you see to Aida getting a voice?

9. I vilka situationer skulle det vara gynnsamt att Aida har en röst / kan förstå röstkommandon?

   What situations would become easier?

10. Hur skulle du vilja att Aidas röst / språk / ordval låter? Hur skulle du vilja uppfatta Aida? (svår dialekt, artig, lat, stöttande, tillrättavisande, uppmuntrande, proffsig osv)

    How would you want Aida to sound? dialect, vocabulary? How do you want Aida to make you feel?

11. Vad är det bästa / sämsta med hur Aida funkar idag?

    What is the best / worst thing about how Aida works today?

Appendix B Instructions script

This script is written to fit the first version of the user study (text first). Welcome... and thank you for participating in this user study!

Firstly:

- Would it be alright if I record this session? Perfect thank you!
  START RECORDING
- All answers and data collected here today will only be shared with my supervisors and will be presented anonymously in my paper.
- All answers and recordings will be deleted by the end of this project at the beginning of June.
- The test will be conducted in English but if you want to answer the questions at the end in Swedish that is fine.
- The tasks in this test are not there to test your abilities or knowledge, they are designed to test the prototype to see if it works as intended.
- Twice during the test I want you to share your screen when you interact with Aida as she works today.
- Feel free to ask questions, but not while talking to Aida because it might affect her performance.

Change to page 2
The aim of this study is to compare users' experience between Aida as she works today and a prototype of how she might work in the future. This all boils down to my research questions "which functions will benefit from becoming voice based".

Change to page 3
The tasks that you will be performing today are

- the classic unlocking your colleagues account and
- when you are working from home and want to know how to connect to the VPN.

Change to page 4
For each task you will go through two scenarios,

- calling service desk and talking to Aida or

- logging on and chatting with Aida as she works today

Change to page 5
So here is what we'll be doing today, you will perform two tasks twice, one with the chat and once with the voice assistant. Between each task you will be asked to fill in a short questionnaire with 10 statements, I want you to rate the statements on a scale between strongly agree or strongly disagree. You will find a link to the questionnaire in the teams chat.

When you speak to Aida, please speak slowly, articulate and be patient, sometimes she takes a long time to figure out what to do next. If she doesn't hear you I will signal you *LIKE THIS SIGNAL USER* to repeat what you've just said.

Alright, are you ready for task one?

Change to page 6
For this task I want you to share your screen and chat to Aida. Because my account isn't actually locked you won't get the desired confirmation, but the conversation structure will be the same and that is what is important for this test.

TASK 1
Goal: unlock the account.

In the chat you will find a link to the questionnaire, please fill in the first part.

TASK 2
Goal: unlock the account by calling to service desk.

Because Teams is so efficient in filtering out sound that comes from it's own speakers to prevent the sound circulating I've had to come up with another way for Aida to hear what you say. So, now I need to mute the microphone on the computer and turn on the microphone and speaker on my phone.

I'll start the application just to make sure you hear it alright and then, I'll ask you to greet Aida to "wake her up".

MUTE COMPUTER AND UNMUTE PHONE MICROPHONE AND SPEAKER.

SIGNAL THE USER TO GREET AIDA

Time to fill in the second part of the questionnaire.

MUTE PHONE MICROPHONE AND SPEAKERS AND UNMUTE COMPUTER.

Change to page 7

Time for the second part, By the end of this task I want you to be able to tell me how to connect to VPN.

TASK 3
Alright, I want you to share your screen again and chat with Aida.

Time to fill in the third part of the questionnaire.

TASK 4
MUTE MAC.
RESTART AIDA AND LET HER RUN ONCE SO SHE IS READY TO BE "WOKEN UP".
UNMUTE MAC.

It's time for the last task, ask Aida how to connect to VPN. In this task you will be given a number of instructions, because you are probably already connecte do the VPN you won't be able to follow them, so instead I ask you to imagine that you are.

But first let me get the technical setup to work.
MUTE COMPUTER AND UNMUTE PHONE MICROPHONE AND SPEAKER.

Greet Aida and find out how to connect to VPN.

Now, as a bonus, try this interaction again but instead of asking to be guided ask Aida to send you the instructions.

MUTE PHONE MICROPHONE AND SPEAKER AND UNMUTE COMPUTER

Time for the last part in the questionnaire.

Change to page 8

ASK INTERVIEW QUESTIONS

That's it for today.

Depending on how much time I have left after analysing these test results I might perform a shorter version of this kind of test later in april, if so, would you like to participate again?

Thanks again for participating!

END RECORDING

Appendix C User study wrap-up questions

In this appendix contains the open ended questions from the user study. The aim with these questions was to collect quantitative data from the users experience of textAdia and voiceAida.

- So, what was that like? What are your initial thoughts?
- Do you think Aida would be a good first line of support when calling service desk?
- Which way of interacting with Aida did you prefer? - and why?
- Which way of interaction do you think was faster?
- What did you think of the amount of information provided by her? - the length of the instructions or confirmation?
- Did you feel like something in the interaction was missing?
- For processes that Aida cannot automatically do, would you prefer her to provide you with instructions verbally (in a dialog kind of manner) or written down and sent to you?
- Would you like to have had other ways to get the written information sent to you?
- Was there a difference in cognitive load between the Aida versions? eg. did you need to think harder while using one or the other?
- Do you have any other comments you'd like to share?

Appendix D Sample dialog

Sample dialog from the three iterations in the prototyping phase.

D.1 First iteration

The first iteration of the dialogue for unlocking an account consists of dialogue taken from the Aida chat.

D.1.1 Unlock Windows account - copy of current Aida's dialogue

> A: Nice to see you Christina! How can I help?
> U: Unlock account
> A: I will help unlock your account. Please provide the system name that you need help with.
> U: Windows
> A: Enter the user ID.
> U: Sxxxxxx
> A: Unlocking the account... Sxxxxxx account has been unlocked.

D.1.2 Aida's responses when asked about VPN

The first iteration of the dialogue for connecting to VPN has been inspired by the Aida chat but several different dialogues have been put together. Different conversational parameters have been considered in the different versions. Version 1:

> A: Alright, to connect to SEBs VPN you need to go to an Edge browser and type go.seb.se. Press the button with a padlock (the second to the left called Remote SEB laptop access VPN). And enter your SEB-ID, password and digipass code and press Submit. Do you need guidance how to get the digipass code?
>
> **Including the following conversation aspects:** Explicit confirmation, long response, formal vocabulary, guiding content, suggesting more verbal information

Version 2:

> A: There are three simple steps to follow when establishing a VPN connection. Step one; open a browser and go to go.seb.se. Step two;

Press the VPN-icon (it looks like a padlock). Step three; Enter your SEB-ID, password and digipass code and press submit. That's it, you are all set to access the SEBs internal systems. Can I help you with anything else?

Including the following conversation aspects: Explicit, long response, chatty vocabulary, guiding content, providing next step in conversation

Version 3:

A: There are three simple steps to follow when establishing a VPN connection. Step one; open a browser and go to go.seb.se. Step two; Press the VPN-icon (it looks like a padlock). Step three; Enter your SEB-ID, password and digipass code and press submit. That's it, you are all set to access the SEBs internal systems. Would you like me to send you written instructions?

Including the following conversation aspects: Explicit, long response, chatty vocabulary, guiding content, suggesting to send information to user

Version 4:

A: Go to go.seb.se and follow the given steps to establish a VPN connection.

Including the following conversation aspects: Implicit confirmation, short response, formal vocabulary, informative content, no follow up

Version 5:

A: Go to go.seb.se and follow the given steps to establish a VPN connection. Can I help you with anything else?

Including the following conversation aspects: Implicit confirmation, short response, formal vocabulary, informative content, suggestion how to move conversation forward

Version 6:

A: To find out how to establish a VPN connection search for "working from home" on MyIT.

Including the following conversation aspects: Explicit confirmation, short response, formal vocabulary, informative content, no follow up

Version 7:

A: Search for “working from home” on MyIT.

Including the following conversation aspects: Implicit confirmation, short response, formal vocabulary, factual content, no follow up

Version 8:

A: Connect to VPN, sure, I can help with that! First, go to go.seb.se and press the VPN button. Second, enter your SEB ID, password and digipass code and press submit. That is it. You are all set. Can I help you with anything else?

Including the following conversation aspects: Explicit confirmation, short response, chatty vocabulary, guiding content, suggesting how the conversation can move forward

Version 9:

A: Aha, yes establishing a VPN connection is important when working from outside the office wifi. I can send you this information to you if you’d like. But first let me tell you how to connect to the VPN. Start by double clicking on the Edge explorer icon, search for go.seb.se in your browser. When you reach the page you will find four icons each representing a way to connect to SEB externally. If you’d like to know more about the other connection options don’t hesitate to ask. To connect to the VPN press the padlock icon (called Remote SEB laptop access VPN). You should now see three input fields: one for your SEB-ID, one for your password and the third one for a digipass code. I hope you know how to get the digipass code, if not I can explain that to you in a minute. When you have entered these three things press submit. Your computer will take a minute to load but when the terminal window has disappeared and your teams connection has been established you are all set to star working. Now, would you like me to send you these instructions via email, teams chat or text message, or would you like me to tell you about the other connection options or how to get the digipass code? If none of the above, can I help you with anything else?

Including the following conversation aspects: Explicit confirmation, long response, chatty vocabulary, guiding content, suggestions on how to move the conversation forward, suggestion to send information to the user, relevant information all over the place.

D.2 Second iteration

Sample dialogue from the second iteration in the prototyping phase.

D.2.1 Greetings

The greetings dialogue was created in the second iteration and is inspired by the Aida chat as well as other voice assistants on the market.

> Welcome to service desk, I'm Aida and I can help with unlocking accounts, answering questions regarding VPN and connecting you to a live agent. What would you like to do?
>
> Welcome to service desk, this is Aida, how can I help?
>
> This is Aida at service desk, I can help you unlock accounts, answer questions about VPN and connect you to one of my colleagues.
>
> This is Aida at service desk, what can I help you with?
>
> This is Aida at service desk, how can I help?
>
> You have called service desk, there is a queue right now, perhaps I can assist you in the meantime, what do you need?
>
> Thank you for calling service desk, I'm Aida, how can I assist you today?
>
> Hello Christina, it's nice to hear from you, how can I help?
>
> Nice to hear from you Chrisitna, what can I do for you?
>
> Hi, I'm Aida, SEBs intelligent voice assistant, would you like to find out what I can help you with or ask me a question or ask me to redirect you to a live agent?

End phrases to the greetings:

- Which one would you like?
- What can I help you with?
- What can I do for you?

- What would you like to do?
- What do you want to do?
- What do you need help with?
- What's next?
- What do you need?
- How can I help?
- How can I assist you today?
- Tell me what to do.
- Tell me what you need help with.
- Pick an option.
- Ask me anything.
- Ask me about accounts or VPN.

D.2.2 Unlock Windows account

The sample dialogue in the second iteration has been changed and put into different versions based on the users response. Version 1:

> U: I need to unlock an account for a colleague
> A: I can help you with that, what's the system you'd like to unlock?
> U: Windows
> A: And what's the ID?
> U: s8697c
> A: Account s8697c has been unlocked. Do you need anything else?
> —end of conversation—

Version 2:

> U: unlock windows account s8697c
> A: Account s8697c has been unlocked. Can I help you with anything else?
> —end of conversation—

D.2.3 Connecting to VPN

The version-numbering in the second iteration has nothing to do with the version-numbering from the first iteration. In this version user responses have been added. Version 1:

> A: It is very easy to log on to the VPN, would you like me to guide you or would you prefer if I'd send the written instructions to you?
> U: Tell me
> A: There are three simple steps, ready?
> U: Yes
> A: Ok, step one: open a browser and go to go.seb.se. Found it?
> U: Yes
> A: Step two; Press the VPN-icon (it looks like a padlock). See it?
> U: Repeat this step
> A: Alright, press the button that looks like a padlock. Found it?
> U: Yes
> A: Great, the final step is to: Enter your SEB-ID, password and digipass code and press submit. Do you need help to get the digipass code?
> U: Yes
> A: When you've entered your pin, and the screen shows APPLI, press 1 and enter the six digits. Did you manage to log on?
> U: Yeah
> A: Alright, can I help you with anything else?
> U: No
> A: Okay, have a nice day, good bye!
> —end of conversation—

Version 2:

> U: Send them
> A: Ok, would you like them sent via email or teams chat?
> U: Send them to me via email please.
> A: Alright, ¡pause¿ an email has been sent to you. Do you need anything else?
> U: Thanks, no.
> A: Alright, good bye for now.
> —end of conversation—

Version 3:

> A: Go to go.seb.se and follow the given steps to establish a VPN connection. Are you experiencing an issue with establishing a con-

nection?
U: Yes
A: I will connect you to a live agent who will help you with this issue.
—end of conversation—

Version 4:

A: To establish a VPN connection visit go.seb.se and follow the steps. Do you have an issue with connecting to your VPN?
U: No
A: Alright, can I help you with anything else?
U: Yes, I still don't know how to connect to VPN.
A: Connecting to VPN is easy, would you like me to guide you though it or send you a written instruction?
U: Tell me
A: Ok, there are three simple steps, step one; open a browser and go to go.seb.se. Found it?
U: Yes
A: Step two; Press the VPN-icon (it looks like a padlock). See it?
U: Yes
A: Great, the final step is to: Enter your SEB-ID, password and digipass code and press submit. Do you need help to get the digipass code?
U: Yes
A: When you've entered your pin, and the screen shows the letters "APPLI", press 1 and enter the six digits that appear. Did you manage to log on?
U: Yeah
A: Alright, can I help you with anything else?
U: No
A: Okay, have a nice day, good bye!
—end of conversation—

D.3 Third iteration:

The third iteration is the dialogue that was later implemented in the Dialogflow prototype.

D.3.1 Greetings

The greetings were changed and shortened.

> Welcome to service desk. My name is Aida and I can unlock accounts, answer questions regarding VPN and connect you to a my human colleague. What would you like to do?
>
> Welcome to service desk, this is Aida, how can I help?
>
> This is Aida at service desk, I can help you unlock accounts, answer questions about VPN and connect you to one of my human colleagues.
>
> This is Aida at service desk, what can I help you with?
>
> You have called service desk, there is a queue right now, perhaps I can assist you in the meantime, what do you need?
>
> Thank you for calling service desk, I'm Aida, how can I assist you today?
>
> Hi, I'm Aida, SEBs digital assistant , would you like to find out what I can help you with or ask me to redirect you to a live agent?

End phrases to the greetings:

- What can I help you with?
- What can I do for you?
- What would you like to do?
- What do you want to do?
- What do you need help with?
- What do you need?
- How can I help?

- How can I assist you today?
- Tell me what you need help with.

D.3.2 Unlock Windows account

A few more versions on the unlock account interaction with additional user-input. Version 1:

> U: I need to unlock an account for a colleague
> A: I can help you with that, what's the system you'd like to unlock?
> U: Windows
> A: And what's the ID?
> U: s8697c
> A: Account s8697c has been unlocked. Do you need anything else?
> —end of conversation—

Version 2:

> U: unlock windows account s8697c
> A: Account s8697c has been unlocked. Can I help you with anything else?
> —end of conversation—

Version 3:

> U: unlock account
> A: To unlock an account I need to know the system and the ID.
> U: windows, s8697c
> A: The account has been unlocked. Is there anything else you need?
> —end of conversation—

Version 4:

> U: unlock account
> A: To unlock an account I need to know the system and the ID.
> U: s8697c
> A: And what's the system?
> U: windows
> A: The account has been unlocked. Is there anything else you need?
> —end of conversation—

D.3.3 Connecting to VPN

Only a few words were changed between these two dialogue versions. Version 1. A: There are three steps you need to follow to log on to the VPN, would you like me to guide you or would you prefer if I'd send you the written instructions?

U: Tell me

A: Ok, step one: open a browser and go to go.seb.se. Found it?
U: Yes
A: Step two: Press the VPN-icon (it looks like a padlock). See it?
U: Repeat this step
A: Alright, press the button that looks like a padlock . Found it?
U: Yes
A: Great, the final step: Enter your SEB-ID, password and digipass code and press submit. Do you need help to get the digipass code?
U: Yes
A: When you've entered your pin, and the screen shows APPLI, press 1. Enter the six digits that appear in your digipass in your browser . Did you manage to log on?
U: Yeah
A: Alright, can I help you with anything else?
U: No
A: Okay, have a nice day. Good bye!
—end of conversation—

U: Send them

A: Ok, would you like them sent by email or in a Teams chat?
U: Send them to me by email please.
A: Alright, ¡pause¿ an email has been sent to you. Do you need anything else?
U: Thanks, no.
A: Alright, good bye for now.
—end of conversation—

Version 2: A: To establish a VPN connection visit go.seb.se and follow the steps. Do you have an issue with connecting to your VPN?

U: No

A: Alright, can I help you with anything else?
U: Yes, I still don't know how to connect to VPN.
A: Would you like me to guide you through the steps or send you a written instruction?
U: Tell me
A: Ok, there are three simple steps. Step one: Open a browser and go to go.seb.se. Found it?
U: Yes
A: Step two: Press the VPN-icon (it looks like a padlock). See it?
U: Yes
A: Great, the final step is to: Enter your SEB-ID, password and digipass code and press submit. Do you need help to get the digipass code?
U: No
A: Alright, can I help you with anything else?
U: No
A: Okay, have a nice day, good bye!
—end of conversation—

U: Yes

A: I will connect you to a live agent who will help you with this issue.
—end of conversation—

Version 3: A: Go to go.seb.se and follow the given steps to establish a VPN connection. Are you experiencing an issue with establishing a connection?

U: Yes
A: I will connect you to a live agent who will help you with this issue.
—end of conversation—

Appendix E Comments from the evaluation of the sample dialogue

In this appendix comments from the around the table reads have been summarised.

E.1 First iteration - Aida's responses when asked about VPN

Version 1: My initial though was: Can you repeat that. A bit too much information presented at once and especially for a novice user.

Version 2: Same feeling of wanting repetition as in version one. There needs to be a pause between the steps to give the user time to process.

Version 3: Same as previous versions, suggest to send instructions in the beginning of the interaction.

Version 4: Short and not very informative. But it might be suitable for an expert user. It is easy to navigate to the suggested page. Might need a follow-up question such as "do you have an issue with establishing a connection?". Try rephrasing it and adding the URL in the beginning of the sentence.

Version 5: Ask a relevant question. "can I do anything else" should appear when the task is completed or at the end of the conversation.

Version 6/7: MyIT provides licenses for establishing VPN access. Perhaps it would be interesting to look at how to fill in a form though voiceAida.

Version 8: Better wording in this version.

Version 9: A good example of what the dialog should not sound like.

E.2 Second iteration - Aida's greeting phrases

Version 1: To much information all at once

Version 2: Good but Aida is not a person and users might get confused by the wording in "I'm your digital assistant"

Version 3: Introducing what she can to must be on a higher level

Version 6: Best one yet. What if there is no queue? Can the user explain the issue?

Version 8: This made me assume Aida is a person, not good

Version 10: like the options, good to explain what she is! Combine asking questions and what Aida can do. I'd guess what she can do.

Randomly select the end-phrases to make the dialogue more alive and give the feeling of a broad vocabulary.

Appendix F Comments from the first pilot test

These are the comments from the first pilot test that led to the redesigning of the structure of the prototype:

It was good but it is hard when you don't see the chat window otherwise it was fine.

I have a bit of experience with speaking to VUI before and I am never quite sure how complex questions I can ask, and it was the same with voiceAida. Which is the main hurdle for me.

It would be helpful to get a ball park understanding of the complexity of the questions that Aida can handle at the start of the conversation. The length of the menu should be short but the users should be given the possibility to hear the full thing if they want to. So a better approach would be to provide the different categories that the system can help with and then allow the users to die deeper if needed.

On the other hand, seeing as this VUI will only be used by SEB employees, they should know what they want and also what voiceAida can do, and then I'd probably just start using her and work my way though her different commands.

I didn't try navigating backwards in the prototype or saying anything outside what i though it could handle, what would have happened if i did?