Designing navigation in smart speakers

Guidelines for designing a voice user interface for navigating radio content using a smart speaker

Master's thesis in Interaction Design and Technologies

PAULINE LORIN & LINN THORSAGER

MASTER'S THESIS 2019

Designing navigation in smart speakers

Guidelines for designing a voice user interface for navigating radio
content using a smart speaker

PAULINE LORIN & LINN THORSAGER

UNIVERSITY OF
GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2019

Designing navigation in smart speakers
Guidelines for designing a voice user interface for navigating radio content using a
smart speaker
PAULINE LORIN & LINN THORSAGER

Designing navigation in smart speakers
Guidelines for designing a voice user interface for navigating radio content using a
smart speaker
PAULINE LORIN & LINN THORSAGER
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

How we interact with our devices seems to be changing, and voice interaction is
gaining ground in a field previously dominated by screen-based interaction. How
to design for voice interaction is however yet in its infancy, and navigating using
voice seems to be unexplored. As of today, multiple problems have arose regarding
voice interaction and designing a voice user interface. Because of the differences
between visual and auditory perception, existing standards for screen based devices
cannot be directly translated for voice based interactions, resulting in no, or few,
standards. There are also challenges regarding understanding and mimic human
conversation, and the technology behind voice interaction is still limited. However,
there are several advantages of using voice interaction instead of visual including
being fast, hands-free, intuitive and empathetic. Designed right and used with the
right sort of content, using voice interaction can be truly favorable.

The study was done in collaboration with Swedish Radio, as implementing voice
interaction to their content would be beneficial, and aimed at exploring how to
design for navigation using voice with the expected outcome of design guidelines. A
literature review, competitive testing, stakeholder interviews, and an ideation session
was conducted and resulted in a first set of design guidelines. The guidelines were
tested and changed through two user tests, which in the end resulted in sixteen
guidelines divided into five themes. The guidelines can be considered as usable but
not as definite. Further testing needs to be done to include testing in a natural setting,
on a broader scope of users, and for long time usage. Also, the guidelines needs to be
tested with designers as to find out how they would interpret and implement these
guidelines when designing.

Keywords: voice user interface, navigation, smart speaker, interaction design, user
experience, design guidelines, radio content.

Acknowledgements

Contents

Contents

List of Figures

List of Tables

List of Tables

1

Introduction

There is said to be an upcoming paradigm shift in how we interact with our devices, moving from screen-based interaction towards voice-based. With this shift, new problems arises. While moving from type-search to voice-search is quite easy, developing navigation through content using voice interaction is a bit trickier (Holoubek & Bowling, n.d.). The main difference between designing navigation for graphical user interfaces (GUIs) and voice user interfaces (VUIs) is the difference between the two senses, visual and auditory. Using the eyes to gain information makes it possible to gain a lot of information during a short period of time by skimming through the content. Our ears are not as fast at perceiving information, e.g. Yang, Sobolev, Tsangouri, & Estrin (2018) previous work has shown that visual content is perceived nine times faster than auditory content.

The use of smart speakers such as Google Home and Amazon Echo has become much more frequent in English-speaking countries and as the virtual assistants start speaking more languages the use of them is predicted to grow in other countries as well (Virji, 2016). In Sweden, the first Swedish-speaking smart speaker launched 24th of October 2018 (Karlsson, 2018).

This thesis will be done in collaboration with Swedish Radio who is currently working on adjusting and adapting their digital channels for voice interaction. Their goal is to make all their digital content adapted for voice interaction with smart speakers such as Google Home and Amazon Echo. Swedish Radio have an existing application for both Google Home and Amazon Echo, where the Google Home application is in Swedish. It is from that application possible to reach all their content by naming what to listen to by its exact name or ask to be surprised and a random episode will be played (Suneson & Björnsson, 2018).

As part of their project they are facing multiple challenges, such as how to change the name of their content to make it adapted for voice search. Another challenge is how a user can navigate through their content using only their voice instead of visually navigate using the mobile application or website. As of today, users not knowing exactly what to listen to have a hard time using the voice-based application (M.

Suneson, personal communication, February 6, 2019). This thesis aims to focus on the latter, to design for navigation of digital content through voice interaction, thus how future users not knowing what to listen to can navigate through the content.

1.1 Research problem

Voice user interfaces have been around for a long time, but it is not until recently that they have become widely commercially available (Dasgupta, 2018). But compared to when designing for graphical user interfaces, designers of voice user interfaces does not have universal standards to rely on. Dahl (2009) also argues that because of the nature of VUIs, it may not even be applicable to have standards but rather design guidelines and best practices. Since Dahl's statement in 2009, several sets of design guidelines for VUIs have been created (Amazon, n.d.-a; Dasgupta, 2018; Google, n.d.-b,-c; Pearl, 2017). But according to the authors own observations, these are far from complete and does not cover the previously stated problem with designing for navigation using speech.

Other problem areas when designing voice user interfaces is the challenge of understanding human conversation and the limitations of the technology. The way humans communicate with each other is a complex system that researchers yet try to fully understand. We do not only communicate using words, but also with how we say those words (Traxler, 2007). Something that can not yet be replicated by todays existing technology. In addition, Pearl (2017) states that even though there is much that VUIs can do, there is yet also a lot that they cannot do.

Given that there are yet no standards as well as the challenge of understanding and technically implement the complex human communication system, it seems like the usability of todays systems is back to the dark ages of interface design (Budiu & Laubheimer, 2018). A test done by Budiu & Laubheimer (2018) indicates that interactions with todays systems that goes beyond a simple question and answer does not meet the requirements for good usability. For the most simple interaction, the usability reaches a minimum level of requirements. Despite this, users still seem to be positive towards using them for what they do good, and the test shows that despite technical limitations, it is possible to improve the user experience with more thorough design work (Budiu & Laubheimer, 2018).

The main problem when designing VUIs is the key difference between auditory and visual perception, creating a change in how to design for them as well. Humans are not able to perceive auditory information as fast as visual information, which puts more demands on the user (Yang et al., 2018). For auditory information, the

user needs to keep more in their head and can not skim or jump back and forth through content as in a GUI (DAC, 2018). Despite all its challenges, there are several advantages using voice interaction instead of visual, such as those mentioned by Pearl (2017): Speed, hands-free, intuitiveness, and empathy. It is fast, does not occupy the users hands, it uses the intuitiveness of speech and can convey more information than written words. Dasgupta (2018) also mentions the advantage of the fact that humans assign personalities to everything around them. This phenomena is especially strong in these systems because of their ability to understand spoken words and speak back, which enables the user to create an even stronger bond with the system.

1.2 Aim

The aim of this thesis is to explore the possibilities and constraints regarding voice user interfaces, and therefore begin the work towards creating standards when designing it. More specifically, the aim is to study how to design for navigation using voice, with an expected end result of design guidelines helpful to Swedish Radio, but also other designers of voice user interfaces focusing on radio content.

1.3 Research question

What should be considered when designing a voice user interface for navigating radio content using a smart speaker?

1.4 Delimitations

As to fit the scope of this project, some delimitations have been set. To stay focused on our research aim, general guidelines for VUIs that already exists in the literature were excluded from the set of guidelines created during this research. These existing general guidelines were however used when designing the prototypes. The guidelines created in this report is a complement to general guidelines for VUIs when designing for navigation through content. It was also decided to not test error handling during this project, as error is a complex matter and could be a project of its own to study.

Swedish was the only language used during testing and in prototypes, due to Swedish Radio's content being mainly in Swedish but also to facilitate for the researchers to find users that could use their native language during testing. Only one smart speaker, Google Home, was used during the project as it is the only speaker which supports Swedish.

1.5 Ethical issues

Smart speakers have faced several ethical issues since they were first introduced, most of them regarding privacy (Dickson, 2018). Potential users have been concerned about whether a smart speaker constantly listens, calling it a "smart spy" (Stokel-Walker, 2018). There is also concerns regarding if the companies behind the smart speaker stores and sells information about the user to other companies (Williams, 2018).

Google claims they do not sell data to other companies. They store information only to "...make our services faster, smarter, more relevant, and more useful to you. Google Home learns over time to provide better and more personalised suggestions and answers." (Google, n.d.-a). Swedish Radio states that their app for Google Home and Amazon Echo only collects anonymous information about the user's behaviour and that their purpose of the collection is to "...improve the experience for the end user as well as making content available based on individual users' needs and interests." (Sveriges Radio, 2017b, 2018a). Since this technology will be used during the project, these ethical issues will have to be taken into consideration. When testing prototypes or the existing application a discussion with the participants about their potential concerns and existing privacy policies should be held.

As always when collecting data during user testing, it has to be handled with care. The tests regarding this thesis will have the possibility of handling personal information anonymous, as real personal information such as name, home address or similar will not be needed. Also, the participant information obtained in this project will not be of sensitive character. Nevertheless, the participants will always be informed about their rights and how collected data will be handled, and consent forms will be used when deemed necessary. This to establish that the participants are aware of the situation that they are put in and that their data will be handled with care.

Another potential ethical issue is the dilemma that arises with personalised content. As mentioned, SR broadcasts a vast amount of hours worth of content each year. To guide a user through all content is problematic in a GUI, and even more in a VUI. All content cannot be presented on an equal basis, which means that SR needs to more or less decide for the user which content they are gonna be presented with (Sveriges Radio, 2018b). This causes a risk of creating filter bubbles, a term coined by Pariser (2011). Internet filters are trying to figure out what you like, places you visit or what people like you like so that they can make more accurate predictions about which content you would like to see (Pariser, 2011). By doing this, the filters creates "...a unique universe of information for each of us - what I've come to call a filter

bubble - which fundamentally alters the way we encounter ideas and information"
(Pariser, 2011, p. 10). Sveriges Radio (2018b) is aware of the problem that arises
with personalisation and filter bubbles, and in a report from 2018 they mention
that they are actively working with trying to avoid filter bubbles for their audience.
During this thesis, personalisation and the potential filter bubbles that it can create
needs to be handle with caution in order to give each user a fair representation of
the available content.

2

Background

This chapter consists of a thorough description of Swedish Radio and their work with smart speakers. Further, the domain and problem area of this project will be elaborated upon, including voice user interfaces, visual and auditory perception, and smart speakers and virtual assistants.

2.1 Swedish Radio

Swedish Radio (SR), founded in 1925, is a public service company in Sweden that works independently and without commercials to create radio content for Sweden's population (Sveriges Radio, 2008, 2011). Their mission is to be the general public's radio and as for their vision, it is to be Sweden's most important media company and leading cultural creators. The core values of Swedish Radio is independent, trustworthy, open, and innovative (Sveriges Radio, 2018b). Swedish Radio broadcasts to five million people everyday and has 40 different channels, whereas four of these are most widely known: P1, P2, P3, and P4 (Sveriges Radio, 2011, n.d.). Every channel has their own programs, which in their turn has corresponding episodes. In total, SR broadcasted 176 304 hours of content in 2017 (Sveriges Radio, 2018b).

During the year of 2017 a new development team was created, which since then is working with making Swedish Radio's content available through voice-based platforms, such as Amazon Echo, Apple TV, Sonos, Google Home, and Windows platform (Sveriges Radio, 2018b). During 2019 Swedish Radio's innovation team was also assigned to work with the development of their voice-based applications (L. Gomes, personal communication, February 6, 2019). Swedish Radio's first voice-based application, for Amazon Echo, was created during a Hackaton late 2016 and in the beginning of 2017 the application was launched to be used by the public (Sveriges Radio, 2017a). Their goal with the development of voice-based applications is to make all SR content available through voice control. They cannot afford missing out on this expected paradigm shift, and claims that being too late in following the technical development may devastate the company as well as their core value of being innovative (M. Suneson, personal communication, February 6, 2019).

As for now, SR has two different teams, The Innovation team and Play-X Voice, which are currently working on the development of making their content available through voice control. The teams have developed a persona based on previously executed focus groups. They are also in the upstart of mapping out the differences between synthetic and natural speech and if these can be used in combination for the intelligent assistant without compromising the user experience (L. Gomes, personal communication, February 6, 2019).

As previously mentioned, they are also working on making the names of their content suitable for voice control. An example is the program *Sommar och vinter i P1*, which is more widely known as Sommar i P1 or Sommarpratarna, but asking for Sommarpratarna would not result in the right program, since it is not the correct name. In order to solve this problem, programs can be assigned multiple names (T. Björnsson, personal communication, February 6, 2019).

Renaming content is a suitable solution for users who knows what they want to listen to but do not know the correct name. But there are also users that do not know what they want to listen to. How can also those users' needs be met? The next step for the two teams at SR is to work on, and in the future release, the next version of their voice-controlled application for Google Home, in which case obstacles like this needs to be solved.

2.2   Voice user interfaces

Dasgupta (2018) explains voice user interface as a model of interaction between a user and a machine using at least partly voice. One part of the technology behind VUIs is speech recognition, which is not a new technology. As early as in the 1950s the first steps was taken towards technology understanding spoken language (Pearl, 2017). In the year of 1961 IBM created Shoebox which understood 16 words. This was the beginning of Automated speech recognition (ASR) and Natural Language Understanding (NLU), but it only dealt with speech recognition and not with the other part of a voice user interface - the creation of a voice. This was experimented on even earlier, beginning in 1939 (Dasgupta, 2018).

During the following decades further developments were made, trying to make it possible to distinguish more words. During the 1980s advances were made and in the 1990s the first system which anyone could talk to was introduced. The first steps toward voice user interfaces was the interactive voice response (IVR) systems. Using only your telephone and your voice, without talking to another human, you could for

instance book a train ticket or ask for information and it was the IVR systems that made it possible by recognising the spoken words from the user on the phone. There is a lot to learn from IVR design when continuing the development (Pearl, 2017). Voice interaction have been researched for many years now and the technology have come so far so that it is time for designers to make it more adapted for the users, just as it was when GUI became mainstream in computers, designers had to make it more user adapted and less clustery (Dasgupta, 2018).

Pearl (2017) calls the current developmental situation "the second era of VUI" with both mobile apps like Siri and Cortana as well as devices controlled only by voice such as Google Home and Amazon Echo under constant development. She continues with describing that we are in the infancy of this era, even though there are several things which can now be done by speaking, there are also many things that are still troublesome. Even so, there are many advantages with using speech instead of the conventional graphical user interface (Pearl, 2017).

Pearl (2017) mentions four important advantages with voice interaction compared to visual: Speed, hands-free, intuitiveness, and empathy. Regarding speed, a recent study made by Ruan, Wobbrock, Liou, Ng, & Landay (2017) indicated that using speech to write text messages were faster than simply texting by hand. Voice interfaces also have the advantage of not requiring the users hands (hands-free), making it possible to use while for example cooking or driving. Most humans are also born with the ability to talk, making it an intuitive interaction mode. Lastly, speech conveys more information than simple written text, such as tone, volume and emphasis (Pearl, 2017). In addition to speed, hands-free, intuitive, and empathy, Dasgupta (2018) also brings up the advantage of persona. Humans have the ability of assigning personalities to everything, whether it is human or not, and this also applies to computers, something which is called anthropomorphism. When making it possible to talk to a computer, this will be even more strengthened, enabling the user to create a stronger bond with its system (Dasgupta, 2018).

Just as there are advantages with voice-interaction, there are also some factors that make users uncomfortable using it and Pearl (2017) mentions the following four: Public spaces, discomfort speaking to a computer, some users prefer texting, and privacy. Speaking to a device in a public space may be awkward, but something which is possible to overcome as the interaction mode will gain common usage. It will however be troublesome if everyone talks to a device, especially in a limited space, easily resulting in a noisy environment. The device could also have trouble distinguishing who it is supposed to listen to in such an environment. It may also be a challenge to recruit users which are fond of typing on their phone or computer to

suddenly shift to another type of interaction-mode. Privacy is a constantly recurring subject regarding VUIs and there are more than one problem regarding this. Just as well as users may not want to talk about private affairs, such as health issues, out loud when others can hear, it is also troublesome what the system does with such information. Another worry may be about the system starting to read something out loud which may embarrass the user (Pearl, 2017). Furthermore, Nass & Brave (2007) brings up the fact that to some extent, users relate to voice interfaces as they relate to other humans. It is hard for humans to disregard our beliefs of a real conversation when speaking with a machine, even if fully aware of it. It is also in our nature to simply understand a conversation's all subtle signs, beyond spoken words. This is very complex for a system to mimic, and a system not understanding it can result in a bad user experience, especially if its goal is to mimic human conversations (Nass & Brave, 2007).

Pearl (2017) is very resolute when conclusively stating that VUIs are not for every occasion and every new application. It is not something which should be used just because its a current trend, the use case should actually benefit from it. One of the advantages mentioned earlier is that of VUIs being hands-free. Listening to radio, podcasts or other auditory material is a hands-free activity, and the only time hands are needed is when to find and start what to listen to. This is why developing a voice based application for Swedish Radio's users is beneficial, enabling them to ask for, start and listen to SR's content only using speech as input.

2.2.1 Command and control and conversational interfaces

It is common to talk about command and control, and conversational interfaces when designing VUIs. Today, most VUIs are command and control, which means that to start talking with the device, the user must indicate it somehow. When this is done, the system gives feedback of it being "awake" by either a non-verbal sound or by a visual cue like lighting up. It often also indicates when it has identified that the user stopped speaking using the same way. When an answer has been given, the conversation is most typically over and to talk to it again the user has to do the procedure all over again and most often the system does not remember the previous conversation (Pearl, 2017).

Conversational VUIs are when an interaction with a VUI goes beyond one turn. Context is most typically included, the system remembers what the turn before was about and keeps history of conversations. This makes use of a more natural way of turn-taking, for longer conversations, and there is no need to use a wake-word for every turn-taking. Conversational VUIs are however more technically demanding, which is why command and control is still the most common implementation (Pearl,

2017).

2.2.2   Navigation in voice user interfaces

In a GUI, an overview of the content is usually provided enabling the user to navigate on their own terms in the content. For a VUI, this is not the case. Indications on where in the navigation the user is, how to move forward or backward as well as which options that are available is more difficult to communicate to the user (Furqan, Myers, & Zhu, 2017). Furqan et al. (2017) refers to this as the "invisible" nature of VUIs which can make it more difficult for users to navigate the system. Navigation is a well-identified problem when designing for VUIs which still stands unsolved.

2.3   Visual and auditory perception

When we eat, read, or listen to a song, our brains receive sensory information. Sensory information can be visual, auditory, touch, taste, or smell. This information are then interpreted by our brain in order for us to be able to understand it. The organization, identification, and interpretation of sensory information is called perception and is vital for our everyday life (Schacter, Gilbert, Wegner, & Nock, 2014).

Just as sensory information is categorized by different senses, so is perception. For this thesis, visual perception, auditory perception, and their differences are relevant. Visual perception is the interpretation of visual information, and an important note is that visual information does not have to be the same as visual perception. An example of this is optical illusions, which is perceiving something that does not mirror the reality (Allwood & Jensen, 2012).

A part of visual perception and optical illusions are scene gist; our ability to understand not only the individual objects that we see but the whole context. Scene gist can even affect our memory, resulting in people remembering that they saw something they did not, for example a binder in an office just because binders usually are found in offices. Studies have shown that only a few amounts of objects is needed for us to interpret an environment as complete and fully detailed, and that we can determine the meaning of a scene within 100-120 ms (Rensink, 2000).

Visual information travels fast, and a study by Yang, Sobolev, Tsangouri, & Estrin (2018) has shown that visual content is perceived nine times faster than auditory content. While we read, we are able to read up to 300 words per minute. It also enables skimming, jumping forward and backward, and creating an overview, which lets us gain knowledge at rates of thousand words per minute. This can

be compared to speech, which uses approximately 60 words per minute while also being serial (DAC, 2018). Visual perception is not only about what we see, and Sutherland (2015) means that we recognize words as pictures building mental images as well as hear them spoken aloud. Evidently, visual sensory information is much more than just the individual objects or words that we perceive and our brains add on what is missing in order for us to interpret it as a whole (Allwood & Jensen, 2012).

Similar can be said for auditory sensory information. Spoken words conveys more than just their literal meaning, and both irony and emotions can be communicated by tone and context (Traxler, 2007). Sound waves can also give us spatial information as well as the ability to recognize and distinguish between different sounds. To be able to recognize different sounds, properties like amplitude and frequency are important but also the sounds' extension in time and timbre (Allwood & Jensen, 2012).

Sound can also have a certain pulse or rhythm (Allwood & Jensen, 2012). Pulse or rhythm for spoken language is called prosody and is the speech's melody. Prosody can be divided into two different categories; linguistic prosody and non-linguistic prosody. Linguistic prosody refers to the aspects of speech that provides cues for how words should be organized into phrases and cues. Non-linguistic prosody is the aspects of speech that refers to the speakers mental state, and is the reason to why we for example can understand that someone is sad not by the words they are using but how they say these words (Traxler, 2007). Prosody plays an important part in spoken language and together with the previously mentioned sound properties they add on dimensions of information in comparison with written words (Allwood & Jensen, 2012; Traxler, 2007).

Visual perception gives the user an overview and visual content is perceived faster than auditory. Auditory content in the form of spoken words has on the other hand multiple dimensions which written words lacks. The differences between visual and auditory senses plays a key role when designing for GUIs versus VUIs as they are reliant on two distinct perceptions.

2.4 Smart speakers and virtual assistants

Amazon Echo and Google Home are the names of two popular devices that are currently in the top of their domain. They are smart speakers, and in 2020 their worldwide spending is expected to hit $2 billion (Lau, Zimmerman, & Schaub, 2018). A smart speaker is in its narrow definition a speaker that does more than just emitting sound. However, when we talk about smart speakers we are usually referring to a speaker with a built-in virtual assistant, Amazon Echo has Alexa and Google Home

has Google Assistant, which can be controlled by voice-commands (Kabir, 2019). To enable voice-commands, a wake word is needed. After the wake word has been uttered, the smart speaker is listening to all the user has to say. Abilities differ from speaker to speaker, but some examples are playing music, setting a timer, or reading the news. A study by Nielsen (2018) shows that users mostly uses their smart speakers to listen to music (90 %), search for real-time information (81 %), search for factual information (75 %), listen to news (68 %), chat with your voice assistant for fun (68 %), and use alarms/timers (68 %).

The development of intelligent assistants and later on smart speakers has not been without obstacles. For an intelligent assistant to work several pieces need to be in place; voice recognition, speech comprehension and speech synthesis are crucial for an interaction to take place. Speech recognition, the ability for an intelligent assistant to convert the users' speech to text, currently has an error rate of about 5.5 percent. Whereas human parity is said to be at 5.1 percent, speech recognition is not far behind (Seon, 2017).

In addition to being able to recognize what the user is saying, the intelligent assistant also has to understand the meaning of the content, which is called speech comprehension. Exactly how intelligent assistants do this differs between brands, but in general the users' words and tone is analyzed by an algorithm connected to servers which then is matched with a corresponding command. All commands do not need information taken from servers, so meanwhile the algorithm is looking for an answer to the command, the device tries to solve the issue on its own. For example, setting a timer is something a smart speaker can handle without the involvement of servers, unlike a request for the next bus departure (Fischer, 2017).

Finally, the intelligent assistant needs to produce a spoken answer to the user. Speech synthesis, or text-to-speech, is used in order for the assistant to be able to talk. Mctear, Callejas, & Griol (2016) divides speech synthesis into two stages: text analysis and waveform synthesis. In the stage of text analysis, the text is fractioned into multiple phonemes and prosodic information. These parts are then converted into a sound wave that lastly can be output as spoken words (Mctear et al., 2016). Several techniques exists for speech synthesis, and companies are continuously competing to create a synthetic speech that sounds like natural, human, speech. Google uses WaveNet, while Apple and Amazon uses concatenative speech synthesis (Vincent, 2018).

Concatenative speech synthesis is based on recordings of human speech collected from a single voice actor. The recordings are cut into tiny chunks which are then combined to create complete utterances. Hence, the voice actor does not have to

record every word in a lexicon, but every word can be created using the speech synthesis (van den Oord, Walters, & Strohman, 2017). A problem with concatenative speech synthesis is that the recordings' prosodic attributes remains, resulting in unnatural sounding speech (Tatham & Morten, 2005). WaveNet instead uses a deep neural network to create individual waveforms from scratch. The neural networks is trained with a large dataset of speech samples to be able to learn the underlying structures of natural speech, resulting in a speech synthesis containing intonation and similar features that is crucial for speech to sound natural (van den Oord et al., 2017).

2.4.1   Google Home and Swedish Radio

Google Home and Amazon Echo are generally much alike, but what makes Google Home special is the fact that it is made by Google, with all that comes with it. Google possesses a great deal of information and knowledge about the world, peoples preferences as well as habits, and Google Home have access to all of it (Seifert, 2016).

Google Home enables others to create content for it by offering them to create so called actions. An action is an interaction that can be developed by others than Google, and can be seen as an application for VUIs. A company that wishes to create an action for Google Home has to get it approved by Google first. All available actions are then accessed by the utterance "Hey Google, talk to..." (Google, 2019).

To access Swedish Radio's action from Google Home, the user has to say "Talk to Swedish Radio". Their action is then accessed, and Swedish Radio's own assistant presents itself and can help the user to listen to a channel, program or episode by their choice. The assistant is also able to help the user by so-called fuzzy match. Fuzzy-match means that the assistant can pick a content for the user to listen to based on different keywords. For example, a user could say "Play something with Avicii" instead of being obligated to name a specific program in order for it to play (Suneson & Björnsson, 2018). Also, the user can ask the intelligent assistant to surprise them in which case the user would be assigned a randomly picked episode from SR's 10 most popular programs (T. Björnsson, personal communication, February 6, 2019).

3

Theory

This chapter aims to bring up relevant research, existing frameworks, and general considerations for the work of this project. Included in this chapter is design practice, designing voice user interfaces, and design guidelines.

3.1   Design practice

Products lacking thorough design work often leads to users experiencing severe frustration. A well designed artefact, on the other hand, will result in users gladly wanting to spend their money on it as well as recommend it to others. This, while also contributing to the satisfaction of actually using something which helps achieving user goals. Therefore, design is an important domain in wanting to succeed with a product (Cooper et al., 2014). The overarching domain of designing products which a user is supposed to interact with is interaction design. Preece et al. (2015) explains interaction design as "designing interactive products to support the way people communicate and interact in their everyday and working lives" (p. 8).

3.1.1   Human-centered design

Since the first rise of technology used by the public, design work have become more and more involved in the overall development, leading to products being more adapted for the attempted user. The field of interaction design has exploded and there are now an endless amount of books and papers on the subject, giving rise to design continuously getting better. However, the rapid pace of new technologies arising constantly sets new demands on design work, and most often the same mistakes are made in the beginning of a new technological field. New studies are required in order for a new field to set its principles for good design, putting designers at constant challenge (Norman, 2013).

Norman (2013) states the solution to this challenge to be human-centered design (HCD). HCD is an approach claimed to contribute to good interaction design, and means involving the user throughout the design process (Preece et al., 2015). Further,

it is a practice where the user is put first, and have the goal of designing to accommodate human needs, behaviour and capabilities (Norman, 2013). To achieve this, evaluation of the design should be done at different developmental stages, preferably involving users whenever suitable, and strive to adapt the product accordingly to user feedback (Preece et al., 2015).

3.1.2 User experience

In the emergence of interaction design, usability was in focus. In todays design practice, good usability is still fundamental, but not everything. In complement, another important concept has come and practically dominated the field, user experience (UX). Usability and user experience are tightly intertwined, but while usability is more pragmatic, user experience goes beyond that. Efficiency, ease-of-use and learnability are often talked about regarding usability while UX is explained as a user's resulting emotions, their experience, after interacting with a product. Usability is an important impact on how the user experiences the interaction, but much more contributes to it, such as usefulness and look and feel (Hartson & Pyla, 2012).

3.2 Designing voice user interfaces

When designing a voice user interface, Cohen et al. (2004) advocates using a human centered design approach. Just as an understanding for human cognitive abilities and restraints are important, it is also important to include users throughout the design process to gain insights about their experience of the interaction. It is never possible to foresee this experience, or every possible outcome of an interaction, therefore testing is crucial (Cohen et al., 2004). Voice user interface design is tricky, since the medium of it, speech, comes with many expectations. An essential rule when designing is to understand and utilize those expectations, to then test if the implementation matches them (Cohen et al., 2004).

Continuing, Pearl (2017) presents four concrete factors which are important when designing VUIs, also making use of human expectation of speech. Those are: conversation, confirmations, conversational markers, and handling errors. These are further explained in the following subsections, aside from handling errors as this is not in the scope of this project.

3.2.1 Designing conversation

For conversation design the first thing mentioned is to not ask questions if the answer is something which will not be understood. The human language consists of lots of

examples where the answer to a question is not the literally correct answer, such as "could you play some music for me" where the literal answer is yes or no, but that is not the underlying meaning of the question. When designing a VUI, avoiding such questions is crucial. Another important guideline is to set the user expectations as early as possible. An example of how to do this is to do a first-time usage tour. People usually do not have a hard time enjoying completing an interaction using voice, whereas that first-time interaction is really important to design well. Also, usually do not tell something like "this has been done successfully", it is unnecessary information as a user expects it to be successful when either confirming that it is being done, or by the desired outcome happening. Rather than over-confirming, say if it did not work and make it possible to cancel. A crucial guideline is that of discoverability, how a user can acknowledge that it is possible to speak with the device and what can be said (Pearl, 2017). Pearl (2017) recommends making use of natural ways of interacting and/or natural phrases as the interaction then may happen by itself. It is also good practice to give examples, such as 25 January 1993, instead of just stating what to say, such as birth date day month year, which makes it easier to grasp for the user (Pearl, 2017).

3.2.2 Designing confirmations

Pearl (2017) also brings up the importance of designing confirmations to make sure the users feel understood or letting them know they were not. When designing confirmations, one important consideration is to not over-confirm, it is unnatural and annoying. When determining the right confirmation strategy, consider what the consequence for getting it wrong would be. If transferring money, the consequences can be quite severe. Playing music, on the other hand, is not as sensitive to errors. What kind of confirmation is the most appropriate should be based on the context of use, there are explicit, implicit and a combination of them both which can be used. Explicit confirmation is basically to force the user to confirm whereas implicit confirmation is letting the user know what was understood, but not asking them to confirm. When using implicit confirmation it is most common to allow to cancel or go back (Pearl, 2017).

For voice only VUIs it is important to craft confirmations carefully. One way is to use confident thresholds which is how the speech recognition engine lets you know how well it thinks its performed. For example if the system recognises "yes please", it may be 45% sure that is what you said. Depending on how severe consequences there might be if the system got it wrong, different thresholds is set with different types of confirmation. It is also favorable to take advantage of existing assets when using voice-only interactions. One example that is easily forgotten is the light ring on Alexa (Pearl, 2017).

There are many different ways to confirm, Pearl (2017) brings up five different ones. The first one is three-tirered confidence. This means that the system will explicitly confirm between a certain threshold, reject anything with a lower confidence and implicitly confirm anything above the threshold. The next one is implicit confirmation which logically enough means to always use implicit confirmations. It may also be cases where only the answer is provided, something which is appropriate to use when confidence is high and when trying to be conversational. Non-speech confirmation is another one. This is when confirmation relies on completing an action, for example turning lights on. It is not necessary for the system to say that it is turning the lights on since the light per se will be the confirmation. Audio confirmation can be used, however, if there is a delay or if the action is not visible like turning on the oven when the user is in another room. Another non-speech confirmation is landmarking, that uses audio but not speech. It can for example be a brief distinctive sound guiding the user as to where they are in a "menu". Not confirming exactly what the user was saying is called generic confirmation and it implies for more conversational systems. It is better for mimicking human-human conversations and when used right, it is okay for computers to use. Lastly, there is also visual confirmations for multimodal systems (Pearl, 2017). As this project only focuses on voice-only systems, this will not be presented.

3.2.3   Conversational markers

To make the conversation with a VUI even more natural it is good to make use of conversational markers. Instead of only asking questions, and saying nothing more than those questions, conversational markers can be used to help the user understand where in the conversation they are and that they are being understood. There are three types of conversational markers: Timelines, acknowledgements and positive feedback. Examples of timelines are "first" and "lastly", and of acknowledgements "got it" "I understand" and "I am sorry". Lastly, examples of positive feedback are "good job", "well done" and "nice" (Pearl, 2017).

It is often quickly understood where the conversation sounds unnatural or confusing for the user if read out loud, where conversational markers can be put to make it sound more natural. The conversational markers should be used with care, though, as to not disturb the user to try to sound human when it is obviously not (Pearl, 2017).

3.3   Design guidelines

Design guidelines are recommendations for good practice when designing. Design guidelines are between principles in design and rules for implementation. For example, the design principle "an interface should be easy to use" can be paired with the design guideline "text should be easy to read" and the rule for implementation "background:white;font-color:black;font-size:20px". Thus, design guidelines should help designers on how to implement a principle without restricting them as a rule for implementation could (Interaction design foundation, n.d.).

The companies behind Amazon Echo and Google Home have presented their own design guidelines for developers of applications to their devices (Amazon, n.d.-a; Google, n.d.-a,-b). There are also general guidelines regarding voice user interfaces, both Pearl (2017) as well as Dasgupta (2018) have presented some of these in their printed works. Also, a similar project to this one by Päärni (2017) investigated how radio listeners wishes to interact with a VUI to listen to on-demand radio content which resulted in seven design guidelines. Below, a summary of the previously mentioned guidelines are presented.

Additionally, several design guidelines for GUIs exist. Shneiderman et al. (2009) as well as Nielsen (1994) studies have resulted in two sets of guidelines which are typically referred to regarding GUIs, but it does not mean that they are irrelevant for designing VUIs. As design gudielines for VUIs are yet not as well-established as those for GUIs, these guidelines are also presented below.

Table 3.1: *Amazon (n.d.-b) design guidelines.*

Guideline	*Explanation*
Persona	Use a persona to give the user a coherent experience.
Be informal	Use an informal language so that the assistant feels like a dear coworker.

Engage with questions	Questions is a natural cue to speak and should be used to help the assistant determine how to assist the user. Avoid asking questions in the middle of a prompt, as it may cause the user to answer immediately. Also avoid directly asking yes/no questions, instead ask the user what they want to do. Avoid either/or questions if there is more than two options. If there is more than two options, present it in a short list and ask the user which one they would like to choose in the end.
Be contextually relevant	Always mind the context in which a conversation is taking place. When presenting different options, always present the most relevant option first.
Use contractions	Use contractions to mimic spoken natural language more than written.
Vary responses	Use a variety of responses for similar interactions to prevent the assistant sounding too repetitive and robotic. This is especially important for answers which are likely to be repeated in a row, like error prompting.
Be brief	Always strive for brief prompts. It is important to be precise as to not confuse the user.
Write for engagement	Conversation should feel engaging whether it is the first interaction or the 100th. Therefore, customize the experience for different types of users, and be able to phase out information that long-term users will learn over time.

Table 3.2: *Google (n.d.-c,-b) design guidelines.*

Guideline	*Explanation*
Expect users to be informative	Users often offer more information than is asked of them. Avoid forcing the user to repeat information they have already given.

Get the dialog back on track	When a user states information in a manner that the assistant cannot understand, the assistant should use lightweight and conversational error handling to get the dialog back on track without drawing attention to the error.
Move the conversation forward	The assistant should be cooperative and informative and striving to advance the conversation and avoid abrupt endings.
Optimize for relevance	Aim for relevance, saying too much should be avoided just as saying too little. Always focus on what is relevant from the user's point of view.
Listen between the lines	Be able to distinguish between what a user says and what they mean as this two can differ greatly.
Speak clearly, in plain English	Use familiar words and phrases to make the conversation sound natural while also reducing cognitive load for the user.
Context	Always have the context in mind when understanding what the user meant. This should also enable conversations to continue without having the user to explicitly state so.
Variation	Use different kinds of prompts to similar questions from the user in order to avoid the assistant from sounding to monotonous or robotic. Prompts that are used often are in higher need for variation than the more seldom ones.
Turn-taking	Use turn-taking in conversations so that the user feels involved while also keeping the conversation in sync.
Focus on the user	The conversational elements should focus on the user and not the assistant.
Don't launch into monologues	Avoid creating too long responses. Let the user take their turn and avoid going into details unless it is clearly beneficial for the user.
Lead with benefits	Give the user a reason to why they should do something, "To get what you want, do this thing".

Avoid niceties	Niceties can feel distant and formal, avoid them to keep the conversation friendly.
Use contractions	Strive for making the conversation as close to natural spoken language as possible. Avoid spelling out words that normally would be used in a contraction.
Don't provide UI-specific directions	Avoid UI-specific directions and refer to actions or concepts instead. As the field of interaction design evolves, different terms can become out of date.

Table 3.3: *Pearl (2017) design guidelines.*

Guideline	*Explanation*
Novice and expert user	Remember if the user is new or have used the function several times. Do not give a description of how to do things every time, shorten it. Refresh their minds every once in a while with the complete description.
Keeping track of context	Remember what was talked about before so that the system understand when referring to e.g."him" or "there". Remember the last mentioned person and/or place mentioned is a good start.
Help and other universals	There is no perfect VUI, always offer some kind of help-function. Goodbye, go back etc. are good universals to use.
Latency	Design for latency if it is known that it will happen. Tell the user e.g. "one moment..." or non-verbally let the user know, such as a sound. If using this and the delay may become short, make it a bit delayed anyhow so that the beforehand mentioned delay was not a lie.
Disambiguation	When a user does not give enough information, the primary solution is to try to understand through context or already known information. Another solution is to ask for clarification, but clearly state what it is that is not understood.

Table 3.4: *Dasgupta (2018) design guidelines.*

Guideline	*Explanation*
Recognize intent	Think about what the user wants, the intent, and not overloading the user with information. Short and effective conversation. Look for hints about intent in accessible places such as calendar or current position.
Leverage context	Understand context in what we speak. Human-human interaction is bound to context and much information is conveyed through it. Humans use this to create shared meaning, but this non-verbal information is nearly impossible for systems to grasp and make sense of. Until further development, keep the conversation direct and simple and be aware of the drawbacks.
Cooperate and respond	There are two types of conversations, intent-based an casual conversations. For intent-based conversations, the user has an intention with the conversation, such as finding an answer to a question or completing a task. Accommodate questions asked by users, do more than what is expected, such as alternatives or information. Casual conversations are conversations without specific intent. In this case, let the user get to know you, they want to do so by nature when conversing with others.
Progressive disclosure	Progressive disclosure us a technique within interaction design often used in GUIs, with the aim of not overloading the user with information. An example from GUIs is pages of objects when shopping, instead of presenting all objects at once. Same goes with VUIs, it should not tell all options possible at once, rather give examples and then ask if the user wants to hear more options. Do not present more than 3-4 options at once, give the option to ask for more. This adds more steps to the design, but does not need to be negative. Do it the right way and you can keep the user motivated.

Variety	Always present more than one answer to a question so that the user can get a variety of answers when asking the same question. One technique is to randomize several phrases answering the same thing.
Give and take	Take turns talking when conversing. To accomplish this task, humans rely on a rich set of cues, including intonation, eye gaze, and body language. For systems, turn taking must be designed.

Table 3.5: *Päärni (2017) design guidelines.*

Guideline	*Explanation*
Enable user profiles for a personalized experience	By enabling user profiles the assistant can base suggestions on previous user behaviour. Also, it gives the user the ability to manage keywords that can facilitate the interaction.
Allow synonyms for material	Allow multiple names for the same content.
Give recommendations based on behaviour and preferences	Suggestions and recommendations from the assistant should be based on user behaviour and preference settings.
Provide a GUI companion application	Interactions that otherwise would be unnatural or tedious through a VUI can be made in an companion application using GUI.
Guide the user, with both an introduction and informative response	Introduction and informative responses should be used to make the user feel comfortable talking with the assistant. The guidance should be adaptive and depending on the user's abilities.
Tag material with keywords	Use keywords to cluster similar content together. This also enables for searches outside strictly program names.

Table 3.6: *Shneiderman et al. (2009) eight golden rules for interface design, adapted for VUIs by the researchers.*

Guideline	*Explanation*
Strive for consistency	Standardize how information is conveyed, so that the user gets familiar with the system, making it easier and easier to use.
Enable users to use shortcuts	For users frequently using the system and being very familiar with its functions it should be possible to do most often used functions faster.
Offer informative feedback	Give insight to the user where they are and what is happening so that they do not feel lost. The system should also provide fast understandable feedback to the user at all times when completing a task.
Design dialogues to yield closure	Always confirm to the user what the conversation has led to, if it is an intent-based conversation. It could be for example telling them what time and date a reminder was set to.
Offer error prevention and simple error handling	When an error occurs it should be possible for the user to recover from it by for instance be informed of where the trouble lies and make it possible to redo.
Permit easy reversal of actions	It should be possible to reverse simple actions, just as well as a whole set of actions. Knowing this is possible encourages the user to try different functions in the system.
Support internal locus of control	Make the user feel as in control, which they will if the system behaves as expected. This is hard regarding VUIs, since the user is usually not feeling as in total control when conversing with smart devices (Dasgupta, 2018). Designing conversations to behave as expected will however increase the chances.

	The short-term memory of humans are restricted to usually about five things at a time. This is especially important regarding VUIs as all information need to be stored in the users memory, and do not have a visual screen as a reminder. Making use of the ability to recognize rather than recall can be a good rule of thumb regarding how to present information verbally without only relying on the memory.
Reduce short-term memory load	

Table 3.7: *Nielsen (1994) ten usability heuristics.*

Guideline	*Explanation*
Visibility of system status	Let the the user know what is going on. This is done through relevant feedback that is given within reasonable time.
Match between system and the real world	The system should be designed so that it matches the users mental model, meaning as the users expect it to work. Words, phrases, order of information etc. should follow conventions which already exists in the real-world.
User control and freedom	Users will make mistakes, do not design so that they are stuck in that mistake. Emergency exit, undo and redo are fundamental functions.
Consistency and standards	Platform conventions should be followed. The user should never be confused whether functions in the system are the same or not.
Error prevention	Design for prevention of errors to occur. Check for errors during the design phase so that they can as much as possible be eliminated.
Recognition rather than recall	Do not rely on users memory capacity, they should never be forced to remember too much information by themselves. The system should rather constantly remind the user.

Flexibility and efficiency of use	Design for both novice and expert users. Make it possible for novice users to learn, as well as for expert users to use accelerators for a more speedy interaction.
Aesthetic and minimalist design	Keep the information presented relevant at all times. For every less relevant information presented, the more relevant will be faded.
Help users recognize, diagnose, and recover from errors	Error messages should be designed so that the user understand them. They should state the problem as well as suggest a solution.
Help and documentation	The optimal state is to not need help and documentation. If the case is that it is needed, it should be presented so that it is easy to grasp. It should also be searchable, concrete and within reasonable size.

4

Methodology

This chapter brings up problems related to the design domain, mentioning wicked problems, the importance of an iterative design, and research through design. The design process *the Wheel* is presented as well as how to choose the right method followed by relevant design methods.

4.1   Wicked problems and iterative design

When tackling design problems, these are needed to be interpreted like wicked problems. Wicked problems, unlike other problems, have no right or wrong solution, instead their solutions should be valued from good to bad (Rittel & Webber, 1973).

Rittel & Webber (1973) describes that for other fields not facing wicked problems, like mathematics or engineering, the problem-solver can run numerous trials without penalty. However, implemented solutions to wicked problems cannot, and are referred to as "irreversible". An example brought up is that of building a freeway, "One cannot build a freeway to see how it works, and then easily correct it after unsatisfactory performance"(p. 163). This poses a challenge for designers trying to find the right solution, as implementation can have unforeseen negative consequences (Rittel & Webber, 1973). A solution to this is using an iterative design process, enabling the designer to explore, fix, and refine a design several times before implementation, which makes it more likely that the final implementation will be successful (Hartson & Pyla, 2012).

Iterative design is when parts of a design process is repeated in order to improve a product or end result. Designers can for example develop a prototype, test the prototype with users, and then change the prototype according to the previous findings. When changed, the prototype can be tested again, and again, and again, until the designers feel like they have reached their goal with the design (Preece et al., 2015). As mentioned, design problems does not have right or wrong solutions, and it is therefor up to each designer to determine when the iterative process is finished and the result is sufficient for the given purpose.

4.2   The design process

The process of this project will be based on research through design, explained by
Gaver (2012) as when insights are gained from actual design work, which can be
by own practice or derived from others' research. While research through design
is progressing, others have raised their voices about a need for an over the field
agreed-upon standards and processes. Gaver (2012), on the other hand, means that
such standards could result in a form of self-policing that would be too restrictive for
the field of design which should be explorative and creative. A common design pro-
cess model is presented below, which intends to guide designers without limiting them.

The design process *the Wheel* consists of; analyze, design, implement, and eval-
uate (figure 4.1). Analysis is about understanding the research domain, previous
work and the users behavior and needs (Hartson & Pyla, 2012). Cooper et al. (2014)
advises designers to start with defining the scope of the project in the analysis phase.
When the scope of the project is clear, the designers should include a review of
existing work and product by market research and literature reviews, stakeholder
interviews to understand product vision and constraints, and user interviews and
observations to further understand user needs and behaviour (Cooper et al., 2014).
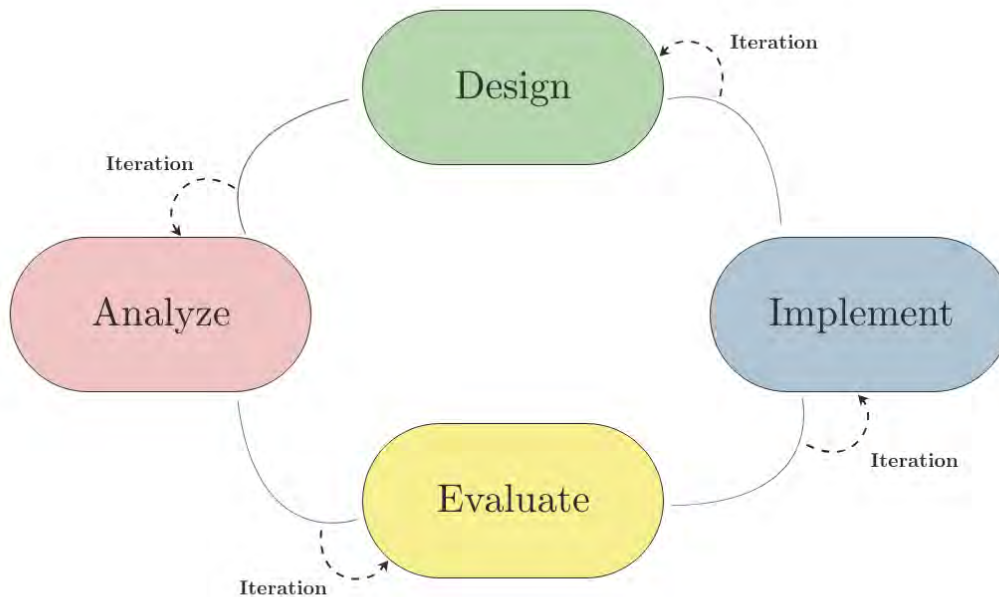
Figure 4.1: *The design process the Wheel with the parts analyze, design, implement,
and evaluate. Illustration made by the researchers.*

The design part represent the process of designing, which could include design ideation, sketching, or creative design thinking. In the implement part, various prototypes of different kinds of fidelity can be created. Evaluation is used to refine the design and to understand if different goals and targets have been achieved. An important note about the design process is that it is not rigorous. As mentioned earlier, it is also important that the process is iterative. The steps of the design process can be walked through several times and in any order (Hartson & Pyla, 2012).

4.3 Design methods

As a designer you may think that you know if a design works or not, but it may not be as obvious to the user and changing things during the development phase is much easier than doing it when it is already launched, as explained in relation to wicked problems. Also, changing the user experience when a product is already out for public use is often complicated. Instead, involving users early as well as often is crucial for a successful design process as well as to a successful end product (Pearl, 2017). To accomplish this, there are an endless amount of design methods out there to fit the purpose of the different design projects. Therefore, choosing which method to use is not a simple task, but still a very important one, since it, together with how it is conducted, decides the outcome of the project (Interaction design foundation, 2016).

There are some guidelines on how to find suitable methods. To begin with, avoid always using the same methods, or the methods that the researcher are already comfortable with. To instead investigate which methods to use, and when they should be used, there is much to gain (Rohrer, 2014). The first thing to consider is according to Hall (2017) "What do I want to know?". The researchers need to thoroughly consider what they want to gain from the method, what the goal is, and what they want to know from the users, how the users can help reaching this goal (Hall, 2017). Continuing, Rohrer (2014) suggests a three-dimensional framework for choosing methods. The axes presented are attitudinal vs. behavioral, qualitative vs. quantitative, and context of use. Considering attitude vs. behavior is about deciding if it is the users own thoughts that will help reaching the goal, or if it is to measure user performance. It is also needed to decide if the outcome of the chosen method should be qualitative or quantitative in order to meet the goal. Context of use refers to where and when the user will use the product being tested, and in line with that compare which methods fits that user context the best (Rohrer, 2014). Considering this framework, comparing methods based on the desired parameter of each axes in accordance with the goal of the research, the researchers are well equipped for choosing a well fitted method.

Further, relevant methods for each step of *The Wheel* will be presented. The methods were picked based on the three-dimensional model presented above to be able to narrow down the amount of methods to be considered. Which methods that were used during this project as well as why will be presented in the process chapter.

4.3.1 Methods for analysis

Ideation session
An ideation session is most commonly when designers work in a group to come up with new ideas to a problem. A number of ideation methods can be used during an ideation session. For an ideation session to be successful, a non-judgemental environment and a clearly defined problem is crucial. The goal of an ideation session should not be to find the best possible solution, but rather to find new angles and generating multiple kinds of possible solutions.

Design workshops
Design workshops usually involves non-designers as participants in a form of co-design session. There are no strict rules to what a design workshop can consist of, but the aim is to gain understanding of the user's world and needs. A workshop could include creation of mock-ups, sketches, or storyboards as well as role-play interaction. A critical part of workshops is the elements of planning and carrying out. A good workshop demands good planning, and usually a lot of design team facilitators relative to the amount of participants (Martin & Hanington, 2012).

Role-playing
In role-playing, designers take on the role of the user to act out different scenarios. It is important that the participants are willing to play along realistically. To be able to carry out this method, no more than the participants themselves are actually needed but props can be used as well if fitting. The participants acts upon a user scenario and is also guided by a general situation or suggestions for actions being performed which should be clearly communicated by the facilitator. It can be difficult for researchers to document all that is happening during a role-play, which is why it is recommended to end each role-playing session with some sort of a review for the participants so that their genuine feelings and thoughts can be shared (Martin & Hanington, 2012).

Affinity diagram
Affinity diagrams is used to draw conclusions based on research and observations. Previously collected qualitative data is written on post-its, one post-it per insight. These notes are then put on a whiteboard or similar so that the design team has an

overview of all pieces. The team then re-organize the notes, trying to create clusters of similar insights. Affinity diagram is an inductive exercise, working from the bottom-up instead of organizing notes in predefined categories (Martin & Hanington, 2012).

Content analysis

Content analysis is an systematic technique for analysing qualitative research, especially for data that consists of rich descriptions which can be challenging and exhaustive to analyse. There are two approaches to content analysis, inductive and deductive where inductive is the most commonly used. When using inductive content analysis deriving categories or codes from a data set is usually the initial action. From those categories and codes, overarching themes are created which divides and explains the content. When using deductive content analysis, the categories or codes are created before analysing often using theoretical frameworks (Martin & Hanington, 2012).

Focus groups

Focus groups are a qualitative method to gather opinions, feelings, and attitudes from participants. A usual setup for a focus group is participants gathered around a table, while having a moderator present but not seated at the table. Moderators plays a crucial part in focus groups as an atmosphere where the participants can quickly accept one another as peers is the core of the method (Martin & Hanington, 2012).

Interview

Interviews are used to gain direct contact with participants, and are best conducted in person so that nuances of personal expressions and body language are captured by the interviewer. Interviews can be structured, semi-structured or unstructured. In structured interviews the interviewer has a set of predefined questions, while in unstructured interviews the interviewer mostly works from a set of topics that the conversation should touch upon (Martin & Hanington, 2012).

When conducting an interview, an interview protocol can be useful. That is, a protocol containing a script for the beginning and end of the interview as well as the interview questions with associated probes to keep the interview on the right track. The initial script should contain informed consent and the goal of the study. Some guidelines when writing interview questions are to keep them open ended, to start with easy questions, such as background, and then advance and to write broad questions so that the interviewee can speak freely. During the interview, using an audio recorder is most often advantageous over taking notes so that the interviewer is sure not to miss important information as well as be fully focused on the interview

(Jacob & Furgerson, 2012).

Brainstorming
Brainstorming is a method used for ideation in groups. It is based on creating a safe space for the participants so that ideas can be generated without a focus on judging the quality of them. By generating ideas in groups, participants can engage with each other, listen, and build on others ideas. To have an effective brainstorming session, a time line and a clearly defined problem statement is crucial. Participants should also be encouraged to build on each others ideas, as brainstorming works well when participants own thinking is elicited by others ideas (Dam & Siang, 2018).

4.3.2 Methods for designing

Scenarios
Scenarios is a written narrative about a future use of a product. The narrative should be written from the users perspective, and can be based on research, personas, or the design team's understanding of the target group. It is recommended to have at least one scenario for every persona. Scenarios can help design teams understand the use of a product in a users life, while also being able to keep the development focused on the user and not technical requirements (Martin & Hanington, 2012).

Sample dialogs
Pearl (2017) suggests beginning designing with the help of sample dialogs. Sample dialogs are, just as it sounds, parts of possible interactions with the intended VUI. It should be written as a movie script, with a dialog going back and forth. Sample dialogs are especially good for getting started with designing the actual conversation, and avoids the technique of instead writing prompts one at a time, resulting in conversations which may be unnatural and stilted. Other advantages with sample dialogs are that they are low tech, a beginning of designing the user experience, and easy to work with as it is something most people can understand and give feedback on. This method is most typically performed by choosing a couple of (most often five) most common user scenarios for the intended VUI. For each user scenario a couple of dialogs should be written based on what is believed to be the most optimal interaction, and also some for when things go wrong. When done with the dialogs, they should be tried by reading them out loud to acknowledge what sounds weird when actually spoken. The next step is to instead read them out loud with someone else. It might also be favourably to read it out loud with the developers as well, as they may give insights in what may be impossible to implement. Lastly, a recording of the dialogs are made, preferably with the intended voice (Pearl, 2017).

Flow

Another valuable method when designing VUIs are according to Pearl (2017) to start sketching on the flow of the VUI consisting of diagrams illustrating every possible path through the VUI system. How detailed it should be, depends on the system. For command and control systems all possible branches of the flow should be stated. This does not mean every possible phrase, but rather that the phrases are well grouped. For more conversational systems like intelligent assistants, it is not possible to state every possible interaction, the flow should rather be grouped into types of interactions, for example calling and texting, searching, and calendar. Not all possible interactions can typically be spelled out either, but it helps having the various intents grouped (Pearl, 2017).

4.3.3   Methods for implementation

Prototyping
Prototyping is the creation of artifacts which can be on different levels of fidelity. The tangible creation of products or interfaces are a crucial part of the design process and are used to test the design "in real life". Low-fidelity prototypes can be used in early stages of the design process as they are easily changed and fast to create. Sketching and paper prototypes are common techniques for low-fidelity prototyping. High-fidelity prototypes are often similar to the end product's look and feel, and sometimes even has a bit of functionality (Martin & Hanington, 2012).

Parallel prototyping
Using parallel prototyping, various prototypes are created in parallel. The purpose of this is to let designers explore a wide variety of designs before selecting one to move on with. Designers quickly and independently create various low-fidelity prototypes, and then let them be tested by users. The users give insights on the strengths and weaknesses of each design and its individual elements. The best qualities from all designs can then be merged into an optimized design. Parallel prototyping especially help designers not to get stuck on one single idea early in the design process (Martin & Hanington, 2012).

4.3.4   Methods for evaluation

Wizard of oz testing
Wizard of oz testing is used when the system does not yet exist. The idea is instead that a researcher controls the system as if it is actually working for the user. Since one of the researchers are busy controlling the "system", at least one more researcher is needed to conduct this method, which will handle interviews and taking notes. As it is not possible to use this method for a complete system, the researchers need to decide what to do based on the intended outcome. It is for example possible to

test just one short task with one utterance from the system, to see how the user expresses themselves (Pearl, 2017).

Usability testing

Usability testing is something which is done when a working system exists (Pearl, 2017). It is done to test how well the system works and if some functions are troublesome or frustrating for the user. A usability test most often consist of tasks which the user shall complete. The tasks should represent typical goals of the end user as well as be straight to the point. Scenarios are often used in combination with tasks to give the user the necessary context of the task (Martin & Hanington, 2012). Pearl (2017) recommends doing test studies before the real testing so that major issues can be fixed beforehand, for example recognition accuracy, since the usability test is not done to recognize such errors.

Questionnaire

Questionnaires can be used both as an analyzing tool as well as an evaluative tool. It is most typically used for quantitative data, but open ended questions can be asked as well. It is very important that the questions are well conducted so that they are actually asking whatever it is that the researchers wants to know (Martin & Hanington, 2012). Using Likert scales in questionnaires are recommended both by Martin & Hanington (2012) and Pearl (2017). Instead of only asking if the user agrees or disagrees, Likert scales provide a scale of five statements ranging from strongly agrees to strongly disagrees. Questionnaires can be used as a method alone or in combination with other methods such as to collect users thoughts after an interaction with a prototype (Martin & Hanington, 2012).

Competitive testing

Most often, a similar product as the one being created already exists, and it does not need to be something negative. Instead, it is possible to make use of already existing, similar products by doing competitive testing. This is simply done by user- and/or usability testing those products to recognise what works and what does not. The test should be designed so that the same one can be used throughout every product. The tasks in the test should therefor test functions that exists across the different products. Competitive testing also gives the advantage of keeping up with the industry (Martin & Hanington, 2012).

Evaluative research

Another name for this much established method is user testing or product testing. The goal of evaluative research is to understand potential users expectations of the system. Usable, useful, and desirable are key concepts which are tested and the

approach is to test prototypes with real users. Research using this method goes beyond only measuring performance and includes user emotions and aesthetics. This method is most often conducted in cycles, refining the prototype after feedback from each test. It is decided by the researchers how the method should be designed and where it should take place, and can vary from very formal research to more fast and informal testing depending on the goal of the research (Martin & Hanington, 2012).

Pearl (2017) explains that what to look for when doing user tests is not only what the user is doing and what it does and does not understand. The reactions of the user are just as interesting. It could be both facial expressions as well as body language. Trying to catch their emotions when doing different tasks is also important. Pearl (2017) also recommends trying to register if the user understand when it is possible to speak and when it is not and when they are confused or hesitates. How long time it takes for the user to complete a task is also essential.

Heuristic evaluation
Heuristic evaluation is, unlike most other methods, not done with users but is instead conducted through inspection with members of the team, expert evaluators, or similar. The inspection is done assessing some beforehand agreed upon guidelines, and each person is then individually going through the system with those guidelines in mind, writing down potential issues. Heuristic evaluation can not be used instead of user testing, but it is a valuable method early in the design process, or simply to identify clear usability problems before testing it on actual users (Martin & Hanington, 2012).

5

Process

The process can be divided into four main phases: Planning, Research and ideation phase, Creation of guidelines phase, and Testing of guidelines phase. Planning includes defining the scope of the project and a literature review. The Research and ideation phase includes competitive testing, stakeholder interviews, and an ideation session. Creation of guidelines phase includes creation of first set of guidelines, first set of guidelines, refining guidelines, and the second set of guidelines. Testing of guidelines phase consists of the first user test, refining guidelines, third set of guidelines, second user test, refining guidelines, and forth set of guidelines. An overview of the design process divided into the four phases can be seen in figure 5.1.

5.1   Planning

To start out the project, the researchers began at the analyze phase of the design process. As adviced by Cooper et al. (2014), the scope of the project was outlined. Numerous discussions were held, both between the researchers themselves but also with the supervisor and advisor concerning the scope of the project. It was from the beginning clear that the project should focus on navigation in a VUI using a smart speaker. The discussions mostly concerned how to limit the project in order for it to fit the estimated time plan. Discussions regarding if the project should only focus on creating guidelines for designing a home screen, carousel, or a help function was held. These are common concepts for GUIs and the project's scope would then be to try to design VUI equivalents. However, the researchers felt that this would limit the research in beforehand, considering that VUIs are different from GUIs and it is not certain that a VUI equivalent is needed or desired. Also, limiting the research in such an early phase by only studying a part of the interaction when VUIs are perceived as a whole seemed to risky as well. Therefore, the researchers decided to study the whole interaction of navigation in VUIs using smart speakers. To be able to fit the time plan delimitations were instead set, which can be found in the introduction chapter.

Following, a literature review was conducted. Firstly, literature was searched via
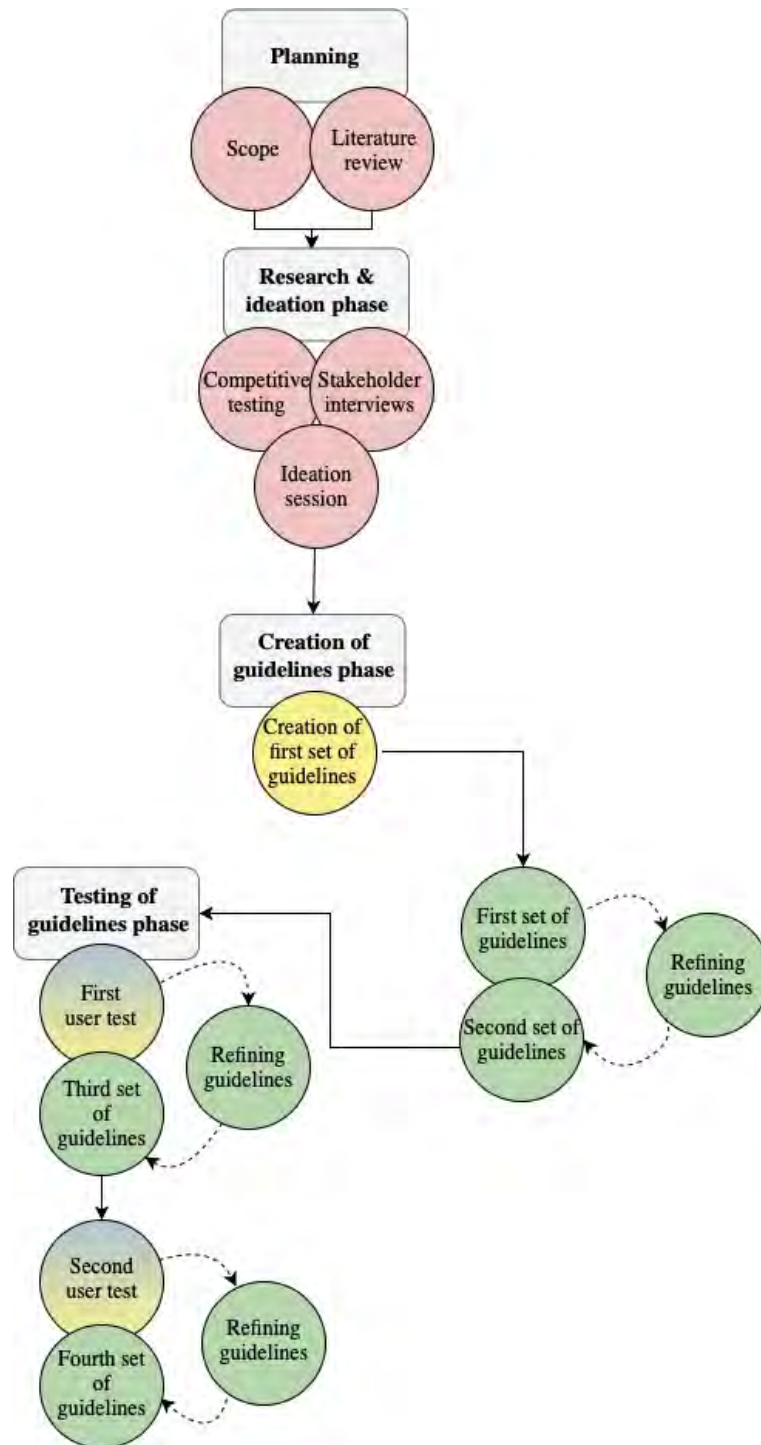
Figure 5.1: *An overview of the design process with the four phases and iterations. The parts from the design process the Wheel, accessed from Hartson & Pyla (2012) can be seen with corresponding colors: Analyze (pink), Design (green), Implement (blue), and Evaluate (yellow).*

Chalmers Library, Gothenburg University Library and Google Scholar. Information that could not be found through the previously stated sources was searched for using

Google. Main keywords that were used when searching for literature were: Voice user interface, Smart speakers, Navigation, Google Home, Intelligent assistants, User experience, Human-centered design, Interaction design, Conversational interfaces, Speech recognition, Perception, Voice interaction design, and Voice browsing. The result of the literature review is presented in previous chapters.

When the literature review had been conducted and a scope for the project had been defined the researchers went on to design the design process. In this early stage, no definitive decisions were made but instead a simplified design process was illustrated. This to structure the work but leave the exact choices to be taken during the actual process, enabling flexibility and creativity during the project. A human-centered design approach and an iterative process was considered as important aspects of the design process from the beginning, making sure that the users were involved during different stages of the process and that the design accommodated their needs, while being able to refine and improve by iteration. The early illustration of the design process consisted of the analyse phase explained by Cooper et al. (2014), to then move on to the wheel explained by Hartson & Pyla (2012) which involved prototyping, testing, and evaluating as an iterative process. Methods were at this stage considered but not chosen, and the three-dimensional model of Rohrer (2014) was used to narrow down the overwhelming amount of methods to a set of methods that seemed suitable for this project. Which methods to use were chosen during the actual process and will be explained further during this chapter.

Continuing the analyze phase recommended by Cooper et al. (2014), a review of existing work and products by market research was carried out by using competitive testing. Stakeholder interviews were then conducted with personnel from Swedish Radio, working closely with the development of their digital content including their VUI action.

The next step recommended by Cooper et al. (2014) is user interviews, however frequent users of Google Home, and especially of Swedish Radios action, are not that common. Considering this, user interviews were not conducted in this stage. Cooper et al. (2014) also recommends to include user observations, but as smart speakers are used in a home environment user observations were considered as too cumbersome in relation to the expected outcome. Instead, the researchers decided to conduct an ideation session with fellow interaction design students in order to gain new insights. The implementation of Cooper et al. (2014) advices for the analyze phase can be seen in figure 5.2. The competitive testing, stakeholder interviews and ideation session are further explained in the next section.

Figure 5.2: *A comparison between recommendations from Cooper et al. (2014) and the implementation made by the researchers.*

5.1.1   Timeplan

The time plan for this project looked as below:

Week 1-5: Literature research
Project proposal and planning report
Gain domain knowledge
Defining scope of project

Week 6: Designing the design process

Week 7-10: Case studies
Competitive testing, interviews, and ideation session
Create guidelines

Week 11-17: Prototyping, user testing, evaluating
Designing and creating prototypes based on case studies
Planning and carrying out user tests
Refining guidelines

Week 18-20: Writing report
Finishing report
Presentation and opposition

5.2 Research and ideation phase

After planning the project, the next step was to research the field and coming up with ideas. This phase involved searching the area of Google Home actions, and testing the relevant ones through competitive testing. Interviews with relevant employees at SR was then conducted, gaining insights on their daily work and thoughts about their content and future within smart speakers. Lastly, an ideation session was held in order to gain new ideas on how to navigate through content using speech.

5.2.1 Competitive testing

Competitive testing can give researchers knowledge about how others have solved similar issues while also keeping them up-to-date about the industry. In this case, it was used to gain knowledge about how other companies have designed their actions for Google Home. By using this method, the researchers hoped to gain understanding for which design solutions that works, and which ones does not. This knowledge could then be used when creating prototypes later in the process.

Seven actions for Google Home were tested by the two researchers of this project. The actions were picked based on similarity to Swedish Radio in content and if they were available in Swedish or not. The actions tested were: Aftonbladet, Expressen, I like radio, Sydsvenskan, Dagens nyheter, and SR. Helsingborgs dagblad's application was used for a pilot test. The pilot test went on as intended and did not result in any changes to the actual test.

Three tasks were used to test the applications. The tasks were constructed so that they could be used for all applications. The tasks were as follows:

1. Play something with the help of the assistant.
2. Explore the possibilities of exploring the content.
3. Ask for help.

The researchers conducted the testing individually using a Google Home mini, meaning every application was tested twice. All interactions were audio recorded so that no note-taking had to take place during the interactions. After each task, the researchers noted positive and negative feedback about the interaction as well as the different interaction steps they had taken to complete the task by using the recordings.

The result of the competitive testing was analysed using the method affinity diagram. Affinity diagram was used because of its ability to help researchers collaborate in clustering and finding themes of qualitative data. The positive and negative at-

tributes written down during the competitive testing were summarized and written down on post-its, one post-it for each attribute. These post-its were then sorted into clusters by placing post-its which were similar, or related, together. This part was done in iterations until the researchers were pleased with the clusters. The clusters were then analysed one and one, where the researchers discussed what the main area is for the clusters trying to find a fitting theme name for them. The clusters were then divided into positive and negative attributes of a VUI and the result can be found in figure 5.3 and table 5.1.
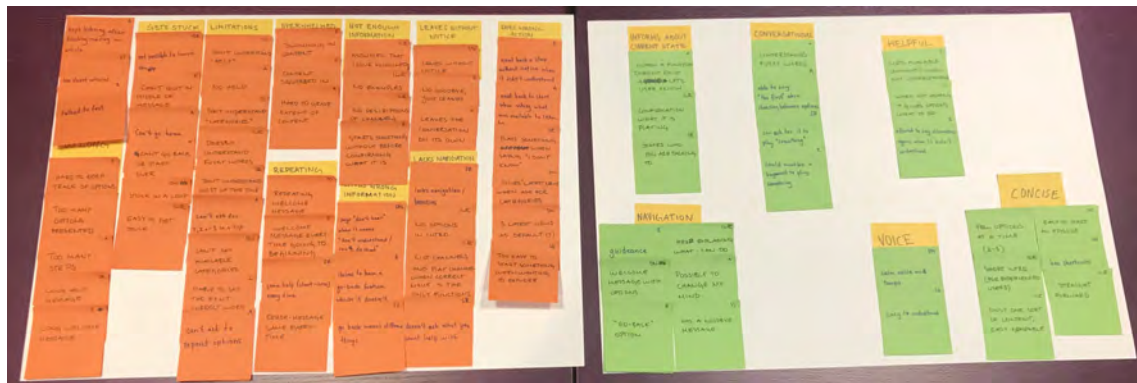
Figure 5.3: *The affinity diagrams created, orange post-it notes represents negative attributes, while green represents positive attributes.*

Table 5.1: *Result from competitive testing, where green text represent positive attributes and red text negative.*

Theme	Explanation
Informs about current state	Tells the user what is going on, and where in the navigation they are.
Conversational	The user can talk to the assistant in a more conversational way, not only saying exact phrases.
Navigation	Possible to navigate in the app with guidance from the assistant.
Helpful	The assistant feels helpful, offering the user solutions when problems arise.
Voice	The voice is perceived as pleasant.

Concise	Gives the user the proper amount of information in a concise manner.
Too long	Presents the user with too much information at once.
Gets stuck	Does not offer the user to go to the beginning, go back, or leave the application. Not possible to talk at the same time as the assistant.
Limitations	The assistant does not understand simple commands or conversational commands, and does not offer any help.
Overwhelmed	The assistant overwhelms the user with the amount of content available, by presenting it in the wrong way.
Repeating	Uses the same phrase without any variation.
Not enough information	Leaves the user without enough information to be able to use the application properly.
Gives wrong information	The system says that it has functions that it does not, or says that it could not hear when in fact, the function does not exist.
Leaves without notice	The assistant leaves the conversation without notice if it does not understand something, most often after three tries.
Lacks navigation	Provides no navigation through the content, but rather assumes that the user knows what it wants to listen to.
Does wrong action	Does something else then the user tells it to do.

5.2.2 Stakeholder interviews

Interviews were conducted with the goal of gaining insight into Swedish Radios work as well as make use of their already gained knowledge in the area. By interviewing members from different teams and with different professional roles at SR, the researchers hoped to gain a deeper understanding of the problem area, the design choices that have been made, is going to be made, and what the teams think will happen in the future. A semi-structured interview was used to be able to ask the actual questions that was of interest, but at the same time enable flexibility during

the interview if interesting topics were mentioned (Rowley, 2012).

Furthermore, the domain of interviewing was studied and the interviews were planned out in detail before they took place. As recommended by Jacob & Furgerson (2012), an interview protocol was written consisting of an introduction, a reminder of the goal with the interview, the questions with prompts, and an ending (see appendix A). The introduction ended with handing out a consent form (see appendix B). The interview questions were organised as also recommended by Jacob & Furgerson (2012), with background questions to begin with, followed by easy questions which over time becomes more advanced. A small amount of questions were used, which were open-ended and quite broad to let the interviewee talk freely (Jacob & Furgerson, 2012). The researchers further studied recommendations for how to be a successful interviewer beforehand, mainly following Turner (2010) tips for interviewing. Turner (2010) also brings up the importance of pilot testing, and in accordance with that a pilot test was conducted before the actual interviews took place. Some minor communication errors that occurred during the pilot test were corrected.

Four people participated in the interview: the product owner of the voice application and head of the play x/voice-team, the design lead of the mobile application, the UX-designer for VUIs in play x/voice-team and the innovation team, and lastly the team leader of the innovation team. The interviews were conducted in a quiet room inside of Swedish Radio's building in Stockholm, Sweden. There were two interviewers and one interviewee present at each interview. Using the interview protocol, the interviewers began with presenting themselves, welcoming the interviewee and explaining the purpose of the interview and what to expect from it, in accordance with Turner (2010). The interviewee was then asked to read through a consent form, and if comfortable with the agreements, signing it. After introducing the interview and signing the consent form, the interview started. With consent from the interviewees, all interviews were audio-recorded.

The interview consisted of six questions, and the interviewers were responsible for every other question. There were three background questions regarding role and work at SR, two main questions regarding Swedish Radios content and guiding a user, and one questions for rounding off regarding the future. When the interview was finished, the interviewers thanked the interviewee for participating and asked if it was okay to make contact again if further questions were to be considered (Turner, 2010).

When interpreting the data gathered from the interview, the researchers followed recommendations given by Rowley (2012). The first step was to listen to the recordings

of the interviews in order to familiarize with the content and gain a first impression and insights. The interviews were then transcribed, divided between the two researchers. Each transcription was then checked for errors by the other researcher. The complete transcriptions were then organised by the questions as to understand in what context the information came from.

Content analysis was then used to analyse the data further, and was chosen because of its systematic guide in analysing big amounts of qualitative data. Question by question the data was analysed by highlighting and picking out relevant phrases. These were then put in a sheet, where each phrase was given keyword(s). Creation of keywords was an iterative process, ending when the researchers felt pleased with the result. The keywords were then grouped which were the beginning of creating themes. When all keywords were desirably grouped and named, each theme was given an explanation. The result was quite extensive, and therefore only the result from the two main questions will be presented below. For the complete result of the content analysis, see appendix C.

Result from question 4
Swedish radio has an extensive amount of content, and it is not possible to present it all at once. How do you reason when presenting your content?
Current
The current status of the platform is that it is designed for making all content available but without any second thoughts about the UX of the interaction.

Usage
The voice application is mostly used at home through Google Home, and not as much through the phone. And when it is used at home, mostly channels are listened to. It is mentioned that this could be the new kitchen radio mainly used as background sound.

Personalised
Personalisation for the voice application is described in conjunction with recommendations based on your profile while combining that with the diversity that SR represents. Usually it is enough to recommend content based on popularity but considering SR and public service this can not be the case. The content could be picked based on the specific user and the listening could be more in the format of a personal radio playlist.

Content solutions
Content solutions refers to future possible solutions for SRs digital content. The content needs to be limited and some kind of prioritisation needs to be established

so the user does not get overwhelmed. Content that is labeled as more important is placed on the home screen in the mobile application, followed by the personalised content. The content could be divided by categories, making the navigation easier. On-boarding is mentioned as a possible solution for being able to use personalisation. Meaning that the user is prompted with questions about their interest the first time they enter the application as to get a better, more personalised, experience later on.

An important note is that even though users is not familiar with SRs content, they should still feel included and be able to use the application.

Content challenges
Challenges with the content is to make everything available without making it overwhelming or tiresome while also not assuming that the users is familiar with the content. And also by this not forcing the user to choose between a lot of different categories that may not say anything to them. A critical challenge is to solve the actual browsing through the content.

Requirements
Requirements that needs to be included accounting that it is SRs application. News is a big part of their agenda and considering that they are a public service company they also need to work towards diversity in their content.

Challenges
Challenges include limiting the content and making it conversational. A challenge mentioned is also how to present content.

Work structure
The work structure refers to the work habits at SR. There are many people involved in different decisions, making it hard to please everyone. SR tries to work data driven while also staying in the forefront. A dedicated personalisation team is also available so that not all teams need to work on those matters.

Result from question 5
How do you reason concerning guiding a user to find something to listen to?
Guide
It is mentioned that there are different views on what guiding means. Explanations should be used both during the guide, as well as for the content. For the mobile application, on-boarding has been used for the home page for first time users, so that the users can get started with suggestions at once. Another suggestion for how

to present content is to make personalised playlists. The guiding, though, can make use of content divided in disciplines and/or categories.

Content

Something which the interviewees have experienced is that the content is not typically known by the user and an explanation of the content is needed. The content is divided into categories, but also more traditionally as radio with their channels such as P1, P2, P3, and their associative content. How the content should be divided is expressed as a big challenge.

Characteristics

The characteristics of the system should according to the interviewees be helpful, pleasing, efficient and initiative. It should give personalised suggestions and result in a good user experience. It should help to find something if that is needed, but also make it possible to effectively start playing something for those who know exactly what they want to listen to. At this point, it could be more important to make it easy for the user to start something, rather than that something being exactly what they wanted.

Users

It is common for the users to simply just listen to radio, and therefore only wanting to start playing one of the live broadcasted channels. As a novice user you do not want to make something wrong and be ridiculed. It is instead important to make it easy to play something the first time, so that they can feel as if they succeeded.

Challenges

Interacting with this kind of device is, as it is now, often unpleasant for the user. It is, and will be, a challenge to change that and instead create a good user experience. The users, especially the younger audience, are commonly not familiar with Swedish radios content, and more weight will then be put on the guiding through the content.

5.2.3 Ideation session

An ideation session was used to gain new insights into the research problem. The researchers had already examined other solutions for VUIs by competitive testing and interviews as well as researched the area and collected existing design guidelines and was therefore well equipped to create basic prototypes with existing knowledge. Beyond this, it is however also needed to try out new solutions that have not yet been tested, and for getting inspiration for this new solutions, an ideation session was conducted with fellow interaction design students. The session was held as a design workshop consisting of two warm-up exercises, role-playing, and a group discussion

(see appendix D for the protocol used during the ideation session). Both researchers worked as facilitators and four students at the master's programme Interaction design and technologies participated in the session. The session was held in a group room at Chalmers University of Technology, Gothenburg.

Beginning the session, the participants were welcomed and informed about the research area and problem. Then, the warm-up exercises Grandma, Tiger, Ninja and a 1000 uses was performed. The warm-up exercises was used to get the participants to loosen up and to sprung their creative thinking. Grandma, Tiger, Ninja is a game used as a warm-up exercise for ideation. The exercise is similar to rock, paper, scissors, but instead of the participants showing their pick by different hand signs, this exercise lets the participants act out their pick with their whole body. The goal of the exercise is to let participants get to know each other while also making them step out of their comfort zone. In 1000 uses, participants are divided into teams which are given pens and paper. Then, the teams are presented with a random everyday object (same object for all teams) and are instructed to write down as many uses as they can think of for that object. In this case the object picked was a folding umbrella. After four minutes the number of uses are counted and the team with the highest amount wins. The purpose of 1000 uses is to make the participants think creatively and outside-the-box (Taylor, 2017).

Following the warm-up exercises the participants were given more information about the role-play and its purpose. Because of the verbal format of VUIs, role-playing seemed as a good fit as it lets the participants act and talk with each other while taking on different roles. When informing the participants about the session, emphasis was put on that there is no right and wrong, and no bad ideas. During the role-playing, the participants took on different roles. One participant acted as the user and one as Swedish Radio's action for Google Home, while the two other waited for their turn. A turn ended when the participant acting as Google Home started "playing" something for the user, and after each turn the participants rotated seats. The set-up for the role-play can be seen in figure 5.4.

The participant playing the user was given a scenario to mind while acting. The scenario was to act as a user asking Google Home for help finding something to listen to. For the role of Google Home, four different characteristics were used which were switched for each turn. It was only the one acting as Google Home who saw the current characteristic. The characteristics were efficient, tedious, helpful, and stupid and were based on previous findings about VUIs. Characteristics were used as to inspire for, and variate the Google Home responses. Also, the participant acting as Google were given a diadem with a cut-out cardboard Google Home as to make them feel more in character. Worth mentioning is that the participants were

Figure 5.4: *The setup for the role play seen on left image, and the role-play in action*

carefully instructed that even though they played the role of Google Home they did not need to act as how they thought that Google would act, and improvisation was encouraged. Each individual acted as the user two times and as Google Home two times with different characteristics. The researchers took notes of important events or citations during the role-playing.

A group discussion took place after the role-playing, where the facilitators used previously made questions combined with findings from the role-playing. Afterwards, the participants were thanked for their participation.

The result of the ideation session was selected important phrases which were used in the next step of the project, creating the first set of guidelines. The result can be seen in appendix E.

5.3 Creation of guidelines phase

This phase was divided into several parts, as the guidelines were refined and rephrased during the process. Initially, the information gathered at this point was summarised and used for creation of the first set of guidelines. The first set was then restructured due to it being deemed as not testable, resulting in the second set of guidelines.

5.3.1 Creation of first set of guidelines

At this point, an extensive amount of data had been collected, and were to be transformed into guidelines. A first set of guidelines were made at this early point as to have time to test and refine them in several iterations.

To create the first set of guidelines based on the information gathered at that point, affinity diagram method was used. Information from existing guidelines mentioned in the theory section, from competitive testing, stakeholder interviews, and the ideation session was put together and made into one line statements. During this process, statements that were irrelevant or not contributing to the research question were removed, as to shorten the overwhelming amount of information. The phrases was then printed and cut into pieces, one statement per piece. The pieces were spread out on a big table and the researchers worked on putting the pieces together in common themes, which can be seen in figure 5.5. When all pieces were put into different clusters, the clusters were studied one at a time to see if there were any misplaced pieces, then the clusters were given a theme name for each one of them. Since there were a lot of themes created, an affinity diagram was again conducted, to put those themes into overarching themes.
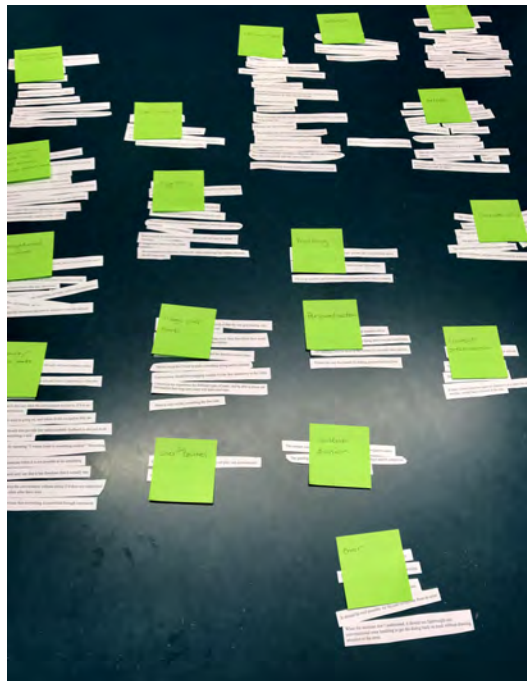
Figure 5.5: *The pieces put into clusters and assigned themes*

To simplify the process of making these into guidelines, all information was structured on big papers, one paper for each overarching theme, which can be seen in figure 5.6. The overarching themes were then studied one at a time, and from the structured information, guidelines were extracted. The guidelines were then given explanations based on the collected information. The guidelines created were therefore divided into an overarching theme, with corresponding guidelines and explanations.

Figure 5.6: *The organised result from the affinity diagram, each paper being the created overarching themes. These were used in the process of creating the first set of guidelines.*

5.3.2 First set of guidelines

These are the first set of guidelines produced, consisting of overarching themes, guidelines and explanations of the guidelines.

How to handle content
- The content should be divided to make navigation easier
  - To make the content more accessible it should be divided, such as into categories, disciplines or other type of division.
- Don't present too many options at once
  - Rather than presenting a lot of options at once, it is instead recommended to present them in multiple interactions and offer a "more options" command. Start broad and then narrow down until the desired content is reached. Generally it is recommended to not present more than three to four options at once, but can vary depending on the complexity of the wording.
- Episodes should be presented with amount and a suggestion
  - When presenting episodes, the user should be informed about the amount of available episodes and offered suggestions for which one to pick.
- The content should be prioritised based on relevance and importance
  - The content should be prioritised based on relevance for the user and importance for the company behind the content. When presenting different

options to the user, begin with the most relevant and important options first.

Error & Feedback
- Error messages should be provided which mirrors the actual problem, are concise, and offers a solution.
    - Error messages should be provided that informs the user about the actual problem, are concise, and offer a solution so that the user can recover from the error. The error messages should also not draw attention to the error, but focus on the solution instead.
- The user should feel as in control
    - To avoid confusion and frustration from the user, the VUI should be designed so that the user feels as they are in control during the interaction. To always inform about the systems current state by letting the user know what is happening, where they are in the navigation and never lie is generally enough. Also, do not decide something for the user without an explanation or a reason and give options rather than automatically play something.
- Use confirmations
    - Confirmations should be used so that the user knows what a conversation has led to. There are different kinds of confirmations, and which ones that should be used depends on the situation. Implicit confirmation is suitable in most situations, but when the consequences of an action is severe explicit confirmation is more fitting.
- It should be possible to use universal navigational terms
    - To be able to navigate in the system, universal navigational terms should be provided. The user should be able to use back, cancel, exit, and home, with correlating synonyms. It should also be possible to reverse an action, or a set of actions.

Users
- Design for both novice and experienced users
    - A novice user needs extra guidance to get started and to understand the interaction, while those help systems may be annoying for a more experienced user. To remember the user's previous usage and adapt for it is a good start. Focus on guiding a novice user so that they don't feel ridicule and supporting an experienced user so that they more efficiently can reach their goals.
- Don't assume that the user knows things
    - Don't assume that the user is familiar with the content or that they know what they are interested in. The system should be able to handle an "I

don't know" from the user.

- Make users understand the system
  - Make the users understand how the system works, not in a technical meaning but rather which interactions that are possible and how these can be executed.
- Always consider the users' perspective
  - This can be done by asking the user what they want to do, focusing on what is relevant for them and what their intent is.

Suggestions

- Provide explanations for content
  - The user should be able to ask what the content is about. Users may not be familiar with the content and only presenting a title may not suffice. Implicit confirmations should be used before playing content.
- Provide samples for content
  - The user should be able to ask for a sample of the content and/or provided suggestions, to make it possible for the user to gain a feeling for it. Whether both samples and explanations should be used requires further testing.
- Use personalisation
  - The system should be able to remember the user and use personalisation for a more customized experience. This can be done by either questioning the user about their preferences during first-time usage or by learning over time. The user should be able to ask the system for personal advice.
- There should be diversity in provided suggestions
  - While personalised suggestions are great, it is also important to provide a diversity in suggestions as well. This is of importance both to not contribute to a filter bubble where the user is only exposed to one type of content, and so that different kinds of content can be highlighted. When giving suggestions to a user, presenting both personalised as well as other content is therefore crucial.

Characteristics

- Design with characteristics in mind
  - The VUI should be designed with characteristics in mind, where which characteristics to use depends on context and usage and requires beforehand testing. Examples of characteristics are efficient, helpful, tedious, and initiative.
- The interaction should be coherent
  - A coherent interaction is crucial, and means that the user should feel as the system works in a similar way throughout the whole interaction. This

results in that the user gets familiar with the system faster, and with more ease.

- Do not overwhelm the user with too much information
  - Avoid overwhelming the user with too much information. The presented information should be relevant and details should only be provided when the user clearly benefits from it. At the same time all content should be available, prioritise the content in order to avoid overwhelming and make use of users recognition rather than recall.
- The interaction should be conversational
  - As far as possible, already existing conversational cues and phrasing should be used. A variation in responses and phrases is crucial, as well as for the system to understand conversational, instead of exact, commands. Questions are fundamental in natural language and should be used in the end of a phrase. The system should also remember what is being talked about, so that it understands what is being referred to when a user mentions e.g. "her".

5.3.3   Refining guidelines

After creating the first set of guidelines, it was planned to move on to testing. This did not happen as the researchers found themselves perplexed with how to test the guidelines. Instead, a meeting with a teacher at the Chalmers division of Interaction design was held in order to get guidance on how to progress. The teacher advised to refine the guidelines making them more pin-pointed to the research question as well as more concrete. Also, the guidelines were compared with existing ones, mentioned in the theory section. Guidelines that had already been developed by other researchers was deleted, as the purpose with this work is not to test already existing guidelines but to develop new, complementing, ones. However, a problem arised during this. Too many guidelines were to be erased if the researchers would strictly erase every guideline that overlapped with an already existing one. At the same time, including all guidelines made the set too extensive. Therefore, a trade-off had to be made where guidelines that was judged as more relevant for the scope stayed while guidelines not as relevant was erased.

An affinity diagram was again conducted to restructure the guidelines, striving towards making them relevant to the research question as well as testable. Both overarching themes as well as guidelines were restructured, followed by some being renamed with the goal of making it clearer. The guidelines "Error messages should be provided which mirrors the actual problem, are concise, and offers a solution", "Always consider the users' perspective", "Design with characteristics in mind" and "The interaction should be conversational" were erased. "Error messages should be

provided which mirrors the actual problem, are concise, and offers a solution" was erased as it was realized that including error handling in the scope would be too extensive and rather a project on its own, as previously mentioned in the delimitation section. "Always consider the users' perspective" was erased considering the existing guideline "Focus on the user" created by Google (n.d.-b). "Design with characteristics in mind" was also deemed as too big of a challenge for this project as well as somewhat included in the already existing guideline "Persona" created by Amazon (n.d.-b). The guideline "The interaction should be conversational" is already supported in existing guidelines "Vary responses", "Use contractions", and "Speak clearly, in plain English" and was therefor erased (Amazon, n.d.-b; Google, n.d.-b,-c).

Apart from excluding already existing guidelines, the restructuring mostly concerned changing the "level" of the guidelines. As mentioned in the theory chapter, guidelines lay between design principles and rules of implementation. A problem with the first set of guidelines was that the guidelines were either too specific or too broad, resulting in some guidelines being incorporated in the explanations and some in the overarching themes for the second set of guidelines. Also, some guidelines were re-phrased as to make them clearer. In the end, the refining resulted in a set of sixteen guidelines with five overarching themes.

5.3.4 Second set of guidelines

The second set of guidelines, presented with overarching theme and guideline. After having written the explanations for the first set of guidelines, it was perceived as unnecessary to create explanations again as the guidelines were still in an early stage, with a possibility of undergoing several revisions. It was therefore decided to not write explanations before having tested the guidelines.

Don't overwhelm the user
- The content should be relevantly sorted
- Present a minimum amount of options at a time

The user should feel as in control
- The user should understand where in the navigation they are
- The user should be the final decision maker
- Confirmations should be used
- Universal navigational functions should be provided
- The user should be able to do an informed choice

Don't assume that the user knows things
- The system should be able to handle "I don't know"

- Samples for content should be provided
- Explanations for content should be provided

Provide what the user wants and/or needs
- Personalisation should be used
- Design for both novice and experienced users
- Diversity should be provided in suggestions
- Content should be prioritized

The system should be intelligible
- Provide a coherent interaction
- Inform about possible actions in the system

5.4 Testing guidelines phase

In this phase, the created guidelines were tested and refined based on results from two user tests. The first user test focused on testing individual guidelines, while the second test aimed at measuring the user experience.

5.4.1 First user test

Based on the second set of guidelines, already existing guidelines when applicable, and recommendations by Pearl (2017), a set of prototypes were created. The aim of this was to enable testing of the guidelines, hence enable to confirm or dismiss them. To create a set of prototypes rather than one was decided after discussing that creating one whole interaction, and hence test all guidelines at once, could result in a difficulty distinguishing what guideline affected what. By creating individual interactions for each guideline, or few guidelines, it would enable pinpointing the user feedback to that guideline(s).

In order to create the prototypes, the researchers were adviced by a teacher at the Chalmers division of Interaction design to look for examples, both good and bad, of how the guidelines were implemented in already existing systems. This made the process of making prototypes easier, as concrete examples of how to implement, and how to not implement, the guidelines were made. Focusing on one guideline at a time, actions were tested in Google Home while looking for how that guideline was, if it was, implemented in that design. These were written down and marked as good or bad examples. The examples of the guidelines were then inspected and narrowed down to the best and the worst examples. Innovative ideas from the ideation session were added as examples as well.

When this was done, it was clear which guidelines were testable in this project, and which were more reasonable to argue for. The guidelines that proved to be more suitable to argue for was "The content should be relevantly sorted", "Confirmations should be used", "Personalisation should be used", "Diversity should be provided in suggestions", "Content should be prioritized", and "Provide a coherent interaction", and these were therefore not explicitly tested.

In total, twenty prototypes were created using sample dialogs, which were divided into eight tests. Each test had the purpose of testing one or two guidelines and could consist of one or several variations, hence eight tests but twenty prototypes. The variations most typically consisted of implementing the guideline versus not implementing the guideline, to be able to explicitly ask the participants for which one of them they preferred. Sample dialogs was used as the prototypes were not intended as a complete interaction including all possible outcomes, but instead one possible interaction. It is also a method that does not demand any work to be put into creating something which technically works, but the focus instead lies on designing a good dialog. Reading the dialogs out loud also helps finding any obvious faults in the interaction at an early stage.

The intention was to use the method parallel prototyping when creating the set of prototypes. In the beginning, the researchers worked individually on separate prototypes but later changed approach and worked together instead. It was realised that following the numerous guidelines left little room for own creative solutions and it was judged as more valuable to work together and making sure that all guidelines were followed correctly. Parallel prototyping also includes making various variations of one design and testing them on users. This part of the method was practiced in the user test.

The user test consisted of an introduction, eight different tests, where some involved several variations resulting in twenty different interactions. Each interaction involved an interaction between Google Home and the participant. To avoid training effects, the order of the variations were randomised when possible. After each test or interaction, this differed between the tests, the participant was asked questions. After completing all tests, the participants were asked if they had any more questions and was then thanked for their participation.

A pilot test with one participant was conducted to establish if the interactions and questions was comprehensible as well as letting the researchers practice their roles. After the pilot test, the test was updated according to the findings and the real testing began. The outlay of the test, with interactions and questions, can be

found in appendix F, while a list of changes made based on the pilot test can be found in appendix G.

The user test was conducted during a period of two days, 12 and 15 of April, testing five participants between 22-30 years old, with a mean of 25 years. The decision of testing with five participants was made in accordance with Nielsen (2000), whose calculation demonstrates that using more than five participants most often does not lead to any more insights. The participants previous experience with smart speakers were evenly distributed, where two participants had tried using one, one participant owned one, and two participants had never used a smart speaker before. All participants were students at the master's programme Interaction design and technology. The participants were instructed to act as users of Swedish Radio's action not knowing what they wanted to listen to.

One of the researchers was the test leader while the other one acted as Google Home. During the test, the participants and the researcher acting as Google Home could not see each other. This to try making the interaction as similar as possible to an interaction with an actual Google Home. The test leader facilitated the test, which included introducing each test, giving clears to both the participant and the other researcher when to start an interaction, asking questions and taking notes. Scripts were used during the tests, one for Google Home and one for the participant to read from. The scripts for the participants were handed to them one at a time to avoid confusion. The scripts could also include a part where the participant were asked to improvise an answer, this was explicitly stated by the test leader and in the script. Improvisation was used to verify if the users had comprehended what Google had said and also to investigate if they would be able to give a suitable answer. In some interactions, the improvised answers from the participants was used in order for the researchers to investigate how a user would like to express themselves when performing a certain action within the system.

The test began with the participants being welcomed, and given an introduction to the test including the purpose of the test and the outlay. The participants were also told that due to the excessive time to complete the test (1 hour), they could help themselves to Swedish chocolate balls and candy during the whole test as to keep them motivated. Before starting, the researcher acting as Google Home was seated on one side of a tilted table and the participant on the other so that they could not see each other. The test leader was seated straight forward to be able to see both the other researcher as well as the participant as to more easily interact with them both (see figure 5.7). The test started with a warm-up exercise so that the participant could get a feeling for how to interact with Google Home using a script. After that,

the eight tests with corresponding variations were performed together with questions from the test leader. The test ended with the participant being thanked for their participation. The result from the first user test is explained in the following section.
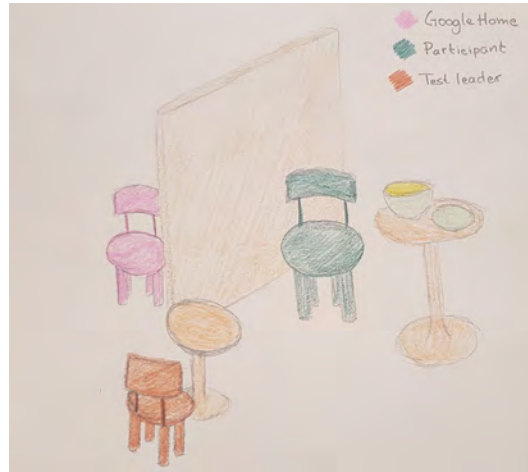

Figure 5.7: *The setup during the user tests with the placement of the participant, the test leader, and Google Home.*

5.4.2 Refining guidelines

The results from the first user test were analyzed and summarized. For each test, a summary was written with a conclusion if the guideline tested had been validated or not (summaries and conclusions can be found in appendix H). Two guidelines were erased, "The user should understand where in the navigation they are" and "Confirmations should be used". "The user should understand where in the navigation they are" gained no support from the participants. The guideline was implemented as the system informing the user about which step in the navigation that they were in, and which step that was next. When compared to an interaction where the system did not inform about the steps, all participants preferred the one not informing about the steps. "Confirmations should be used" was erased due to it being judged as too general and already being described in detail by Pearl (2017) how to design for.

During the refining, most guidelines were re-phrased and given an updated or more elaborated explanation. Also, three new guidelines were added; "Keep the interaction concise but still adequate", "The system should support making decisions for the user", "Design for both the patient and impatient user" due to findings during the user test.

In accordance with the analyze, the third set of guidelines were refined and resulted in 16 guidelines with five overarching themes. As the guidelines now had been

tested, explanations for each guideline were also created.

5.4.3 Third set of guidelines

The third set of guidelines is presented below, with overarching themes, guidelines and corresponding explanations.

Don't overwhelm the user
- Keep the interaction concise but still adequate
    - The interaction should be as short as possible with only relevant information presented while at the same time not excluding important information. To add something that makes the interaction longer needs to be truly valuable for the user.
- Presented options should support making an informed choice while not being overwhelming
    - The amount of options presented at a time should be such as the user should not feel overwhelmed but at the same time be enough for the user to make an informed choice.

The user should feel as in control
- The user should be the final decision maker
    - The user wants to be given options rather than the system automatically playing something. Regardless if it was the user or the system that made beforehand decisions, the user always wants to be the one making the call if something should be played or not.
- The system should support navigational utterances
    - Start over and go back one step are two basic navigational functions which are needed where the system should support utterances like "go back to the start" or "start over", and "go back" or "go back one step". It should also be possible for the user to refer to previous conversational subjects when wanting to navigate.
- The user should be able to do an informed choice
    - Present the user with enough information so that they feel comfortable in making a decision. Present an explanation for suggestions given, such as what they are based on.
- The sorting of the content should be comprehensible
    - The content should be sorted in such a way that when presented to the user, the user should easily be able to make a decision whether it is something they want or not without further explanation. This should be possible regardless if the user have been presented a few or many alternatives.

Don't assume that the user knows things
- The system should be able to handle "I don't know"
  - When a user expresses "I don't know" the system should be more initiative and provide the user with suggestion(s) that lets the user move forward more quickly.
- Descriptions for content should be provided
  - It should be possible for the user to ask about what the content is about and be given a description.
- The system should support making decisions for the user
  - When a user does not know what they want, they should be able to ask the system to make a decision for them. This lets indecisive users play something faster instead of forcing them to make choices.

Provide what the user wants and/or needs
- Personalisation should be used
  - The system should be able to remember the user and use personalisation for a more customized experience. By using personalisation, shorter interactions can be provided, as well as better suggestions, and better explanations for the suggestions. Usage without personalisation should also be designed for.
- Design for both novice and experienced users
  - Provide novice users with guidance so that they feel comfortable using the system, and provide experienced users with support so that they can reach their goals more efficiently. The interaction for a novice user is typically longer than for an experienced user.
- Variation should be provided in options given to the user
  - Make sure the user gets different types of options and suggestions, to prevent filter bubbles and letting them explore more of the content.
- Content should be prioritized
  - When presenting different suggestions to the user, begin with the most relevant and important options first. Users expect that suggestions given are based on something and not handed out randomly by the system.
- Design for both the patient and impatient user
  - Users of VUIs can be both patient and impatient with the system. The standard interaction should therefor be such as it fits them both. For the patient user, provide more functionality where explanations for it can be asked for. For the impatient user, provide a short-cut to let them play something fast without having to make several decisions.

The system should be intelligible
- Provide a coherent interaction

– When users feel like the system consistently behaves in the same way they
are able to learn how to use it, also leading to the possibility of becoming
experienced users.
- Inform about possible actions in the system
    – Users like to be informed about what they can do in the system, but they
    dislike too much talk more. Therefore, this information is needed, but
    should not be explicitly stated at all times. Instead, make it possible to
    ask for this information, and explicitly state how to do that to a novice
    user and give reminders once in a while after some usage.

5.4.4 Second user test

Based on the prototypes created in the first user test, two refined prototypes were
created according to the results. For the second user test, the researchers aimed at
creating one prototype but as some guidelines were not suitable to include in one
interaction, for instance "Design for both novice and experienced users", two different
interactions were designed. As mentioned in the theory chapter, UX is an important
part of interaction design. The aim of the second user test was to investigate how the
third set of guidelines contributed to the user experience when used in combination,
hence the strive for creating as few interactions as possible including as many of the
third set of guidelines as possible.

The second user test consisted of an introduction, a warm-up interaction, test-
ing the prototype using Wizard of Oz, a questionnaire measuring the user experience,
and an interview discussing the answers of the questionnaire (see appendix I for the
outlay of the user test and the interview and appendix J for the questionnaire). The
test participants were then encourage to ask questions followed by being thanked for
their participation.

As for the first user test, the prototypes were created using sample dialogs be-
cause of the purpose being that only one possible interaction per prototype were to
be made, and not all possible outcomes of it. The dialogs were written based on
the third set of guidelines, as well as using general guidelines and recommendations
for designing VUIs. When done, the dialogs were read out loud, to allow for any
unnatural phrasings to be found. This was done in iterations until the dialogs were
assessed as well formulated.

The method Wizard of Oz was used to be able to test the prototypes in a more
natural setting. The participants interacted with the prototypes using a Google
Home mini, creating a stronger feeling for how the end product will behave. Also,
the researchers argued that a natural setting would give a more fair result of the

user experience measured. It does exist prototype programs for creating VUIs and by using them Wizard of Oz would not have been needed. However, these were seen as too complex for the time frame, and did not support Swedish.

Using a questionnaire after the interaction was decided in discussion together with the Chalmers supervisor of the project with the reason of gaining an indication of quantitative data, as well as for the ease of analysing. It was also decided for interviews to be held afterwards to make it possible to gain insights as to why eventual extreme values in the questionnaire was chosen, and to enable the participants to explain their overall experience of the interactions. The researchers discussed if the participants should be informed about the interview being a discussion about their questionnaire responses before filling it out or not. Not informing the participants beforehand would probably lead to more honest responses, but the researchers deemed it unethical to deliberately make the participants think that their responses would be anonymous. Therefore, the participants were informed when given the questionnaire that the test leader would look at their answers afterwards and holding an interview about them.

As the user test aimed at investigating how the the third set of guidelines contributed to the user experience, the questionnaire was based on the Subjective assessment of speech system interfaces (SASSI). SASSI is a standardized questionnaire for voice interaction design that is suitable for measuring user experience. SASSI consists of 34 statements divided in five categories: System response accuracy, Likeability, Cognitive demand, Annoyance, Habitability, and Speed (Hone & Graham, 2000). The questionnaire was adapted to fit the current user test, resulting in 17 statements from all categories except Speed. Two statements, "I felt overwhelmed using the system" and "To use this system, I need to know things in beforehand" were added to the questionnaire by the researchers due to them being relevant for the guidelines overarching themes "Do not overwhelm the user" and "Do not assume that the user knows things in beforehand" and not being covered in the SASSI-statements. The questionnaire used in the user test consisted of 19 statements with corresponding Likert scales from one (Strongly disagree) to five (Strongly agree). The questionnaire was translated into Swedish by the researchers.

In this test, no guidelines were explicitly tested. Instead, as many guidelines as possible were included in the created interactions. "The system should support navigational utterances", "The system should be able to handle "I don't know"", "Personalisation should be used", and "Variation should be provided in options given to the user" were however not included in the interactions. The researchers judged that incorporating the guidelines "The system should support navigational utter-

ances" and "The system should be able to handle "I don't know"" as redundant when paired with other guidelines and if included anyway would not result in any new insights during this test, and was therefore excluded. "Personalisation should be used" were excluded since no data about the participants were accessible making personalisation impossible. The interactions created were deemed too short for the guideline "Variation should be provided in options given to the user" to be able to work.

A pilot test with one participant was conducted in beforehand. The pilot test resulted in re-phrasing the translation of two statements in the questionnaire. The pilot test also revealed that Google Home's microphone had to be turned off during the test preventing the actual Google Assistant to respond to the users utterances. This, on the other hand, resulted in the lights on Google Home to turn orange (see figure 5.8). For an experienced user, this would make it obvious that they were not interacting with the actual system. Therefore, the participants were told in the introduction that the system they were gonna interact with was a prototype and that it may not behave as they were used to. This would also make experienced users not pay attention to details in the interactions that may not be as they are used to.

Figure 5.8: *Google Home mini with the microphone turned off causing the lights to turn orange.*

The user test was conducted during the 29th of April at Swedish Radio's office in Stockholm. Seven participants with an age range between 25-46 years, with a mean of 35 years, participated in the user test. Their experience with smart speakers were evenly distributed, where two participants had tried using one, two participants had one themselves and three participants had never used a smart speaker before.

All participants were instructed to act as users of Swedish Radio's action for Google Home not knowing what they wanted to listen to.

The test consisted of two interactions. The phrases used by Google was recorded using one of the researchers voice and played through a Google Home mini. Because of the phrases being pre-recorded, the participants had to follow a script for the whole interaction leaving no room for improvised responses. Even though the final response theoretically can be improvised by the participant this would lead to the users understanding that it was not a working product that they were testing. Therefore, the researchers decided that it was more important to keep the illusion of the users actually interacting with a system than to include improvised answers.

The same researcher that acted as test leader in the first user test and also had not recorded the responses for Google Home acted as test leader in the second user test as well. The other researcher said as few words as possible during the test in order to avoid participants from recognising the voice.

The test leader welcomed the participants and gave an introduction to the test. During the introduction the participants were given information about the researchers, the project, and the test, as well as about their rights during the test. The participants were told that the other researcher was taking notes during the test. This was true for the interview but during the warm-up task and the two following interactions the researcher played the pre-recorded responses using a computer.

Following the introduction was performing the warm-up task so that the participant could try out interacting with Google Home using a script. The participants were asked if the interaction was understood and were then to carry out the two main interactions with corresponding scripts handed out by the test leader (see figure 5.9). After the test, the participants were asked to fill out the questionnaire in another room so that the participant would not feel observed. This was followed by the test leader holding a unstructured interview with the participant based on the answers given in the questionnaire. Answers being extreme values of the questionnaire was in focus, as to understand why they were picked.

The results from the questionnaire and the interview were summarized and analyzed. The statements from the questionnaire was divided into positively and negatively phrased statements which are presented in figure 5.10 as to gain an insight of the overall experience of the system. Results for each individual statement in the questionnaire as well as a summary of the interviews can be found in appendix K.

For the positively phrased statements, one answer was placed at 1 (Strongly disagree) while the rest was between 3 and 5. The answer at 1 belonged to the statement "I
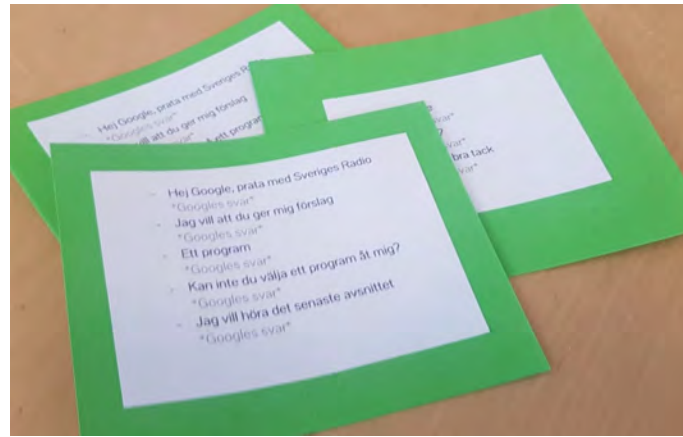
Figure 5.9: *The manuscripts used by the participants during the second test.*
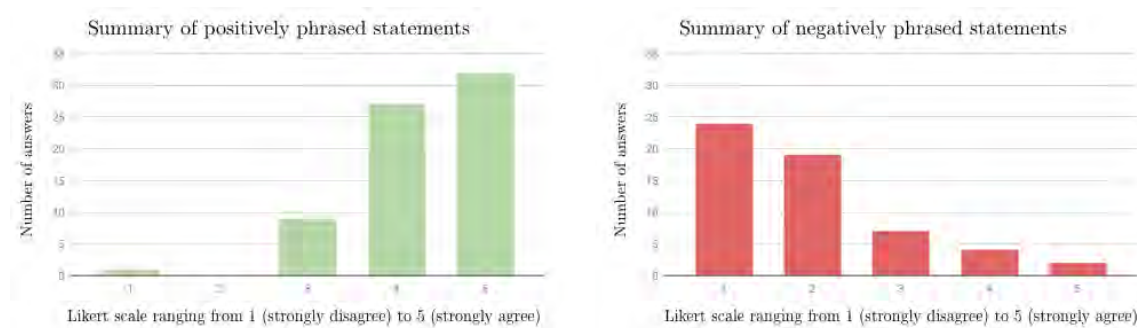

Figure 5.10: *The overarching result from the second test. The result from the questionnaires were divided according to the positively and negatively phrased statements, and summarised in these two charts.*

would use this system". The participant answering this explained that looking for something to listen to is something which is seldom done. Instead, the participant claims usually knowing what to listen to in beforehand.

For the negatively phrased statements, two answers were placed at 5 (Strongly agree), one belonging to the statement "To use this system, I need to know things beforehand" and one to "I felt tense using the system". The participant answering a 5 to "To use this system, I need to know things beforehand" explained that the reason for this is that it is not obvious that the action of Swedish Radio exists in Google Home and how to access it. But while in the action, it did not feel like any beforehand knowledge or information was needed. The participant answering a 5 to "I felt tense using the system" stated that there had been some troubles when using Google Home before, and that it contributed to the feeling of being tense during the test. Also the fact that the two researchers were present while the participant interacted with it contributed. When asked about the system being the cause of the feeling of tension, the participant dismissed that interpretation.

There were four answers placed at 4 (Agree). Two of the answers belonging to the statement "A high level of concentration is required when using the system", one to "The interaction with the system is frustrating" and one to "To use this system, I need to know things beforehand". Regarding "A high level of concentration is required when using the system" the participants answering this explained that you always need a higher concentration level when using these kinds of systems because of the linear interaction and the need of answering the systems questions within a given time. Regarding the statement "The interaction with the system is frustrating", the participant explained that systems like this always feels a bit frustrating but that this system was perceived as good and with fairly short interactions in comparison. Regarding the statement "To use this system, I need to know things beforehand", the participant claims that to be able to use this system, a basic knowledge about how to interact with Google Home is needed. Continuing, the participant explained that it felt like you had to be familiar with Swedish Radio's content in beforehand, and that the interaction in itself did not contribute to this knowledge.

5.4.5 Refining guidelines

The second user test did not explicitly test the individual guidelines but instead the overall user experience. The result from the test did not result in any changes to the actual guidelines but instead contributed to more elaborated motivations. However, the phrasing of the guidelines were changed. In the third set of guidelines, some guidelines were more urging whereas others were more definitive. Also, some guidelines focused on the user and some on the system without any obvious distinction. For the fourth set of guidelines, this was changed in order to create a more coherent set of guidelines. The fourth set of guidelines, with corresponding motivations are presented in the results chapter.

6
Results

This chapter presents the fourth and for this project, final set of guidelines. Sixteen guidelines divided into five themes were created during this project. The guidelines aim to help designers design a VUI for navigating radio content, and are made to be used in combination with existing general guidelines for designing a VUI. In the first section, the guidelines are presented as cards with overarching themes and explanations. In the following section, each guideline is presented with a corresponding motivation.

6.1 Design guidelines

The guidelines are presented as cards in order to improve readability. The cards are color coded in the overarching themes and one theme is presented at a time. However, the order of the presented guidelines or themes has no value to their relevance. Following, themes and their corresponding colors are presented: "Don't overwhelm the user" with the color red (figure 6.1), "The user should feel as in control" with the color yellow (figure 6.2), "Don't assume that the user knows things" with the color green (figure 6.3), "Provide what the user wants and/or needs" with the color blue (figure 6.4), and "The system should be intelligible" with the color purple (figure 6.5).
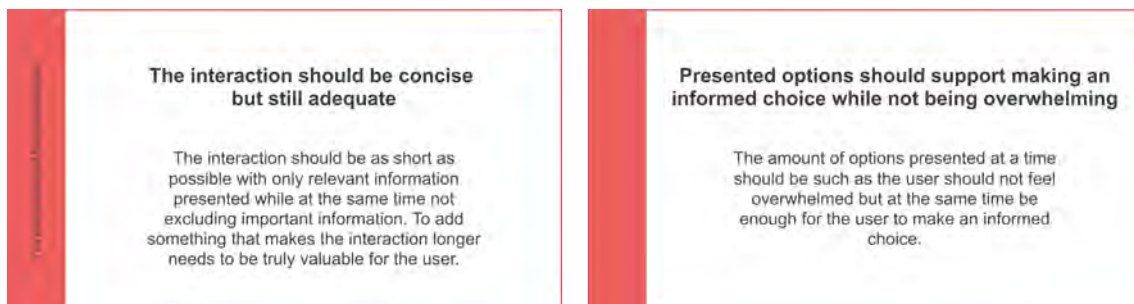

Figure 6.1: *The theme "Don't overwhelm the user" with the two corresponding guidelines "Keep the interaction concise but still adequate" and "Presented options should support making an informed choice while not being overwhelming".*
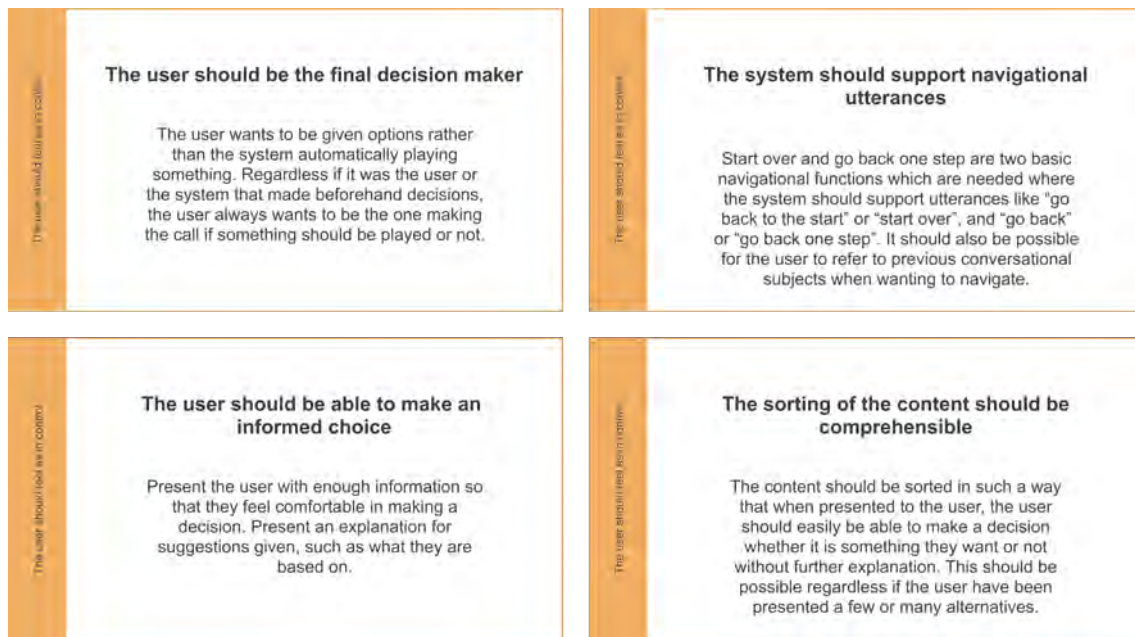
Figure 6.2: *The theme "The user should feel as in control" with the four corresponding guidelines "The user should be the final decision maker", "The system should support navigational utterances", "The user should be able to do an informed choice", and "The sorting of the content should be comprehensible".*
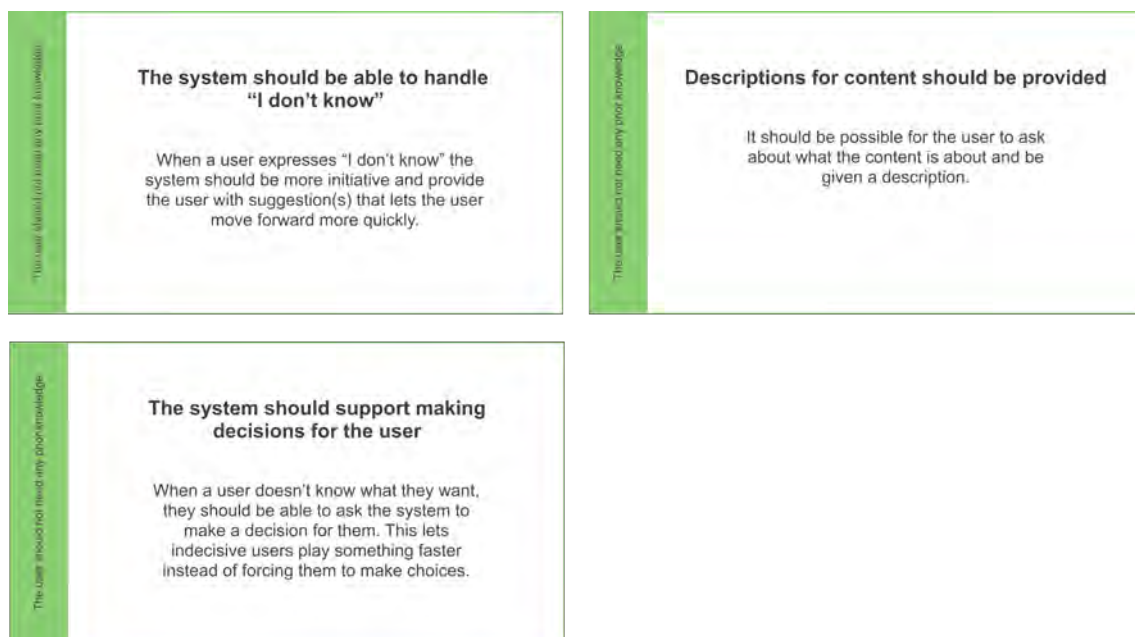

Figure 6.3: *The theme "Don't assume that the user knows things" with the three corresponding guidelines "The system should be able to handle "I don't know"", "Descriptions for content should be provided", and "The system should support making decisions for the user".*

Figure 6.4: *The theme "Provide what the user wants and/or needs" with the five corresponding guidelines "Personalisation should be used", "Design for both novice and experienced users", "Variation should be provided in options given to the user", "Content should be prioritized", and "Design for both the patient and the impatient user"*
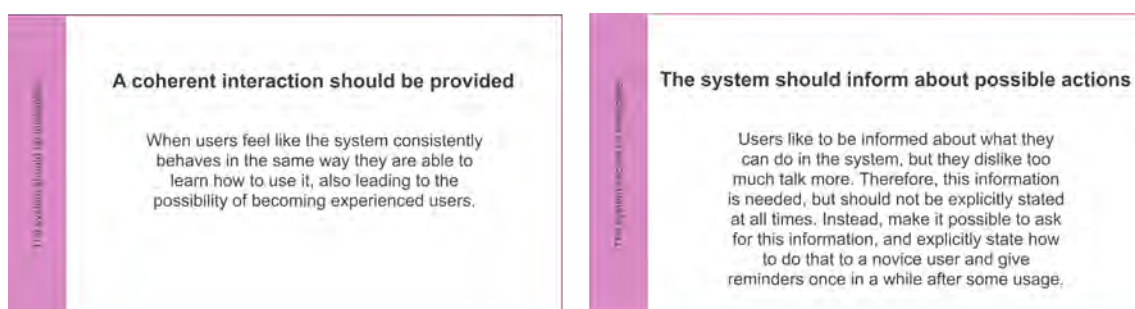

Figure 6.5: *The theme "The system should be intelligible" with the two corresponding guidelines "Provide a coherent interaction" and "Inform about possible actions in the system"*

6.2 Motivations for guidelines

Below, each guideline will be presented in combination with a motivation for their presence in the final set of guidelines. Overarching themes and explanations have been excluded from this section as the focus is on the motivations.

- **Keep the interaction concise but still adequate**
  - *Motivation*: The first set of guidelines included the guideline "Do not overwhelm the user with too much information", which involved keeping the information given to a minimum. During the restructuring of the guidelines, this instead became an overarching theme. During the first test, the perceived length of the interaction was a prominent subject and mentioned by all participants several times, most of them stating that it was perceived as too long. They wanted the interaction to be as short as possible, and it was observed that as soon as the system brought up by them irrelevant information, it was mentioned as too long of an interaction. When asked about choosing between a short interaction and relevant information, it was still wished for keeping information necessary for the usage of the system. As for this, this guideline was recreated, but phrased according to results from the first test. During the second test, the prototype was redone, making it shorter, cutting information that seemed as not completely necessary for the usage. Five participants of the second test stated, without the test leader asking for it, that there was relevant information presented and that it felt short, which was stated as a positive experience. These findings led to the keeping of the design guideline.

- **Presented options should support making an informed choice while not being overwhelming**
  - *Motivation*: This guideline originally appeared in the first set of guidelines created based on an extensive amount of data gathered at that point. The guideline was then phrased as "Don't present too many options at once". In the second set of guidelines it was instead phrased as "Present a minimum amount of options at a time" as "too many options" was believed as unclear and missing the point of it being as few as possible, a minimum amount. When testing this guideline in the first user test, it turned out to not be completely true. Instead, it became clear that it is more a question about not presenting too much as to not overwhelm, but at the same time present enough options so that the user feels like they can make an informed choice. When very few options were given, the participants felt as they were not given enough information. As for this, the guideline was

changed for the third set of guidelines. No more information regarding this guideline was collected during the second test, and the guideline was kept.

- **The user should be the final decision maker**
  - *Motivation*: This guideline originates from the first set of guidelines where it was a part of the explanation of the guideline "The user should feel as in control", where it stated to "Give options rather than play something automatically". During the restructuring of the guidelines it was extracted and put into an own guideline in the second set of guidelines "The user should be the final decision maker". During the first test, this guideline was one of the most prominent ones regarding confirmation of the guideline. All participants stated during several parts of the test that they never wanted the system to play something without asking first. They also wanted the default to be that they as a user should make all of the decisions and if they feel as they need more help, they should be able to ask for it, but it needs to be their decision. And before the system plays it, it needs to present the suggestion and ask if it should be played. No more information regarding this guideline was collected during the second test, and the guideline was kept.

- **The system should support navigational utterances**
  - *Motivation*: This guideline was already a part of the first set of guidelines, where it was phrased as "It should be possible to use universal navigational terms". In the second set of guidelines it was rephrased, but containing the same meaning, to "Universal navigational functions should be provided". During the first test, this was tested by letting the participants express how they would like to state commands as to start form the beginning and go back to the previously mentioned categories. Afterwards questions regarding navigational terms were asked. All participants expressed that navigational utterances should be supported and that it is obvious that it should be possible to speak with the system in those terms. No more information regarding this guideline was collected during the second test, and the guideline was kept.

- **The user should be able to do an informed choice**
  - *Motivation*: This guideline originates from "Episodes should be presented with amount and a suggestion" from the first set of guidelines. During the restructuring, it seemed to be too specific as a guideline, and the implication of it was believed to be enabling the user to do an informed choice. During the first test, this guideline was tested by presenting the

participants with information such as amount of episodes and several suggestions with corresponding information versus be given no further information about the choice to be made. The test supported this guideline, but not the part about presenting amount of episodes. All participants liked to know which episode that would be played, and suggestions for which episode that could be played (in this case first, latest, most listened). No more information regarding this guideline was collected during the second test, and the guideline was kept.

- **The sorting of the content should be comprehensible**
  - *Motivation*: This guideline was a part of the first set of guidelines, where it was phrased as "The content should be divided to make navigation easier". During the restructuring of the guidelines, this was rephrased due to it being somewhat unclear, especially the meaning of divided. The new phrasing was "The content should be relevantly sorted". During the first test, while not being explicitly tested, several participants expressed that it was hard making a decision based on the unclear sorting of the content (which was traditional categories such as documentaries, sports, humour et cetera). The phrasing of the guideline was therefore changed as to pinpoint the need of the sorting being comprehensible. During the second test, a new kind of sorting was tested where more precise categories were given. While this was liked by some, and disliked by some, the problems expressed by those disliking it was because of them being to narrow rather than them not being comprehensible. This did therefore not give any further indications for this guideline and it was kept.

- **The system should be able to handle "I don't know"**
  - *Motivation*: This guideline also originates from the first set of guidelines where it was part of the explanation for the guideline "Don't assume that the user knows things". During the restructuring, that guideline became an overarching theme while handling "I don't know" seemed important enough for being its own guideline. This was then tested during the first user test. All participants expressed it as a matter of course that the system in some way should be able to handle the answer "I don't know". As it is about a user that really does not know, this answer may be common and it should therefore be designed for, and not treat it as an indecisive user and repeat the options to give it another shot at making the decision. No more information regarding this guideline was collected during the second test, and the guideline was kept.

- **Descriptions for content should be provided**

– *Motivation*: This guidelines originates from two guidelines in the first set, "Provide explanations for content" as well as "provide samples for content". As a matter of wording, they were rephrased for the second set of guidelines to "Explanations for content should be provided" and "Samples for content should be provided". In the first user test, both explanations for programmes as well as samples for episodes was given, which resulted in the participants preferring explanations. Because of the difficulty in pinpointing the preference being actually explanations or if the sample used was not good enough, the guidelines were rephrased into a common guideline being "Descriptions for content should be provided" which can involve both explanations as well as samples. During the second test, explanations were used which were stated as appreciated by the participants. It was also mentioned during the second test that samples may be even more informative, also creating an understanding of the mood of the episode. As for this, the guideline is kept.

- **The system should support making decisions for the user**
  – *Motivation*: This guideline was created after the first test, as it was mentioned by several participants that this was something desirable. During the second test, this guideline was implemented in one of the two interactions, which was stated as appreciated by the participants. As for this, the guideline stays.

- **Personalisation should be used**
  – *Motivation*: This guideline originates from the first set of guidelines, where it was phrased as "Use personalisation". As of the second set of guidelines, the phrasing have been the same: "personalisation should be used". This guideline was never explicitly tested in any of the user tests, as it was perceived as troublesome in the test settings conducted. Personalisation is however believed to contribute to presenting better suggestions to the user as it is based on their own usage as well as others that have similar usage. During the first test, it was however mentioned by four out of five participants that it should be based on previous usage. Also, during the second test all participants expressed that personalisation was desired, or even expected. Therefore, the guideline stands supported.

- **Design for both novice and experienced users**
  – *Motivation*: This guideline has remained unchanged since the first set of guidelines. The guideline was tested during the first user test, resulting in the guideline being confirmed. During the test, participants expressed that they want more information during their first interaction and also more

guiding in the beginning of their usage. Later, all participants expressed that they appreciated the possibility of faster interactions and a change of greeting and interaction techniques within the system, when becoming more experienced with the system. During the second test, even if not explicitly asked for, the participants stated that they appreciated the different greetings, and the possibility of getting a more thorough greeting at first, to then only getting the short one. It was also mentioned by four participants that the system needs to change its behaviour somehow if using it often, so that it is not needed to go through all the steps in order to get suggestions. These findings resulted in the guideline being kept.

- **Variation should be provided in options given to the user**
  - *Motivation*: This guideline was never tested in any of the user tests. It is instead reasoned that variation should be provided to let users explore different kinds of content and prevent filter bubbles. This is especially important for Swedish Radio, who has included in their public service-accounting that they aim at preventing filter bubbles. (Sveriges Radio, 2018b).

- **Content should be prioritized**
  - *Motivation*: This guidelines was part of the first set of guidelines, then being phrased as "The content should be prioritized based on relevance and importance". For the second set of guidelines, it was rephrased as "Content should be prioritised". This guideline was never explicitly tested, reasons being the difficulty in making an appropriate test for it. During both tests it was however mentioned that the suggestions given are expected to be based on something, and that something being the system believing it to be relevant for the user. This guideline is believed to be fundamental to presenting suggestions, and is kept as a constant reminder to always present suggestions believed to be relevant to the user or important for the company behind it. As for this, the guideline is kept.

- **Design for both the patient and impatient user**
  - *Motivation*: This guideline was created due to an insight during the first test. During the test, it was perceived that some users have patience when interacting with these system, and some users have not, and both type of users needs to be designed for. This was implemented during the second test by making the main interaction as concise as possible, while also presenting an alternative for more exploration, something which was expressed as appreciated by the participants.

- **Provide a coherent interaction**

– *Motivation*: This guideline was created during the first set of guidelines, and were then phrased as "The interaction should be coherent". For the second set of guidelines, the wording was reversed, resulting in the phrasing "Provide a coherent interaction". This guideline was never explicitly tested, it was however implemented by making all interactions in both first and second test behave in the same way, thus coherent. During the first test, all participants were perceived as learning the system during usage. During the last interaction of the test, being an interaction for experienced users, all participants stated that they felt as the interaction for experienced users were intelligible and expressed appreciation for it. As coherency provides the user with the possibility of more smoothly becoming an experienced user, this was interpreted as an indication for confirming the guideline. During the second test, the questionnaire involved a statement regarding coherency. All partitipants expressed the interaction as coherent, answering a four (agree) or a five (strongly agree) on the questionnaire. This again can be seen as an indication for the guideline, and it is therefore kept.

- **Inform about possible actions in the system**
  – *Motivation*: This guideline originates from the first set of guidelines, and is derived from "Make users understand the system". During the restructuring, the guideline was rephrased to its current phrasing as to better pinpoint what it is the user should understand. During the first test, this guideline was tested and resulted in that the users likes being informed, but dislikes too much talk more. The conclusion made was therefore that it does not need to be stated explicitly at all times for everyone, but rather state the possibility of having possible actions explained, and/or once in a while remind users about it. This became the explanation of the guideline to state for users of these guidelines that explicitly informing about actions at all times may result in impatient users being both frustrated as well as irritated. During the second test, it was expressed that the interaction was perceived as easier than expected and one participant stated that: "My mom could have used this, who usually have troubles with technology", something which this guideline may be part of creating.

7

Discussion

This chapter will present a discussion of the project, its results, process, ability of generalisation, ethical issues, and future work.

7.1   Result and Findings

The project resulted in sixteen guidelines, sorted into five themes. The guidelines were first created based on existing guidelines, competitive testing, stakeholder interviews, and an ideation session. These guidelines were then during several iterations restructured, renamed, deleted and added. The process of creating and refining the guidelines have been both complex but also instructive. Even so, the result have some room for improvements. The guidelines created during this project should at this point be usable, but not seen as definite. More iterations are needed partly as to involve a broader scope of users, long time usage, natural setting, but also because of several other factors which will be further explained during this chapter. It is also needed to test with designers as to find out how they would interpret and implement these guidelines when designing.

During the process of creating these guidelines, there exists possible external variables which could have affected the results. One of those things are the prototypes made for testing the guidelines. The prototypes was made mainly based on the created guidelines, but also existing guidelines and general recommendations for designing a VUI. It can however not be completely excluded that other things may have affected the creation of the prototypes, such as gained knowledge during the project not resulting in guidelines which unintentionally affected the outcome anyhow. We as creators of the guidelines also knew exactly what we meant with them, but it is questionable however others do. Our project lacks an external creator or analyst of prototypes, something which preferably would be used if done again.

The guidelines per se turned out quite general, and there is a possibility that they can be used outside of the scope of this project, both outside of radio content as well as outside of VUIs. This was a recurrent subject for discussion during the

project, and something which at first was seen as a limitation of the project. Later, it was rather seen as a result of not wanting them to be too specific as to tie designers to specific solutions rather than opening up for innovative ideas while also following the guidelines. The focus during the project has always been on navigation, and resulted in these guidelines. The guidelines have been tested for this cause, and even if they may be applicable for other causes as well, that does not make it less of guidelines for this.

There have also been several discussions regarding similarity to other already existing guidelines during this project. Most guidelines being the same or very similar to others were excluded during the process, but there are some guidelines still kept. To begin with, existing guidelines that have been found during this project have most of them brought up similar areas, but phrased it differently. It is troublesome creating a whole new set of guidelines, never seen before, but the different guidelines still contributes. The guidelines in this project have the advantage of having tested an area within VUIs which have been seen as challenging by others and therefore not being studied in any great extent, if at all. The guidelines created during this project being similar to existing ones were kept because of being fundamental for the complete set of guidelines for this cause. To not include how options should be presented, or that both novice and experienced users should be designed for felt as excluding a fundamental part of designing for a successful navigation.

The similarity between different sets is also the reason behind us only using the guidelines presented in the background section. It would not be possible to consider all guidelines and these were chosen based on relevance as well as being the, in the sake of the guidelines for GUIs, most accepted ones.

In conclusion, the result of this project has its limitations and could benefit from further exploration as to state as definite, if definite is ever possible in the field of design. It has however emerged from an extensive amount of previous research as well as testing and should therefore contribute to the field of voice interaction and VUI.

7.2 Process

The process as a whole is in detail presented in a chapter of this report. It is however not discussed potential limitations of it. This, and how that may have impacted the result, is further explained and discussed below.

In the two user tests, manuscripts were used. The first user test included some

improvised answers from the participants to capture how they reacted to the question asked by the system, but apart from this the participants only said what was stated by the manuscript. The use of manuscripts may have had an impact on the result, leading to only a result of how the participants reacted to the system and not how the participants would have preferred to interact with it. Also, when participants of the tests stated how they perceived the interaction may have resulted in a more positive outcome due to the fact that they used manuscripts and therefore knew what to say, rather than being the one having to come up with it. What to say to the system can be troublesome when interacting with VUIs, something which was not captured during our tests. A solution to this would have been to let the user speak freely with the system, and having designed for several possible interactions. Making these types of designs is however very time-consuming and would not have been possible due to the time frame of this project. Furthermore, it can be claimed that the participants reactions to the system was the intended outcome, rather than how the user would want to express themselves when speaking with the system. We wanted to capture how the system should behave in order for the users to feel as it was advantageous to use, and the users behaviour when interacting with the system should be left for future research.

Not all guidelines were tested during the two user tests, and some guidelines were never tested but were even so included in the final set of guidelines. The guidelines "Variations should be provided in options given to the user" and "Content should be prioritized" were never tested but are still a part of the final set. In the process chapter, their relevance are argued for as well as to why they were not tested, but it could still be perceived as questionable to include guidelines that never have been tested. They have however emerged from research and are not made up from thin air. "Variations should be provided in options given to the user" is based on the guideline "Optimize for relevance" created by Google (n.d.-c) as well as input from the stakeholder interviews and the ideation session. "Content should be prioritized" is based on Amazon (n.d.-b) guideline "Be contextually relevant" and input from the stakeholder interviews.

During the process, users were only included during the second user test. Claiming to use a human-centered design approach, it could be argued for that users should have been included in more steps. The intention was to include users in both user tests, but due to an availability selection all participants in the first user test were design students. A problem encountered during the project, and which Swedish Radio also informed about, was that users to Google Home are difficult to find and users to Swedish Radio's action is nearly non-existing. At the same time, since Swedish Radio has Sweden's population as a target group, it can be argued for that also the design

students can be seen as users even though their background can have effected their answers during the tests.

7.2.1 Competitive testing

During the competitive testing, only actions that support Swedish were included in the test. Our argument for this was that the project was about Swedish radio content and that the prototypes to be created would be in Swedish. However, this is not a valid argument and a limitation in the study. The market for smart speakers in Sweden has not boomed yet, and only Google Home supports Swedish. A result of this is that it exists few actions in Swedish and the ones that exist are more or less deficient. By including actions supporting English, more insights and inspiration could have been gained.

7.2.2 Stakeholder interviews

Stakeholder interviews were conducted based on recommendations from Cooper et al. (2014). However, too much time was spent on this step compared to the outcome. We spent a huge amount of time on researching how to conduct interviews, how to be successful interviewers, and how to analyze interviews. When conducting and analyzing the interviews, recommendations from Jacob & Furgerson (2012); Rowley (2012); Turner (2010) were followed, leading to the interviews being audio-recorded, transcribed and analyzed by using content analysis. Transcribing and using content analysis took a vastly amount of time and in the end we realized that we would probably have gotten enough information from simply taking notes during the interviews and summarizing the notes. A positive note from this is that the lesson was learned, and neither audio-recording in combination with transcribing, or content analysis was used again during the project saving a lot of time.

7.2.3 Ideation session

The aim of the ideation session was to gain new insights in how to design for VUIs, and not being stuck in what have already been invented. During the session, the participants acted as Google Home and a user trying to find something to listen to. The participant acting as Google Home had different characteristics to follow, but apart from that no restrictions for the interaction were given. The purpose of this was to not limit the creative thinking with what is feasible right now. Having few restrictions may have positioned the findings from the interview further into the future then was intended. For example, the user was able to interrupt Google, and register facial and emotional expressions. However, we found it more important to value the need for innovation than to make the environment more similar to an

actual interaction, by for example prohibit interruptions, making the user say "Hey Google" in every other turn, and not having the participants see each other.

7.2.4   First user test

During the first user test, one of the researchers acted as Google Home and read from a script when interacting with the participants. By reading out loud variations in speed, pitch, and emotional expressions can differ between different tests and maybe impact the outcome of the test. For example, participants could have perceived the interaction as too long because of Google not talking fast enough. But by letting Google read out loud the test could be conducted in an earlier stage allowing for more iterations then if time would have been spent on recording or developing a hi-fi prototype. Also, this likely contributed to participants experiencing that the interaction was a prototype that easily could be changed, maybe resulting in more critical and innovative answers during the interview.

To test the second set of guidelines, different interactions were created based on the guidelines, existing guidelines and recommendations from Pearl (2017). When presented with different variations of an interaction, the participants were asked to choose which one they preferred. This does however not validate that the variation picked the most is the best solution, only that it is the best one from the variations created. Also, if a guideline was validated or not is heavily influenced by how we decided to implement that guideline. It cannot be ruled out that a guideline that did not get supported during the user test may only have lacked support due to a poor implementation. An example of this is the implementation of the guideline "Samples for content should be provided", where the sample used in the user test was criticised by the participants probably influencing their answers about if samples in general are desirable. However, in most cases participants expressed that they did like one of the variations and regardless of if a better solution is out there these variations are appreciated by the participants which in turn gives support to their corresponding guidelines.

7.2.5   Second user test

In the second user test, participants interacted with a Google Home mini by reading from a script. One of the researchers were controlling the Google Home without the participants' knowledge and played pre-recorded responses. In this test, only two short interactions were used. The interactions together included the majority of the guidelines from the third set, but some participants expressed after filling out the questionnaire that it is difficult to gain a full comprehension of the system from such short interactions. Short interactions were used to keep the sessions within the

set time frame, but may have resulted in participants not being able to answer the questionnaire as intended.

The second user test did not involve any improvised answers. This could have resulted in participants not paying attention to the interaction and what was actually being said and solely reading from the script. As mentioned in the process chapter, improvised answers were deliberately excluded from the scripts and it was argued that it was of greater importance to keep the illusion of the participants actually interacting with a system than to include improvised answers.

Using pre-recorded answers and having the interaction via Google Home took focus from the interaction per se. Participants had a hard time looking beyond how the voice sounded, that no errors occurred, and in general that the interaction went on smoothly. These factors were not in the scope of the test and made participants lose focus of the interaction. We wanted to test the interactions in a more natural setting, but this may have affected the results making participants more positive due to the interactions being smoothly rather than them being good.

7.2.6 Excluding errors

Another factor that could have made participants from both user tests to be more positive in their answers was the exclusion of errors. Errors were early on excluded from the scope of the project since it was deemed as to big of a matter for the project and that detailed previous work about how to design for errors in VUIs exists (Pearl, 2017). But since it does exist previous work about how to design for errors, they could have been included without testing if they were good or bad. Including errors could however make participants lose focus on the other parts of the interactions.

7.2.7 Creating and refining guidelines

A big part of the project was the creation and refining of guidelines. This process could have benefit from a more structured approach. Worth repeating is however that all information that were used came from literature reviews, competitive testing, stakeholder interviews, ideation session, and the two user tests. But it would have been favorable if a pre-defined structure had been used while refining and formulating the guidelines.

7.3 Generalisation

Regarding the generalisation of this project, it may be troublesome to answer whether the result can be generalised or not as testing has not been conducted on a great

amount of users. It can however be claimed that changes based on the first test, where the participants were students of the Interaction design programme, seemed to be appreciated during the second test, where the participants were collected by Swedish Radio and belonging to another user group. This may give an indication of the generalisability, but further testing including probability sampling is needed. The biggest issue regarding this is the fact that because of being public radio, Swedish Radio focuses on including everyone in their design, making a probability sampling nearly impossible.

Continuing, the field of navigating using voice interaction lacks existing research, resulting in a hard time finding guidelines as well as other relevant information to base our research on. This led to a more extensive work, but also to the research being a start of something, rather than being fully generalisable. As the project was done in collaboration with Swedish Radio, their content as well as their knowledge was used during the project. This makes it questionable as to if it is possible to generalise it to all radio content. To ensure this, testing using other content should be conducted. Also, because of Google Home being the only smart speaker understanding Swedish, it was the only smart speaker used. This also makes it questionable if it is applicable to all smart speakers. However, this may be argued for because of the smart speakers of today having very similar functions. When implementing a design for navigating radio content the result in different smart speakers does not differ remarkably, if at all.

Conclusively, there are potential weaknesses regarding the generalisability of the resulted guidelines of this project. However, this may, hopefully, only be the beginning of discovering possibilities for navigation using speech, where more research results in greater possibilities of generalisation.

7.4 Ethical issues

Several decisions and actions were made in regards to ethical considerations when involving participants during the process. Participants were always told of their rights in beforehand, including their right to cancel at any point and that the data collected would at all times be handled anonymously. Collecting sensitive information was avoided, and it was never asked for more information than was needed. It was also ensured to make the participants feel as comfortable as possible, avoiding any possible discomfort such as anxiety.

During the interviews, a consent form was used informing the participants about the test, how the data will be used, and that the interview will be recorded if approval was given. The consent form was signed by both the researchers as well as by the

participant, symbolising an equal understanding of the rights and obligations in regards to the interview.

For the first and second user test, filter bubbles as well as privacy concerns was considered. Regarding filter bubbles, it was not seen as an extensive concern because of the limited interactions with the prototype, meaning not enough suggestions were to be given to result in any filter bubbles. However, the suggestions given were even so considered and varied as to not initially contribute to any filter bubbles. During the first user test, one of the researchers acted as Google Home, something which the participants were fully aware of. Therefore, privacy concerns during interacting with a Google Home was not an issue. During the second test, however, a Google Home was used and even if the assistant was turned off leaving room for the prototype to act as one, the participants were under the impression that the system was working. The participants were told in beforehand that their results would be anonymous, however, it was never explicitly stated that Google would not take part of any data. Since the participants were not aware of this, it left room for uncertainty which was stated by one participant during the second test. It was never explicitly stated that the concern was in regard to this test, it was however mentioned that the reason for not owning or wanting a Google Home was because of concerns of Google handling their data. Even if it was not perceived as so, it can not be excluded that that participant, or any of the others, were concerned about their data privacy, resulting in a possible limitation.

During the second test, the questionnaire was discussed together with the participant after them having answered it. This was a subject for discussion between the researchers and the supervisor, as it may result in the participants feeling discomfort in regards to feeling as they will be held accountable for their answers. The discussion resulted in that the test leader explained that the questionnaire will be used during the interview both during the introduction and when handing out the questionnaire, as to prepare the participants. It was also clearly explained that their answers were of true value, both good and bad, and that it should not feel as they needed to please us. Even if this potentially could affect the results of the questionnaire, as the participant may feel obliged to answer more positively, not telling them could result in a severe ethical flaw.

In conclusion, there have been several ethical considerations when involving participants during the process. However, one possible limitation lies in concerns for Google collecting data during the second test, even if not stated explicitly. For future work, better care for people being worried about the companies behind the smart speakers collecting data should be taken.

7.5 Future work

As been mentioned earlier, more iterations are beneficial and the final set presented in the result section is only a final set in the aspect of this project. During the project, the guidelines relevance have been tested, but since the intention of the guidelines are that they should be used by other designers for VUIs also this needs to be tested. The guidelines should thus in future works be tested with other designers as subjects in order to investigate how they interpret the guidelines and how they would use them.

In the user tests, an availability selection was used in order to select participants. Further testing would benefit from using a probability sampling making sure that the participants are a representative sample of the target group. This includes narrowing down the target group, as it is of today everyone in Sweden.

The project aimed at investigating what should be considered when designing a VUI for navigating radio content using a smart speaker. In this project, only the radio content from Swedish Radio was used. Future studies should include other radio content as well, increasing the generalisation of the results.

As an ending note, more testing concerning navigation in VUIs is needed. The field still has many unexplored parts and especially navigation is seen as an unsolved issue (Furqan et al., 2017). Testing and exploring the area of navigation in VUIs without the constraints set by this project would also contribute to the field of VUIs moving forward.

8
Conclusion

An upcoming paradigm shift in how we interact with our devices is said to occur, moving from screen-based interaction towards voice-based. The field of voice user interfaces is yet in its infancy, and how to navigate using voice seems to be unexplored. A study with the aim of exploring how to design for navigation using voice was therefore conducted. The study was done in collaboration with Swedish Radio and aimed at answering the research question:

What should be considered when designing a voice user interface
for navigating radio content using a smart speaker?

The research question was answered by creating design guidelines. The study resulted in a set of sixteen design guidelines, sorted into five themes. The themes are as follows: "Don't overwhelm the user", "The user should feel as in control", "Don't assume that the user knows things", "Provide what the user wants and/or needs", and "The system should be intelligible".

The guidelines created during this project should at this point be usable, but not seen as definite. More iterations are needed partly as to involve natural setting, a broader scope of users, and long time usage. It is also needed to test with designers as to find out how they would interpret and implement these guidelines when designing. The project has however initiated design standards in the complex field of finding solutions for navigation using speech.

References

Allwood, J., & Jensen, M. (2012). *Kognitionsvetenskap: En introduktion.* Lund: Studentlitteratur.

Amazon. (n.d.-a). *Alexa, Echo Devices, and your privacy.* Retrieved from https://www.amazon.com/gp/help/customer/display.html/ ?nodeId=GVP69FUJ48X9DK8V

Amazon. (n.d.-b). *Voice.* Retrieved from https://developer.amazon.com/docs/alexa-design/design-voice.html

Budiu, R., & Laubheimer, P. (2018, 7). *Intelligent assistants have poor usability: A user study of Alexa, Google Assistant, and Siri.* Retrieved from https://www.nngroup.com/articles/intelligent-assistant-usability/

Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). *Voice user interface design.* Boston: Addison-Wesley.

Cooper, A., Reimann, R., Cronin, D., & Noessel, C. (2014). *About face: The essentials of interaction design* (4th ed.). Indianapolis: John Wiley & Sons.

DAC. (2018). *The UX of voice user interface (VUI): What it means for brands.*

Dahl, D. (2009, 10). Standards in the voice user interface. *Speech Technology*, 41.

Dam, R., & Siang, T. (2018). *Learn How to Use the Best ideation methods: Brainstorming, braindumping, brainwriting, and brainwalking.* Retrieved from https://www.interaction-design.org/literature/article/learn-how-to-use-the-best -ideation-methods-brainstorming-braindumping-brainwriting-and-brainwalking

Dasgupta, R. (2018). *Voice user interface design: Moving from GUI to mixed modal interaction* (1st ed.). Apress. doi: https://doi.org/10.1007/978-1-4842-4125-7

Dickson, B. (2018). *Beware the privacy and security risks of smart speakers | TechTalks.* Retrieved from https://bdtechtalks.com/2018/06/05/ google-home-amazon-echo-privacy-security-risks/

References

Fischer, J. (2017, 7). How does a virtual assistant work? *Gear patrol*. Retrieved from https://gearpatrol.com/2017/07/19/how-does-a-virtual-assistant-work/

Furqan, A., Myers, C., & Zhu, J. (2017). Learnability through adaptive discovery tools in voice user interfaces. *CHI 2017*, 1617–1623.

Gaver, W. (2012). What should we expect from research through design? In *Chi'12* (pp. 937–946). Austin: ACM Press.

Google. (n.d.-a). *Data security & privacy on Google Home.* Retrieved from https://support.google.com/googlehome/answer/7072285?hl=en

Google. (n.d.-b). *Language.* Retrieved from https://designguidelines.withgoogle.com/conversation/style-guide/language.html

Google. (n.d.-c). *Learn about conversation.* Retrieved from https://designguidelines.withgoogle.com/conversation/conversation-design/ learn-about-conversation.html#

Google. (2019). *Actions on Google.* Retrieved from https://developers.google.com/actions/extending-the-assistant

Hall, T. (2017). *How to choose a user research method.* Retrieved from https://uxplanet.org/how-to-choose-a-user-research-method-985112051d84

Hartson, R., & Pyla, P. S. (2012). *The UX book: Process and guidelines for ensuring a quality user experience.* Waltham: Morgan Kaufmann.

Holoubek, S., & Bowling, E. (n.d.). *Voice technology isn't just a trend; it's a paradigm shift.* Retrieved from https://www.luminary-labs.com/insight/voice-technology-paradigm-shift/

Hone, K. S., & Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering*, *6*(3-4), 287–303.

Interaction design foundation. (n.d.). *Design guidelines.* Retrieved from https://www.interaction-design.org/literature/topics/design-guidelines

Interaction design foundation. (2016). *How Do You Choose the Right UX Research Technique?* Retrieved from https://www.interaction-design.org/literature/article/ how-do-you-choose-the-right-ux-research-technique

Jacob, S. A., & Furgerson, P. S. (2012). Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research. *The*

Qualitative Report, *17*(42), 1–10. Retrieved from
http://www.nova.edu/ssss/QR/QR17/jacob.pdf

Kabir, K. (2019). *Smart speakers - everything you need to know.* Retrieved from
https://www.whathifi.com/advice/smart-speakers-everything-you-need-to-know

Karlsson, M. (2018, 10). *Google Home till Sverige den 24 oktober.* Retrieved from
https://omni.se/google-home-till-sverige-den-24-oktober/a/L0Ev7Q

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening? Privacy
perceptions, concerns and privacy-seeking behaviors with smart speakers.
Proceedings of the ACM on human-computer interaction, *2*(CSCW), 1–102.
Retrieved from https://doi.org/10.1145/3274371   doi: 10.1145/3274371

Martin, B., & Hanington, B. (2012). *Universal methods of design: 100 Ways to
research complex problems, develop innovative ideas, and design effective solutions.*
Beverly: Rockport publishers.

Mctear, M., Callejas, Z., & Griol, D. (2016). *The Conversational interface: Talking
to smart devices.* Basel: Springer.

Nass, C., & Brave, S. (2007). *Wired for speech: How voice activates and advances
the human-computer relationship.* MIT.

Nielsen. (2018). *(Smart) Speaking my language: Despite their vast capabilities,
smart speakers are all about the music.* Retrieved from
https://www.nielsen.com/us/en/insights/news/2018/smart-speaking-my
-language-despite-their-vast-capabilities-smart-speakers-all-about-the-music.html

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. In
*Proceedings of the sigchi conference on human factors in computing systems
celebrating interdependence - chi '94* (pp. 152–158). New York: ACM Press. doi:
10.1145/191666.191729

Nielsen, J. (2000). *Why you only need to test with 5 users.* Retrieved from
https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Norman, D. (2013). *The design of everyday things* (2nd ed.). New York: Basic
Books.

Päärni, A. (2017). *Designing a voice controlled interface for radio guidelines for the
first generation of voice controlled public radio* (Tech. Rep.). Retrieved from
http://umu.diva-portal.org/smash/get/diva2:1114580/FULLTEXT01.pdf

Pariser, E. (2011). *The Filter bubble: What the internet is hiding from you.* New

York: The Penguin Press.

Pearl, C. (2017). *Designing voice user interfaces: Principles of conversational experiences* (1st ed.). Sebastopol: O'Reilly Media, Inc.

Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction design: Beyond human-computer interaction* (4th ed.). New York: John Wiley & Sons Inc.

Rensink, R. A. (2000). *Scene Perception* (Vol. 7). Oxford University Press. Retrieved from https://www2.psych.ubc.ca/~rensink/publications/download/RR-EncycPsych.pdf

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169.

Rohrer, C. (2014, 10). *When to Use Which User-Experience Research Methods.* Retrieved from https://www.nngroup.com/articles/which-ux-research-methods/

Rowley, J. (2012). Conducting research interviews. *Management Research Review*, *35*(3/4), 260–271. Retrieved from https://www.emeraldinsight.com/doi/pdfplus/10.1108/01409171211210154 doi: 10.1108/01409171211210154

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., & Landay, J. A. (2017, 12). Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. In *Proceedings of the acm on interactive, mobile, wearable and ubiquitous technologies* (Vol. 1, pp. 1–159). ACM. doi: 10.1145/3161187

Schacter, D. L., Gilbert, D. T., Wegner, D. M., & Nock, M. K. (2014). *Psychology* (Third ed.). New York: Worth Publisher.

Seifert, D. (2016, 11). *Google Home review: Home is where the smart is.* Retrieved from https://www.theverge.com/2016/11/3/13504658/google-home-review-speaker-assistant-amazon-echo-competitor

Seon, G. (2017). *Reaching new records in speech recognition.* Retrieved from https://www.ibm.com/blogs/watson/2017/03/reaching-new-records-in-speech-recognition/

Shneiderman, B., Plaisant, C., Cohen, M., & Jacobs, S. (2009). *Designing the user interface: Strategies for effective human-computer interaction* (5th ed.). Addison-Wesley.

Stokel-Walker, C. (2018, 11). *Will you be getting a smart home spy for Christmas?*

Suneson, M., & Björnsson, T. (2018, 8). *Styr radion med rösten.* Retrieved from https://sverigesradio.se/sida/artikel.aspx?programid=4282&artikel=7019412

Sutherland, S. (2015, 7). When we read, we recognize words as pictures and hear them spoken aloud. *Scientific American.* Retrieved from https://www.scientificamerican.com/article/ when-we-read-we-recognize-words-as-pictures-and-hear-them-spoken-aloud/

Sveriges Radio. (n.d.). *Alla kanaler.* Retrieved from https://sverigesradio.se/sida/allakanaler.aspx

Sveriges Radio. (2008). *Radiohistoria.* Retrieved from https://sverigesradio.se/sida/ artikel.aspx?programid=3113%5C&artikel=1971599

Sveriges Radio. (2011). *Fakta om Sveriges Radio.* Retrieved from https://sverigesradio.se/sida/artikel.aspx?programid=3113&artikel=1971296

Sveriges Radio. (2017a, 8). *Pratradio – Sveriges Radio på Amazon Alexa.* Retrieved from https://sverigesradio.se/sida/ artikel.aspx?programid=2370%5C&artikel=6762272

Sveriges Radio. (2017b). *Privacy Policy for Swedish Radio's Google Home and Google Assistant app.* Retrieved from https://sverigesradio.se/sida/artikel.aspx?programid=2370&artikel=6974108

Sveriges Radio. (2018a). *Privacy Policy for Swedish Radio's Amazon Alexa app.* Retrieved from https://sverigesradio.se/sida/artikel.aspx?programid=2370&artikel=6869248

Sveriges Radio. (2018b). *Public service redovisning 2017 - Ljudet i fokus* (Tech. Rep.). Sveriges Radio. Retrieved from https://sverigesradio.se/diverse/appdata/ isidor/files/3113/1d418606-1096-4dee-8440-2603a760f027.pdf

Tatham, M., & Morten, K. (2005). *Developments in speech synthesis.* Chichester: Wiley.

Taylor, K. (2017). *Five warm-ups to ignite your design thinking workshop.* Retrieved from https://smashingideas.com/five-warm-ups-ignite-design-thinking-workshop/

Traxler, M. J. (2007). *Introduction to psycholinguistics: Understanding language science.* Chichester: Wiley-Blackwell. doi: 10.1016/b978-155860702-6/50002-8

Turner, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report*, *15*(3), 754–760. Retrieved from

http://www.nova.edu/ssss/QR/QR15-3/qid.pdf

van den Oord, A., Walters, T., & Strohman, T. (2017). *Wavenet launches in the Google Assistant.* Retrieved from https://deepmind.com/blog/wavenet-launches-google-assistant/

Vincent, J. (2018, 5). *Google launches more realistic text-to-speech service powered by DeepMind's AI.* Retrieved from https://www.theverge.com/2018/3/27/17167200/google-ai-speech-tts-cloud-deepmind-wavenet

Virji, P. (2016, 9). *The evolution of voice search.* New York.

Williams, A. (2018, 2). *How Amazon, Google and Apple use your smart speaker data.* Retrieved from https://www.the-ambient.com/features/how-amazon-google-apple-use-smart-speaker-data-338

Yang, L., Sobolev, M., Tsangouri, C., & Estrin, D. (2018). Understanding user interactions with podcast recommendations delivered via voice. In *Proceedings of the 12th acm conference on recommender systems - recsys '18* (pp. 190–194). New York: ACM Press. doi: 10.1145/3240323.3240389

Appendices

References

A

Interview protocol

The interview protocol was originally in Swedish, and have been translated for the sake of this report.

Intro Linn
Thank you so much for taking the time to participate in this interview. We are Pauline and Linn, and are students at the master program interaction design and technologies at Chalmers in Gothenburg.

We are currently doing our master's project in collaboration with you here at SR, where we look at how to improve the voice-controlled app. A little more specifically, we will examine how to help a user who does not know what they want to listen to, to find this in the voice-controlled app. The interviews we do now will hopefully help us understand your work with Swedish Radio's digital content, and how we can continue this work of making it more accessible through smart devices and voice control.

During this interview we will ask some questions about your work with the voice-controlled app / mobile app. We expect the interview to take about 30 minutes.

Before we get started, we want to ask if it's okay to record the interview? It is only for ourselves so that we can analyze the answers, and no one else will take part in the recordings. We have created a consent form that you can read through and if everything feels okay, sign. Do you have any questions before we start?

Goal
Helps us begin our work on developing prototypes to test, where we can use their already found knowledge and strategies around the area.

Questions
(Start recorder)
We'll get started with some background information.

Q1 Pauline
Could your start by giving us a timeline of your time at Swedish Radio?
PROBE How long have you been working at SR?
PROBE What is your role today?
PROBE How long have you been working on what you do today?

Q2 Linn
Tell us what you and your team are working on today
PROBE Challenges
PROBE Solutions to it / how do you work on solutions?

More general questions completed, now we want to talk a little more about SRs digital content.

Q4 Linn
Swedish Radio has a lot of content and everything is not possible to present at the same time. How do you reason when you present your content?
PROBE How to prioritize content - where it is placed and what content is placed there.
PROBE Visible/hiding

Q5 Pauline
How do you reason about guiding a user to find something to listen to?
PROBE Planned for how such a user interacts?
PROBE Categories, "you might also like this / selected for you", popular

Q6 Linn
Concluding question, what do your hopes about the future look like?
PROBE Dreaming, a little unrealistic

(Turn of recorder)
Is it okay to ask more questions later on, if any turns up?

Ending Pauline
Thanks again for taking the time to let us interview you, it is very helpful for our continued work. We will now process and analyze the answers we received from the interview. If you wish, we can of course send our essay to you when it is finished, is it something you are interested in? Our contact details are on the consent form. If there is anything you are wondering about, do not hesitate to contact us. Do you have any questions?

B

Interview consent form

The original consent form was in Swedish, and has been translated for this report.

CHALMERS

Consent of participation in interview

Chalmers Tekniska Högskola | Interaction design and technologies

Pauline Lorin & Linn Thorsager

We, Pauline and Linn, are two students at Chalmers University of Technology who are studying our last year in the master program Interaction design and technologies. Since the beginning of the year, we have worked with our masters project which we write in collaboration with Sveriges Radio.

Purpose
The interview is held as part of a thesis at Chalmers University of Technology. The degree project is carried out in collaboration with Swedish Radio and the purpose is to investigate how a user can find something to listen to in the Swedish Radio's range using voice control.

Participation
The interview is expected to take about 30 minutes.

Use of information
Only the responsible students, Pauline Lorin and Linn Thorsager, will have access to the recorded interview. The interview will only be presented in text format in the final essay. The recording will be deleted after completion of the project, however, no later than 2019-09-31.

In the final essay you will never be named by name but we will mention your professional role. As the paper will be published, we ask you to consider not sharing sensitive information.

Data and privacy management
Data from the interview will be used as part of the thesis work and as a basis for the development of the Swedish Radio's app for voice control. If we would like to use the information for any other purpose, we will contact you and ask for your consent again.

Voluntariness
Participation in the interview is completely voluntary. You can cancel the interview at any time or choose not to answer a question without justification.

Contakt
If you have any questions or want to get in touch with us or any of our supervisors:
Pauline Lorin [student], tel. 0707 48 47 38, email: pauline.lorin@hotmail.com
Linn Thorsager [student], tel. 0703 08 42 04, email: linn.thorsager@gmail.com
Linda Gomes [Advisor SR], tel. 0736 55 87 83, email: linda.gomes@sr.se
Palle Dahlstedt [Supervisor Chalmers], tel. 0703 08 82 07, email: palle@chalmers.se

CHALMERS

Consent

Is it okay to audio record the interview?   Ja □   Nej □

I hereby give my consent to participate in the interview:

_____   _____
Date and signature participant   **Name in block letters**

_____   _____
Date and signature student   **Name in block letters**

_____   _____
Date och signature student   **Name in block letters**

C

Interview complete result

The results from the interview questions presented with corresponding themes and description.

Question 1

Profession

The interviewees all mentioned their current roles and previous ones. Mentioned professions were UX, Journalism, Facilitator, Developer, Product owner, and Concept.

Team

The teams that the interviewees worked in were Innovation team, Play/X Voice, and App team.

Work structure

The work structure includes how the interviewees describes their daily work, including making different platforms coherent and making content available. "Nu är det tydligt röst-fokus".

Question 2

Goal

The interviewees mentioned some goals with their current work. The goals mentioned were to make the Google home application more usable (usability), to make the experience more pleasant (UX) especially regarding the voice used and to not forget the future wow-factor while focusing on what is feasible right now. More goals mentioned were to make users actually understand the system, to make it more simple to use, and for the system to be helpful so that it is not overwhelming to choose something to listen to.

Current

The field of voice user interfaces is mentioned to be currently very unexplored, and the platform is yet immature, making most persons unaware of it. "De vet att det handlar om att man kan sätta på musik, typ, och så kan man fråga lite frågor,

sådär" translation "They know that it is about that it is possible to turn on music, kind of, and that you can ask questions". This means that user testing in this area requires very specific functions and that users most often have not tried the technology before.

As of the immature platform, the focus lies in making whatever is feasible right now, and doing it as best as possible while also preparing for future improvements. Currently, the application is in the forefront technically, but lacks user adaptation. The first step they have taken towards better usability in their application was to make a persona for the assistant.

Future
The future is believed to be very uncertain at this point regarding VUIs. How will this service work in a couple of years? Where will it be used, where will it contribute? It is said to be all around us at a later point, but that is something which is not yet known.

Work structure
When working on this at SR they work in different teams, some of which are more technically focused and some of which are more user focused. They work on doing whatever is possible right now, and also look at what others are doing and have done, and try to make things testable. Overall there is also a work in looking into the different platforms at SR, and if they are supposed to be more coherent.

Uncertainty
In addition to the uncertainties mentioned for the future, the interviewees also feel uncertain about how the interaction with the user is supposed to work, how the user actually wants it to work.

Users
The current users of SRs application in Google home are very technically skilled and have patience, since it can all be very overwhelming at times.

Question 3
Current
Currently the field of VUI is very novel, meaning no standardization is yet implemented and the possibility to influence the future is bigger.

Users
VUI:s is a rather unknown and novel platform, causing users to generally have little

or no experience with the domain.

Challenge
Challenges with VUI:s are designing comprehension, giving explanations, naming content, creating guidning, UX, and that the platform is relatively new and therefore limiting in itself. The technology isn't there yet.

Question 4
Current
The current status of the platform is that it's designed for making all content available but without any second thoughts about the UX of the interaction.

Usage
The voice application is mostly used at home through Google Home, and not as much through the phone. And when it's used at home, mostly channels are listened to. It's mentioned that this could be the new "köksradio" used for "tillbakalutad lyssning".

Personalised
Personalisation for the voice application is described in conjunction with recommendations based on your profile while combining that with the diversity that SR represents. Usually it's enough to recommend content based on popularity but considering SR and public service this can't be the case. The content could be picked based on the specific user and the listening could be more in the format of a personal radio playlist.

Content solutions
Content solutions refers to future possible solutions for SR:s digital content. The content needs to be limited and some kind of prioritization needs to be established so the user doesn't get overwhelmed. Content that is labeled as more important is placed on the home screen in the mobile application, followed by the personalized content. The content could be divided by categories, making the navigation easier. On-boarding is mentioned as a possible solution for being able to use personalisation. Meaning that the user is prompted with questions about their interest the first time they enters the application as to get an better, more personalised, experience later on.

An important note is that even though users isn't familiar with SR:s content, they should still feel included and be able to use the application.

Content challenges
Challenges with the content is to make everything available without making it
overwhelming or tiresome while also not assuming that the users is familiar with the
content. And also by this not forcing the user to choose between a lot of different
categories that may not say anything to them. A critical challenge is to solve the
actual browsing through the content.

Requirements
Requirements that needs to be included accounting that it's SR:s application. News
is a big part of their agenda and considering that they are a public service company
they also need to work towards diversity in their content.

Challenges
Challenges include limiting the content and making it conversational. A challenge
mentioned is also how to present content.

Work structure
The work structure refers to the work habits at SR. There is many people involved
in different decisions, making it hard to please everyone. They try to work data
driven while also staying in the forefront. A dedicated personalization team is also
available so that not all teams needs to work on those matters.

Question 5
Guide
It is mentioned that there are different views on what guiding means. Explanations
should be used both during the guide, as well as for the content. For the mobile
application, on-boarding has been used for the home page for first time users, so
that the users can get started with suggestions at once. Another suggestion for how
to present the user the content is to make personalised playlists. The guiding,
though, can make use of content divided in disciplines and/or categories.

Content
Something which the interviewees has experiences is that the content is not typically
known by the user and an explanation of the content is needed. The content is
divided into categories, but also more traditionally as radio with their channels such
as P1, P2, P3, and their associative content. How the content should be divided is
expressed as a big challenge.

Characteristics
The characteristics of the system should according to the interviewees be helpful,

pleasing, efficient and initiative. It should give personalised suggestions and result in a good user experience. It should help to find something if that is needed, but also make it possible to effectively start playing something for those who know exactly what they want to listen to. At this point, it could be more important to make it easy for the user to start something, rather than that something being exactly what they wanted.

Users
It is common for the users to simply just listen to radio, and therefore only wanting to start playing one of the live broadcasted channels. As a novice user you do not want to make something wrong and be ridicule. It is instead important to make it easy to play something the first time, so that they can feel as if they succeeded, "jag lyckades, jag startade något" translation "I succeeded, I started playing something".

Challenges
Interacting with this kind of device is, as it is now, often unpleasant for the user. It is, and will be, a challenge to change that and instead create a good user experience. The users, especially the younger audience, are commonly not familiar with Swedish radios content, and more weight will then be put on the guiding through the content.

Question 6
Future
Refers to ideas and solutions that the interviewees wish to see in the future. The applications should be coherent, seamless, intuitive, and situational while also been able to use personalisation. One interviewee also mentioned that they wanted the voice application to pick up on tone and emotions in voice.

Challenges
Challenges for the future mostly evolve around making the actual voice better but also making the application's interaction work well independent of the content.

Domain
Domains that the voice based application can be seen, besides the home, in the future are cars, elevators, and as traditional radio.

Characteristics
Characteristics for the personal assistant in the future is helpful, trustworthy, pleasing, and appealing. A wish for the assistant to have its own personality but also more use as a tool than company.

Usage

Major usage for the voice application in the future is while the user is occupied. Also, be able to rewind, fast forward and play/pause. The application should also be able to use as a complete experience, "med rösten också skulle kunna styra det mer än att man bara ah men nu får du den här och när det är slut då får du börja från början".

D

Ideation session protocol

This protocol was originally written in Swedish, and have been translated for the sake of this report.

1. **Introduction: domain and setup** Pauline
   Hello and welcome to our ideation session! Thank you for taking the time to come here and participate. As you may know, our essay is about Voice user interfaces, VUI. We use Google Home as it is the only assistant in Swedish. What we want to investigate is how a user who does not know what they want to listen to can find something to listen to in a VUI. We have investigated how others have designed their voice interactions and we have become a bit locked in solutions that already exist. And this is where you can help us, because you are not so familiar with this but at the same time are experienced in design work!

   During this ideation session, we are going to start with two warm-up games, then move on to the actual ideation part, which will be concluded with a discussion. Then, you will be provided with fika!

2. **Warmup 1: Grandma, Tiger, Ninja** Pauline
   Ninja beats Tiger
   Tiger eats Grandma
   Grandma beats Ninja
   Best out of three, two versus two. The winners battle and the losers battle.

3. **Warmup 2: 1000 uses - Umbrella** Linn Groups of two, 4 minutes to write down as many uses as possible. Afterwards the uses are counted.

4. **Introduction: ideation session** Linn
   Now we are going to switch to the actual ideation part and today we are going to conduct some role playing. The idea is that you act out various interactions like Google Home and users. For each round, one person is Google Home and one person a user. As Google Home, you are given an attribute that you

should act from. Two people are waiting. Remember that it is very free, and to have fun! There are no faults or bad ideas.

Scenario: You are a user of the Swedish Radio's content, both radio and podcasts. When you use the app you usually scroll through the content for something that sounds interesting to listen to. You have just bought a Google home and have found out that Swedish Radio has an app that you can talk to. To talk to the app, you say "Hello Google, talk to Swedish Radio", whereupon the Swedish Radio application welcomes you. You should now use this application to find something to listen to.

As Google Home, you should act on the attribute provided to the extent possible, and interact with the user who needs help. You do not need to know the Swedish Radio's content, it is possible to make up radio channels and podcast names. As Google Home you do not have to act as you think it would, think new and think freely!

5. **Discussion** Linn (notes Pauline)
   Discussion subjects:
   How did it feel to act as Google Home?
   How did it feel to be the user?
   What was good during the interaction?
   What was bad during the interaction?
   How would you as a user want to be guided to get help finding something to listen to?
   Can you come up with some crazy ideas that can be fun to use?
   (Discussion about findings during the session)

E
Ideation session result

The analysis of the ideation session resulted in phrases which were to be used during the creation of the first set of guidelines. The phrases have been translated from Swedish to English for the sake of this report.

Explanation of what the episode is about
Perplexed when to answer, and not prepared for it
Perplexed by too much talking
Welcome to the program itself to then ask about which episode
User behavior is mirrored by Google homes behavior
Confirmation by repetition "I want to listen to something unexpected" "something unexpected, yes ..."
Play snippets of episodes as a sample
Tell what is being played before it is played
Efficiency can result in inefficiency
Can always cancel
The user must know that it is possible to cancel
Be able to choose from all programs
Long interaction good if relevant
Effectiveness good if right
The users' preferences for the interaction differ
The assistant should have the attribute helpful
Be able to ask what the programs are about
Start wide, then narrow down
Don't usually know what you are interested in
Provide options even if the user has no answer to what they want
A risk of being trapped if only options based on previous usage are presented.
Mix suggestions and other options
Suggestions based on time of day
Give something other than what can be achieved via mobile
Google home as a search function and personal advisor
Play info / snippets from episodes
Communicate clearly if it is something it cannot do anything

Choosing episodes can be presented with the most popular, the latest, the first, and search.

Enter how many episodes there are

Short error messages

When using randomisation, an explanation must be given.

Can give suggestions based on activity

Suggestions based on parable words

Be able to say an interest and then offer programs that are suitable for it.

Zapp function, x seconds of different episodes with stop function.

Not important that it is quite right, better to be proposing something

Not to be clear that everything is command-controlled.

More initiative

F

First user test protocol

The protocol was originally written in Swedish, but have been translated for this report. The interactions will be presented both in Swedish, the original language, and translated into English.

Intro:
Hello and welcome to our test! Thank you for taking the time to come here and participate. Our essay is about Voice user interfaces that we write in collaboration with Swedish Radio. We use Google Home because it is the only assistant in Swedish. In Google Home, Swedish Radio has an action that can be opened by saying "Talk to Sweden's Radio". An action can be seen as an application that can be accessed in Google Home. During the essay we have worked out a number of design guidelines that we now need to test.

So, the test is divided into 8 small parts, we will test different interactions, sometimes you will read from the script and sometimes you will improvise, but the focus is on the interaction. You will also have to answer some questions. Before each part I will of course explain to you what applies. You will have the role of a user of Swedish Radios action that wants to listen to something in from their content, but do not really know what. Behind the screen here is your Google Home that you will interact with during the test. Does it sound good? I am here and will help you throughout the test. I will also take notes during the test. Some questions? Yes, to get started, we have planned a small warm-up exercise!

Warm-up exercise
The warm-up exercise consists of two interactions. Follow the script, *spontaneous response* means that you should answer the question without a script.

Swedish
- – Hej Google!
- – Hejhej
- – Vet du vilken månad det är?
- – Det är april

– Okej, tack!

– Hej Google!
– Hej igen!
– Jag skulle vilja spela lite musik
– Vilken genre vill du lyssna på?
– *spontant svar*

English

– Hi Google!
– Hello!
– Do you know what month it is?
– It is April
– Okay, thank you!

– Hi Google!
– Hello again!
– I would like to play some music
– What genre would you like to listen to?
– *spontaneous response*

Good! Do you feel warmed up?
Have you used Google Home or any other smart speaker before?
Let's start the tests!

Test 1: inform about possible actions in the system.
You have just asked Google home to talk to Sweden's radio for the first time and have now entered their action. Answer the greeting.

Swedish

– Hej! Välkommen till Sveriges radio, vad kul att du har hittat hit! Vet du vad du vill lyssna på redan nu, eller vill du att jag ger dig några förslag?
– *spontant svar*

English

– Hey! Welcome to Swedish Radio, I am glad to see you here! Do you already know what you want to listen to, or do you want me to give you some suggestions?
– *spontaneous response*

Does it feel like you know what to do in the system, what?
Do you feel comfortable using this further?

XVI

Then we will try another alternative, it is still the first time you talk to Swedish Radio. Again, answer the greeting.

Swedish
  – Hej! Välkommen till Sveriges radio, vad kul att du har hittat hit! Här kan du spela allt vårt innehåll bara genom att be om det. Vårt innehåll består av radiokanaler, nyheter och program. Om du skulle vilja avsluta konversationen med mig, går det alltid bra att säga "avsluta". Så, vet du vad du vill lyssna på redan nu, eller vill du att jag ger dig några förslag?
  – *spontant svar*

English
  – Hey! Welcome to Swedish Radio, I am glad to see you here! Here you can play all our content just by asking for it. Our content consists of radio channels, news and programs. If you would like to end the conversation with me, it is always possible to say "quit". So, do you already know what you want to listen to, or do you want me to give you some suggestions?
  – *spontaneous response*

Did you gain another experience about what you can do in the system now? What? Do you (still) feel comfortable using it further?
Which of the alternatives would you rather use for the first time you talked to Swedish Radio?

Test 2: Present a minimum amount of options at a time
Now you are going to test the Swedish Radio's action again, remember that you are a user who wants help finding something to listen to. You will have to test three different variants. Read from the script in front of you. But the last answer from you should be a spontaneous answer.

Three at a time:
Swedish
  – Hej! Vad kul att du har kommit tillbaka! Vill du bli påmind om vad du kan göra här så går det alltid bra att säga "hjälp". Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
  – Jag vill att du ger mig förslag
  – För att kunna ge dig några förslag, vill jag att du väljer ut en kategori som du tycker låter intressant. Vill du lyssna på något inom dokumentär, sport, eller vill du höra fler kategorier?
  – Jag vill höra fler kategorier

– Inga problem, vill du lyssna på något inom samhälle, humor eller vill du höra
  fler kategorier?
– *Spontant svar*

English

– Hey! Welcome back! If you want to be reminded of what you can do here, it is
  always possible to say "help". Do you know what you want to listen to, or do
  you want me to give you some suggestions?
– I want you to give me suggestions
– In order to give you some suggestions, I want you to select a category that you
  think sounds interesting. Do you want to listen to something regarding
  documentary, sports, or do you want to hear more categories?
– I want to hear more categories
– No problem, do you want to listen to something within society, humor or do
  you want to hear more categories?
– *spontaneous response*

Spontaneously, how did this feel?
Why do you think you answered as you did?
Was it difficult to make a choice?
Now we are going to test another variant, you must not choose the same category
again

Six at a time:
Swedish

– Hej! Vad kul att du har kommit tillbaka! Vill du bli påmind om vad du kan
  göra här så går det alltid bra att säga "hjälp". Vet du vad du vill lyssna på,
  eller vill du att jag ger dig några förslag?
– Jag vill att du ger mig förslag
– För att kunna ge dig några förslag, vill jag att du väljer ut en kategori som du
  tycker låter intressant. Vill du lyssna på något inom dokumentär, sport,
  kultur, samhälle, humor eller vill du höra fler kategorier?
– Jag vill höra fler kategorier
– Inga problem, vill du lyssna på något inom drama, livsstil, livsåskådning,
  musik, nyheter eller vill du höra fler kategorier?
– *Spontant svar*

English

– Hey! Welcome back! If you want to be reminded of what you can do here, it is
  always possible to say "help". Do you know what you want to listen to, or do
  you want me to give you some suggestions?
– I want you to give me suggestions

XVIII

– In order to give you some suggestions, I want you to select a category that you think sounds interesting. Do you want to listen to something regarding documentary, sports, cultures, society, humour or do you want to hear more categories?
– I want to hear more categories
– No problem, do you want to listen to something within drama, lifestyle, philosophical, music, news or do you want to hear more categories?
– *spontaneous response*

How did it feel now?
Why do you think you answered as you did?
Was it difficult to make a choice?
Last variant, remember that you must not choose the same category again

Rattle:

Swedish

– Hej! Vad kul att du har kommit tillbaka! Vill du bli påmind om vad du kan göra här så går det alltid bra att säga "hjälp". Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
– Jag vill att du ger mig förslag
– För att kunna ge dig några förslag, vill jag att du väljer ut en kategori som du tycker låter intressant. Jag kommer nu snabbt att läsa upp alla kategorier. Försök komma ihåg den kategori du tycker låter intressant. Då sätter vi igång Barn 3-8 år, Barn 9-13 år, Dokumentär, Drama, Humor, Kultur/Nöje, Livsstil, Livsåskådning, Musik, News in other languages, Nyheter, Samhälle, Sport, Vetenskap/Miljö. Vilken kategori är du intresserad av?
– *Spontant svar*

English

– Hey! Welcome back! If you want to be reminded of what you can do here, it is always possible to say "help". Do you know what you want to listen to, or do you want me to give you some suggestions?
– I want you to give me suggestions
– IIn order to give you some suggestions, I want you to select a category that you think sounds interesting. I will now quickly read all categories. Try to remember the category you think sounds interesting. Lets get started Children 3-8 years, Children 9-13 years, Documentary, Drama, Humor, Culture / Pleasure, Lifestyle, Philosophical, Music, News in other languages, News, Society, Sport, Science / Environment. Which category are you interested in?
– *spontaneous response*

How did it feel now?

Why do you think you answered as you did?
Was it difficult to make a choice?
Which alternative did you like the most? Why?

Test 3: The user should understand where in the navigation they are
In this test you will have to test two variants of being offered suggestions to listen to.
You should only read from your script in this test.

Swedish
– Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några
  förslag?
– Förslag, tack
– Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
– Jag vill lyssna på ett program
– Självklart! För att kunna ge dig några förslag på program behöver jag ta reda
  på lite mer information om vad du är intresserad av. Vi kommer nu
  tillsammans att gå igenom tre steg, sen kan jag ge dig några väl valda förslag.
  Låter det bra?
– Ja
– Första steget är att hitta en kategori som intresserar dig. Vill du lyssna på
  något inom dokumentär, sport eller vill du höra fler kategorier?
– Jag vill lyssna på en dokumentär
– Absolut. Då går vi vidare till steg 2 som är att välja ett program. De tre mest
  populära programmen inom dokumentärer är P3 dokumentär, P3 dystopia och
  verkligheten i P3. Vill du lyssna på någon av dessa?
– Ja, jag vill lyssna på P3 dystopia
– Bra val! Då har vi bara tredje och sista steget kvar, att välja avsnitt. P3
  dystopia består av 11 avsnitt, vill du att jag spelar det första, senaste eller det
  mest lyssnade avsnittet?
– Jag vill höra det mest lyssnade
– Kul! Då spelar jag "När maten tar slut" av P3 dystopia...

– Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några
  förslag?
– Förslag, tack
– Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
– Jag vill lyssna på ett program
– Självklart! För att kunna ge dig några förslag på program behöver jag ta reda
  på lite mer information om vad du är intresserad av. Låter det bra?
– Ja

XX

– Vill du lyssna på något inom dokumentär, sport eller vill du höra fler kategorier?

– Jag vill lyssna på en dokumentär

– Absolut. De tre mest populära programmen inom dokumentärer är P3 dokumentär, P3 dystopia och verkligheten i P3. Vill du lyssna på någon av dessa?

– Jag vill lyssna på P3 Dystopia

– Bra val! P3 dystopia består av 11 avsnitt, vill du att jag spelar det första, senaste eller det mest lyssnade avsnittet?

– Jag vill höra det mest lyssnade

– Bra val! Då spelar jag "När maten tar slut" från P3 Dystopia

English

– Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?

– Suggestion, please

– Absolutely! Do you want to listen to a radio channel, news or maybe some program?

– I want to listen to a program

– Of course! In order to give you some suggestions for programs, I need to find out more information about what you are interested in. We will now go through three steps together, then I can give you some well-chosen suggestions. Does it sound good?

– Yes

– The first step is to find a category that interests you. Do you want to listen to something in documentary, sports or do you want to hear more categories?

– I want to listen to a documentary

– Absolutely. Then we proceed to step 2 which is to select a program. The three most popular programs in documentaries are P3 dokumentär, P3 dystopia and verkligheten i P3. Do you want to listen to any of these?

– Yes, I want to listen to P3 dystopia

– Good choice! Then we only have the third and last step left, to choose the episode. P3 dystopia consists of 11 episodes, do you want me to play the first, most recent or the most listened episode?

– I want to hear the most listened

– Fun! Then I play "När maten tar slut" by P3 dystopia ...

– Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?

– Suggestion, please

– Absolutely! Do you want to listen to a radio channel, news or maybe some program?
– I want to listen to a program
– Of course! In order to give you some suggestions for programs, I need to find out more information about what you are interested in. Does it sound good?
– Yes
– Do you want to listen to something in documentary, sports or do you want to hear more categories?
– I want to listen to a documentary
– Absolutely. The three most popular programs in documentaries are P3 dokumentär, P3 dystopia and verkligheten i P3. Do you want to listen to any of these?
– I want to listen to P3 Dystopia
– Good choice! P3 dystopia consists of 11 episodes, do you want me to play the first, most recent or the most listened episode?
– I want to hear the most listened
– Good choice! Then I play "När maten tar slut" from P3 Dystopia

What difference did you experience between the two variants?
Which one addressed you the most? Why?

What we are testing here is to notify the user where in the navigation you are. Our solution to this was to announce the number of steps and the step you are in. How did you experience it? Did you find that information helpful? Do you think it is better to know where in the navigation you are, even if our solution did not appeal to you?

Test 4: The user should be the final decision maker
The user should be able to do an informed choice This test will consist of three parts, with two variants per part. Read from the script for all variants. Chooses for you directly:

Swedish
– Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
– Jag vill att du ger mig förslag
– Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
– Jag vill lyssna på ett program
– Perfekt! Här kommer P3 dystopia avsnitt "När havet kommer"...

English

– Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
– I want you to give me suggestions
– Absolutely! Do you want to listen to a radio channel, news or maybe some program?
– I want to listen to a program
– Perfect! Here comes P3 dystopia episode "När havet kommer" ...

What did you think about this variant? Chooses for you with an explanation:

Swedish
– Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
– Jag vill att du ger mig förslag
– Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
– Jag vill lyssna på ett program
– Perfekt! Då spelar jag P3 dystopias senaste avsnitt "När havet kommer"...

English
– Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
– I want you to give me suggestions
– Absolutely! Do you want to listen to a radio channel, news or maybe some program?
– I want to listen to a program
– Perfect! Then I play P3 dystopia's latest episode "När havet kommer" ...

What did you think about this variant?
Did you perceive any differences between the two variants?

Chooses for you with after choice:
Swedish
– Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
– Jag vill att du ger mig förslag
– Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
– Jag vill lyssna på ett program
– Självklart! För att kunna ge dig några förslag på program behöver jag ta reda på vilken kategori du är intresserad av. Låter det bra?
– Ja
– Toppen! Vill du lyssna på något inom dokumentär, sport eller vill du höra fler kategorier?

   – Jag vill lyssna på en dokumentär
   – Okej, här kommer P3 dokumentär om "Galna ko-sjukan"

English
   – Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
   – I want you to give me suggestions
   – Absolutely! Do you want to listen to a radio channel, news or maybe some program?
   – I want to listen to a program
   – Of course! In order to give you some suggestions for programs, I need to find out which category you are interested in. Does it sound good?
   – Yes
   – Awesome! Do you want to listen to something in documentary, sports or do you want to hear more categories?
   – I want to listen to a documentary
   – Okay, here comes P3 dokumentär about "Galna ko-sjukan"

What did you think about this variant?

Chooses for you after choice with an explanation:
Swedish
   – Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
   – Jag vill att du ger mig förslag
   – Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
   – Jag vill lyssna på ett program
   – Självklart! För att kunna ge dig några förslag på program behöver jag ta reda på vilken kategori du är intresserad av. Låter det bra?
   – Ja
   – Toppen! Vill du lyssna på något inom dokumentär, sport eller vill du höra fler kategorier?
   – Jag vill lyssna på en dokumentär
   – Okej, här kommer P3 dokumentärs senaste avsnitt "Galna ko-sjukan"

English
   – Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
   – I want you to give me suggestions
   – Absolutely! Do you want to listen to a radio channel, news or maybe some program?
   – I want to listen to a program

– Of course! In order to give you some suggestions for programs, I need to find out which category you are interested in. Does it sound good?
– Yes
– Awesome! Do you want to listen to something in documentary, sports or do you want to hear more categories?
– I want to listen to a documentary
– Okay, here comes P3 dokumentärs latest episode "Galna ko-sjukan"

What did you think about this variant?
Did you experience any difference between the variants?
Do you find that you are more okay with the system choosing for you if it gives an explanation for it?

Let the user make all decisions:
Swedish

– Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
– Jag vill att du ger mig förslag
– Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
– Jag vill lyssna på ett program
– Självklart! För att kunna ge dig några förslag på program behöver jag ta reda på lite mer information om vad du är intresserad av. Vi kommer nu tillsammans att gå igenom tre steg, sen kan jag ge dig några väl valda förslag. Låter det bra?
– Ja
– Första steget är att hitta en kategori som intresserar dig. Vill du lyssna på något inom dokumentär, sport eller vill du höra fler kategorier?
– Jag vill lyssna på en dokumentär
– Absolut. Då går vi vidare till steg 2 som är att välja ett program. Vill du lyssna på P3 dokumentär, verkligheten i P3 eller höra fler alternativ?
– Jag vill lyssna på verkligheten i P3
– Bra val! Då har vi bara tredje och sista steget kvar, att välja avsnitt. Vill du att jag spelar det första, senaste eller det mest lyssnade avsnittet?
– Jag vill höra det mest lyssnade
– Kul! Då spelar jag "Tom brände ner sin lägenhet" av verkligheten i P3. . .

English

– Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
– I want you to give me suggestions

- Absolutely! Do you want to listen to a radio channel, news or maybe some program?
- I want to listen to a program
- Of course! In order to give you some suggestions for programs, I need to find out more information about what you are interested in. We will now go through three steps together, then I can give you some well-chosen suggestions. Does it sound good?
- Yes
- The first step is to find a category that interests you. Do you want to listen to something in documentary, sports or do you want to hear more categories?
- I want to listen to a documentary
- Absolutely. Then we proceed to step 2 which is to select a program. Do you want to listen to P3 dokumentär, verkligheten i P3 or hear more options?
- I want to listen to verkligheten i P3
- Good choice! Then we only have the third and last step left, to choose the episode. Do you want me to play the first, latest or most listened episode?
- I want to hear the most listened
- Fun! Then I play "Tom brände ner sin lägenhet" by verkligheten i P3...

What did you think about this variant?

Let the user make all decisions with explanation:
Swedish

- Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
- Jag vill att du ger mig förslag
- Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
- Jag vill lyssna på ett program
- Självklart! För att kunna ge dig några förslag på program behöver jag ta reda på lite mer information om vad du är intresserad av. Vi kommer nu tillsammans att gå igenom tre steg, sen kan jag ge dig några väl valda förslag. Låter det bra?
- Ja
- Första steget är att hitta en kategori som intresserar dig. Vill du lyssna på något inom dokumentär, sport eller vill du höra fler kategorier?
- Jag vill lyssna på en dokumentär
- Absolut. Då går vi vidare till steg 2 som är att välja ett program. De två mest populära programmen inom dokumentärer är P3 dokumentär och verkligheten i P3. Vill du lyssna på någon av dessa eller vill du höra fler alternativ?
- Jag vill lyssna på verkligheten i P3

– Bra val! Då har vi bara tredje och sista steget kvar, att välja avsnitt.
  Verkligheten i P3 består av 500 avsnitt, vill du att jag spelar det första,
  senaste eller det mest lyssnade avsnittet?
– Jag vill höra det mest lyssnade
– Kul! Då spelar jag "Tom brände ner sin lägenhet" av verkligheten i P3...

English

– Hello again! Do you know what you want to listen to, or do you want me to
  give you some suggestions?
– I want you to give me suggestions
– Absolutely! Do you want to listen to a radio channel, news or maybe some
  program?
– I want to listen to a program
– Of course! In order to give you some suggestions for programs, I need to find
  out more information about what you are interested in. We will now go
  through three steps together, then I can give you some well-chosen suggestions.
  Does it sound good?
– Yes
– The first step is to find a category that interests you. Do you want to listen to
  something in documentary, sports or do you want to hear more categories?
– I want to listen to a documentary
– Absolutely. Then we proceed to step 2 which is to select a program. The two
  most populär programs within documentary is P3 dokumentär and
  verkligheten i P3. Do you want to listen to any of those, or do you want to
  hear more options?
– I want to listen to verkligheten i P3
– Good choice! Then we only have the third and last step left, to choose the
  episode. Verkligheten i P3 consists of 500 episodes, do you want me to play
  the first, latest or most listened episode?
– I want to hear the most listened
– Fun! Then I play "Tom brände ner sin lägenhet" by verkligheten i P3...

What did you think about this variant?
Did you experience any difference between the variants? What did you think of that
difference?
If you compare all six variants, which one of these would you prefer to continue
using? Why?
How did you like when the system was more initiative? Do you prefer to make the
decision yourself or not?
In which step (if at any time) do you want the system to make the choice for one?
Would you have been willing to go through one of the longer versions to get

suggestions?

Test 5: Universal navigational functions should be provided
This test consists of two variants. Follow the script's instructions.

Swedish
– Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
– Jag vill att du ger mig förslag
– Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
–
– Jag vill lyssna på ett program
– Självklart! För att kunna ge dig några förslag på program behöver jag ta reda på lite mer information om vad du är intresserad av. Vi kommer nu tillsammans att gå igenom tre steg, sen kan jag ge dig några väl valda förslag. Låter det bra?
– *Du har ångrat dig och vill ta dig tillbaka till start, uttryck det*

– Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
– Jag vill att du ger mig förslag
– Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
– Jag vill lyssna på ett program
– Självklart! För att kunna ge dig några förslag på program behöver jag ta reda på lite mer information om vad du är intresserad av. Vi kommer nu tillsammans att gå igenom tre steg, sen kan jag ge dig några väl valda förslag. Låter det bra?
– Ja
– Första steget är att hitta en kategori som intresserar dig. Vill du lyssna på något inom dokumentär, sport eller vill du höra fler kategorier?
– Jag vill höra fler kategorier
– Inga problem. Vill du lyssna på något inom kultur, drama eller vill du höra fler kategorier?
– *Du vill höra de föregående kategoriförslagen igen, uttryck det*

English
– Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
– I want you to give me suggestions
– Absolutely! Do you want to listen to a radio channel, news or maybe some program?

- I want to listen to a program
- Of course! In order to give you some suggestions for programs, I need to find out more information about what you are interested in. We will now go through three steps together, then I can give you some well-chosen suggestions. Does it sound good?
- *You have changed your mind and want to get you back to start, express it*

- Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
- I want you to give me suggestions
- Absolutely! Do you want to listen to a radio channel, news or maybe some program?
- I want to listen to a program
- Of course! In order to give you some suggestions for programs, I need to find out more information about what you are interested in. We will now go through three steps together, then I can give you some well-chosen suggestions. Does it sound good?
- Yes
- The first step is to find a category that interests you. Do you want to listen to something in documentary, sports or do you want to hear more categories?
- I want to hear more categories
- No problem. Do you want to listen to something in culture, drama or do you want to hear more categories?
- *You want to hear the previous category suggestions again, express it*

What do you think of being able to navigate i the system like this?

Test 6: The system should be able to handle "I don't know"
Next test, here we have three different variants, follow the script's instructions.

Swedish
- Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
- Jag vill att du ger mig förslag
- Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
- Jag vet inte
- Okej, välj mellan radiokanal, nyheter eller program.
- *Spontant svar*

English
- Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?

- I want you to give me suggestions
- Absolutely! Do you want to listen to a radio channel, news or maybe some program?
- I do not know
- Okay, choose between radio channel, news or program.
- *Spontaneous response*

What did you think about this variant?

Swedish

- Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
- Jag vill att du ger mig förslag
- Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
- Jag vet inte
- Okej, vill du istället att jag ska välja något från topp 10 mest lyssnade?
- *Spontant svar*

English

- Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
- I want you to give me suggestions
- Absolutely! Do you want to listen to a radio channel, news or maybe some program?
- I do not know
- Okay, do you want me to choose something from the top 10 most listened?
- *Spontaneous response*

What did you think about this variant?

Swedish

- Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
- Jag vill att du ger mig förslag
- Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
- Jag vet inte
- Inga problem, då väljer jag åt dig. Jag tror att vår radiokanal P3 skulle passa dig, som just nu sänder talkshowen PP3.

English

- Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?

XXX

– I want you to give me suggestions
– Absolutely! Do you want to listen to a radio channel, news or maybe some program?
– I do not know
– No problem, then I choose for you. I think our radio channel P3 would suit you, who is currently sending the talk show PP3.

What did you think about this variant?
Which one of the variants would you rather have kept using? Why?

Test 7: Samples for content should be provided
Explanations for content should be provided
Next last test now! You know this now, one variant, follow the script.

Swedish
– Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
– Jag vill att du ger mig förslag
– Absolut! Vill du lyssna på en radiokanal, nyheter eller kanske något program?
– Jag vill lyssna på ett program
– Självklart! För att kunna ge dig några förslag på program behöver jag ta reda på lite mer information om vad du är intresserad av. Vi kommer nu tillsammans att gå igenom tre steg, sen kan jag ge dig några väl valda förslag. Låter det bra?
– Ja
– Första steget är att hitta en kategori som intresserar dig. Vill du lyssna på något inom dokumentär, sport eller vill du höra fler kategorier?
– Jag vill lyssna på en dokumentär
– Absolut. Då går vi vidare till steg 2 som är att välja ett program. De tre mest populära programmen inom dokumentärer är P3 dokumentär, P3 dystopia och verkligheten i P3. Vill du lyssna på någon av dessa?
– Vad handlar P3 dokumentär om?
– P3 dokumentär är programmet som blickar bakåt, gräver i arkiven och pratar med dem som var med när det hände. Låter det bra?
– Jaha, då kan jag lyssna på det
– Bra val! Då har vi bara tredje och sista steget kvar, att välja avsnitt. P3 dokumentär består av 319 avsnitt, vill du att jag spelar det första, senaste eller det mest populära avsnittet?
– Vad handlar det senaste avsnittet om?
– Jag spelar upp en trailer till senaste avsnittet av P3 dokumentär "galna ko-sjukan". . . *spelar sample* Låter det bra?

– Ja, det låter kul, sätt igång det
– Okej, då spelar jag "galna ko-sjukan" från P3 dokumentär...

English

– Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
– I want you to give me suggestions
– Absolutely! Do you want to listen to a radio channel, news or maybe some program?
– I want to listen to a program
– Of course! In order to give you some suggestions for programs, I need to find out more information about what you are interested in. We will now go through three steps together, then I can give you some well-chosen suggestions. Does it sound good?
– Yes
– The first step is to find a category that interests you. Do you want to listen to something in documentary, sports or do you want to hear more categories?
– I want to listen to a documentary
– Absolutely. Then we proceed to step 2 which is to select a program. The three most popular programs in documentaries are P3 dokumentär, P3 dystopia and verkligheten i P3. Do you want to listen to any of these?
– What is P3 dokumentär about?
– P3 dokumentär is the program that looks backwards, digs into the archives and talks to those who was there when it happened. Does it sound good?
– Well, then I can listen to it
– Good choice! Then we only have the third and last step left, to choose the episode. P3 dokumentär consists of 319 episodes, do you want me to play the first, latest or most popular episode?
– What is the latest episode about?
– I play a trailer to the latest episode of P3 dokumentär "galna ko-sjukan"... * playing sample * Does it sound good?
– Yes, it sounds fun, turn it on
– Okay, then I play the "galna ko-sjukan" from P3 dokumentär ...

What do you think about that you can get explanations for programs?
Do you think this is something you would ask for?
What do you think about being able to listen to trailers on episodes?
How did it feel that both explanations and trailers were used?
If you had to choose, explanation or trailer?

Test 8: Design for both novice and experienced users

Now we are on the last test and we think you have become an experienced user, congratulations! Follow the script, one variant.

Swedish
- Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?
- Jag vill att du ger mig förslag på ett program
- Självklart! Vill du lyssna på något inom dokumentär, sport eller vill du höra fler kategorier?
- Jag vill lyssna på en dokumentär
- Absolut. De tre mest populära programmen inom dokumentärer är P3 dokumentär, P3 dystopia och verkligheten i P3. Vill du lyssna på någon av dessa?
- Ja, jag vill lyssna på första avsnittet av P3 dystopia
- Kul! Då spelar jag "Du kan inte gömma dig" av P3 dystopia. . .

English
- Hello again! Do you know what you want to listen to, or do you want me to give you some suggestions?
- I want you to give me suggestions for a program
- Of course! Do you want to listen to something in documentary, sports or do you want to hear more categories?
- I want to listen to a documentary
- Absolutely. The three most popular programs in documentaries are P3 dokumentär, P3 dystopia and tverkligheten i P3. Do you want to listen to any of these?
- Yes, I want to listen to the first episode of P3 dystopia
- Fun! Then I play "Du kan inte gömma dig" by P3 dystopia ...

How did this feel? Did you think it was good or bad that the system got more straightforward?
Did you as experienced user estimate that the interaction became more concise after a while?

Did you notice that the greeting phrase has changed during the test? (if not, explain the changes)
What do you think of the different greetings?
Would you like to have suggestions from Google Home? Is it something you would make use of?
Do you have any final thoughts or more input that you have thought of during the test

Ending:

Thank you so much for taking the time to test our guidelines! Now you will get rewarde with some fika!

G

First user test pilot test changes

The following was changed during the pilot test of the first user test:

- Say the number of variants each test has instead of the number of interactions.
- Change phrasing on repetitive questions.
- During test 1, emphasize that it is the first time the participant interacts with Swedish Radio on the question being asked.
- During test 2, reminder that the participant is a user who wants help finding something to listen to, and say that they should not say the same category twice.
- During test 5, if the participant expresses it in the same way as the script says, ask for another way of expressing it.
- In test 6, say "something from the top 10 most listened" instead of "the top 10 programs".
- In test 8, it did not work to have open dialogue with Google, change to script instead.

H

First user test complete results

The result was originally written in Swedish, below a translation into English is presented.

Age range: 22-30
Mean: 25
(Median: 23)

Smart speaker usage:
Never used any smart speaker: 2
Have one themselves: 1
Have tried it: 2

Test 1
All participants answered a "correct answer" to the first and most simple greeting. All participants except one answered correctly to the second more elaborated greeting. After the first greeting, all participants said that they got a general understanding of what they can do in the system and listed the two choices that were given (say what to listen to or get suggestions). All said to be comfortable in keep using the system after the first greeting. Four out of five participants felt that the second greeting was too long and gave too much information.

Conclusion: The short one gives too little information while the second one is too daunting to listen to. It needs to be informative but still short and concise.

Inform about possible actions in the system
The test shows that the users likes being informed, but dislikes to much talk more. We therefore make the conclusion that it does not need to be stated explicitly at all times for everyone, but rather state possibility of having possible actions explained, and/or once in a while remind users about it.
Guideline stays

Test 2

3 alternatives: all answers are "correct" and 4/5 answered a category, the 5th person answers "more categories".

Conclusion: two categories at a time is easy to understand and to answer to, but it is expressed that it is Google that has the control, and that it feels like maybe to few as it may be boring to keep on answering "more categories" many times, and that you as a user have no control over which alternatives really exists.

6 alternatives: all answers are "correct" and 3/5 answered a category, the other two answered "more categories" and "can I hear it again?".

Conclusion: A little hard to make a choice when this many categories are presented, but at the same time it is expressed that it gave a better understanding of the available categories and made it easier to do an informed choice.

Rattle: Due to the wrong script being handed out to one participant, that person couldn't answer Googles question. For the rest of the participants, all answered correctly and 3/5 answered a category and one person said "i don't know".

Conclusion: To rattle all possible categories needs the users full attention, which may be hard to ask for. Not any of the participants were totally against the idea of this solution and expressed that it is nice to know all alternatives, but at the same time they felt stressed and overwhelmed.

When asked to choose between the three alternatives, three participants chose 6 alternatives and two participants chose 3 alternatives. Few options makes it easier to choose from but it doesn't give the user a good insight in the available options and would quickly get tedious asking for more categories. Six alternatives requires more attention from the user but lets them make an informed choice and is more suitable for multiple usage.

Present a minimum amount of options at a time
During the test it became clear that it's a question about not presenting too much as to not overwhelm but at the same time enough options so that the user feels like they can make an informed choice.

One problem that became clear regarding this is the categories used as they are very broad and hard to grasp. With this categories, the participants felt like they needed to hear all categories as to understand the division of content. Our interpretation is that if they were to be presented with a more obvious division, like "murders", the

user should more easily be able to make the decision if that is something they wants to listen to or not.

Re-phrase the guideline and include the balance between overwhelmed/informed choice

Test 3

3/5 did not like the implementation of the steps to present where in the navigation they are. The other two said they liked the steps and that it felt nice to know, so that you know how long the interaction is. The ones that did not like the steps, explicitly stated so repeatedly during the test, and that it was annoyingly much talk about it. They also stated that they do not need to know where in the navigation they are, and rather prefer a more concise interaction.

The user should understand where in the navigation they are

Even though some participants appreciated the navigational steps provided in the test, it most likely will be tiresome and unnecessary in the long run. It was also expressed that it should not feel as a navigation, but rather a natural conversation, which also is the overarching goal with VUIs. All participants most of all appreciated the more short and concise conversation in test 8, where no steps or similar approach were given. Therefore:

Delete this guideline or rephrase into a guideline stating that it should not feel as a navigation

Test 4

Making a choice for the user directly: Three participants expressed that they did not like this interaction, it was too fast and too much room for error. They didn't feel like they had a part in the choice-making. The two remaining participants expressed that they liked this function, especially that it was a very fast interaction to play something. But on the other hand, both these participants and the ones disliking this idea of an extremely initiative assistant mentioned that this function could be useful, sometimes you just want someone to choose for you, but not as a standard solution for suggestions.

Making a choice for the user after choosing category: The participants expressed similar thoughts as when the system chooses directly for them, but are a bit more positive against initiatives. It is mentioned that this kind of behaviour would be much more appreciated if it was based on personalisation. They also mentions that they still wants to be the decision maker, and that it should not play automatically.

The user makes all the choices: No one were completely against this longer

interaction, but experienced that it was too much irrelevant information in it. When the system stated how many episodes a certain program contained, participants was puzzled as to why because they didn't feel like they could choose between all episodes, just first, latest or most listened to.

Willingness to go through all steps in order to get a good suggestions: All participants expressed a willingness to go through the whole interaction to get a good suggestions, but with some modifications. Not so much talk as it is now, and there should be another options as well to get a faster suggestion, if you are not completely sure what it is you want and like.

Conclusion: It should be possible to both go through all steps to find something to listen to and for the system to make choices for the user, but in that case always ask for confirmation first so that the user makes the final decision. To go through all steps, the user needs to have much preferences as well as have a clue about what they wants, otherwise the system should be able to provide suggestions earlier on in the conversation. All participants appreciated when the system stated which episode it started playing (the latest) instead of just presenting its title.

The user should be the final decision maker
The test supported this guideline. All participants stated during several parts of the tests that they never wanted the system to play something without asking first. They also wanted the default to be that they as a user should make all of the decisions (based on systems decisions) and if they feel as they need more help, they should be able to ask for shuffle, but it needs to be their decision. And before the system plays it, it needs to present the shuffled outcome and ask if it should be played.
Guideline stays

The user should be able to do an informed choice
The test supported this guideline, but not the part about presenting amount of episodes for a program. All participants liked to know which episode that would be played, and suggestions for which episode that could be played (first, latest, most listened)
Guideline stays with some modification of the implementation

Test 5
Navigate to previous categories: "Go back one step", "could you repeat the first categories" x2, "I would like to hear the first categories again".

Go back to the start: "Go back to the start"/"start over" x5, "cancel" x2

Conclusion: All participants expressed that they would like to be able to navigate in the system, especially to be able to go back and to start over.

Universal navigational functions should be provided
The test resulted in that it may be possible to find and implement universal navigational functions in a VUI because of the participants using similar expressions for the same navigational functions. All participants expressed it as it is obvious that it should be possible to speak with the system in those terms which implies that the functions doesn't need to be explicitly stated for the users.
Guideline stays but re-phrased "The system should support (universal) navigational functions"

Test 6
Repeats options: This made the participants feel stupid and irritated, they did not feel as the system answered their questions but rather was a bit cocky.

Play something from top 10: All participants appreciated that the system could handle their answer and give them a counteroffer, and appreciated that the system asked before making the decision. One participants said that they wanted to know what was in top 10.

Chooses something for the user: Participants liked the idea of being given a suggestion if they don't know themselves but everyone added that they want to say if the suggestion is acceptable or not before the system starts to play it.

Conclusion: When asked to choose between the different solutions, all participants preferred the option with playing something from top 10 and that they would like to be the one deciding before starting to play.

The system should be able to handle "I don't know"
All participants expressed it as a matter of course that the system in some way should be able to handle the answer "I don't know". As it is about a user that really does not know, this answer may be common and it should therefore be designed for, and not treat is as an indecisive user and repeat the options to give it another shot at making the decision.
Guideline stays with added description of implementation

Test 7

Explanations: All participants appreciated to be given explanations of the content when asking for it. They also believed that it is something which is natural to ask for and does not need to be stated. Two mentions that to make sure that the feature is possible it may be good to remind the user once in a while.

Samples: The main opinion about playing a sample is that it is good for understanding the mood of the episode, which can be hard with just explanations, but that it can be tiring to listen to and hard to make so that they are relevant and still short.

Conclusion: Explanations are something which everyone is positive to, and are also harder to get wrong. Samples are much appreciated if done right, but if done wrong they are tiresome and 3/5 were against using it and instead simply stick with explanations. 2/5 stated that they preferred samples over explanations, but still liked the idea of explanations.

Samples for content should be provided
Explanations for content should be provided
Everyone that liked samples also liked explanations, but not the other way around. Therefore, sample may be seen as unnecessary to implement in a system when explanations would be sufficient. All participants expressed it as explanations were a contributing function in the system and something which they intuitively would ask for.
Guideline for explanations stays
Delete guideline for samples

Test 8
Experienced interaction: All participants appreciated the interaction for experienced users and that it was intuitive for them to speak with the system in that way after having used it for several interactions during the test. They liked that it became more to straight to the point and more efficient. Some mentioned that this interaction requires some experience because of the lack of guiding, but that it wasn't an issue, rather the opposite, if you are used to the system.

Design for both novice and experienced users
During the test, participants expressed that they want more information during their first interaction and also more guiding in the beginning of their usage. Later, all participants expressed that they appreciated the possibility of faster interactions and a change of greeting and interaction techniques within the system, when becoming more experienced with the system.

Guideline stays but needs more elaborate explanation of implementation

Overall

Different greetings: The long greeting were expressed as only acceptable the first time using the system, as it was a bit too long. But at the same time they appreciated that some functionality was stated in the beginning, but only in the beginning. They liked that the greetings became shorter during usage. The second greeting were criticised because of the ambiguity of the word "help", what does it mean in this context? The short greeting were the most appreciated greeting.

Suggestions as a function: All participants appreciated that they could get suggestions and that it felt like a intuitive function that they would most likely use. Either you know what you want to listen to, or you would like suggestions.

Other thoughts: One participant expressed that they wanted to be able to shuffle/"slumpa" in every step of the interaction. More everyday language, and that it doesn't need to be so polite, while another participant appreciated the politeness. To make the system more efficient and concise was also mentioned by several participants. Otherwise the participants was happy with the previous interactions and that using Google Home for suggestions was desirable.

Other comments during test:
- Liking of the explanation of content in greeting
- A suggestion to ask for possible actions and possible moves etc. rather than saying them in the greeting.
- It is easy for the user to feel as they have missed something when it is a too long message. Rather converse than monologue.
- "let's do this together" conversation is nice/and not nice according to another (as therapy session)
- Positive feedback can feel unnecessary at points, could be used for more useful information.
- Do not want to know that it is a navigation, but rather that it should be a simulated conversation.
- After having done a choice of program, it should be short until it is actually played something aka not too much talk around choosing episode.
- Wants it to be fast but still be the decision maker.
- Base suggestions on an account or previous knowledge of user, then initiative is more accepted x3
- Most listened to is better than most popular.
- Liked the choice of first, latest or most listen to when presenting episodes.

- To ask for confirmation "yes/no" in the interaction is unnecessary.
- Divide content into subjects based on episodes rather than programmes, since that is more relevant.
- Politeness can be tedious in the long run
- Assumes that when few options is presented those are the most popular ones
- When categories are rattled it also results in no categories being highlighted
- Good with information in the greeting about help-function
- The participant wishes to be able to state a category instead of the assistant handing him/her them
- (classic) Categories are better suitable for GUI
- When navigating in the system the user should be able to reference to their previous utterances and from that navigate
- If the system needs to repeat something for the user, it should repeat with more information or rephrase it in some way
- Offering to play something from top 10 is nice, but maybe that top 10 is the same kind of content for a long time
- The user will start to distrust when the system is more initiative taking if the suggestions provided is bad
- When acting as an experienced user, a participant expressed that it felt like he/she had adjusted to the system and not the other way around
- Same "level" of greeting should have multiple variations

I

Second user test protocol

The protocol was originally in Swedish but has been translated for the sake of the report. The interactions are however presented in both Swedish, the original language, and English.

Introduction

Thank you for taking time to participate in this test. My name is Linn and I'm doing this test together with my thesis partner Pauline who is sitting over there and will take notes during the test. Me and Pauline are students at the programme Interaction design and technologies at Chalmers in Gothenburg and are currently writing our master thesis in collaboration with Swedish Radio.

We are investigating how Swedish Radio's voice-based application for Google Home can be improved. - Do you know what Google Home/smart speakers are? - Have you used Google Home/a smart speaker before? Swedish Radio has an action in Google Home which can be reached by saying "Talk with Swedish Radio". An action is like an app but for Google Home.

In more detail, we are going to investigate how Google Home can help with finding something to listen to in Swedish Radio's content, and the purpose of this test is for us to understand your experience doing this. So there is no right or wrong, we just want to know what you think.

So, the test consists of three parts and will take approximately 30 minutes. In the first part you will interact with Google Home, and I will give you scripts to read from while doing this. Afterwards, you will fill out a questionnaire about how you experienced the interactions. When you have filled out the questionnaire, you and I will together walk through the questionnaire with your answers, just so we get an explanation to why you answered as you did. But as mentioned earlier, there is no wrong answers and you definitely does not need to feel that you can't answer negatively, its very valuable for us to get both positive and negative feedback.

And of course, everything that you say in here will be handled anonymously. Do you

have any questions before we start? Yes, we have prepared a warm-up task so you get to try to talk to Google before. So you just follow the script, and start whenever you feel ready.

Warm-up task
Swedish
- Hej Google!
- Hejhej
- Vem är du?
- Jag är din assistent. Vill du ha hjälp med något?
- Nej, det är bra tack
- Det var så lite så

English
- Hi Google!
- Hey there
- Who are you?
- I'm your assistant. Do you want help with anything?
- No, it's fine thank you
- No problem

Good! The focus during the test will be on the interaction, and you will play a user who does not know what you want to listen to.

Interaction - script 1
Swedish
- Hej Google, prata med Sveriges Radio
- Javisst, jag hämtar Sveriges Radio *blip*
- Hej! Välkommen till Sveriges radio! Här kan du spela allt vårt innehåll, både radiokanaler, nyheter och program. Om du vill veta mer om hur det här funkar, kan du alltid säga "hjälp". Vet du vad du vill lyssna på, eller vill du att jag ger dig förslag?
- Jag vill att du ger mig förslag
- Absolut! Vill du lyssna på en radiokanal, nyheter eller ett program?
- Jag vill lyssna på ett program
- Självklart! För att kunna ge dig förslag på program behöver jag veta lite mer.
- Vill du lyssna på något som handlar om true crime, odling, relationer, eller vill du höra fler alternativ?
- Jag vill lyssna på något om true crime
- Visst! De tre mest lyssnade avsnitten som handlar om true crime är Mordet på

Nils Horner, Avrättad av IS och Örebromannen. Vill du lyssna på någon av dessa?

- Vad handlar mordet på Nils om?

- Mordet på Nils Horner är ett avsnitt av P3 dokumentär, som handlar om utrikeskorrespondenten Nils, som blev skjuten i bakhuvudet i Afghanistan. Vill du lyssna på det?

- Aa

- Okej, då spelar jag det

English

- Hi Google, talk with Swedish Radio

- Sure, I'll get Swedish Radio *blip*

- Hi! Welcome to Swedish Radio! Here you can play all of our content, meaning radio channels, news, and programs. If you would like to know more about how this works, just say "help". Do you know what you wanna listen to, or do you want me to give you suggestions?

- I want you to give me suggestions

- Sure! Do you wanna listen to a radio channel, news, or a program?

- I wanna listen to a program

- Of course! To be able to give you suggestions for programs I need to know a bit more. Do you wanna listen to something about true crime, gardening, relationships, or do you wanna hear more suggestions?

- I wanna listen to something about true crime

- Sure! The three most listened episodes about true crime is Mordet på Nils Horner, Avrättad av IS and Örebromannen. Do you wanna listen to one of these?

- What is the murder of Nils about?

- Mordet på Nils Horner is an episode by P3 dokumentär, it's about the foreign correspondent Nils, who was shot in the back of the head in Afghanistan. Do you wanna listen to that?

- Mm

- Okay, I'll play that then

Interaction - script 2

Swedish

- Hej Google, prata med Sveriges Radio

- Javisst, jag hämtar Sveriges Radio *blip*

- Hej igen! Vet du vad du vill lyssna på, eller vill du att jag ger dig några förslag?

- Jag vill att du ger mig förslag

- Jajjamän! Vill du lyssna på en radiokanal, nyheter eller ett program?

- Ett program

- Okej! För att kunna ge dig förslag på program behöver jag veta lite mer. Vill du lyssna på något som handlar om true crime, odling, relationer, eller vill du höra fler

alternativ?

- Kan inte du välja ett program åt mig?

- Javisst, då kör vi på Hungriga hjärtan. Vill du att jag spelar det första, senaste eller det mest lyssnade avsnittet?

- Jag vill höra det senaste avsnittet

- Okej, då spelar jag "Pamela Andersson - Dödligt kär i min PT" av Hungriga hjärtan

English

- Hi Google, talk with Swedish Radio

- Sure, I'll get Swedish Radio *blip*

- Hi again! Do you know what you wanna listen to, or do you want me to give you suggestions?

- I want you to give me suggestions

- Yesbox! Do you wanna listen to a radio channel, news, or a program?

- A program

- Okay! To be able to give you suggestions for programs I need to know a bit more. Do you wanna listen to something about true crime, gardening, relationships, or do you wanna hear more suggestions?

- Can't you choose a program for me?

- Sure, then we'll take Hungriga hjärtan. Do you want me to play the first, latest or the most listened episode?

- I wanna hear the latest episode

- Okay, then I'll play "Pamela Andersson - Dödligt kär i min PT" by Hungriga hjärtan

Questionnaire

So here you get a questionnaire to fill out about your experience of the interactions. Take your time and just ask if you wonder about something! Regardless, we will look through the questionnaire together afterwards and discuss your answers a bit.

Interview

Focus on:

Overall impression

Extreme values from the questionnaire

Do you have any additional comments?

Would you please describe your experience of the interaction?

What do you think made you feel like that?

Anything that you want to change or improve?

I see that you have answered...what made you do that?

What would you change to go from negative -> positive?

The interaction with the system is repetitive

What did you feel was repetitive? Good/bad?

Ending

That was it! Thanks again for taking your time to participate in our test, it is of great value for our project. If there is anything that you wonder about don't hesitate to contact us, here are our contact information. Do you have any questions now? Okay, bye and thanks again!

L

J

Second user test questionnaire

The questionnaire used in the second user test in its original language Swedish and a translated version in English.

Utvärdering av Google Home

Nedan följer ett antal påståenden om de interaktioner du nyss haft. Fyll i det alternativ för varje påstående som du tycker stämmer bäst överens med din upplevelse.

1. Jag uppskattade att använda systemet

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

2. Jag var inte alltid säker på vad systemet gjorde

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

3. Systemet är användbart

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

4. Systemet gjorde alltid det jag förväntade mig

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

5. Interaktionen med systemet är frustrerande

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

6. Systemet är pålitligt

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

7. Jag måste koncentrera mig mycket när jag använder systemet

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
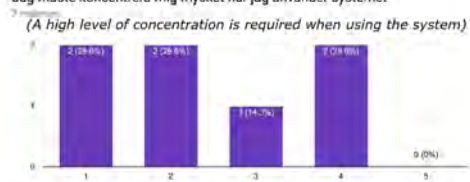
8. Interaktionen med systemet är konsekvent

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
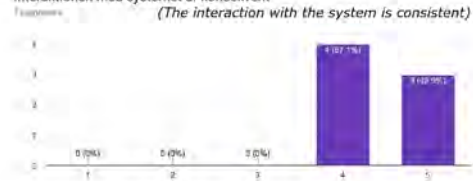
9. Interaktionen med systemet är upprepande

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
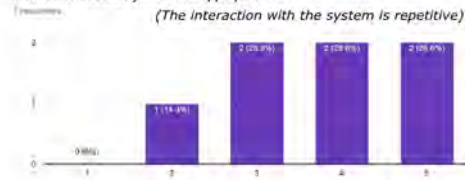
10. Interaktionen med systemet är effektiv

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
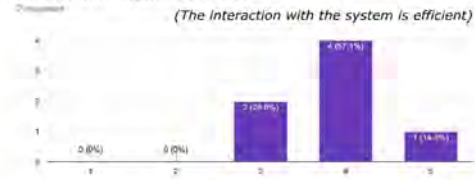
11. Interaktionen med systemet är irriterande

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
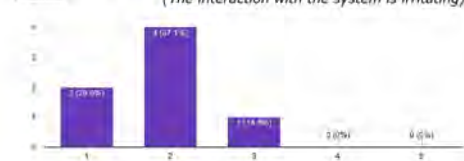
12. Interaktionen med systemet är tråkig

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
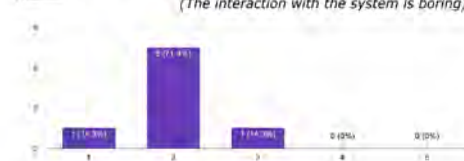
13. Systemet är behagligt

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
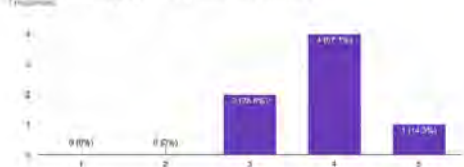
14. Jag kände mig spänd när jag använde systemet

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
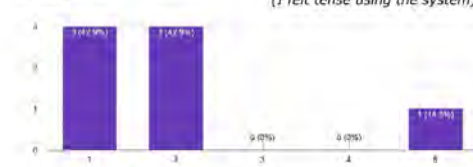
15. Systemet är enkelt att använda

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
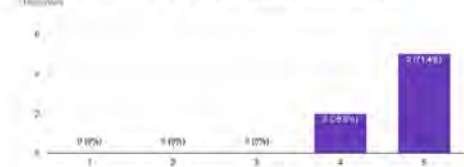
16. Jag behöver veta saker i förväg för att kunna använda systemet

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
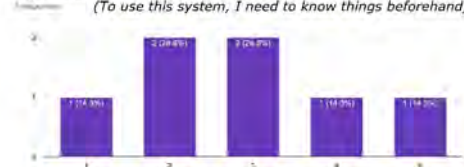
17. Jag kände mig trygg när jag använde systemet

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
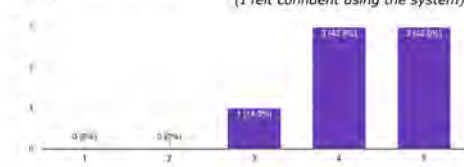
18. Jag kände mig överväldigad när jag använde systemet

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med
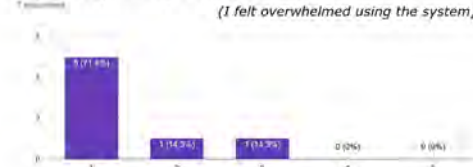
19. Jag hade velat använda mig av det här systemet

	1	2	3	4	5	
Håller inte alls med	◯	◯	◯	◯	◯	Håller helt med

20. Ålder:

Evaluation of Google Home

Below are a number of statements about the interactions you just had. Fill out the alternative for each statement which you feel corresponds best with your experience.

1. I enjoyed using the system

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

2. I was not always sure what the system was doing

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

3. The system is useful

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

4. The system always did what I expected

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

5. The interaction with the system is frustrating

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

6. The system is dependable

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

7. A high level of concentration is required when using the system

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

8. The interaction with the system is consistent

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

9. The interaction with the system is repetitive

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

10. The interaction with the system is efficient

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

11. The interaction with the system is irritating

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

12. The interaction with the system is boring

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

13. The system is pleasant

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

14. I felt tense using the system

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

15. The system is easy to use

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

16. To use this system, I need to know things beforehand

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

17. I felt confident using the system

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

18. I felt overwhelmed using the system
    Mark only one oval.

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

19. I would use this system
    Mark only one oval.

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

20. Age

    _____

K

Second user test complete results

A summary of the interviews and individual results for each questionnaire statement.

Summary of interviews
Age range: 25-46
Mean: 35
(Median: 36)

Smart speaker usage:
Never used any smart speaker: 3
Have one themselves: 2
Have tried it: 2

Personalised suggestions:
All mentioned that this is wanted, or even expected, even if it was never a specific question in the questionnaire or during the interview. Also, one participant mentioned that it would be appreciated if the system would ask if the user wishes to continue where they left off when returning to the system. It would feel more like a personal advisor if the suggestions were actually based on the users previous activity.

Overwhelming:
Not too much information presented, 5 participants mentioned that the amount of information felt relevant but still short which was nice.

Easy to follow:
In the context of a voice-based interaction, all participants stated that the interaction felt easy to follow. "My mom could have used this, who usually have troubles with technology" ("det här hade min mamma kunnat använda, som vanligtvis har problem med tekniken").

Knowledge of SR:
Not needed (5 persons), basic knowledge (2 persons)

Wanted to use the system:
Yes, if it weren't for Google. All participants, except for one, expressed that they would have liked to use the system and that the feature to be able to get help is desirable. One mentioned that their usage of swedish radios content would increase with such a system. One mentioned that they would use it only if doing an activity requiring the users hands, that these kinds of systems feels limiting because of the serial interaction.

Categories:
One person believed that the categories were too narrow/specific, one person thought that the categories were hard to grasp.

Descriptions for content:
Three participants mentioned that it is desirable to be able to ask for descriptions about content.

Novice and experienced user:
The test participants liked the different greetings, that you can get a more thorough greeting at first, to then only getting the short one. It was also mentioned by four participants that the system needs to change its behaviour somehow if using it often, so that it is not needed to go through all the steps in order to get suggestions.